

Lawrence Berkeley National Laboratory

LBL Publications

Title

Flame (v2.0): advanced integration and interpretation of functional enrichment results from multiple sources.

Permalink

<https://escholarship.org/uc/item/3df580s5>

Journal

Computer applications in the biosciences : CABIOS, 39(8)

Authors

Karatzas, Evangelos

Baltoumas, Fotis

Aplakidou, Eleni

et al.

Publication Date

2023-08-01

DOI

10.1093/bioinformatics/btad490

Peer reviewed

Genome analysis

Flame (v2.0): advanced integration and interpretation of functional enrichment results from multiple sources

Evangelos Karatzas ^{1,*†}, Fotis A. Baltoumas ^{1,*†}, Eleni Aplakidou¹, Panagiota I. Kontou^{2,3}, Panos Stathopoulos^{4,5}, Leonidas Stefanis⁴, Pantelis G. Bagos ³, Georgios A. Pavlopoulos ^{1,6,7,*}

¹Institute for Fundamental Biomedical Research, BSRC "Alexander Fleming", Vari (Athens), 16672, Greece

²Department of Mathematics, University of Thessaly, Lamia, 35100, Greece

³Department of Computer Science and Biomedical Informatics, University of Thessaly, Lamia, 35131, Greece

⁴1st Department of Neurology, Eginition Hospital, Athens, 11528, Greece

⁵School of Medicine, National and Kapodistrian University of Athens, Athens, 11527, Greece

⁶Center of Basic Research, Biomedical Research Foundation of the Academy of Athens, Athens, 11527, Greece

⁷Hellenic Army Academy, Vari, 16673, Greece

*Corresponding authors. Institute for Fundamental Biomedical Research, Biomedical Sciences Research Center "Alexander Fleming", 34 Fleming St. Vari (Athens), 16672, Greece. E-mails: pavlopoulos@fleming.gr (G.A.P), karatzas@fleming.gr (E.K), and baltoumas@fleming.gr (F.A.B.)

† = equal contribution.

Associate Editor: Tobias Marschall

Abstract

Summary: Functional enrichment is the process of identifying implicated functional terms from a given input list of genes or proteins. In this article, we present Flame (v2.0), a web tool which offers a combinatorial approach through merging and visualizing results from widely used functional enrichment applications while also allowing various flexible input options. In this version, Flame utilizes the aGOTool, g:Profiler, WebGestalt, and Enrichr pipelines and presents their outputs separately or in combination following a visual analytics approach. For intuitive representations and easier interpretation, it uses interactive plots such as parameterizable networks, heatmaps, barcharts, and scatter plots. Users can also: (i) handle multiple protein/gene lists and analyse union and intersection sets simultaneously through interactive UpSet plots, (ii) automatically extract genes and proteins from free text through text-mining and Named Entity Recognition (NER) techniques, (iii) upload single nucleotide polymorphisms (SNPs) and extract their relative genes, or (iv) analyse multiple lists of differentially expressed proteins/genes after selecting them interactively from a parameterizable volcano plot. Compared to the previous version of 197 supported organisms, Flame (v2.0) currently allows enrichment for 14 436 organisms.

Availability and implementation: Web Application: <http://flame.pavlopouloslab.info>. Code: <https://github.com/PavlopoulosLab/Flame>. Docker: <https://hub.docker.com/r/pavlopouloslab/flame>.

