

# MyGeneset.info: an interactive and programmatic platform for community-curated and user-created collections of genes

Ricardo Avila<sup>1,†</sup>, Vincent Rubinetti<sup>2,†</sup>, Xinghua Zhou<sup>1</sup>, Dongbo Hu<sup>2</sup>,  
 Zhongchao Qian<sup>1</sup>, Marco Alvarado Cano<sup>1</sup>, Everaldo Rodolpho<sup>1</sup>, Ginger Tsueng<sup>1</sup>,  
 Casey Greene<sup>2,\*</sup> and Chunlei Wu<sup>1,\*</sup>

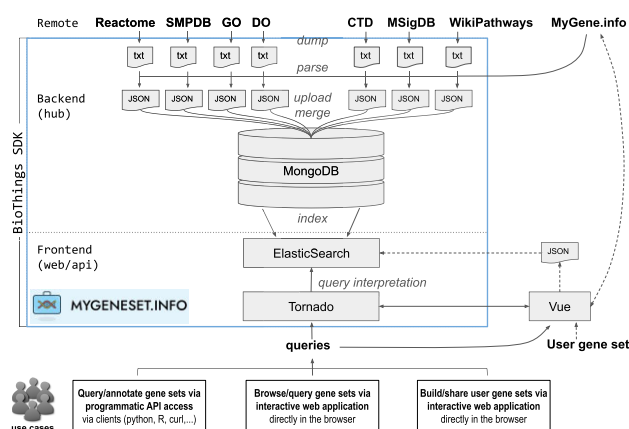
<sup>1</sup>Department of Integrative Structural and Computational Biology, The Scripps Research Institute, La Jolla, CA, USA and <sup>2</sup>Department of Biochemistry and Molecular Genetics, Center for Health AI, University of Colorado School of Medicine, Aurora, CO, USA

Received February 06, 2023; Revised March 28, 2023; Editorial Decision April 04, 2023; Accepted April 13, 2023

## ABSTRACT

Gene definitions and identifiers can be painful to manage—more so when trying to include gene function annotations as this can be highly context-dependent. Creating groups of genes or gene sets can help provide such context, but it compounds the issue as each gene within the gene set can map to multiple identifiers and have annotations derived from multiple sources. We developed MyGeneSet.info to provide an API for integrated annotations for gene sets suitable for use in analytical pipelines or web servers. Leveraging our previous work with MyGene.info (a server that provides gene-centric annotations and identifiers), MyGeneSet.info addresses the challenge of managing gene sets from multiple resources. With our API, users readily have read-only access to gene sets imported from commonly-used resources such as Wikipathways, CTD, Reactome, SMPDB, MSigDB, GO, and DO. In addition to supporting the access and reuse of approximately 180k gene sets from humans, common model organisms (mice, yeast, etc.), and less-common ones (e.g. black cottonwood tree), MyGeneSet.info supports user-created gene sets, providing an important means for making gene sets more FAIR. User-created gene sets can serve as a way to store and manage collections for analysis or easy dissemination through a consistent API.

## GRAPHICAL ABSTRACT



## INTRODUCTION

Different types of gene-centric information can be accessed through numerous resources, many of which have their own identifiers for accessing information on each gene or gene product. For example, the identifiers for ‘N-glycanase 1’ include 55768 (NCBI Gene) (1), 610661 (OMIM) (2), Q96IV0 (UniProt) (3), etc. Further, each resource may include different annotations describing this gene or gene product such as information on its molecular function (Gene Ontology) (4), chromosomal location (UCSC Genome Browser) (5), clinical significance (ClinGen) (6), and more (7–11). Ensuring up-to-date information on a specific gene can be time-consuming as users would need to either continuously download and merge data from different resources or ensure up-to-date mappings of resource-specific identifiers in their maintenance pipelines.

