SUPPLEMENTAL MATERIAL

TABLES

Supplemental Table 1. Description of Drug-Gene Interaction Sources in DGIdb 3.0

Data Type	Source	Description of Sources and Imported Data	3.0 Status
Drug Gene Interactions	CancerCommons (1)	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.	P
	Cancer Genome Interpreter (CGI) (https://www.cancergenomeinterpreter.org)	CGI is a database that provides information about identified alterations and currently available treatment treatment. The *biomarkers per variant* file was downloaded and parsed for import into the DGIdb.	N, A
	ChEMBL: Interactions (2)	ChEMBL is a database of small molecules capable of bioactivity that have been annotated with metadata such as 2-D structures and calculated properties (e.g. Molecular Weight). Drug-gene interactions were pulled from the database.	A, U
	CIViC (3)	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.	A, U
	Clearity Foundation: Biomarkers	The Clearity 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.	P
	Clearity Foundation: Clinical Trials	124 curated clinical trial records were provided by The Clearity Foundation based on their relevance to breast and ovarian cancer. The 356 interactions from these trials were imported.	P
	DoCM (4)	DoCM is a manually curated database of mutations associated with cancer progression. Drug-gene interaction data were pulled via the API and imported into the DGIdb.	A, U
	DrugBank (5)	DrugBank is a large community resource detailing drug to drug target information. The DrugBank database was downloaded as XML and parsed for import into the DGIdb.	A, U
	FDA Biomarkers (https://www.fda.gov/Drugs/ScienceResearch/ResearchAreas/Pharmacogenetics/)	The FDA provides drug-gene interactions in their *Pharmacogenomic Biomarkers in Drug Labeling* table. 248 drug-gene interactions were pulled from this table for various diseases.	N, S
	Guide To Pharmacology: Interactions (6)	The Guide to Pharmacology is an extensive resource detailing pharmacological targets and their corresponding drugs. The *interaction* and *target and family* files were downloaded and parsed for import into the DGIdb.	A, U
	Jax-Clinical Knowledgebase (CKB) (7)	CKB is a knowledge base that provides drug-gene interactions based on a tumor's genomic profile as well as drug efficacy and resistance evidence. Drug-gene interactions and metadata were pulled via the API and imported into the DGIdb.	N, A
	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.	U
	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 the DGIdb.	P, S
	National Cancer Institute (NCI) Cancer Gene Index (https://wiki.nci.nih.gov/display/cageneindex)	This resource was curated through scraping interactions from publications followed by manual curation. The *Gene-Compound xml* file was downloaded and parsed for inclusion into the DGIdb.	N, S

	Precision Oncology Knowledge Base (OncoKB) (9)	OncoKB is a knowledgebase that provides information about the specific effects of a somatic molecular alteration and potential treatment options with predicted drug response information. Drug-gene interactions and metadata were pulled via the API and imported into the DGIdb.	N, A
	PharmGKB (10)	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 the DGIdb.	P
	Targeted Agents in Lung Cancer (TALC) (11)	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 the DGIdb.	U
	TDG Clinical Trials (12)	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 the DGIdb.	P, S
	Trends in the Exploitation of Novel Drug Targets (TEND) (13)	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 the DGIdb.	P
	Therapeutic Target Database (TTD) (14)	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 the DGIdb.	P

A: Denotes source that is automatically updated, N: Denotes source that is new in the DGIdb 3.0, P: Denotes source that was previously in the DGIdb 2.0 and has not been updated, S: Denotes source that is static (e.g. a table from a paper or website or a data download from a paper), U: Denotes source that has been updated in the DGIdb 3.0.