1 Introduction

Functional enrichment is the process of identifying biological functions, pathways, diseases, or phenotypes where groups of genes or proteins are involved. To this end, several applications have been developed (van den Berg *et al.*, 2009, Huang *et al.* 2009, Maleki *et al.* 2020, Chicco and Jurman 2022, Wijesooriya *et al.* 2022). DAVID (Sherman *et al.* 2022), Panther (Mi *et al.* 2021), Metascape (Zhou *et al.* 2019), WebGestalt (Wang *et al.* 2017), Enrichr (Xie *et al.* 2021), aGOTool (Schölz *et al.* 2015), AllEnricher (Zhang *et al.* 2020), GO-Elite (Zambon *et al.* 2012), MSigDB (Liberzon *et al.* 2011), WEGO (Ye *et al.* 2018), KOBAS (Bu *et al.* 2021), clusterProfiler (Yu *et al.* 2012), modEnrichr (Kuleshov *et al.* 2019), agriGO (Tian *et al.* 2017), DOSE (Yu *et al.* 2015), GeneTrail (Gerstner *et al.* 2020), GOrilla (Eden *et al.* 2009), ToppGene (Chen *et al.* 2009), PANGEA (Hu *et al.* 2023), and Enrichr-KG (Evangelista *et al.* 2023) are a few of the widely used applications which are offered either as web

tools or as software packages while others such as BiNGO (Maere *et al.* 2005), stringApp (Doncheva *et al.* 2023), or ClueGO (Bindea *et al.* 2009) are offered as plugins (Saito *et al.* 2012) in larger frameworks such as Cytoscape (Otasek *et al.* 2019).

While this variety of applications allows focusing on different aspects such as enriching for (i) biological pathways [e.g. KEGG (Kanehisa and Goto 2000), WikiPathways (Slenter *et al.* 2018), Reactome (Fabregat *et al.* 2018)], (ii) Gene Ontology [GO (Aleksander *et al.* 2023)] terms such as biological processes, molecular functions and cellular components, (iii) diseases [e.g. OMIM (Amberger *et al.* 2015), DisGeNet (Piñero *et al.* 2020)], (iv) protein complexes [e.g. CORUM (Giurgiu *et al.* 2019)], (v) protein domains [e.g. Pfam (Finn *et al.* 2016)], (vi) phenotypes [e.g. HPO (Robinson *et al.* 2008)], or (vii) regulatory motifs [e.g. TRANSFAC (Matys *et al.* 2003), miRTarBase (Huang *et al.* 2020)], an obvious selection of the right tool to cover certain needs is most of the

Received: February 17, 2023. Revised: May 31, 2023. Editorial Decision: July 15, 2023. Accepted: August 3, 2023

© The Author(s) 2023. Published by Oxford University Press.

This is an Open Access article distributed under the terms of the Creative Commons Attribution License (<https://creativecommons.org/licenses/by/4.0/>), which permits unrestricted reuse, distribution, and reproduction in any medium, provided the original work is properly cited.

times not straightforward. Moreover, while most of these tools come with a simple interface, little emphasis has been given on the visualization and combination of results, id conversion, parameter selection, result prioritization, record association, support of multiple lists and versatile input options. Therefore, usage and interpretation of results in a more streamlined and combinatorial aspect still remain difficult tasks for the average user.

In this article, we present Flame (v2.0), an advanced service for addressing the aforementioned issues. Flame focuses on usage simplicity, offering many degrees of freedom regarding importing and handling multiple lists, running advanced state-of-the-art functional enrichment analysis pipelines as well as combining and prioritizing results through advanced interactive visualizations. Flame is able to report results as lists but also associate them at a network level (Koutrouli *et al.* 2020), while it offers major integration of resources (Baltoumas *et al.* 2021a) such as aGTool, g: Profiler, WebGestalt, Enrichr, and STRING (Szklarczyk *et al.* 2023); thus referring to their broad audience while taking full advantage of their strengths.

2 Methods and results

2.1 Input options

Compared to its predecessor (Thanati *et al.* 2021), Flame (v2.0) comes with a variety of newly introduced input options (Fig. 1). These are mainly: (i) support of simple gene lists, (ii) support of variants and polymorphisms, (iii) support of text-mining and Named Entity Recognition (NER) techniques to process free text and mine it for genes and proteins, and (iv) support of gene expression data derived from experiments.