\*To whom correspondence should be addressed. Tel: +1 858 784 2111; Email: cwu@scripps.edu  
 Correspondence may also be addressed to Casey Greene. Tel: +1 303 724 0343; Email: casey.s.greene@cuanschutz.edu  
<sup>†</sup>The authors wish it to be known that, in their opinion, the first two authors should be regarded as joint First Authors.

We have previously reported the creation of MyGene.info, an up-to-date RESTful gene annotation as a service API (Application Programming Interface), which helps simplify the maintenance of pipelines that pull and/or analyze data from gene-specific resources (12,13). However, the functional effects of a gene are often context-dependent, and context can be difficult to capture, let alone make Findable, Accessible, Interoperable, and Reusable (FAIR) (14). A previous server, Tribe, allowed the capturing and sharing of collections of genes (gene sets) (15). However, it supported a subset of identifiers and did not plug into a widely used ecosystem of APIs. To create a FAIR gene-function-focused web server, we use the BioThings SDK (16) to create MyGeneset.info, which supports collaborative, context-storing, user-friendly access to collections of genes both via web and application programming interfaces.

MyGeneset.info can link both software and web-based analytical pipelines. A user might run an analysis on a BioThings-supported webserver (16,17) to generate a set of genes and then access that set through an R analysis. For example, a user could annotate a set of genes that are frequently mutated in a rare cancer. They could then query MyChem.info (<https://mychem.info>) (16) to identify drugs that target those genes. In parallel, they could analyze RNA-seq data using an R script and then rapidly examine the expression levels of genes within the gene set. Developers of curated resources can use the MyGeneset.info infrastructure to distribute collections within a well-used and highly interoperable ecosystem.

## MATERIALS AND METHODS

### Public gene set data sources

MyGeneset.info aggregates multiple publicly available gene sets and provides one-stop access to these built-in, curated gene sets. The current list of data sources, as of March 2023, include CTD (Comparative Toxicogenomics Database) (18), DO (Disease Ontology) (19), GO (Gene Ontology) (4), MSigDB (Molecular Signatures Database) (20), Reactome (7), SMPDB (Small Molecule Pathway Database) (21,22) and WikiPathways (23). There are 188650 (as of February 2023) curated gene sets in total collected from seven data sources. Similar to other biomedical web service APIs we have built (e.g. the MyGene.info API), the BioThings SDK (16) package was used to build a data source ‘hub’ to monitor, download and aggregate underlying data (Figure 1). All curated gene sets and their associated gene information are updated in MyGeneset.info on a weekly to monthly basis based on their respective sources. During each data update, individual gene annotations from each associated gene set are normalized using the MyGene.info API (12,13), with commonly used identifiers included. In the case of retired gene identifiers used in the original data source, it will be either replaced when a gene entity is replaced by a new entity or removed when a gene entity has been retired permanently.

### User submitted gene sets and annotations

MyGeneset.info also allows authenticated users, via their existing Github (<https://github.com>) or ORCID (24) ac-

count, to submit their own gene sets, e.g. from their own research studies or relevant literature. Users may set their gene sets to be either private or public, so that they can be shared when they are ready. The MyGeneset.info web application provides a user interface to help users to annotate and standardize the individual gene annotations within a user gene set using the MyGene.info API. This is the same process by which curated public gene sets are annotated and updated so that users do not need to handle the identifier conversion and maintenance of up-to-date gene annotations.

### Programmatic access via a web-service API

MyGeneset.info provides two primary ways to interact with the service to read and write data: a web-service API and its web application. The API is intended for users who require or prefer programmatic access to the service. The API can be used from a command-line interface or any programming language capable of standard HTTP requests. API access is useful for performing large-scale, batch, or complex analyses in an automated way. Developers may also utilize the API to build web applications that import gene sets and metadata automatically. The same API is powering MyGeneset.info’s own web application. The MyGeneset.info API is built upon the BioThings SDK package (16), which allows us to quickly build a set of web API endpoints with rich query features and scalable query performance. The BioThings SDK uses Tornado web framework (<http://tornadoweb.org>) to create API query endpoints and process the queries passed to an underlying Elasticsearch (<https://www.elastic.co/elasticsearch/>) cluster, where all gene set data are stored and indexed.

