

2015

# DGIdb 2.0: Mining clinically relevant drug-gene interactions

Alex H. Wagner

*Washington University School of Medicine in St. Louis*

Adam C. Coffman

*Washington University School of Medicine in St. Louis*

Benjamin J. Ainscough

*Washington University School of Medicine in St. Louis*

Nicholas C. Spies

*Washington University School of Medicine in St. Louis*

Zachary L. Skidmore

*Washington University School of Medicine in St. Louis*

*See next page for additional authors*

Follow this and additional works at: [http://digitalcommons.wustl.edu/open\\_access\\_pubs](http://digitalcommons.wustl.edu/open_access_pubs)

---

## Recommended Citation

Wagner, Alex H.; Coffman, Adam C.; Ainscough, Benjamin J.; Spies, Nicholas C.; Skidmore, Zachary L.; Campbell, Katie M.; Krysiak, Kilannin; Pan, Deng; McMichael, Joshua F.; Eldred, James M.; Walker, Jason R.; Wilson, Richard K.; Mardis, Elaine R.; Griffith, Malachi; and Griffith, Obi L., "DGIdb 2.0: Mining clinically relevant drug-gene interactions." *Nucleic Acids Research*,. 1-9. (2015). [http://digitalcommons.wustl.edu/open\\_access\\_pubs/4379](http://digitalcommons.wustl.edu/open_access_pubs/4379)

This Open Access Publication is brought to you for free and open access by Digital Commons@Becker. It has been accepted for inclusion in Open Access Publications by an authorized administrator of Digital Commons@Becker. For more information, please contact [engeszer@wustl.edu](mailto:engeszer@wustl.edu).

---

**Authors**

Alex H. Wagner, Adam C. Coffman, Benjamin J. Ainscough, Nicholas C. Spies, Zachary L. Skidmore, Katie M. Campbell, Kilannin Krysiak, Deng Pan, Joshua F. McMichael, James M. Eldred, Jason R. Walker, Richard K. Wilson, Elaine R. Mardis, Malachi Griffith, and Obi L. Griffith

# DGIdb 2.0: mining clinically relevant drug–gene interactions

Alex H. Wagner<sup>1</sup>, Adam C. Coffman<sup>1</sup>, Benjamin J. Ainscough<sup>1,2</sup>, Nicholas C. Spies<sup>1</sup>, Zachary L. Skidmore<sup>1</sup>, Katie M. Campbell<sup>1</sup>, Kilannin Krysiak<sup>1</sup>, Deng Pan<sup>3</sup>, Joshua F. McMichael<sup>1</sup>, James M. Eldred<sup>1</sup>, Jason R. Walker<sup>1</sup>, Richard K. Wilson<sup>1,2,3,4</sup>, Elaine R. Mardis<sup>1,2,3,4</sup>, Malachi Griffith<sup>1,2,4,\*</sup> and Obi L. Griffith<sup>1,2,3,4,\*</sup>

<sup>1</sup>McDonnell Genome Institute, Washington University School of Medicine, St. Louis, MO 63108, USA, <sup>2</sup>Siteman Cancer Center, Washington University School of Medicine, St. Louis, MO 63110, USA, <sup>3</sup>Department of Medicine, Washington University School of Medicine, St. Louis, MO 63110, USA and <sup>4</sup>Department of Genetics, Washington University School of Medicine, St. Louis, MO 63110, USA

Received August 31, 2015; Revised October 08, 2015; Accepted October 16, 2015

## ABSTRACT

The Drug–Gene Interaction Database (DGIdb, [www.dgldb.org](http://www.dgldb.org)) is a web resource that consolidates disparate data sources describing drug–gene interactions and gene druggability. It provides an intuitive graphical user interface and a documented application programming interface (API) for querying these data. DGIdb was assembled through an extensive manual curation effort, reflecting the combined information of twenty-seven sources. For DGIdb 2.0, substantial updates have been made to increase content and improve its usefulness as a resource for mining clinically actionable drug targets. Specifically, nine new sources of drug–gene interactions have been added, including seven resources specifically focused on interactions linked to clinical trials. These additions have more than doubled the overall count of drug–gene interactions. The total number of druggable gene claims has also increased by 30%. Importantly, a majority of the unrestricted, publicly-accessible sources used in DGIdb are now automatically updated on a weekly basis, providing the most current information for these sources. Finally, a new web view and API have been developed to allow searching for interactions by drug identifiers to complement existing gene-based search functionality. With these updates, DGIdb represents a comprehensive and user friendly tool for mining the druggable genome for precision medicine hypothesis generation.

## INTRODUCTION

With the increasing availability and decreasing cost of molecular profiling methods, growing attention has been paid to the use of these technologies for characterizing the mechanisms of human disease at the cohort—and more recently, the individual—level (1–3). Improved understanding of such molecular characteristics has been demonstrated to impact the diagnostic, prognostic, and treatment decisions made by clinicians, yielding positive results, less toxicity and improved quality of life (4,5). However, finding well-annotated drug–gene interactions relevant to medical decision-making remains an ongoing challenge. Several resources exist that characterize both established and experimental drug–gene interactions but these resources vary widely in data structure and complexity. Many of these databases are missing useful features provided by DGIdb, such as tools for querying lists of terms (6,7), filtering based on interaction type (6–9) or gene class (7–9), and a RESTful API (7–10). Others lack a search interface entirely, and instead have only a browsing interface (8) or are available only as static supplementary documents published online (11–20).

