FAIR Functional Enrichment: Assessing and Modelling Provenance in Omics Results

Yi Chen[1], Fons.J.Verbeek[1] and Katherine.J. Wolstencroft[1,*]

[1]*Leiden Institute of Advanced Computer Science, Leiden 2333CA, NL*

Abstract

Functional enrichment analysis is an essential downstream process in high throughput omics studies, such as transcriptomics and proteomics. By using the Gene Ontology (GO) and its annotations (GOA), underlying functional patterns of over-representation can be identified, leading to better interpretation of the omics data and new biological insights. However, GO reflects the current understanding of gene product function and evolves with our changing biological knowledge. When performing such analyses, it is therefore crucial to record GO version provenance, together with related parameters, such as statistical cut-offs and annotation sources. Surveying the literature on functional enrichment results reveals provenance information is rarely available, reducing the reproducibility and interpretation of results and preventing objective comparisons between related studies. In this work, we propose minimal metadata requirements for functional enrichment reproducibility. Our model complies with the FAIR principles and is based on the provenance ontology (PROV-O). We demonstrate the scale of the problem and the utility of our solution with data from SARS-CoV-2.

Keywords

enrichment analysis, reproducibility, provenance, Gene Ontology, FAIR, PROV-O

1. Introduction

Functional enrichment analysis has been widely used in biomedical research, to interpret high throughput data[1][2] or to discover underlying mechanisms of diseases[3]. These analyses are dependent on biological knowledge-bases, such as the Gene Ontology (GO)[4], or Kyoto Encyclopedia of Genes and Genomes (KEGG)[5], that capture and structure our biological understanding. However, knowledge-bases are not static. Instead, they are frequently updated to depict the latest biological knowledge in the science community[6]. The Gene Ontology, for example, is updated monthly. Updates may include changes to the hierarchical structure and the conceptualization of our knowledge about gene functions, and changes to Gene Ontology annotation, which describes the associations between genes and GO terms, including the evidence for associations. The KEGG pathway knowledge-base is updated quarterly. Differences between knowledge-base versions can strongly affect the outcome of functional enrichment analyses. Tomczak et al[7] showed the extent to which the consistency, significance scores and

SWAT4HCLS 2023: The 14th International Conference on Semantic Web Applications and Tools for Health Care and Life Sciences

*Corresponding author.

y.chen@liacs.leidenuniv.nl (Y. Chen); f.j.verbeek@liacs.leidenuniv.nl (Fons.J.Verbeek);
k.j.wolstencroft@liacs.leidenuniv.nl (Katherine.J. Wolstencroft)

the interpretation of enrichment analysis results are changed by using different versions of GO and GOA. Another study[8] showed that 74% of enriched terms were changed between 2010 and 2016. If versioning information is not recorded, results from different studies are less comparable, making previously published studies less re-usable. To confound this problem further, a range of software tools[9][10][11] are available for functional enrichment. Each tool has its own update schedule, which may not follow the update schedules of the underlying knowledge-bases. Using the latest version of a functional enrichment tool does not guarantee that the latest version of underlying knowledge-bases are available[8]. In addition to problems with versioning, different functional enrichment applications implement different statistical tests, different background gene sets, and different default values for statistical significance and multiple testing corrections.

Wijesooriya et al (2022)[12] showed the extent of the differences in significant pathways and ontology terms for different methods in functional enrichment and multiple-testing correction. These findings together demonstrate the importance of capturing provenance to enable the interpretation of enrichment analysis results, but this information is seldom found in publications. Here, we demonstrate the current state of enrichment analysis reproducibility by studying data from SARS-Cov-2. The international response to the virus resulted in the generation and publication of large amounts of data and fast evolution of our collective knowledge. By surveying this data, we show that research conclusions were frequently based on outdated versions of available knowledge-bases, potentially missing the inclusion of new insights as they were discovered and shared. In this work, we randomly selected and manually inspected the metadata and provenance provided in research output from the PubMed Central[13] identifying common reporting practices and the most common tools[10][14][11][15][16][17][9][18] and methods used for analysis. In addition, we compared the versions of tools and underlying knowledge-bases between 2020 and 2022, revealing a lack of consistency in knowledge used to interpret experimental results. To address these problems, we propose minimal metadata and provenance requirements to improve comparison between functional enrichment experiments.