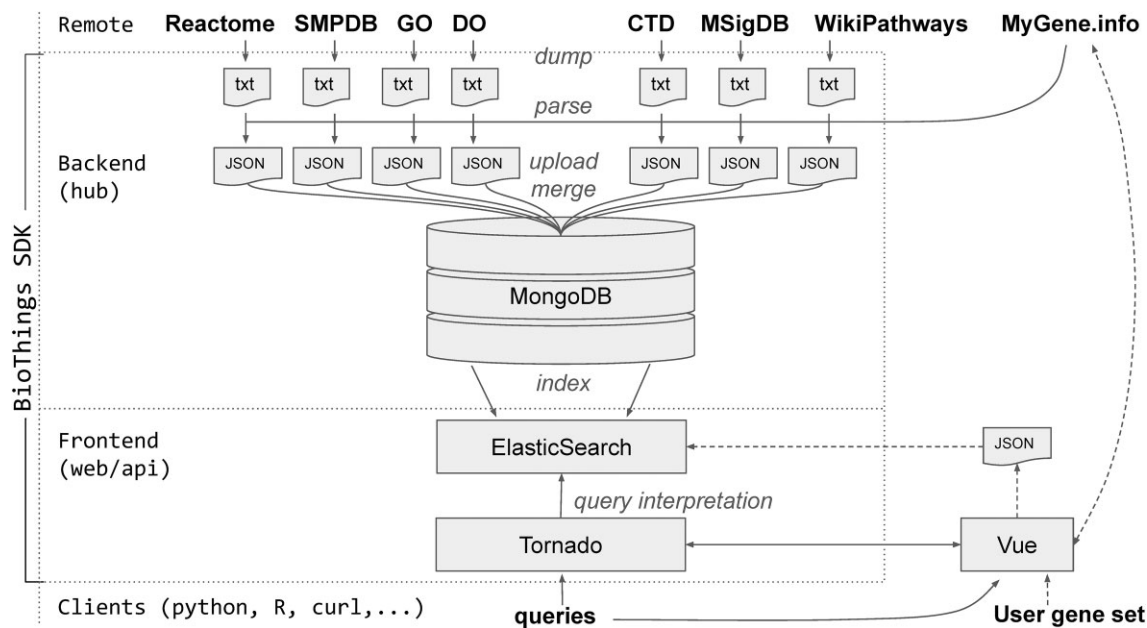
### Web application

The MyGeneset.info web application is intended for users who prefer web-based access to the service. It provides all the same features as the API, but with a user-friendly graphical interface. It allows users to quickly and conveniently search, view and manage gene sets from any connected device. The web application is built as a standard single-page application (SPA) in Vue 3 (<https://vuejs.org>) and TypeScript (<https://www.typescriptlang.org>). Its implementation follows modern web development best practices, such as responsive design (i.e. works on desktop and mobile devices) and accessibility, as well as modern Vue-specific best practices, such as using Composition API and script setup syntax.

## RESULTS

### Collecting and sharing publicly available annotated gene sets

MyGeneset.info currently provides read-access for about 180k gene sets, improving the findability and reusability of gene sets from humans, common and uncommon model organisms. As seen in Table 1, different resources have grouped genes into sets based on which chemicals they interact with (CTD), their involvement in specific diseases (DO), their co-localization (GO), their contribution or exclusion in biological processes or molecular functions (GO), their membership in pathways (Reactome, Wikipathways,



**Figure 1.** Publicly-accessible, curated gene sets are structured as resource plugins in MyGeneset.info, which was built using the BioThings SDK (dotted box). Each plugin retrieves gene annotations from MyGene.info before uploading and merging the data. Users can submit queries via clients or RESTful queries to the API, or via the Vue-based browser interface. Users can build and submit gene sets via the browser interface (dashed arrows).