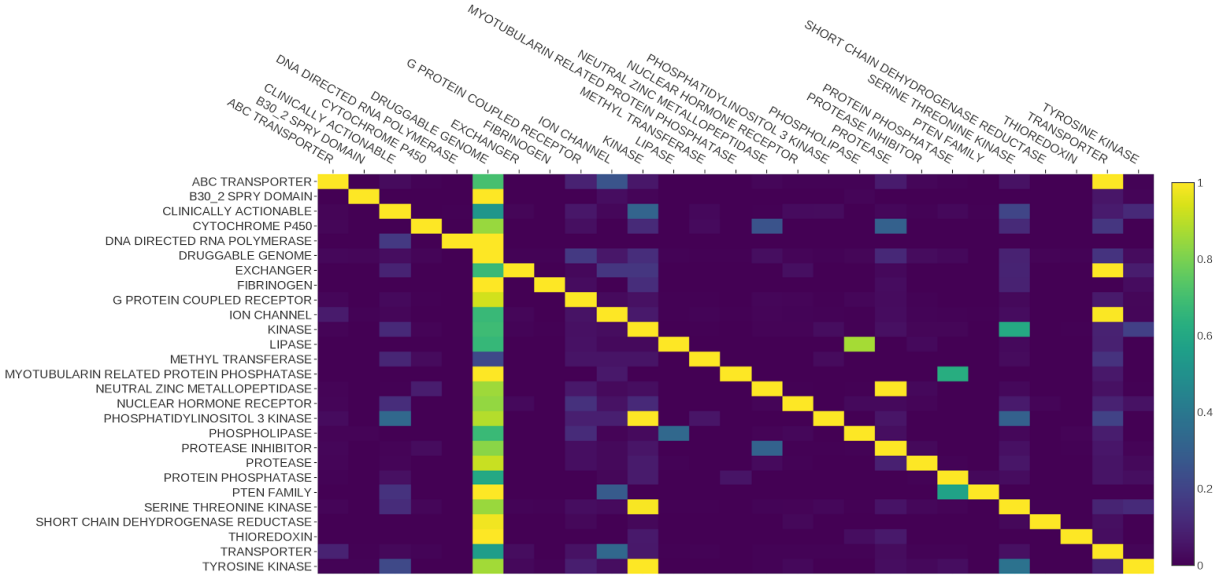
Supplemental Table 2. Description of Drug, Gene, and Druggable Gene Category Sources in DGIdb 3.0

Data Type	Source	Description of Sources and Imported Data	3.0 Status
Gene Definitions	NCBI Entrez Gene (14)	Entrez gene serves as the canonical concept for genes in the DGIdb. Entrez Gene records are imported from NCBI using the online *gene_info* file.	A, U
	Ensembl (15)	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 *gene_info* file.	A, U
Drug Definitions	PubChem (16)	PubChem serves as the canonical concept for drugs in the DGIdb. PubChem records for drugs are selectively added, corresponding to imported drug-gene interactions.	A, U
	ChEMBL: Molecules (1)	ChEMBL is a database of small molecules capable of bioactivity that have been annotated with metadata such as 2-D structures and calculated properties (e.g. Molecular Weight).	A, U
Druggable Gene Categories	Bader Lab (17)	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 the DGIdb.	P, S
	Caris Molecular Intelligence	We extracted clinically actionable gene lists from the biomarker and NGS panels provided by Caris Life Sciences online at http://www.carismolecularintelligence.com/.	P
	dGENE (18)	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 the DGIdb.	P
	Guide To Pharmacology: Genes (5)	The Guide to Pharmacology is an extensive resource detailing pharmacological targets and their corresponding targets. The *interaction* and *target and family* files were downloaded and parsed for import into the DGIdb.	A, U
	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 the DGIdb.	P, S
	GO (20)	Multiple gene categories were expert selected and subsequently imported from the Gene Ontology (GO) for import into the DGIdb.	A, U