Our model complies with the Findability, Accessibility, Interoperability, and Reusability (FAIR) principles[19] and builds on established methods and standards. We adopt the PROV Ontology (PROV-O)[20] to describe a set of classes, properties, and restrictions to capture experimental parameters sufficiently and demonstrate the utility of the model using data from the SARS-CoV-2 literature.

2. Methods

To determine the variation of tools and knowledge-base versions utilized by functional enrichment studies in SARS-CoV-2, we counted the frequency of tools used from a random sample of PubMed Central (PMC)[13] publications, and examined the additional information provided on functional enrichment. For the most commonly used tools, we investigated the update schedule for underlying knowledge-bases and present the results for the Gene Ontology. From these results, we propose a PROV-O based model to capture functional enrichment provenance.

2.1. Functional Enrichment Literature Survey

2.1.1. Data collection

SARS-CoV-2 was selected for this investigation due to the a large number of publications that were produced on this topic over a short period of time. Understanding the virus required data and knowledge sharing on a large scale, and having access to the most recent insights was essential for comparing studies. To investigate the functional enrichment provenance typically provided, we first identified SARS-CoV-2 publications in PMC. The search terms can be found in supplementary file 1.

A total of 3206 publication identifiers were retrieved (the full list of PMC identifiers can be found in supplementary file 2) and full papers were retrieved using the PubMed Central Open Access API in BioC format PMC BioC API (accessed in Nov 23th, 2022).

Based on the occurrence of the search term 'enrichment analysis' in the methods section of each paper, we selected the top 100 papers and manually inspected them. Articles were excluded if they described enrichment analysis methods, instead of presenting analysis results. In total, 92 articles were retained, containing 135 enrichment analyses.

For each enrichment analysis described in the cohort, we identified the following information: 1) The name of the tool and method used for enrichment 2) the version of the tool 3) The name(s) of the knowledge-bases used for enrichment 4) The version of knowledge-bases used

Data was collected and visualized by python package Matplotlib.pyplot (v3.5.2)[21].

2.1.2. Survey of knowledge-base consistency

Tomczak et al [7], demonstrated that variation in the version of GO and GO annotation (GOA) affects the interpretation, p-value and the consistency of enrichment analysis. For the tools identified with the highest frequency of use in the literature survey, we investigated their update schedule and the update schedule for the underlying knowledge-bases. This investigation was to determine if researchers were using the most recent knowledge and if there was consistency of underlying knowledge across studies published at similar times.

For the top 8 most frequently identified tools, we manually examined the metadata available. The version and the released date information for associated knowledge-bases were recorded. If there were multiple releases of a tool between 2020 and 2022, every release was inspected. In addition, some tools used knowledge-bases derived from primary sources, but further processed and integrated them into other systems. The Molecular Signatures Database, for example, (MSigDB)[14][22], integrates multiple knowledge-bases, but without clear versioning information. By manually inspecting the release notes, we examined the availability of knowledge-base metadata. If the metadata was not provided, we recorded the metadata of the knowledge-base version closest to the releases of database.

2.2. Metadata and Provenance Requirements

The minimal metadata and provenance requirements we propose are based on the recommendations from Jauer for minimal provenance[23], the FAIR principles, and key factors described by Wijesooriya[12] that affect the reproducibility of functional enrichment. We propose metadata for four aspects; input data, knowledge-base data, enrichment analysis methods conducted, and output data. Key factors identified by Wijesooriya[12] included, multiple-testing correction methods, statistical cut-offs and background gene sets. Knowledge-bases like GO, should be available with the version and the source(s) of data for annotation. Following the FAIR principles[19], persistent identifiers (PIDs) and standard gene identifiers should be used, as should timestamps and references to individuals or institutions who hold responsibility for the experiments 1.