**Table 1.** Built-in gene sets from publicly available resources (as of 2023.03.24)

Resource name	URL	license	Gene set type	# of gene sets
Comparative Toxicogenomics Database (CTD)	<a href="https://ctdbase.org/">https://ctdbase.org/</a>	(see site)	chemical interactions	22690
Disease Ontology (DO)	<a href="https://disease-ontology.org/">https://disease-ontology.org/</a>	CC0 1.0	disease involvement	4610
Gene Ontology (GO)	<a href="http://geneontology.org/">http://geneontology.org/</a>	CC BY 4.0	cellular location, biological process, molecular function	183 833
Molecular Signatures Database (MSigDB)	<a href="https://www.gsea-msigdb.org/gsea/msigdb">https://www.gsea-msigdb.org/gsea/msigdb</a>	CC BY 4.0	positional, regulatory, cell type and more	33 673
Reactome	<a href="https://reactome.org/">https://reactome.org/</a>	CC0 1.0	pathway	2593
WikiPathways	<a href="https://www.wikipathways.org/">https://www.wikipathways.org/</a>	CC0 1.0	pathway	1711
Small Molecule Pathway Database (SMPDB)	<a href="https://smpdb.ca/">https://smpdb.ca/</a>	(see site)	pathway	48 686

SMPDB), and more. Users can search for gene sets based on keywords describing the gene or context (Figure 2A), or species from which the gene set was derived (Figure 2B). Although all publicly available gene sets are included by default, the user can also limit their search by type of gene sets to include user-generated gene sets, anonymously submitted gene sets, or select from curated gene sets by source (Figure 2C). The resulting gene sets can be sorted based on the author, source, or number of genes in the gene set (Figure 2D). Users can further filter for gene sets by number of genes in the gene set (Figure 2E) or by source by selecting ‘Curated’ (Figure 2C) and then unchecking sources to be removed from the results (Figure 2F). Hovering over a gene within a gene set will display additional information on the gene including links out to other resources.

**Collecting and sharing user-annotated gene sets**

MyGeneset.info’s user-friendly, browser-based gene set builder improves the accessibility of the web server to researchers with less programming experience while ensuring

interoperability of the underlying data. The builder page guides users to generate important metadata for making their gene set findable (Supplemental Figure 1A), and researchers must search for genes to add enforcing mapping of the genes included in the user-generated gene set (Supplemental Figure 1B). The interface allows users to search a batch of up to 100 genes at once for inclusion in the gene set. If a user chose to log in before submitting their gene set, the user would be able to edit their gene set later on. Otherwise, a user may submit gene sets without logging in (anonymously), or simply download their gene set locally (if they wish to keep the gene set private) (Supplemental Figure 1C).

**Value proposition of MyGeneset.info for targeted user types**

MyGeneset.info was first designed to address the needs of three types of users: biomedical researchers, bioinformatics (core) analysts, and bioinformatics resource developers (Table 2). MyGeneset.info’s user-friendly, web browser is meant to improve the findability, accessibility, and re-

Figure 2 shows the search and browse interface of MyGeneset.info. The interface includes a navigation bar with links to Browse, Build, API, About, and BioThings, a logo for MYGENESET, and a Log In button. Below the navigation bar is a 'Browse' section with a search icon. The search section contains three input fields: (A) 'Search genesets by keyword', (B) 'Search genesets by species', and (C) 'Search by type: All' with dropdown menus for 'Min genes: 0' and 'Max genes: 999999'. Below these is a table (D) showing search results with columns: Name, Author, Source, #, Species, and Genes. The table lists three gene sets: '3,4-di-O-caffeoylquinic acid interactions', 'ERB\_041 interactions', and '(1,2-bis(1,2-benzisoxselenazolon-3(2H)-ketone)ethane interactions'. Below the table is a section for filtering results, including (E) 'Search by type: Curated' with dropdown menus for 'Min genes: 0' and 'Max genes: 999999', and (F) 'Source' with checkboxes for WIKIPATHWAYS, CTD, REACTOME, SMPDB, MSIGDB, GO, and DO. Below this is another table showing the filtered results for the '3,4-di-O-caffeoylquinic acid interactions' gene set.