DGIdb 1.0 (21) was first introduced as a novel resource to enable mining of multiple existing sources of drug–gene interactions and druggable gene categories. *drug–gene interactions* are the observed or inferred interaction between gene products (*targets*) and drug compounds (*ligands*) obtained by literature mining or by parsing publicly available databases. Many of these interactions include information on *interaction types* that specify how the drug interacts with the gene. The curation of DGIdb sources for interaction type data resulted in 33 distinct interaction type terms. In simple cases, these terms were harmonized where the interaction types were unquestionably equivalent (e.g. ‘inhibitor’

\*To whom correspondence should be addressed. Tel: +1 314 747 9248; Fax: +1 314 286 1810; Email: ogriffit@genome.wustl.edu  
Correspondence may also be addressed to Malachi Griffith. Tel: +1 314 286 1274; Fax: +1 314 286 1810; Email: mgriffit@genome.wustl.edu

= ‘inhibitors’). However, in most cases we felt it best to preserve any subtleties of the interaction type according to the source. *Druggable gene categories* include genes that currently may not be targeted therapeutically but are potentially druggable according to their membership in gene categories associated with druggability (e.g. kinases) (21). These categories for the most part were defined by the original seminal paper by Hopkins and Groom (12) who first defined the ‘druggable genome’ as genes belonging to over 130 specific biological types (kinases, nuclear hormone receptors, etc.) that were thought to be druggable, and our curation of these types reduced that definition to 21 distinct types that were incorporated into DGIdb. Russ and Lampel (16) published an update of the ‘Druggable Genome’ concept but without breaking down genes into specific categories. The dGene resource (13) was the most recent update to the concept of a ‘Druggable Genome’ list. These lists, while widely cited and used, were not available in any conveniently usable form. Therefore, we created a generic ‘Druggable Genome’ category, comprised of the union of these three sources to act as a convenient label for these definitive lists. Other categories from other sources are more specific. The ‘Clinically Actionable’ category is a new concept with the DGIdb 2.0 release describing lists of genes that are actively being used in targeted clinical sequencing panels for precision medicine in cancer. The druggable gene categories of DGIdb are by no means mutually exclusive, and in fact often overlap extensively if not entirely (serine threonine kinases, for instance, are a subset of all kinases).

Populating DGIdb 1.0 was accomplished through an extensive curation effort. This process typically included manually extracting and reviewing records, converting them to represent the concepts stored in the DGIdb, and writing custom importers to load the data provided by each source. Thirteen distinct sources were curated into a broad resource that is simple to query and explore. Druggable gene categories could be browsed without any structured query, and the interface accepted most common gene nomenclatures as identifiers for searching drug–gene interactions. Data describing interactions or potentially druggable categories (referred to as interaction or category *claims*) from each of these sources were linked to corresponding drug and gene concepts. Each such source typically included important metadata characterizing aspects of the interaction, such as if a particular gene confers resistance or sensitivity to the drug.

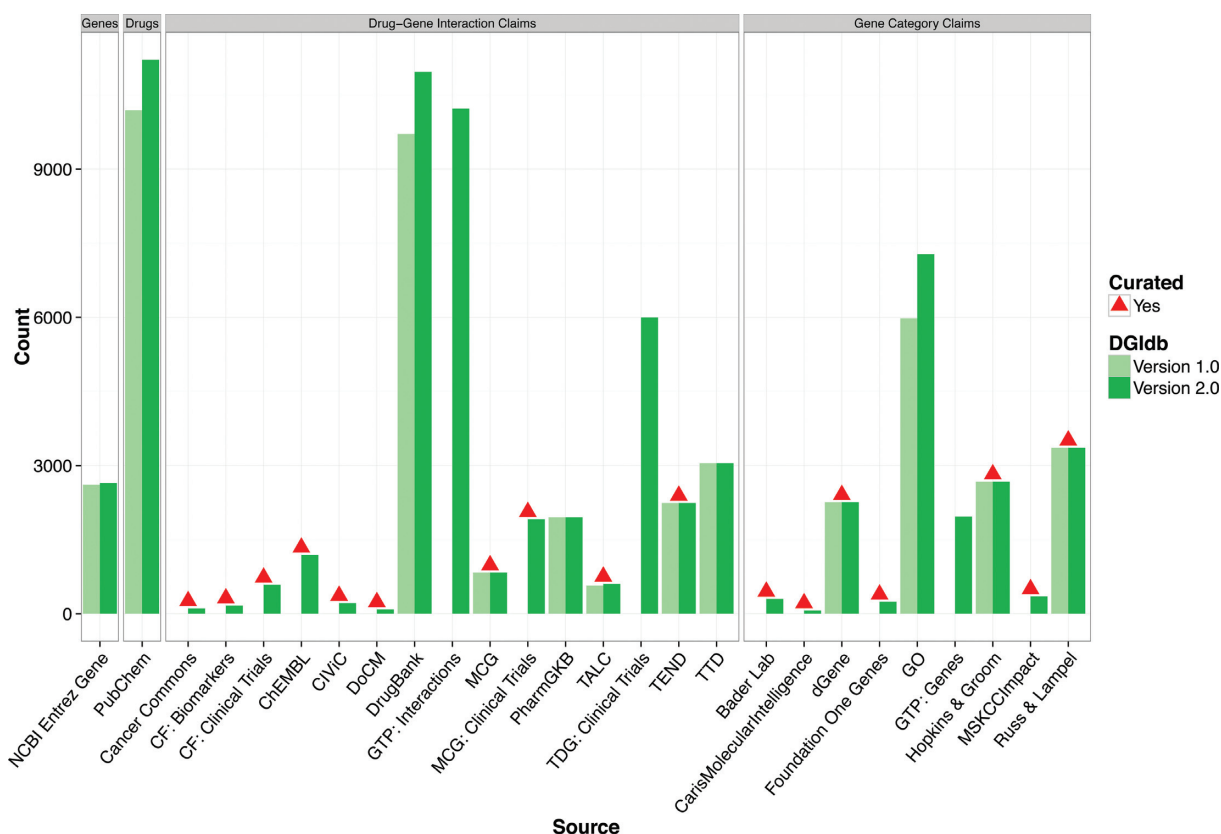
The breadth, quality, and ease-of-access enabled by this resource is reflected in the substantial web traffic DGIdb receives, with ~1500 unique users and ~2300 sessions a month, with an average duration of ~4.4 min/session. In addition, a number of other bioinformatics tools have used DGIdb in developing their own platforms including PANDA (22), iCAGES ([icages.usc.edu](http://icages.usc.edu)), BioGPS (23), OmicsPipe (24), GEMINI (25), StationX ([www.stationxinc.com](http://www.stationxinc.com)) and IHLDB.rf ([www.lungcancerdatabase.com](http://www.lungcancerdatabase.com)), highlighting the broad usefulness of the DGIdb API. The first version of DGIdb was released in October 2013, and has since been cited in over 50 publications.