2.3. Representing Provenance in PROV-O

The PROV-O ontology encodes the PROV data model in OWL (web ontology language)[24]. The core elements of PROV-O are: 1) Entities, which can be any real or conceptual objects, 2) Activities, something that occurs over a given time period, and 3) Agents, which hold the responsibility for activities and the existence of entities. Seven properties, (e.g. 'wasGeneratedBy'), describe the relationships between these core elements. Here, we propose a PROV-O model to describe the entities and activities involved in a functional enrichment analysis, showing how our proposed metadata elements could be used to represent the provenance of the experiment, to improve comparability and reproducibility.

3. Results

Here, We present the results of the literature survey on enrichment analysis metadata, followed by an investigation into the consistency of versions of knowledge-bases in the most frequently used tools. Finally, we propose the minimum metadata required for functional enrichment and a provenance model to address the comparability problems identified by the literature survey. During the survey period, 30 versions of the Gene Ontology were released, showing an overall reduction from 44,700 terms and 92230 edges in 2020 and 43272 terms and 85618 edges in 2022. 935 new terms were added, 524 were merged, and 1417 terms were made obsolete.

3.1. Survey of Enrichment Analysis Results Metadata

Figure 1 shows the results of surveying 135 enrichment analyses from 92 publications, published between 2020 and 2022. Through manual inspection, 25 different tools and 28 knowledge-bases and databases were identified. The largest proportion of analyses was conducted using R, with 'ClusterProfiler'[15], 'FGSEA'[9] and 'GSVA'[16] accounting for 47 of 135 analyses. The GSEA platform was the second largest, in which 21 analyses were conducted. Web-based tools like 'Metascape'[11] were also frequently used, with more than 20 analyses in our survey. In 23 analyses, authors did not report which tool(s) were used. GO and KEGG were the most frequently used knowledge-bases, with 46 and 35 analyses respectively. Other knowledge-bases like Reactome[25][26] and Wikipathway[27] were less common in the collection. 37 analyses did not provide any information on which knowledge-base was used. Our findings showed large variations in the tools and knowledge-bases in functional enrichment analysis. The use of different tools should not prevent the comparison of results, but experimental parameters, source data and version information are required to interpret those differences. Our survey