**Figure 2.** The search and browse interface of MyGeneset.info allows users to search for gene sets by keyword (A) or species (B). Users can also search by type (C) including, gene sets generated by the User, publicly accessible Curated gene sets filterable by source, Anonymously submitted gene sets, and All gene sets. The resulting gene set table can be sorted by author, source, and number of genes in the gene set (D). Gene sets can be filtered by the number of genes in the set (E) or by the source (F) if 'Curated' is selected in (C).

usability of public gene sets; while the gene set builder is meant to improve the interoperability of user-generated gene sets and provide a GUI-based approach for sharing and maintaining user-generated gene sets. For example, a biomedical researcher may be interested in key genes suspected to be important to a biological process but lack the technical expertise to analyze these genes in the context of high-throughput genomic data. To ensure that their collaborator or bioinformatics core analyzes the correct genes, the researcher can leverage the MyGeneset.info website to search for genes (and filter by an organism of interest), add those genes into a collection, and share that collection either by submitting it to MyGeneset.info or downloading the gene set file in a format acceptable by the collaborator or bioinformatics core. By building the gene set via MyGe-

neset.info, the gene set will include metadata describing the gene set, an identifier for the gene set itself, and identifiers for each gene in the gene set. Because the gene set builder leverages the MyGene.info API, users can build gene sets from any species with genes annotated by NCBI Gene and Ensembl.

In contrast, a bioinformatics (core) analyst may be tasked with summarizing information for collections of genes from different researchers studying different biological processes in different animal models. Mapping and keeping track of each gene in each collection can be time-consuming, especially if different biomedical researchers provide these gene sets with different types of identifiers (or no identifiers altogether), and in different formats. By having the biomedical researchers use MyGeneset.info, the bioinformatics (core)

**Table 2.** Feature list and value proposition of each feature for different types of users

Feature	Value of the feature if the user is a:		
	Biomedical researcher	Bioinformatics analyst	Bioinformatics resource developer
User-friendly GUI	Search, create, or share gene sets without the need to program. Easily maintain user-generated gene sets	Remove identifier guess work by having researchers generate machine-friendly gene sets	
Identifier mapping	Ensure each gene in the gene set is mapped to an identifier. Ensure each gene set itself has a referable identifier	Search for gene sets by identifier.	
Built-in gene sets	Find and share existing gene sets	Search for context/gene sets from existing publicly accessible resources	
Gene set creation	Find and add organism-specific genes (with identifiers) into a gene set	Refer less-technical collaborators User-friendly GUI for gene set creation to ensure interoperability	
Gene set download	Download gene sets in machine-friendly format for downstream use by technical collaborators	Find and download gene sets in machine-friendly format for downstream processing	
Gene set submission	Publicly store gene sets for sharing with others. Edit (maintain) submitted gene sets via user-friendly GUI. Make the gene set publicly available with its own identifier	Publicly store pipeline-generated gene sets for sharing with others.	
Gene set API	Refer technical collaborators to API for programmatic access to publicly available gene sets	Programmatically access up-to-date gene sets, off-loading gene set maintenance and storage	Offload gene set and user storage and maintenance

analyst will be able to receive the gene sets in a consistent, easy-to-read, machine-friendly format either via a file sent from the biomedical researcher or by leveraging the API. For example, if a bioinformatics analyst needs Gene Ontology annotations for all genes in a gene set, they could easily access the gene set via MyGeneset.info’s RESTful API using simple python code, pull the identifiers for each gene in the gene set, and then query MyGene.info for additional gene-specific annotations (like GO annotations) (Supplemental Figure 2). Analysts concerned about version changes from the weekly updates of MyGeneset.info can cache the results locally using the BioThings Python client (Supplemental Figure 3).