DGIdb 1.0 also provided a notion of ‘expert-curated’ and ‘non-curated’ resources as a search filter. ‘Expert-curated’ resources, such as dGene (13), MyCancerGenome

(8), TALC (18), and TEND (14), were primarily the result of expert knowledge, or expert-driven curation of the literature. Due to their smaller size, these resources were more closely vetted while imported into DGIdb, and thus the ‘expert-curated’ status also represents the focused curation and literature review effort for these resources by the DGIdb curators. In contrast, the ‘non-curated’ sources, such as GO (26), PharmGKB (9), DrugBank (10) and TTD (7), are generally more inclusive of putative interactions and more comprehensive. For these reasons, these sources do not meet the same standard of trust as the ‘expert-curated’ resources. This is not intended to lessen the value of these excellent resources. In fact, a notable amount of expert curation has gone into each of the ‘non-curated’ resources, and has contributed greatly to the content included therein. The distinction made here is only in the specific context of therapy-relevant drug–gene interactions and/or druggability with a bias toward cancer therapies, and is meant as a sorting tool for our specific purposes. This coarse distinction may be used by clinically-focused researchers to limit their search to the ‘expert-curated’ category for higher standard results.

Despite its popularity, DGIdb 1.0 was limited in several ways. The gene-centric search interface allowed the user to find known drug interactions for a set of gene targets, but did not provide a straightforward way of querying known gene targets for a set of drugs. In addition, several popular resources describing the druggable genome and drug–gene interactions were missing from DGIdb. Finally, some of the resources parsed by DGIdb are continually updated, creating a need for repeated manual updates to keep the associated data current.

With this update, we address these shortcomings. DGIdb 2.0 includes 14 new sources added since version 1.0 for a cumulative total of 27 sources describing drug–gene interactions and druggable gene categories (Tables 1 and 2). These include nine new expert-curated sources describing 10,102 drug–gene interactions potentially relevant to clinical actionability (e.g. relevant clinical trials, etc.). With this focus, DGIdb is better positioned as a resource for drug–gene interaction hypothesis generation in precision medicine efforts. By reviewing drug–gene interactions that are clinically relevant, basic researchers and physician-scientists using DGIdb have a more useful reference for hypothesizing which treatments might provide benefits for individual patients. Of special note are the addition of drug–gene interactions from the Clinical Interpretations of Variants in Cancer (CIViC, [www.civicdb.org](http://www.civicdb.org)) database and the Database of Curated Mutations (DoCM, [www.docm.info](http://www.docm.info)), each of which are a curated source of clinically relevant cancer genes and their therapeutic interactions. DGIdb 2.0 also contains three new sources of potentially druggable gene category annotations, with 2,928 putatively druggable gene category assignments between them (Figure 1). These additions enhance the existing capacity of DGIdb as a tool for prioritizing development of new targeted therapies. Selected sources of gene, drug, interaction, and druggable categories are also now updated automatically, keeping DGIdb current as more information is gathered and deposited in the various sources aggregated by DGIdb. In addition to content updates, DGIdb 2.0 also supports a major new feature—searching interactions by drug identifier—which



**Figure 1.** DGIdb 1.0 and 2.0 content by source. Here, the number of genes with interactions (first panel), drug claims (second panel), drug-gene interaction claims (third panel), and druggable gene categories (fourth panel) are contrasted between DGIdb 1.0 and 2.0. In total, there are currently 2,644 (33 new to DGIdb 2.0) genes with interactions, 11,215 (1,023 new) drug claims, 40,017 (21,624 new) drug-gene interaction claims, and 18,500 (4,224 new) gene category claims. Most of the new sources are considered to be high-quality, 'expert-curated' sources, and are indicated as such by the red triangle symbol. Abbreviations: CF = Clarity Foundation, GTP = Guide To Pharmacology, MCG = My Cancer Genome, TALC = Targeted Agents in Lung Cancer, TTD = Therapeutic Target Database, TEND = Trends in the Exploration of Novel Drug targets, and GO = Gene Ontology MSKCC = Memorial Sloan Kettering Cancer Center.

allows users to find drug-gene interactions by querying the database with drug identifiers in addition to gene identifiers.

These additions improve the use of DGIdb as a tool for precision medicine hypothesis generation. Clinical researchers that wish to make sense of lengthy gene lists may use DGIdb to address numerous seemingly simple questions. One natural question is to ask if there is any information on interactions between drugs and their candidate gene list. Alternatively one might ask if any of the genes in that list belong to potentially druggable categories, such as kinases, G-protein coupled receptors or transporters. DGIdb provides an easy-to-use interface for seeking answers to these types of questions. In the process, DGIdb handles the complexity of gene and drug name ambiguity, provides informative results that detail what resources supplied the interaction or druggable gene category, and summarizes key metadata from each resource along with hyperlinks to the original drug and gene concepts (as applicable). This reduces the burden of needing to search and decipher the results of numerous resources (many of which are lists from supplementary tables in papers or other static datastores). Using DGIdb, the clinical researcher may instead focus their attention on reviewing the interesting results of

these queries instead of exhaustively searching the numerous resources aggregated and curated by the DGIdb.

## EXPANDED CONTENT

DGIdb 2.0 adds significantly to the existing content, providing a broader view of drug-gene interactions and potentially druggable genes from a large number of sources (Figure 1). Most of this new content (89.5%) comes from sources not present in DGIdb 1.0 and the remainder from those original sources that have been updated with DGIdb 2.0. The top priority for new additions was that they describe high-quality, clinically actionable drug-gene interactions. First among these is the Clinical Interpretation of Variants in Cancer (CIViC) web resource: an open-source, community-driven site for the collaborative curation of clinically actionable variants in cancer. At time of last import, CIViC contained 189 evidence statements from 112 published sources describing genes with variants that predict sensitivity or resistance to specific drugs. In total 218 drug-gene interactions were imported into DGIdb from these evidence statements. A second, similar resource, is the open-source Database of Curated Mutations (DoCM). Unlike CIViC which is limited to mutations with proven clinical relevance, DoCM's scope includes any highly curated variant