Gene lists: Similarly to Flame v1.0, users can directly paste or upload multiple lists by using a plethora of gene identifiers

or chromosomal intervals. Gene and protein names can be imported in the Ensembl (Cunningham *et al.* 2022), Entrez (Sayers *et al.* 2022), UniProt (UniProt Consortium 2021), RefSeq (Li *et al.* 2021), EMBL (Arita *et al.* 2021), ChEMBL (Mendez *et al.* 2019), and WikiGene (Maier *et al.* 2005) namespaces, or be later converted into another namespace during the functional enrichment analysis procedure.

Variants and Polymorphisms: In addition to directly uploading gene lists, Flame is capable of parsing and annotating lists of single nucleotide polymorphisms (SNPs), through the “SNP” input option. The input in this case is a list (white-space or comma-delimited) of SNP rs-codes in the dbSNP (Sherry *et al.* 2001) format. The ids are mapped to their corresponding genes and associated Ensembl identifiers through the g:SNPense functionality of g:Profiler and are then searched against dbSNP to retrieve additional metadata, including the chromosome number, strand, genomic coordinates, and the variant’s reported effects. The annotated SNPs can be downloaded, while their associated genes can be added to the list of Flame inputs, using either their Entrez gene name or their Ensembl ids. Since dbSNP has dropped support for nonhuman variants in recent versions, this functionality is currently available only for *Homo sapiens* (human).

Text-Mining and NER: In this version, users can provide free text and utilize the EXTRACT (Pafilis *et al.* 2016) tagger to perform Named Entity Recognition (NER) to mine it for genes and proteins for an organism of preference. Upon completion, Flame will report the identified genes/proteins in an interactive and easy-to-filter table as well as an annotated text with the identified terms highlighted. In mouse-hover, a popup window will appear, providing information about the entity as well as direct links to the STRING and Ensembl databases. The annotated terms can be added to the list of Flame inputs, using their associated Ensembl ids.