Similar to a bioinformatics (core) analyst, a bioinformatics resource developer can offload the gene search and identifier mapping of their resource to MyGene.info. With MyGeneset.info, the resource developer can enable users to query for collections of genes, and allow users to create, append, and maintain these collections of genes. Further, if the resource is implemented with OAuth (25), the bioinformatics resource developer could largely offload the implementation of storing and maintaining gene sets and users.

**DISCUSSION AND CONCLUSION**

Biomedical researchers are constantly studying the functionality of genes and sets of genes, which require context to interpret; however, sharing context has traditionally been done via free text descriptions (i.e. publications). Given the current volume of scientific literature and its rate of growth, it is difficult to make gene function annotations Findable, Accessible, Interoperable, and Reusable. To do so, some resources have captured contextual information in the form of pathways or as collections of genes. Traditional resources that capture this context (like Wikipathways, CTD, DO and others) improve the FAIRness of gene function an-

notation by making their annotations available and/or allowing community curation of gene function annotations. Sacrificing some FAIRness for quality, community curation of gene function annotations by different resources have been limited in scope/topic and hampered by the curation process. For example, Wikipathways enables anyone to submit pathway-based gene sets, which may be useful for researchers studying specific pathways, but less useful for researchers studying polygenic diseases where a common pathway has not yet been elucidated. Other resources may curate based on other features but have a lengthier or more restrictive curation process. MyGeneset.info empowers users to share the context of their research on gene function as gene sets, improving the FAIRness of functional gene annotation.

In contrast to community curation, resources have done a phenomenal job making curated gene sets available; however, improvements in FAIRness can still be made. For example Disease Ontology does not directly provide gene sets on its site; rather, DO curates diseases and xrefs to other resources for diseases. The disease-based gene sets generated based on Disease Ontology require traversing from Disease Ontology to OMIM which can be challenging without programmatic expertise. Furthermore, resources that do provide gene sets may use different identifiers for the genes in the gene sets. For example, SMPDB uses UniProt IDs, while Wikipathways uses NCBI Gene IDs. By providing a unified and harmonized resource for gene sets, MyGeneset.info makes it easier to Find gene sets (only have to search one resource as opposed to several), Accessible (availability of a single user-friendly interface for interacting with gene sets), Interoperable (the genes in the gene set are mapped and linked to multiple commonly used identifiers and allow exporting to commonly used identifiers). By improving the FAIRness of gene sets, MyGeneset.info aims to increase their Reusability; thereby making gene sets more FAIR.

MyGeneset.info is integrated with the BioThings API ecosystem (<https://biothings.io>), which has served >45 million requests in the past 30 days and allows for programmatic traversal from gene sets to genes to variants in those genes or drugs targeting those genes. For example, a user can now easily retrieve a DO-sourced gene set from MyGeneset.info, use the identifiers for the genes in the gene set to pull Gene Ontology annotations from MyGene.info, and perform an overrepresentation analysis using their preferred libraries to investigate processes involved in disease pathology. The user can also use MyGene.info to map the genes to PharmGKB identifiers to find chemicals which might perturb each gene allowing the user to search for compounds that might treat the disease. These compounds could be searched in MyChem.info for additional properties which could affect their appeal as a potential treatment candidate. While not designed for extensive analysis on its own, MyGeneset.info's integration in the BioThings API ecosystem (including the BioThings Python client) makes it suitable for use in downstream analytical workflows or other web servers. Using the client, users can quickly fetch all gene sets available from MyGeneset.info for downstream processing.

MyGeneset.info is fully functional, has documentation, and both the API and front end are fully open source under a permissive Apache 2.0 license. Currently, the gene set creation interface supports the search/upload of up to 100 genes (via the multi-line input mode) in the gene addition process. While this may be sufficient for many biomedical researchers, increasing the batch size and including a means of uploading a list of >100 genes will improve the value of the resource for biomedical researchers and bioinformaticians alike. In the future, we hope to improve the gene set creation process by integrating similarity scores with existing gene sets in the gene set creation process, to minimize redundancy in user-submitted gene sets. Furthermore, we hope to incorporate features for discussing user-submitted gene sets—enhancing collaboration within teams of users.