**Table 1.** Description of drug-gene interaction sources in DGIdb 2.0

Data Type	Source	Description of Sources and Imported Data
Drug Gene Interactions	CancerCommons*+(17)	CancerCommons provides a number of drugs that are approved or undergoing clinical trials for use in lung, prostate, and skin cancer. Drug-gene interactions were extracted from their web pages describing these diseases and imported.
	ChEMBL*(28)	ChEMBL is a database small molecules capable of bioactivity that have been annotated with metadata such as 2-D structures and calculated properties (e.g. Molecular Weight).
	CIVIC*+^	CIVIC is a community-driven platform for identifying actionable variants in cancer. CIVIC genes were pulled via the provided API and drug-gene interactions of those genes were imported.
	Clarity Foundation: Biomarkers*	The Clarity Foundation has analyzed thousands of tumors from ovarian cancers and retrospectively identified biomarkers that predict treatment response to select drugs in patient ovarian tumors. These interaction data were extracted from their web page and imported.
	Clarity Foundation: Clinical Trials*+	124 curated clinical trial records were provided by The Clarity Foundation based on their relevance to breast and ovarian cancer. The 356 interactions from these trials were imported.
	DoCM*+^	DoCM is a manually curated database of mutations associated with cancer progression. Drug-gene interaction data were pulled via the API and imported into DGIdb.
	DrugBank*(10)	DrugBank is a large community resource detailing drug to drug target information. The DrugBank database was downloaded as XML and parsed for import into DGIdb.
	Guide To Pharmacology: Interactions*(6)	The Guide to Pharmacology is an extensive resource detailing pharmacological targets and their corresponding drugs. The <i>interaction</i> and <i>target and family</i> files were downloaded and parsed for import into DGIdb.
	My Cancer Genome(8)	My Cancer Genome provides information on how mutations drive cancer, and the implication of those mutations for treatment, including linking interactions between specific mutations and therapies. These interactions were extracted from their website with permission.
	My Cancer Genome: Clinical Trials*+(8)	My Cancer Genome previously supported the searching of clinical trials to support their interactions. When this was available, these data were extracted from their website with permission and imported into DGIdb.
	PharmGKB(9)	PharmGKB is a knowledge resource that collects information about potentially clinically actionable gene-drug associations. These associations were downloaded with permission from PharmGKB, for parsing and import into DGIdb.
	Targeted Agents in Lung Cancer(18)	This 2012 publication presented a comprehensive survey of targeted agents in lung cancers. The data were provided as pdf tables within the publication and were manually reviewed for import into DGIdb.
	TDG Clinical Trials*+(15)	This 2014 publication evaluated drug-target interactions in the CenterWatch Drugs in Clinical Trials Database. The aggregated drug-gene interactions were provided as a supplementary table, which was manually reviewed for import into DGIdb.
	Trends in the Exploitation of Novel Drug Targets (TEND)(14)	This 2011 publication was an extensive manual curation of FDA approved drugs and their targets from DrugBank. The results of this effort are in a supplementary table, which was manually reviewed for import into DGIdb.
	Therapeutic Target Database (TTD)(7)	The periodically updated TTD resource provides information about therapeutic targets and their corresponding drugs. This information was downloaded from the TTD website as raw text, and then parsed for import into DGIdb.

\* Denotes new source, + Denotes source of clinical evidence interactions, ^ Denotes source with automatic updates.

**Table 2.** Description of drug, gene and druggable gene category sources in DGIdb 2.0

Data Type	Source	Description of Sources and Imported Data
Gene Definitions	NCBI Entrez Gene*(31)	Entrez gene serves as the base concept for genes in DGIdb. Entrez Gene records are imported from NCBI using the online <i>gene_info</i> file.
	Ensembl(33)	Ensembl gene IDs were imported and linked to Entrez gene records. These were imported from Ensembl's ftp site using the transcript GTF file, and also the NCBI <i>gene_info</i> file.
Drug Definitions	PubChem*(32)	PubChem serves as the base concept for drugs in DGIdb. PubChem records for drugs are selectively added, corresponding to imported drug-gene interactions.
Druggable Gene Categories	Bader Lab*(11)	A publication detailing four large protein families. These gene category claims are available from the supplementary data hosted online, which were downloaded for import into DGIdb.
	Caris Molecular Intelligence	We extracted clinically actionable gene lists from the biomarker and NGS panels provided by Caris Life Sciences online at <a href="http://www.carismolecularintelligence.com/">http://www.carismolecularintelligence.com/</a> .
	dGENE(13)	An annotation tool for checking if genes are members of one of ten druggable gene families. The lists of these gene families were provided by the authors and imported into DGIdb.
	Guide To Pharmacology: Genes*(6)	The Guide to Pharmacology is an extensive resource detailing pharmacological targets and their corresponding targets. The <i>interaction</i> and <i>target and family</i> files were downloaded and parsed for import into DGIdb.
	Foundation One Genes*(19)	Foundation One diagnostic test focuses on clinically actionable genes, which are available via an online table on their website. These data were extracted from the site for import into DGIdb.
	GO*(26)	Multiple gene categories were expert selected and subsequently imported from the Gene Ontology (GO) for import into DGIdb.
	Hopkins & Groom(12)	Considered the 'original' druggable genome by DGIdb 1.0, information on genes that were predicted to make good drug targets were extracted from the paper for import into DGIdb.
	MSK IMPACT*(20)	Clinically actionable genes were extracted from the supplementary tables of the Memorial Sloan Kettering IMPACT paper.
	Russ & Lampel(16)	Considered the 'updated' druggable genome by DGIdb 1.0, information on genes that were predicted to make good drug targets were sent by the authors for import into DGIdb.

\* Denotes new source, + Denotes source of clinical evidence interactions, ^ Denotes source with automatic updates.

with a demonstrated role in cancer etiology. A key dataset included in DoCM is the collected annotations of clinically actionable mutations as reported by the Gene Drug Knowledge Database (27).