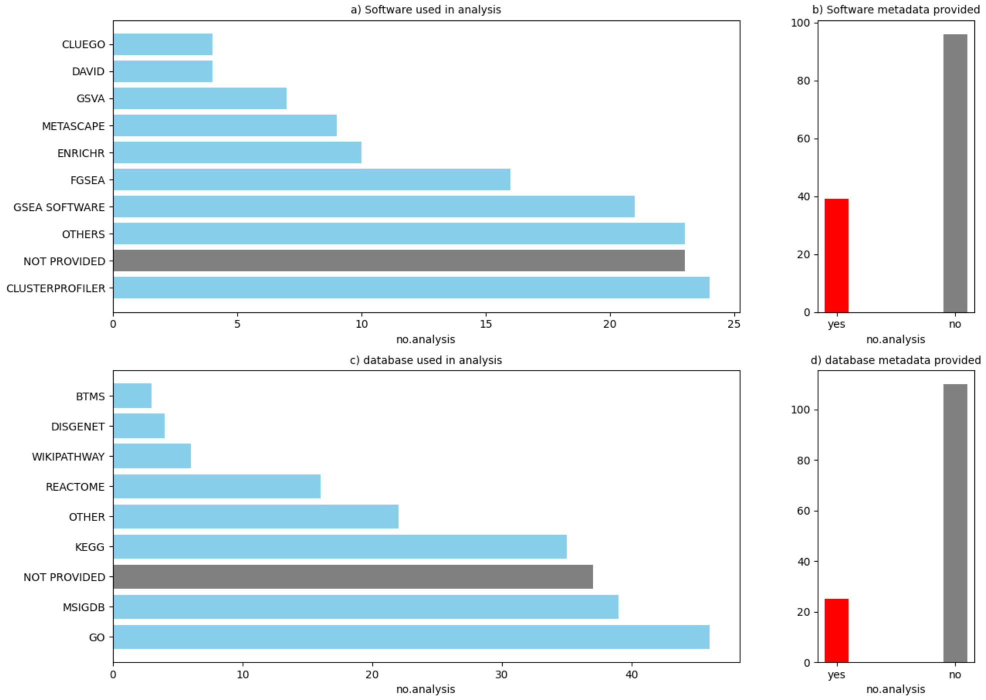

Figure 1: Functional enrichment survey results. a) Software used in functional enrichment b) Frequency of analyses reporting metadata of software c) Knowledge-bases used and their frequencies. d) Frequency of analyses reporting knowledge-base metadata.

showed that metadata relating to the parameters or tool versions were not provided in 96 out of 135 analyses. Versioning information of knowledge-bases was omitted in a further 110 out of 135 analyses. Taken together, these data show a lack of metadata relating to function enrichment analyses in SARS-Cov-2 studies.

3.2. Knowledge-Base Versioning in Enrichment Tools

For the top 8 most frequently used tools from our survey, we identified the versions of the GO knowledge-base in use, as described by the tool providers. Figure2 shows the results.
As we can see, some tools, such as DAVID 6.8[18], used versions of GO that predated the SARS-CoV-2 pandemic. From December 2021, David began quarterly updates of its software, although how these updates tracked GO updates is not transparent. For EnrichR[10], a 2018 version of GO was in use until early 2021. In contrast, ClueGO[17] and other Bioconductor-based tools provided more frequent updates, but only Metascape updated GO monthly and remained up to date with the Gene Ontology. For tools such as, GSEA[14], the GO knowledge-base version depended on the GSEA version. These findings show that at any given time in the pandemic, the choice

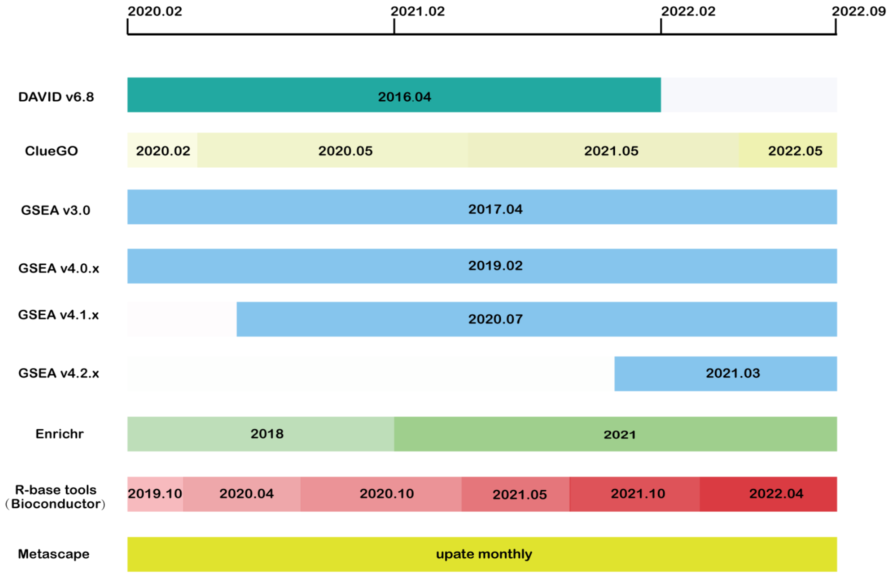

Figure 2: A timeline of GO knowledge-base updates in functional enrichment tools. The timestamp inside the boxes represents the date of the last GO update. At any given time point, multiple GO versions are in-use across the tools, resulting in difficulties in data comparison.

of enrichment analysis software dictated how up-to-date the underlying biological knowledge was for analysing enrichment results. Consequently, papers published at similar times were not necessarily basing analysis conclusions on the same collective understanding of biology. Re-analysing these studies may therefore yield new insights with our recent accumulation of knowledge. These results highlight the necessity of recording version information and more extensive provenance data.