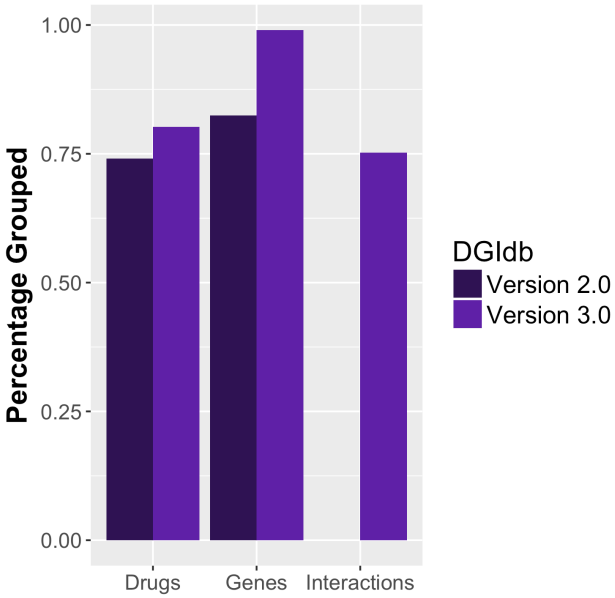
	HingoraniCasas (21)	A publication from the Casas lab that used computational approaches to identify druggable genes from genome-wide association studies. Interaction claims were curated from this paper's supplementary information.	N, S
	Hopkins & Groom (22)	Considered the 'original' druggable genome by the DGIdb 1.0, information on genes that were predicted to make good drug targets were extracted from the paper for import into the DGIdb.	P, S
	MSK IMPACT (23)	Clinically actionable genes were extracted from the supplementary tables of the Memorial Sloan Kettering IMPACT paper.	P, S
	Russ & Lampel (24)	Considered the 'updated' druggable genome by the DGIdb 1.0, information on genes that were predicted to make good drug targets were sent by the authors for import into the DGIdb.	P, S

A: Denotes source that is automatically updated, N: Denotes source that is new in the DGIdb 3.0, P: Denotes source that was previously in the DGIdb 2.0 and has not been updated, S: Denotes source that is static (e.g. a table from a paper or website or a data download from a paper), U: Denotes source that has been updated in the DGIdb 3.0.