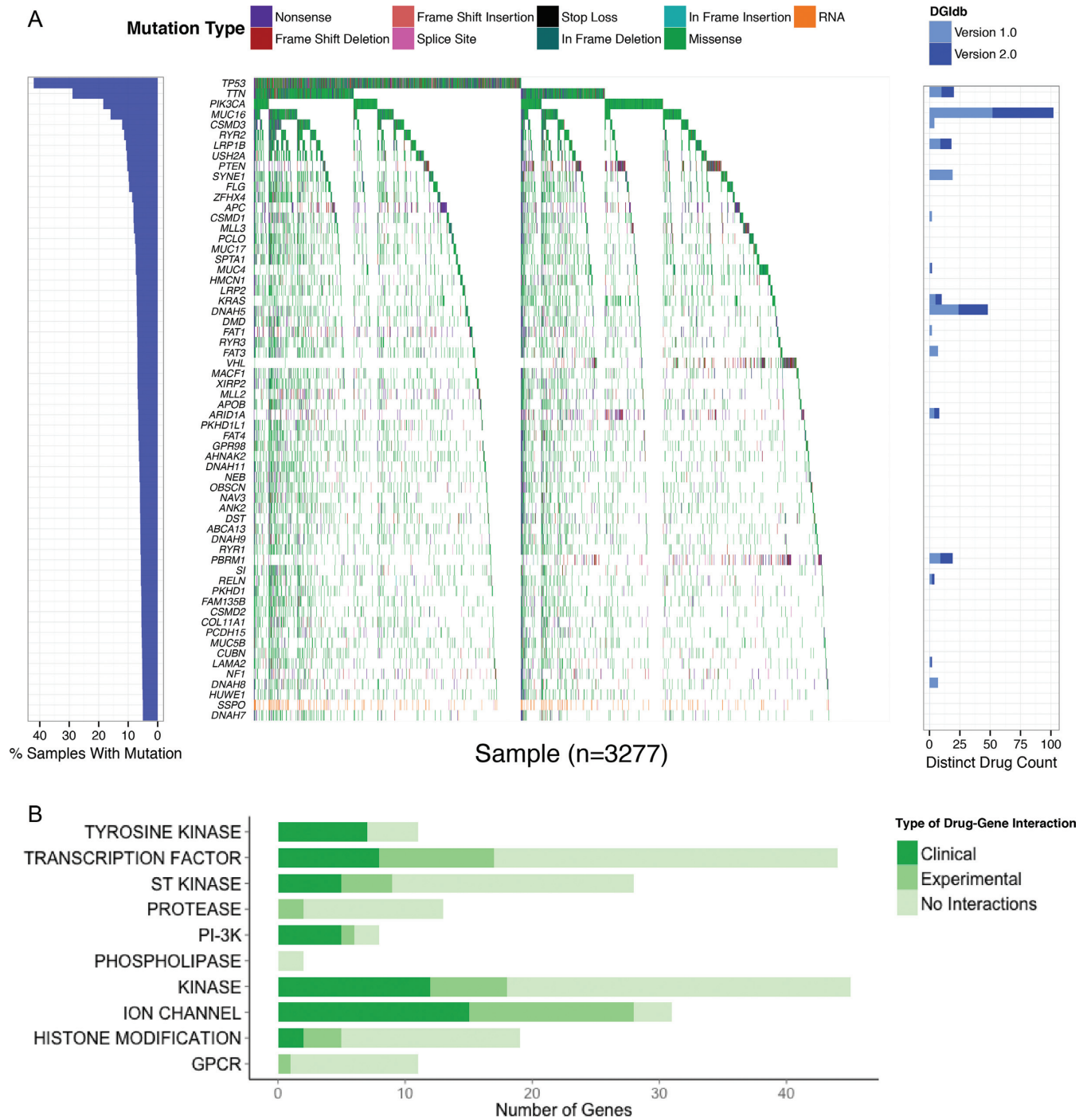
In addition to CIViC and DoCM, six other notable, curated sources of clinical evidence-based drug–gene interactions were integrated in the DGIdb 2.0 release. Rask-Andersen *et al.* (15) released a carefully curated list of putatively novel clinical trial drug targets, based upon mining CenterWatch’s Drugs in Clinical Trials Database ([centerwatch.com](http://centerwatch.com)). Their list of clinical drug–gene interactions is available as a supplementary table in their publication, and was imported into the DGIdb as the ‘Druggable Genome: Clinical Trials’ (‘TdgClinicalTrials’) resource (5,998 claims). My Cancer Genome (MCG) (8) is an original resource included in DGIdb 1.0 (21). However, since the DGIdb 1.0 release, a My Cancer Genome clinical trials search function was introduced. This search tool was used to manually collect information for any clinical trials involving drug–gene interactions. The results were recorded in a table and subsequently imported into DGIdb (1913 claims). Unfortunately, this search functionality is no longer available on the My Cancer Genome site. The Clarity Foundation ([clarityfoundation.org](http://clarityfoundation.org)) works with patients with ovarian cancer to identify therapeutic strategies guided by biomarkers. To support this effort they have created a database of actionable biomarkers through analysis of thousands of patient tumors. These biomarkers, and their effect on drug sensitivity, constitute drug–gene interactions. With permission from Clarity Foundation, these interactions were extracted from their website and imported into DGIdb (165 claims). In addition, Clarity provided us with 124 manually curated clinical trial records with at least one drug–gene interaction for review and import into DGIdb as the Clarity Foundation Clinical Trial resource (585 claims). DGIdb also includes drug–gene interactions from Cancer Commons ([cancercommons.org](http://cancercommons.org)), a group that formerly enrolled patients with lung cancer, skin cancer, and prostate cancer to profile their tumors and help connect patient tumor profiles with potential therapies, in a self-described ‘macro-scale N-of-1 adaptive trial’. This group curated clinical trial records for genes targeted by FDA approved drugs in these three diseases. The resulting lists of drug–gene interactions, annotated with existing or ongoing collection of clinical data, are available online as HTML tables, and were extracted for import into DGIdb. Finally, DGIdb now includes 1,192 high-quality drug–gene interactions involving FDA approved drugs with single gene targets from the ChEMBL 20 database (28).

The largest source of new content was imported from the IUPHAR/BPS Guide To Pharmacology (GTP) (6), accounting for 10 225 interaction claims and 1,969 druggable gene category claims. GTP includes a detailed arrangement of both genes (targets) and compounds (ligands) into functional categories. GTP is the only resource added in DGIdb 2.0 that is considered ‘non-curated’, as the results of DGIdb queries of GTP more closely resemble the specificity of other large resources such as DrugBank. GTP also provides groupings of gene targets into families, which we mapped to existing druggable gene categories through an extensive curation effort (Supplementary Methods).