3.3. Minimum Metadata Requirements

To increase the reproducibility and comparability of enrichment analysis results, in line with the FAIR principles, we propose minimum metadata requirements for enrichment analysis in four aspects; input data, knowledge-base and data sources, enrichment analysis execution, and output data Table1. These recommendations are based on previous work to define minimum provenance and reproducible enrichment analysis, as well as on the results of our literature survey. Example annotations are provided to show what should be recorded for each metadata element.

3.4. Proposed Provenance Model

The proposed minimum metadata requirements from the previous section form the core components of an enrichment analysis provenance model. Figure3 shows the relationships between these metadata elements, formally modelling the provenance of an enrichment analysis and expressed using PROV-O. The example instances represented in the model are the same as the example annotations from table 1, showing how each element is necessary for capturing sufficient information for comparison and interpretation. Where provenance information is incomplete, anonymous nodes can be used to highlight what is unknown in an experiment.

Table 1
Minimal metadata requirements for Functional enrichment analyses. The relation to specific FAIR principles is shown by F(Findable), A (Accessible), I (Interoperable) and R (Reusable).

Criteria	**Recommendations**	**Input data**	**Knowledge-base/Data sources**	**Enrichment Execution**	**Output data**	**FAIR**
Persistent ID (PID)	A persistent identifier should be assigned	prov:Collection; PID	GO URI / gene sets URI	tool URI	prov:Collection; PID	F,A
Standard Identifiers	Gene products and knowledge-base terms should be described using persistent identifiers from a recognised source	ENSG0000012584	GO:0006915	enrichr	GO:0071375	F,I
Creator	An Institution or person bears the responsibility	researcher orcidID	maayanlab	maayanlab	researcher orcidID	A, R
Timestamp	The time when data was generated/enrichment analysis was conducted		2021-03-01			R
Origin	A description of the source of the original data	Differentially Expressed RNA-seq	GO and GOA	Ensembl		F,I,R
Extraction Method	How the data was obtained from the source of the original data	significant up-regulated	biological process terms	reviewed human genes		R
Versioning	The version of tools/knowledge-bases used in enrichment analysis		GO(v2021) and GOAv2021/ release 108	release 2021-03-29		A, I, R
Background Gene sets	Gene sets used as backgrounds in enrichment analysis			prov:Collection; PID		I, R
Statistical test	Statistical test used in enrichment analysis.			fisher exact test		R
multi-test correction method	The methods used for multi-test correction.			Benjamini-Hochberg procedure		R
Cut-off	The cut-off used in enrichment analysis (p and q)			0.05; 0.01		R