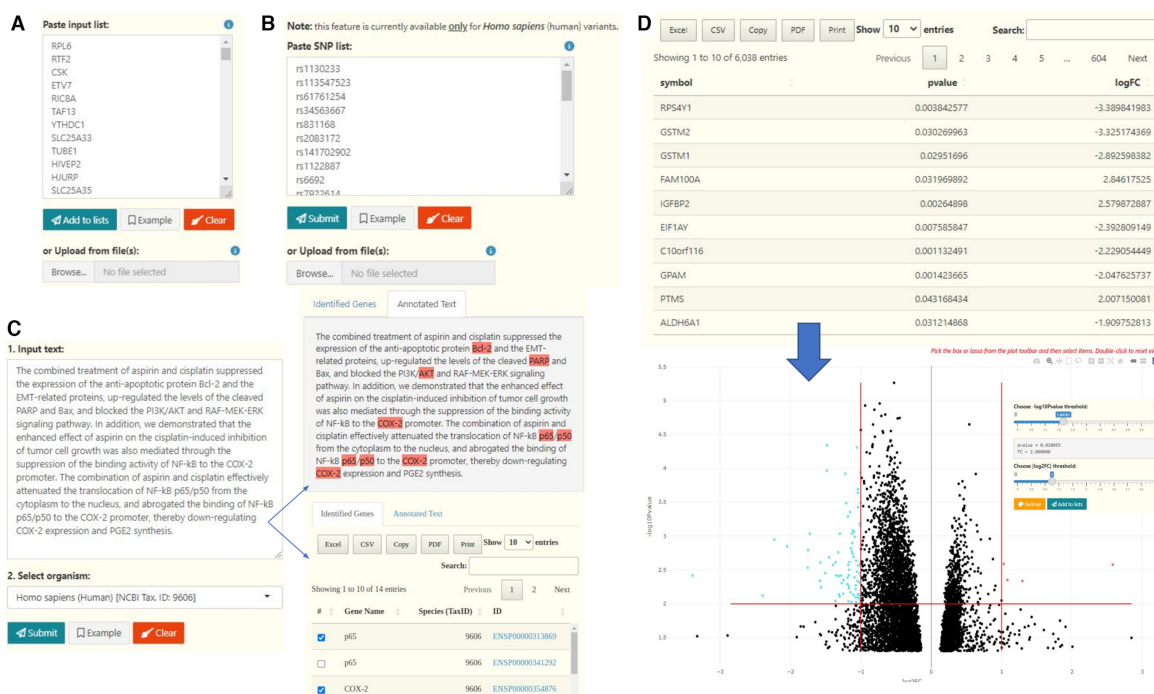


Figure 1. Flame input options. (A) A simple gene list using any of the widely used identifiers (e.g. HUGO names, ENSEMBL, Entrez, UniProt). (B) A list with Human SNPs. (C) A free text which can be further mined for genes and proteins using NER for an organism of preference. The identified genes and proteins can be then used for functional enrichment analysis. (D) Input of gene expression data (gene name, P -value, \logFC) to generate a Volcano plot. Users can manually apply thresholds on the P -value and the FC axes to define the differentially over- and under-expressed genes of significance and interactively select them from the plot for further analysis

This functionality is supported for all currently available organisms in Flame (~14 000 taxa). To this end, it is worth mentioning that the EXTRACT tagger is already being used by a plethora of established text-mining based applications (Pletscher-Frankild *et al.* 2015, Baltoumas *et al.* 2021b, Zafeiropoulos *et al.* 2022, Szklarczyk *et al.* 2023).

Gene Expression data: Defining the significant differentially over- and under-expressed genes is most times an empirical task performed by experimentalists, who usually apply thresholds within an accepted range for parameters such as the fold change (e.g. absolute FC > 2) and the statistical significance (e.g. P -value < 0.05). Therefore, slightly tuning the parameters may result in a more complete biological story. For this purpose, Flame offers the ability to upload expression data in tab delimited or comma separated format (gene name—log fold change—statistical significance) and interactively select sets of genes on a generated Volcano plot. Users can adjust both the P -value and FC thresholds and redraw the plot, highlighting the overexpressed genes in red and the under-expressed genes in blue. Following, users may apply a selection lasso or rectangle to mark gene sets of interest directly on the Volcano plot, and append them to the Flame lists which can then be further processed and enriched.

2.2 Integration and functional enrichment for multiple organisms from various sources

Due to the lack of standards in the field (Wijesooriya *et al.* 2022), functional enrichment applications often differ on the databases they support, the way of processing data, the organisms they cover, their ranking and scoring schemes, their reporting and visualizations as well as the input options they offer. While such differences may hinder usage, a combination of such tools as a means of complementing each other's weaknesses is always an option. However, familiarity with each tool of choice is necessary but often a time consuming process. In Flame (v2.0), we overcome this issue by bringing together four established applications, namely *aGOTool*, *gProfiler*, *WebGestalt*, and *enrichR* under the same interface and making them available, in a consistent and unified manner. These four tools were selected based on the following criteria: (i) availability of programmatic access, either in the form of a package or through an Application Programming Interface (API), (ii) the number and diversity of supported databases for enrichment, (iii) the number of supported organisms, and (iv) the ability to handle large requests.

In detail, *aGOTool* (and subsequently Flame) currently supports 14 436 organisms corresponding to UniProt reference proteomes, including animals (mammals, insects, fish, etc.), plants, fungi, protozoa and bacteria. The tool offers functional enrichment for biological processes, molecular functions, and cellular components through Gene Ontology, pathway enrichment through KEGG, Reactome and WikiPathways, disease enrichment through the Disease Ontology (Schriml *et al.* 2019), annotation enrichment for proteins through UniProt, Pfam and InterPro (Blum *et al.* 2021) as well as tissue enrichment through the Brenda Tissue Ontology (Gremse *et al.* 2011). It also offers literature enrichment through PubMed. Regardless of the input identifiers, gene/protein names are converted to the Ensembl protein namespace (ENSP) through the STRING API before the enrichment process. The ENSP namespace returns the most accurate results for the *aGOTool* enrichment pipeline.