Beyond those added from the GTP, druggable gene category claims were added from four additional sources. The Bader Laboratory (11) druggable gene category claims describe four protein families: Nuclear Hormone Receptors, Kinases, Ion Channels, and Methyltransferases, long considered to be druggable (16). Prior to import into DGIdb, these claims were only available online as supplementary tables (11). The Foundation One (19) diagnostic test focuses on clinically actionable genes, and DGIdb 2.0 uses the data from this source as a new ‘clinically actionable’ druggable gene category. As with the Bader Lab resource, these genes were obtained from an online table on their website ([foundationone.com/genelist1.php](http://foundationone.com/genelist1.php)). The ‘Clinically Actionable’ gene category also had additions from cancer-focused genetic testing panels from Memorial Sloan Kettering (20) and Caris Life Sciences (<http://www.carismolecularintelligence.com/profilemenu>).

With the content from these new sources, there is an abundance of clinical evidence data to support drug–gene interaction queries. Currently, DGIdb contains 26,298 unique drug–gene interaction claims involving 2,644 genes and 7,569 drugs, and 7,524 unique genes belonging to one or more of 41 potentially druggable gene categories. A total of 8,419 unique genes either have known or potential druggability. We are still far from completely illuminating the ‘dark matter’ of the druggable genome (Isserlin *et al.*, <http://arxiv.org/abs/1102.0448>); of the genes in potentially druggable gene categories, only 26.4% (1,983) actually have a known drug–gene interaction (Supplementary Figure S1). Moreover, it is likely that only a fraction of these represent effective targeted therapies. Conversely, only 75.0% (1,983/2,644) of unique genes with at least one drug–gene interaction belong to an existing druggable gene category. The latter illustrates our potentially still inadequate definition of the ‘druggable genome’, but does represent a modest improvement over the 65.2% (1,704/2,611) coverage observed in DGIdb 1.0.

To demonstrate an application of the expanded content of DGIdb 2.0, we evaluated pan-cancer data from The Cancer Genome Atlas (TCGA) (29), and used the R package GenVisR (<https://github.com/griffithlab/GenVisR>) to illustrate the potential druggability of highly recurrent mutated genes in this cohort (Figure 2A). Of the top 62 recurrently mutated genes (those mutated in at least 5% of the evaluated tumors), 16 have known drug–gene interactions in DGIdb 2.0. Moreover, multiple drugs are available for 14 of these. This is a notable improvement from DGIdb 1.0, with nearly a three-fold (193.6%) overall increase in the number of drug–gene interactions for this set. A search of the top 488 recurrently mutated genes (those genes mutated in at least 2.5% of the evaluated tumors) shows that 137 genes (28%) are present in at least one potentially druggable gene category (Figure 2B). Of these, 57 genes (42%) have known drug–gene interactions, and 30 genes (22%) have drug–gene interactions that are based on data describing clinical evidence. These results demonstrate the usefulness of DGIdb as a tool to explore potential hypotheses for clinically actionable drug interactions with implicated cancer genes.



**Figure 2.** DGIdb analysis of TCGA pan cancer recurrently mutated genes. (A) The 62 genes recurrently mutated in at least 5% of pan-cancer tumors, and the corresponding mutations observed in each tumor. (B) The number of potentially druggable genes among the 488 genes recurrently mutated in at least 2.5% of tumors, grouped by druggable gene categories. Colored bars indicate the fraction of each such category with interactions in a clinical evidence source, a non-clinical evidence source, or without interactions. GPCR = G-Protein Coupled Receptor, PI-3K = Phosphatidylinositol 3 Kinase, ST Kinase = Serine Threonine Kinase



## Search Interactions search for drug-gene interactions by gene or drug names

A

Genes
Drugs

Genes

FLT1

FLT2

FLT3

STK1

MM1

AQP9

LOC100508755

FAKE1

B

Replace Identifiers with Demo List
Clear All Identifiers

Source Databases 15 of 15

Source Trust Level 2 of 2

Gene Categories 41 of 41

Interaction Types 33 of 33

Anti-Neoplastic Drugs Only

Find Drug-Gene Interactions

### Interaction Search Results drug-gene interactions for your genes

Results Summary
Detailed Results
By Gene
Search Term Summary
Statistics
D

**Results Enumerated by Source**

Drug-gene interactions broken down by the source(s) that reported them.

Download as TSV

50 records per page Filter results:

E	Interaction	CI	CC	CE	CF	DC	DB	GT	MC	MC	PG	TA	TE	TT	TC	Sources	PMID	Score
	<a href="#">FLT3 and SUNITINIB</a>						✓			✓		✓	✓		✓	5	6	11
	<a href="#">FLT3 and SORAFENIB</a>	✓				✓	✓					✓	✓		✓	6	4	10
	<a href="#">FLT1 and AXITINIB</a>		✓				✓		✓	✓	✓	✓			✓	7	1	8
	<a href="#">FLT1 and PAZOPANIB</a>				✓		✓	✓	✓	✓		✓			✓	7	1	8
	<a href="#">FLT1 and SUNITINIB</a>						✓			✓		✓	✓		✓	5	2	7
	<a href="#">FLT1 and TIVOZANIB</a>		✓		✓			✓	✓			✓			✓	6	0	6
	<a href="#">FLT1 and VATALANIB</a>						✓	✓	✓			✓			✓	5	1	6
	<a href="#">FLT3 and QUIZARTINIB</a>					✓	✓	✓	✓						✓	5	1	6

**Figure 3.** The drug search interface. (A) Switching between gene and drug searches is as simple as selecting the mode with this button. (B) A search field that accepts a variety of drug or gene identifiers, and provides autocompletion suggestions for terms in the database. (C) Filters allow search results to be limited by the source database, curation level, and interaction type. (D) Users may select to review the results summary or more detailed views. (E) The results summary links genes and drugs by interaction and shows which of the queried resources contain interaction data. (F) The unique sources and PMIDs associated with each interaction are displayed to the user. These are combined to provide a score value that may be used for ranking the interactions.