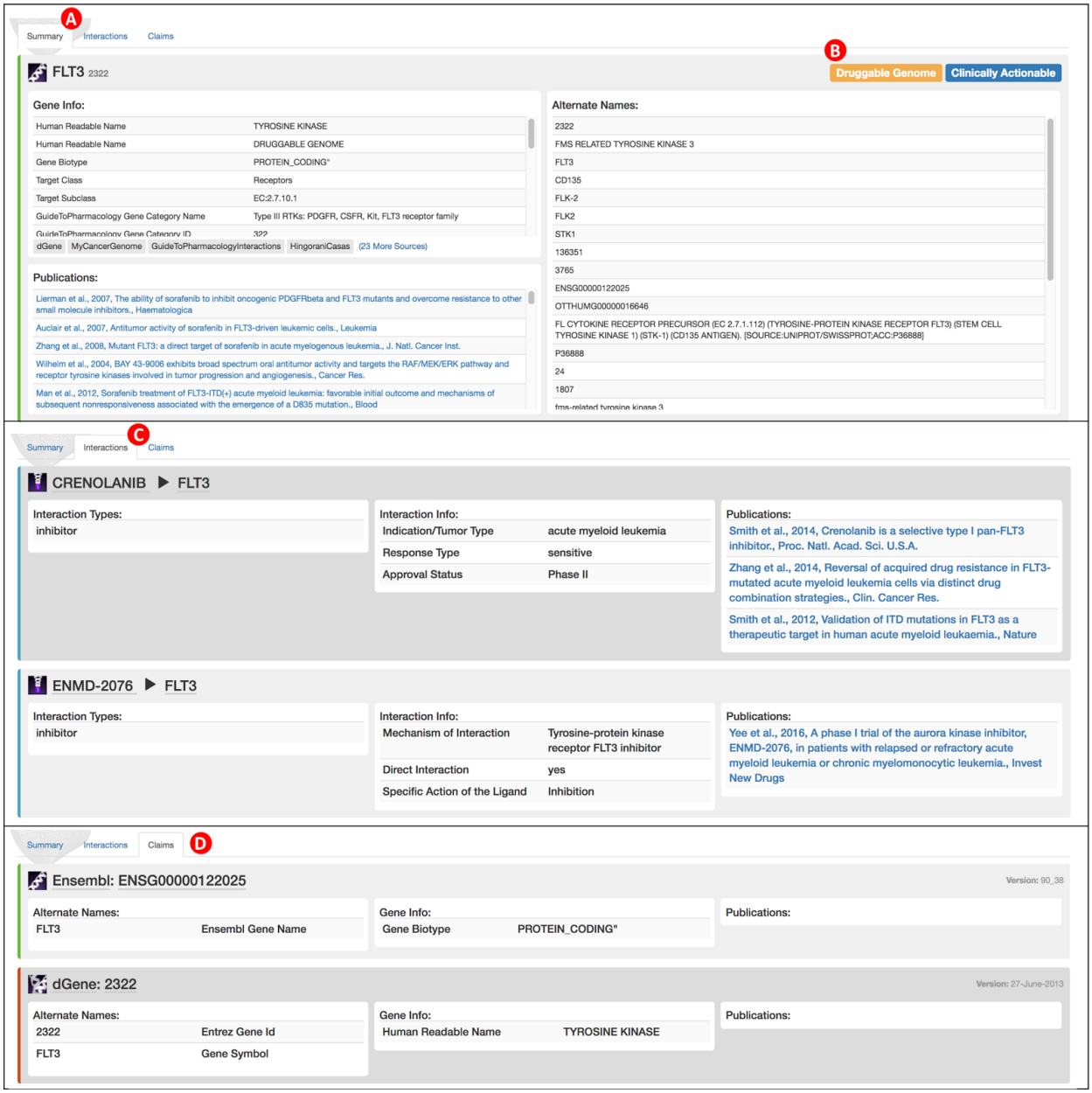
FIGURES



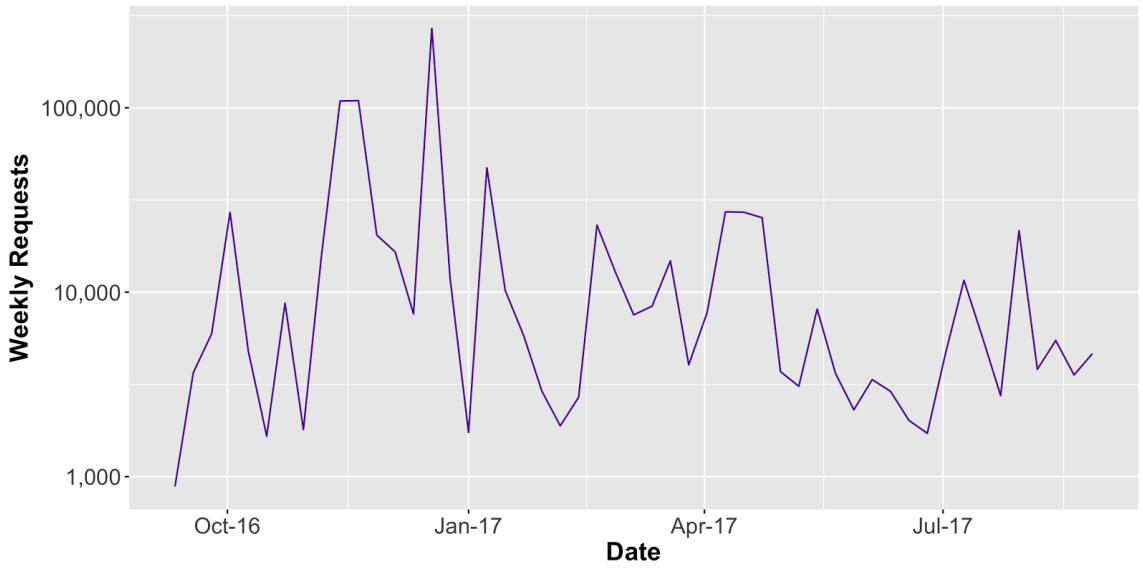
Supplemental Figure 1: DGIdb 3.0 Druggable Genome Categories. This figure shows the percent overlap of the various gene categories of the DGIdb 3.0, with the overlap expressed as a percentage of the category along the Y-axis. For example, the Druggable Genome row has mostly low values due to only a small percentage of the genes within it matching each of the smaller gene categories (e.g. 12% of Druggable Genome genes in the Kinase category). However, the Druggable Genome column has mostly high values due to most of the categories along the y axis being near-subsets of the Druggable Genome category (68% of Kinase genes are in the Druggable Genome category).