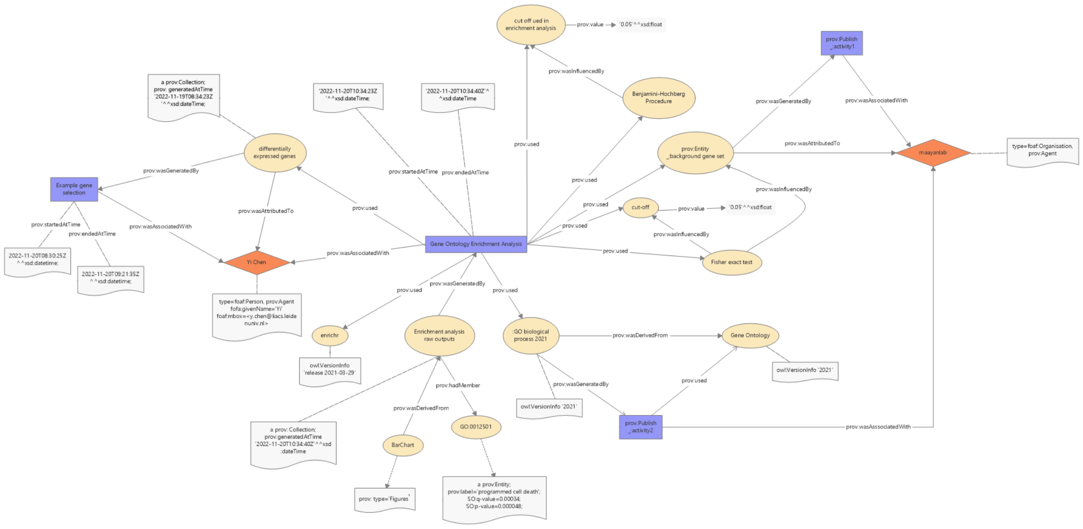
cut off ued in enrichment analysis
prov:value
'0.05'^^xsd:float
prov:Publish _:activity1
prov:wasInfluencedBy
a prov:Collection; prov: generatedAtTime '2022-11-19T08:34:23Z '^^xsd:dateTime;
'2022-11-20T10:34:23Z '^^xsd:dateTime;
'2022-11-20T10:34:40Z'^ ^xsd:dateTime
Benjamini-Hochberg Procedure
prov:wasGeneratedBy
prov:wasAssociatedWith
prov:used
prov:Entity _:background gene set
prov:wasAttributedTo
maayanlab
type=foaf:Organisation, prov:Agent
differentially expressed genes
prov:wasGeneratedBy
Example gene selection
prov:startedAtTime
prov:endedAtTime
prov:wasInfluencedBy
prov:used
prov:used
prov:wasAttributedTo
prov:used
cut-off
prov:value
'0.05'^^xsd:float
prov:startedAtTime
prov:endedAtTime
prov:wasAssociatedWith
prov:used
prov:wasInfluencedBy
Gene Ontology Enrichment Analysis
prov:used
prov:wasAssociatedWith
Fisher exact test
2022-11-20T08:30:25Z ^^xsd:datetime;
Yi Chen
2022-11-20T09:21:35Z ^^xsd:datetime;
prov:used
prov:wasGeneratedBy
prov:used
type=foaf:Person, prov:Agent fofa:givenName='Yi' foaf:mbox=<y.chen@liacs.leideuniv.nl>
enrichr
Enrichment analysis raw outputs
:GO biological process 2021
prov:wasDerivedFrom
Gene Ontology
owl:VersionInfo 'release 2021-03-29'
owl:VersionInfo '2021'
prov:hadMember
prov:used
prov:wasGeneratedBy
prov:wasDerivedFrom
owl:VersionInfo '2021'
a prov: Collection; prov:generatedAtTime '2022-11-20T10:34:40Z'^^xsd :dateTime
GO:0012501
prov:Publish _:activity2
prov:wasAsssociatedWith
BarChart
prov: type='Figures'
a prov:Entity; prov:label='programmed cell death'; SO:q-value=0.00034; SO:p-value=0.000048;

Figure 3: Prov-O representation for an enrichment analysis experiment on differentially expressed RNA-Seq data, analysed with Enrichr and 2021 version of GO. Rectangle represents Activity, eclipse represents entity and rhombus represents agents. Details can be seen atFAIRDOMHUB

4. Discussion

This study highlights problems of reproducibility and comparability in functional enrichment analyses. We showed there was little consistency in the information reported about such experiments and revealed a large proportion of the studies we surveyed were not being conducted using the latest versions of biological knowledge-bases. Structured knowledge resources, such as GO, allow us to identify patterns in complex, high-throughput omics data, enabling new insights from our collective knowledge. However, as our knowledge changes, these supporting knowledge resources also change. This should be an advantage, allowing scientists to benefit from the work of others. However, our survey showed a large range of enrichment analysis tools are in common use (Figure1), but that each has its own update schedule for underlying knowledge (Figure2). The result is that different studies, conducted at similar times, use different versions of knowledge-bases, and therefore different uderlying knowledge. If we know where the differences lie, comparison is still possible, but 110/135 enrichment analyses did not provide information on knowledge-base versioning, and only 39/135 reported the version of the enrichment analysis tool that was used. From the literature survey, and previous studies on minimal provenance [23] and reproducible enrichment analyses[12], we propose a minimum set of metadata to combat the problems described above. In addition, we present a PROV-O based model for expressing enrichment analysis results, with an example of an enrichment analysis experiment run using Enrichr[10]. Minimum metadata guidelines for upstream analyses, describing the generation and statistical analysis of high throughput omics data have long been established[28]. By implementing similar paradigms for downstream analyses, we can improve the FAIRness of studies overall and enable FAIRer comparison and reuse of important data sets.