## DATA AVAILABILITY

The MyGeneset.info web application is available at <https://mygeneset.info>, and its API can be found at: <https://mygeneset.info/api> (API documentation at <https://docs.mygeneset.info>). The source code is available in the GitHub repositories under the Apache 2.0 open-source license: <https://github.com/biothings/mygeneset.info> (API) and <https://github.com/biothings/mygeneset.info-website> (web application). Source code files are also on Zenodo: <https://doi.org/10.5281/zenodo.7803782> and <https://doi.org/10.5281/zenodo.7803811>.

## SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online.

## FUNDING

US National Institute of Health (<https://www.nih.gov/>) [OT2TR003445, R01GM083924 to C.W., R01HG010067, R01CA237170, R01HD109765 to C.S.G.]. Funding for open access charge: National Institutes of Health.

Conflict of interest statement. None declared.

## REFERENCES

- 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.
- Amberger, J.S. and Hamosh, A. (2017) Searching Online Mendelian Inheritance in Man (OMIM): a knowledgebase of human genes and genetic phenotypes. *Curr. Protoc. Bioinformatics*, **58**, 1.2.1–1.2.12.
- UniProt Consortium (2021) UniProt: the universal protein knowledgebase in 2021. *Nucleic Acids Res.*, **49**, D480–D489.
- Gene Ontology Consortium (2021) The Gene Ontology resource: enriching a GOLD mine. *Nucleic Acids Res.*, **49**, D325–D334.
- Kent, W.J., Sugnet, C.W., Furey, T.S., Roskin, K.M., Pringle, T.H., Zahler, A.M. and Haussler, D. (2002) The human genome browser at UCSC. *Genome Res.*, **12**, 996–1006.
- Rehm, H.L., Berg, J.S., Brooks, L.D., Bustamante, C.D., Evans, J.P., Landrum, M.J., Ledbetter, D.H., Maglott, D.R., Martin, C.L., Nussbaum, R.L. *et al.* (2015) ClinGen – the clinical genome resource. *N. Engl. J. Med.*, **372**, 2235–2242.
- Jassal, B., Matthews, L., Viteri, G., Gong, C., Lorente, P., Fabregat, A., Sidiropoulos, K., Cook, J., Gillespie, M., Haw, R. *et al.* (2020) The reactome pathway knowledgebase. *Nucleic Acids Res.*, **48**, D498–D503.
- Karczewski, K.J., Weisburd, B., Thomas, B., Solomonson, M., Ruderfer, D.M., Kavanagh, D., Hamamsy, T., Lek, M., Samocha, K.E., Cummings, B.B. *et al.* (2017) The ExAC browser: displaying reference data information from over 60 000 exomes. *Nucleic Acids Res.*, **45**, D840–D845.
- Thomas, P.D., Ebert, D., Muruganujan, A., Mushayahama, T., Albou, L.-P. and Mi, H. (2022) PANTHER: making genome-scale phylogenetics accessible to all. *Protein Sci.*, **31**, 8–22.
- Finn, R.D., Bateman, A., Clements, J., Coggill, P., Eberhardt, R.Y., Eddy, S.R., Heger, A., Hetherington, K., Holm, L., Mistry, J. *et al.* (2014) Pfam: the protein families database. *Nucleic Acids Res.*, **42**, D222–D230.
- Berman, H.M., Westbrook, J., Feng, Z., Gilliland, G., Bhat, T.N., Weissig, H., Shindyalov, I.N. and Bourne, P.E. (2000) The Protein Data Bank. *Nucleic Acids Res.*, **28**, 235–242.
- Wu, C., Macleod, I. and Su, A.I. (2013) BioGPS and MyGene.info: organizing online, gene-centric information. *Nucleic Acids Res.*, **41**, D561–D565.
- Xin, J., Mark, A., Afrasiabi, C., Tsueng, G., Juchler, M., Gopal, N., Stupp, G.S., Putman, T.E., Ainscough, B.J., Griffith, O.L. *et al.* (2016) High-performance web services for querying gene and variant annotation. *Genome Biol.*, **17**, 91.
- Wilkinson, M.D., Dumontier, M., Aalbersberg, I.J.J., Appleton, G., Axton, M., Baak, A., Blomberg, N., Boiten, J.-W., da Silva Santos, L.B., Bourne, P.E. *et al.* (2016) The FAIR Guiding Principles for scientific data management and stewardship. *Sci Data*, **3**, 160018.
- Zelaya, R.A., Wong, A.K., Frase, A.T., Ritchie, M.D. and Greene, C.S. (2017) Tribe: the collaborative platform for reproducible web-based analysis of gene sets. bioRxiv doi: <https://doi.org/10.1101/055913>, 09 October 2017, preprint: not peer reviewed.
- Lelong, S., Zhou, X., Afrasiabi, C., Qian, Z., Cano, M.A., Tsueng, G., Xin, J., Mullen, J., Yao, Y., Avila, R. *et al.* (2022) BioThings SDK: a toolkit for building high-performance data APIs in biomedical research. *Bioinformatics*, **38**, 2077–2079.
- Xin, J., Afrasiabi, C., Lelong, S., Adesara, J., Tsueng, G., Su, A.I. and Wu, C. (2018) Cross-linking BioThings APIs through JSON-LD to facilitate knowledge exploration. *BMC Bioinf.*, **19**, 30.
- Davis, A.P., Wiegiers, T.C., Johnson, R.J., Sciaky, D., Wiegiers, J. and Mattingly, C.J. (2023) Comparative Toxicogenomics Database (CTD): update 2023. *Nucleic Acids Res.*, **51**, D1257–D1262.
- Schriml, L.M., Munro, J.B., Schor, M., Olley, D., McCracken, C., Felix, V., Baron, J.A., Jackson, R., Bello, S.M., Bearer, C. *et al.* (2022) The Human Disease Ontology 2022 update. *Nucleic Acids Res.*, **50**, D1255–D1261.
- Subramanian, A., Tamayo, P., Mootha, V.K., Mukherjee, S., Ebert, B.L., Gillette, M.A., Paulovich, A., Pomeroy, S.L., Golub, T.R., Lander, E.S. *et al.* (2005) Gene set enrichment analysis: a