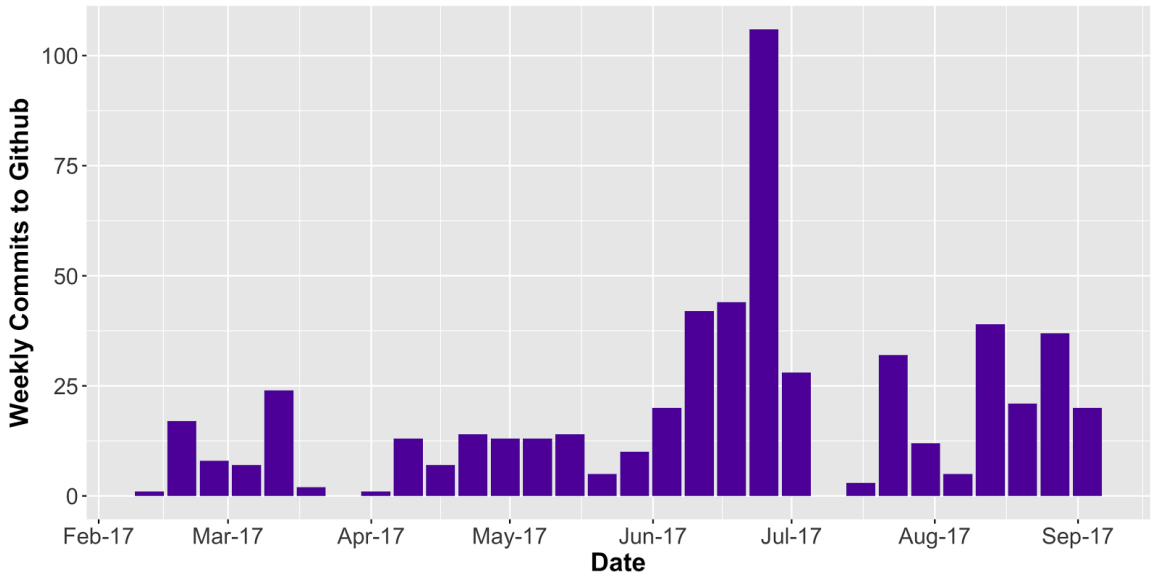
Supplemental Figure 2: DGIdb 2.0 and 3.0 claim grouping. A comparison of the percentage of drug and gene claims successfully grouped between versions 2.0 and 3.0 and an evaluation of the percentage of grouped interaction claims in 3.0, a concept that was present in previous versions.



Supplemental Figure 3: A representative drug group record in the DGIdb. **(A)** The summary tab reports information about the drug metadata, drug aliases, and any publications associated with that drug. **(B)** Tags denoting commonly requested information are now featured on the summary tab. These tags align with our new preset filters. **(C)** The interactions tab reports each interaction claim that the drug is involved in along with interaction metadata and publications supporting the interaction. **(D)** The claim view shows each source that provides a claim for the drug along with the specific information and publications that resource provides.



Supplemental Figure 4: DGIdb Usage. The API requests for each week from September 2016 to August 2017 are shown.



Supplemental Figure 5: DGIdb Github Commit History. The commit history for the DGIdb Github repository (https://github.com/griffithlab/dgi-db) from February 2017 to September 2017 is shown.