While using the *g:Profiler* pipeline, extra features such as enriching for tissues through the Human protein Atlas (HPA) (Karlsson *et al.* 2021), protein annotations through CORUM, phenotypes through the Human Phenotype Ontology (HPO) (Robinson *et al.* 2008), or DNA regulatory motifs through TRANSFAC and miRTarBase are made available. It is worth mentioning that *g:Profiler* comes with an advanced id conversion function which enables namespace and orthology conversion for 821 organisms, while it provides great flexibility on the identifiers which it accepts as input.

When opting for *WebGestalt*, users can further enrich their lists for OMIM, DisGeNet and GLAD4U (Jourquin *et al.* 2012) diseases, PANTHER pathways (Mi *et al.* 2021) as well as DrugBank and GLAD4U drug annotations (Wishart *et al.* 2018) while genes/proteins are automatically converted to Entrez accession identifiers. Currently *WebGestalt* supports twelve organisms, namely, *Homo sapiens* (human), *Mus musculus* (mouse), *Rattus norvegicus* (rat), *Bos taurus* (bovine), *Arabidopsis thaliana* (thale cress), *Drosophila melanogaster* (fruit fly), *Caenorhabditis elegans* (nematode worm), *Saccharomyces cerevisiae* (yeast), *Danio rerio* (zebrafish), *Sus scrofa* (pig), *Canis lupus familiaris* (dog), and *Gallus gallus* (chicken).

Finally, the *enrichR* library offers functional enrichment for seven organisms, namely *Homo sapiens* (human), *Mus musculus* (mouse), *Bos taurus* (bovine), *Drosophila melanogaster* (fruit fly), *Caenorhabditis elegans* (nematode worm), *Saccharomyces cerevisiae* (yeast), and *Danio rerio* (zebrafish), preferably in the Entrez gene namespace and supports extra enrichment options for phenotypes through the KOMP2 Mouse Phenotypes, WormBase (Davis *et al.* 2022), and MGI Mammalian Phenotype (Ringwald *et al.* 2022) databases.

Through Flame, users can perform enrichment using each individual tool, or can select to combine multiple tools. The combination of tools that can be used depends on their supporting the organism chosen for the analysis. Enrichment can be performed using the entire organism's genome as the reference with any of the four tools; alternatively, users can also upload a custom reference background of their own and perform enrichment with *aGOTool*, *g:Profiler*, and *WebGestalt*.

A summary and a direct comparison of the supported features, cutoff types and allowed namespaces are shown in Table 1. Regarding execution times, *aGOTool* is the default selected enrichment tool of Flame and performs the fastest while *WebGestalt*'s API executes the slowest and may affect Flame's performance.

2.3 Visualization options

Flame extends its functionality by offering a plethora of visualizations and network analysis options for the interpretation of enrichment results. Besides simple searchable and interactive tables, it offers several visual alternatives for presenting the reported results from each enrichment tool (Fig. 2). Flame provides four different main visualization options to represent the top enriched terms. These are: (i) networks, (ii) heatmaps, (iii) barcharts and (i) scatter plots. For each category, one can adjust the top hits interactively by setting up filters and parameters such as datasources, number of visualized top terms and scoring metric. In the case of networks and heatmaps, three different modes of association are offered, namely *Functions* versus *Genes*, *Functions* versus *Functions*, and *Genes* versus *Genes*. In the first mode, the top functions and their associated genes are presented either as a network

Table 1. Comparison of the four different enrichment tools integrated in Flame.