- knowledge-based approach for interpreting genome-wide expression profiles. *Proc. Natl. Acad. Sci. U.S.A.*, **102**, 15545–15550.
21. Frolkis, A., Knox, C., Lim, E., Jewison, T., Law, V., Hau, D.D., Liu, P., Gautam, B., Ly, S., Guo, A.C. *et al.* (2010) SMPDB: the Small Molecule Pathway Database. *Nucleic Acids Res.*, **38**, D480–D487.
  22. Jewison, T., Su, Y., Disfany, F.M., Liang, Y., Knox, C., Maciejewski, A., Poelzer, J., Huynh, J., Zhou, Y., Arndt, D. *et al.* (2014) SMPDB 2.0: big improvements to the Small Molecule Pathway Database. *Nucleic Acids Res.*, **42**, D478–D484.
  23. Slenter, D.N., Kutmon, M., Hanspers, K., Riutta, A., Windsor, J., Nunes, N., Mélius, J., Cirillo, E., Coort, S.L., Digles, D. *et al.* (2018) WikiPathways: a multifaceted pathway database bridging metabolomics to other omics research. *Nucleic Acids Res.*, **46**, D661–D667.
  24. Haak, L.L., Fenner, M., Paglione, L., Pentz, E. and Ratner, H. (2012) ORCID: a system to uniquely identify researchers. *Learn. Publ.*, **25**, 259–264.
  25. Boyd, R. (2012) *Getting Started with OAuth 2.0*. 'O'Reilly Media, Inc.