5. Appendices

Supplementary files can be found at https://fairdomhub.org/investigations/583

Acknowledgments

Here we thank the support from the Chinese Scholarship Council through Leiden University.

References

[1] A. Conesa, P. Madrigal, S. Tarazona, D. Gomez-Cabrero, A. Cervera, A. McPherson, M. W. Szcześniak, D. J. Gaffney, L. L. Elo, X. Zhang, et al., A survey of best practices for rna-seq data analysis, Genome biology 17 (2016) 1–19.

[2] P. Krishnamoorthy, A. S. Raj, S. Roy, N. S. Kumar, H. Kumar, Comparative transcriptome analysis of sars-cov, mers-cov, and sars-cov-2 to identify potential pathways for drug repurposing, Computers in biology and medicine 128 (2021) 104123.

[3] P. Gollapalli, S. B. S, H. Rimac, P. Patil, S. K. Nalilu, S. Kandagalla, P. Shetty, Pathway enrichment analysis of virus-host interactome and prioritization of novel compounds targeting the spike glycoprotein receptor binding domain–human angiotensin-converting enzyme 2 interface to combat sars-cov-2, Journal of Biomolecular Structure and Dynamics 40 (2022) 2701–2714.

[4] The gene ontology resource: enriching a gold mine, Nucleic acids research 49 (2021) D325–D334.

[5] M. Kanehisa, S. Goto, Kegg: kyoto encyclopedia of genes and genomes, Nucleic acids research 28 (2000) 27–30.

[6] G. O. Consortium, Expansion of the gene ontology knowledgebase and resources, Nucleic acids research 45 (2017) D331–D338.

[7] A. Tomczak, J. M. Mortensen, R. Winnenburg, C. Liu, D. T. Alessi, V. Swamy, F. Vallania, S. Lofgren, W. Haynes, N. H. Shah, et al., Interpretation of biological experiments changes with evolution of the gene ontology and its annotations, Scientific reports 8 (2018) 1–10.

[8] L. Wadi, M. Meyer, J. Weiser, L. D. Stein, J. Reimand, Impact of outdated gene annotations on pathway enrichment analysis, Nature methods 13 (2016) 705–706.

[9] G. Korotkevich, V. Sukhov, N. Budin, B. Shpak, M. N. Artyomov, A. Sergushichev, Fast gene set enrichment analysis, BioRxiv (2021) 060012.

[10] M. V. Kuleshov, M. R. Jones, A. D. Rouillard, N. F. Fernandez, Q. Duan, Z. Wang, S. Koplev, S. L. Jenkins, K. M. Jagodnik, A. Lachmann, et al., Enrichr: a comprehensive gene set enrichment analysis web server 2016 update, Nucleic acids research 44 (2016) W90–W97.

[11] Y. Zhou, B. Zhou, L. Pache, M. Chang, A. H. Khodabakhshi, O. Tanaseichuk, C. Benner, S. K. Chanda, Metascape provides a biologist-oriented resource for the analysis of systems-level datasets, Nature communications 10 (2019) 1–10.

[12] K. Wijesooriya, S. A. Jadaan, K. L. Perera, T. Kaur, M. Ziemann, Urgent need for consistent standards in functional enrichment analysis, PLoS computational biology 18 (2022) e1009935.

[13] NCBI, Pubmed central, 1999. URL: https://www.ncbi.nlm.nih.gov/pmc/.