Downloaded from <http://nar.oxfordjournals.org/> at Washington University, Law School Library on November 10, 2015

## NEW FUNCTIONALITY

Drug-based search is a new and often-requested feature introduced in DGIdb 2.0, and is a major improvement to the existing search capabilities. DGIdb 1.0 provided a search form for entering mixed gene identifiers (IDs) including gene symbols, Entrez IDs, UniProt symbols, etc., to get a list of drug–gene interactions (Figure 3A). We have expanded the functionality of this form to alternatively use drug IDs for the search process (Figure 3B). These may take the form of any of the varied IDs in the database, including PubChem Compound and Substance IDs, Drugbank IDs, common drug names, brand names, CAS numbers, etc. The drug search results work in much the same way as the gene search results (21). Drug terms are first matched against PubChem primary names, then against any other identifiers. If an exact match is found to a DGIdb drug, DGIdb searches for drug–gene interactions involving that drug. Filters allow the search space to be constrained by source, interaction type, trust level or gene category (Figure 3C). If any search terms are ambiguous (i.e. the search term matches more than one drug concept) or no matches are found, this information is summarized in the ‘Search Term Summary’ tab and in detail in the ‘Detailed Results’ tab (Figure 3D). On the default ‘Search Term Summary’ tab, interactions are grouped together and the sources supporting each interaction are indicated (Figure 3E). Finally, an indication of the number of distinct sources and distinct PubMed IDs (PMIDs) supporting each interaction are provided, along with a score that accounts for both of these values. The results are sortable by inclusion in any particular database or by the described metrics (Figure 3F).

Finally, DGIdb 2.0 now features a codebase supporting the automatic update from sources for each of the core concepts in DGIdb: genes, drugs, drug–gene interactions and druggable gene categories. These updaters have been built for the most actively updated existing sources (Figure 1), including DrugBank (10) (drug–gene interactions), Gene Ontology (26) (druggable gene categories), Entrez (30) (genes), and PubChem (31) (drugs). For sources updated from version 1.0 to 2.0, the smallest increase we observed was for PubChem, at 1,023 additional drugs (10%). The largest increase was for Entrez Gene, which increased by 13,348 genes (31%), and precisely reflects the corresponding increase in gene identifiers as reported by Entrez (Supplementary Figure S2). These large changes highlight the importance of a continually updated resource, particularly for resources that are applicable to precision medicine hypothesis generation.

## CONCLUSION AND FUTURE DIRECTIONS

DGIdb 2.0 has greatly expanded in content for both drug–gene interactions and druggable gene categories. The numerous sources describing clinical evidence of drug–gene interactions have created a useful resource for reviewing the nature of these interactions in a clinical setting. These additions have been made with the express purpose of improving DGIdb for use in clinically relevant applications. DGIdb remains a powerful tool for web-based searches, and also as an add-on service for clinical informatics pipelines through the well-documented API. This aspect of DGIdb has been

improved with the addition of drug-based searches of drug–gene interactions, enhancing the facility with which DGIdb can mine the druggable genome. Automated content updates for large, regularly updated sources now ensures that DGIdb 2.0 will provide users with the most current knowledge of clinically actionable drug–gene interactions.

As DGIdb continues to grow, we anticipate additional improvements to accommodate the complexity of representing drug–gene interactions and druggable gene category assignments. A future goal is to better represent many-to-one relationships between genes and drugs for interactions. This would include both multi-drug interactions with individual gene targets, and drugs that simultaneously interact with multiple genes. These types of data will become increasingly available for aggregation and representation as clinical data describing these events are entered as evidence in resources such as CIViC. We also plan a corresponding expansion of our API, adding options for queries that take advantage of these new data relationships. We also wish to study a more sophisticated scoring and ranking system that takes into account other metrics, such as source quality and FDA drug information. Additionally, we plan to investigate better strategies for the grouping of related drug concepts. Finally, we plan to continue adding new sources of drug–gene interactions as identified by our group or suggested by users of the resource. Several features and sources of the DGIdb were developed through feedback from this community, for which we express our gratitude.

## SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online.

## FUNDING

National Institutes of Health [U54HG003079 to R.W., K99HG007940 to M.G., K22CA188163 to O.G.]. Funding for open access charge: Departmental funding (to O.L.G.). *Conflict of interest statement.* None declared.

## REFERENCES

- Collins, F.S. and Varmus, H. (2015) A new initiative on precision medicine. *N. Engl. J. Med.*, **372**, 793–795.
- Gonzalez-Garay, M.L. (2014) The road from next-generation sequencing to personalized medicine. *Per Med.*, **11**, 523–544.
- Gargis, A.S., Kalman, L., Berry, M.W., Bick, D.P., Dimmock, D.P., Hambuch, T., Lu, F., Lyon, E., Voelkerding, K.V., Zehnbauser, B.A. *et al.* (2012) Assuring the quality of next-generation sequencing in clinical laboratory practice. *Nat. Biotechnol.*, **30**, 1033–1036.
- Adams, V.R. and Harvey, R.D. (2010) Histological and genetic markers for non-small-cell lung cancer: customizing treatment based on individual tumor biology. *Am. J. Health Syst. Pharm.*, **67**, S3–S9, quiz S15–16.
- Lamparella, N., Barochia, A. and Almokadem, S. (2013) Impact of genetic markers on treatment of non-small cell lung cancer. *Adv. Exp. Med. Biol.*, **779**, 145–164.
- Pawson, A.J., Sharman, J.L., Benson, H.E., Faccenda, E., Alexander, S.P., Buneman, O.P., Davenport, A.P., McGrath, J.C., Peters, J.A., Southan, C. *et al.* (2014) The IUPHAR/BPS Guide to PHARMACOLOGY: an expert-driven knowledgebase of drug targets and their ligands. *Nucleic Acids Res.*, **42**, D1098–1106.
- Zhu, F., Shi, Z., Qin, C., Tao, L., Liu, X., Xu, F., Zhang, L., Song, Y., Liu, X., Zhang, J. *et al.* (2012) Therapeutic target database update 2012: a resource for facilitating target-oriented drug discovery. *Nucleic Acids Res.*, **40**, D1128–1136.