REFERENCES

1. Shrager,J., Tenenbaum,J.M. and Travers,M. Cancer commons: biomedicine in the internet age. *Collab. Comput. Technol. Biomed. Res.*
2. Bento,A.P., Gaulton,A., Hersey,A., Bellis,L.J., Chambers,J., Davies,M., Krüger,F.A., Light,Y., Mak,L., McGlinchey,S., *et al.* (2014) The ChEMBL bioactivity database: an update. *Nucleic Acids Res.*, **42**, D1083-90.
3. Griffith,M., Spies,N.C., Krysiak,K., McMichael,J.F., Coffman,A.C., Danos,A.M., Ainscough,B.J., Ramirez,C.A., Rieke,D.T., Kujan,L., *et al.* (2017) CIViC is a community knowledgebase for expert crowdsourcing the clinical interpretation of variants in cancer. *Nat. Genet.*, **49**, 170–174.
4. Ainscough,B.J., Griffith,M., Coffman,A.C., Wagner,A.H., Kunisaki,J., Choudhary,M.N., McMichael,J.F., Fulton,R.S., Wilson,R.K., Griffith,O.L., *et al.* (2016) DoCM: a database of curated mutations in cancer. *Nat. Methods*, **13**, 806–807.
5. 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.
6. Southan,C., Sharman,J.L., Benson,H.E., Faccenda,E., Pawson,A.J., Alexander,S.P.H., Buneman,O.P., Davenport,A.P., McGrath,J.C., Peters,J.A., *et al.* (2016) The IUPHAR/BPS Guide to PHARMACOLOGY in 2016: towards curated quantitative interactions between 1300 protein targets and 6000 ligands. *Nucleic Acids Res.*, **44**, D1054–D1068.
7. Patterson,S.E., Liu,R., Statz,C.M., Durkin,D., Lakshminarayana,A. and Mockus,S.M. (2016) The clinical trial landscape in oncology and connectivity of somatic mutational profiles to targeted therapies. *Hum. Genomics*, **10**, 4.
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. Off. J. Am. Assoc. Cancer Res.*, **19**, 1894–1901.
9. Chakravarty,D., Gao,J., Phillips,S., Kundra,R., Zhang,H., Wang,J., Rudolph,J.E., Yaeger,R., Soumerai,T., Nissan,M.H., *et al.* (2017) OncoKB: A Precision Oncology Knowledge Base. *JCO Precis. Oncol.*, 10.1200/PO.17.00011.
10. 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.
11. Simon,G.R. and Somaiah,N. (2014) A tabulated summary of targeted and biologic therapies for non-small-cell lung cancer. *Clin. Lung Cancer*, **15**, 21–51.
12. Rask-Andersen,M., Masuram,S. and Schiöth,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.
13. Rask-Andersen,M., Almén,M.S. and Schiöth,H.B. (2011) Trends in the exploitation of novel drug targets. *Nat. Rev. Drug Discov.*, **10**, 579–590.
14. 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.
15. Yates,A., Akanni,W., Amode,M.R., Barrell,D., Billis,K., Carvalho-Silva,D., Cummins,C., Clapham,P., Fitzgerald,S., Gil,L., *et al.* (2016) Ensembl 2016. *Nucleic Acids Res.*, **44**, D710–D716.
16. Kim,S., Thiessen,P.A., Bolton,E.E., Chen,J., Fu,G., Gindulyte,A., Han,L., He,J., He,S., Shoemaker,B.A., *et al.* (2016) PubChem Substance and Compound databases. *Nucleic Acids Res.*, **44**, D1202-1213.
17. 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.
18. 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.
19. Wagle,N., Berger,M.F., Davis,M.J., Blumenstiel,B., DeFelice,M., Pochanard,P., Ducar,M., Van Hummelen,P., MacConaill,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. Gene Ontology Consortium: going forward (2015) *Nucleic Acids Res.*, **43**, D1049–D1056.

21. Finan,C., Gaulton,A., Kruger,F.A., Lumbers,R.T., Shah,T., Engmann,J., Galver,L., Kelley,R., Karlsson,A., Santos,R., *et al.* (2017) The druggable genome and support for target identification and validation in drug development. *Sci. Transl. Med.*, **9**, eaag1166.
22. Hopkins,A.L. and Groom,C.R. (2002) The druggable genome. *Nat. Rev. Drug Discov.*, **1**, 727–730.
23. 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. JMD*, **17**, 251–264.
24. Russ,A.P. and Lampel,S. (2005) The druggable genome: an update. *Drug Discov. Today*, **10**, 1607–1610.