[14] A. Subramanian, P. Tamayo, V. K. Mootha, S. Mukherjee, B. L. Ebert, M. A. Gillette, A. Paulovich, S. L. Pomeroy, T. R. Golub, E. S. Lander, et al., Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles, Proceedings of the National Academy of Sciences 102 (2005) 15545–15550.

[15] G. Yu, L.-G. Wang, Y. Han, Q.-Y. He, clusterprofiler: an r package for comparing biological themes among gene clusters, Omics: a journal of integrative biology 16 (2012) 284–287.

[16] S. Hänzelmann, R. Castelo, J. Guinney, Gsva: gene set variation analysis for microarray and rna-seq data, BMC bioinformatics 14 (2013) 1–15.

[17] G. Bindea, B. Mlecnik, H. Hackl, P. Charoentong, M. Tosolini, A. Kirilovsky, W.-H. Fridman, F. Pagès, Z. Trajanoski, J. Galon, Cluego: a cytoscape plug-in to decipher functionally grouped gene ontology and pathway annotation networks, Bioinformatics 25 (2009) 1091–1093.

[18] G. Dennis, B. T. Sherman, D. A. Hosack, J. Yang, W. Gao, H. C. Lane, R. A. Lempicki, David: database for annotation, visualization, and integrated discovery, Genome biology 4 (2003) 1–11.

[19] M. D. Wilkinson, M. Dumontier, I. J. Aalbersberg, G. Appleton, M. Axton, A. Baak, N. Blomberg, J.-W. Boiten, L. B. da Silva Santos, P. E. Bourne, et al., The fair guiding principles for scientific data management and stewardship, Scientific data 3 (2016) 1–9.

[20] K. Belhajjame, J. Cheney, D. Corsar, D. Garijo, S. Soiland-Reyes, S. Zednik, J. Zhao, PROV-O: The PROV Ontology, Technical Report, 2012. URL: http://www.w3.org/TR/prov-o/.

[21] J. D. Hunter, Matplotlib: A 2d graphics environment, Computing in science & engineering 9 (2007) 90–95.

[22] A. Liberzon, C. Birger, H. Thorvaldsdóttir, M. Ghandi, J. P. Mesirov, P. Tamayo, The molecular signatures database hallmark gene set collection, Cell systems 1 (2015) 417–425.

[23] M.-L. Jauer, T. M. Deserno, Data provenance standards and recommendations for fair data, Digital Personalized Health and Medicine (2020) 1237–1238.

[24] S. Bechhofer, F. van Harmelen, J. Hendler, I. Horrocks, D. McGuinness, P. Patel-Schneijder, L. A. Stein, OWL Web Ontology Language Reference, Recommendation, World Wide Web Consortium (W3C), 2004. See http://www.w3.org/TR/owl-ref/.

[25] J. Griss, G. Viteri, K. Sidiropoulos, V. Nguyen, A. Fabregat, H. Hermjakob, Reactomegsa-efficient multi-omics comparative pathway analysis, Molecular & Cellular Proteomics 19 (2020) 2115–2125.

[26] M. Gillespie, B. Jassal, R. Stephan, M. Milacic, K. Rothfels, A. Senff-Ribeiro, J. Griss, C. Sevilla, L. Matthews, C. Gong, et al., The reactome pathway knowledgebase 2022, Nucleic acids research 50 (2022) D687–D692.

[27] M. Martens, A. Ammar, A. Riutta, A. Waagmeester, D. N. Slenter, K. Hanspers, R. A. Miller, D. Digles, E. N. Lopes, F. Ehrhart, et al., Wikipathways: connecting communities, Nucleic acids research 49 (2021) D613–D621.

[28] C. F. Taylor, D. Field, S.-A. Sansone, J. Aerts, R. Apweiler, M. Ashburner, C. A. Ball, P.-A. Binz, M. Bogue, T. Booth, et al., Promoting coherent minimum reporting guidelines for biological and biomedical investigations: the mibbi project, Nature biotechnology 26 (2008) 889–896.