8. Yeh,P., Chen,H., Andrews,J., Naser,R., Pao,W. and Horn,L. (2013) DNA-Mutation Inventory to Refine and Enhance Cancer Treatment (DIRECT): a catalog of clinically relevant cancer mutations to enable genome-directed anticancer therapy. *Clin. Cancer Res.*, **19**, 1894–1901.
9. McDonagh,E.M., Whirl-Carrillo,M., Garten,Y., Altman,R.B. and Klein,T.E. (2011) From pharmacogenomic knowledge acquisition to clinical applications: the PharmGKB as a clinical pharmacogenomic biomarker resource. *Biomark Med.*, **5**, 795–806.
10. Law,V., Knox,C., Djoumbou,Y., Jewison,T., Guo,A.C., Liu,Y., Maciejewski,A., Arndt,D., Wilson,M., Neveu,V. *et al.* (2014) DrugBank 4.0: shedding new light on drug metabolism. *Nucleic Acids Res.*, **42**, D1091–1097.
11. Edwards,A.M., Isserlin,R., Bader,G.D., Frye,S.V., Willson,T.M. and Yu,F.H. (2011) Too many roads not taken. *Nature*, **470**, 163–165.
12. Hopkins,A.L. and Groom,C.R. (2002) The druggable genome. *Nat. Rev. Drug Discov.*, **1**, 727–730.
13. Kumar,R.D., Chang,L.W., Ellis,M.J. and Bose,R. (2013) Prioritizing potentially druggable mutations with dGene: an annotation tool for cancer genome sequencing data. *PLoS One*, **8**, e67980.
14. Rask-Andersen,M., Almen,M.S. and Schioth,H.B. (2011) Trends in the exploitation of novel drug targets. *Nat. Rev. Drug Discov.*, **10**, 579–590.
15. Rask-Andersen,M., Masuram,S. and Schioth,H.B. (2014) The druggable genome: evaluation of drug targets in clinical trials suggests major shifts in molecular class and indication. *Annu. Rev. Pharmacol. Toxicol.*, **54**, 9–26.
16. Russ,A.P. and Lampel,S. (2005) The druggable genome: an update. *Drug Discov. Today*, **10**, 1607–1610.
17. Shrager,J., Tenenbaum,J.M. and Travers,M. (2011) Cancer commons: biomedicine in the internet age. *Collab. Comput. Technol. Biomed. Res.*, 161–177.
18. Somaiah,N., Simon,N.G. and Simon,G.R. (2012) A tabulated summary of targeted and biologic therapies for non-small-cell lung cancer. *J. Thorac. Oncol.*, **7**, S342–368.
19. Wagle,N., Berger,M.F., Davis,M.J., Blumenstiel,B., Defelice,M., Pochanard,P., Ducar,M., Van Hummelen,P., Macconail,L.E., Hahn,W.C. *et al.* (2012) High-throughput detection of actionable genomic alterations in clinical tumor samples by targeted, massively parallel sequencing. *Cancer Discov.*, **2**, 82–93.
20. Cheng,D.T., Mitchell,T.N., Zehir,A., Shah,R.H., Benayed,R., Syed,A., Chandramohan,R., Liu,Z.Y., Won,H.H., Scott,S.N. *et al.* (2015) Memorial sloan kettering-integrated mutation profiling of actionable cancer targets (MSK-IMPACT): a hybridization capture-based next-generation sequencing clinical assay for solid tumor molecular oncology. *J. Mol. Diagn.*, **17**, 251–264.
21. Griffith,M., Griffith,O.L., Coffman,A.C., Weible,J.V., McMichael,J.F., Spies,N.C., Koval,J., Das,I., Callaway,M.B., Eldred,J.M. *et al.* (2013) DGIdb: mining the druggable genome. *Nat. Methods*, **10**, 1209–1210.
22. Hart,S.N., Moore,R.M., Zimmermann,M.T., Oliver,G.R., Egan,J.B., Bryce,A.H. and Kocher,J.P. (2015) PANDA: pathway and annotation explorer for visualizing and interpreting gene-centric data. *PeerJ*, **3**, e970.
23. Wu,C., Orozco,C., Boyer,J., Leglise,M., Goodale,J., Batalov,S., Hodge,C.L., Haase,J., Janes,J., Huss,J.W.3rd *et al.* (2009) BioGPS: an extensible and customizable portal for querying and organizing gene annotation resources. *Genome Biol.*, **10**, R130.
24. Fisch,K.M., Meissner,T., Gioia,L., Ducom,J.C., Carland,T.M., Loguercio,S. and Su,A.I. (2015) Omics Pipe: a community-based framework for reproducible multi-omics data analysis. *Bioinformatics*, **31**, 1724–1728.
25. Paila,U., Chapman,B.A., Kirchner,R. and Quinlan,A.R. (2013) GEMINI: integrative exploration of genetic variation and genome annotations. *PLoS Comput. Biol.*, **9**, e1003153.
26. Gene Ontology, C. (2015) Gene Ontology Consortium: going forward. *Nucleic Acids Res.*, **43**, D1049–D1056.
27. Dienstmann,R., Dong,F., Borger,D., Dias-Santagata,D., Ellisen,L.W., Le,L.P. and Iafrate,A.J. (2014) Standardized decision support in next generation sequencing reports of somatic cancer variants. *Mol. Oncol.*, **8**, 859–873.
28. Bento,A.P., Gaulton,A., Hersey,A., Bellis,L.J., Chambers,J., Davies,M., Kruger,F.A., Light,Y., Mak,L., McGlinchey,S. *et al.* (2014) The ChEMBL bioactivity database: an update. *Nucleic Acids Res.*, **42**, D1083–1090.
29. Kandoth,C., McLellan,M.D., Vandin,F., Ye,K., Niu,B., Lu,C., Xie,M., Zhang,Q., McMichael,J.F., Wyczalkowski,M.A. *et al.* (2013) Mutational landscape and significance across 12 major cancer types. *Nature*, **502**, 333–339.
30. Brown,G.R., Hem,V., Katz,K.S., Ovetsky,M., Wallin,C., Ermolaeva,O., Tolstoy,I., Tatusova,T., Pruitt,K.D., Maglott,D.R. *et al.* (2015) Gene: a gene-centered information resource at NCBI. *Nucleic Acids Res.*, **43**, D36–D42.
31. Bolton,E.E., Wang,Y., Thiessen,P.A. and Bryant,S.H. (2008) PubChem: integrated platform of small molecules and biological activities. *Annu. Rep. Comput. Chem.*, **4**, 217–241.
32. Cunningham,F., Amode,M., Barrell,D., Beal,K. *et al.* (2015) Ensembl 2015. *Nucleic Acids Res.*, **43**, D662–D669.