	aGOtool	g: Profiler	WebGestalt	enrichR
Organisms	14 436	821	12	7
Namespaces	Ensembl IDs, UniProt ACs, STRING IDs	Ensembl, Entrez, UniProt, RefSeq, EMBL, ChEMBL, WikiGene	Entrez gene accession identifiers	Entrez gene names
Significance metrics	<i>P</i> -value, false discovery rate (FDR)	g: SCS (adjusted <i>P</i> -value), FDR, Bonferroni	Bonferroni and five FDR types (BH, BY, Holm, Hochberg, Hommel)	Adjusted <i>P</i> -value
Option to use custom background	Yes	Yes	Yes	No
Supported databases				
Gene ontology	Biological processes, molecular functions, cellular components	Biological processes, molecular functions, cellular components	Biological processes, molecular functions, cellular components	Biological processes, molecular functions, cellular components
Biological pathways	KEGG, REACTOME, WikiPathways	KEGG, REACTOME, WikiPathways	KEGG, REACTOME, WikiPathways, PANTHER	KEGG, REACTOME, WikiPathways, PANTHER
Diseases	Disease ontology	N/A	OMIM, DisGeNET, GLAD4U	Disease ontology
Drugs	N/A	N/A	DrugBank, GLAD4U	N/A
Protein families, domains and complexes	InterPro, Pfam, UniProt	CORUM	N/A	N/A
Tissues	BRENDA tissue ontology	Human protein atlas	N/A	WormBase
Phenotypes	N/A	Human phenotype ontology (HPO)	Human phenotype ontology (HPO)	Human phenotype ontology (HPO), KOMP2 mouse phenotypes, WormBase, MGI mammalian phenotype
Regulatory DNA motifs	N/A	TRANSFAC, miRTarBase	N/A	N/A

(nodes linked with edges), or as a heatmap (associations are shown as colored cells/blocks). In the second mode, functions are connected with other functions based on the number of their shared genes, whereas in the third mode, genes are connected based on the common functions they are involved in. In the cases of networks, bar charts and scatter plots, one can adjust the top hits by combining multiple types of resources (e.g. pathways from KEGG, Reactome and WikiPathways). In addition, while most of the network associations are shown in 2D, Flame has the option to call Arena3D^{web} (Karatzas *et al.* 2021, Kokoli *et al.* 2023) in order to visualize heterogeneous information in a 3D view as a multi-layered graph (each biomedical entity type corresponding to a different layer).

All aforementioned plots are also available for the literature enrichment pipeline, where instead of functions, PubMed articles are returned. A Manhattan plot can also be generated for g: Profiler exclusively. Finally, as an extra analysis step, Flame utilizes the STRING API to generate protein-protein interaction networks, which can subsequently be used to also perform enrichment analysis.

2.4 Use of UpSet plots for comparing input lists and reported results from various sources

Flame (v1.0) uses UpSet plots, as a replacement for Venn diagrams, in order to show unions, intersections and distinct combinations among the input gene lists. Following the same approach, in this version, Flame (v2.0) extends this functionality to show how the four supported back-end pipelines (aGOtool, gProfiler, WebGestalt, enrichR) agree on the results they report (Fig. 3).

To this end, when more than one functional enrichment tool has been executed, an extra tab named “Combination” is

generated automatically. In this tab, distinct intersections between the four resources are summarized in a searchable and easy-to-filter table as well as in an interactive UpSet plot. The table reports the aggregated enriched terms along with the tool(s) that generated the term and its respective rank (number of tools that returned it). Rank 1 means that a biomedical entity (e.g. an enriched pathway) was detected by one resource only, while rank 4 means that all four tools fetched the same result. In addition to the above, the statistical significance of the combined results is estimated by applying Fisher’s combined probability test (Mosteller and Fisher 1948), represented by X^2 and combined *P*-value metrics for each enrichment term identified by two or more tools.

In parallel, Flame (v2.0) generates an interactive UpSet plot where users can click on a bar to generate a table which corresponds to the subset of shared terms shown in the corresponding UpSet column. In this implementation, the UpSet plot consists of two to four rows (one for each enrichment tool) while columns describe the distinct intersections among results. Both the table and the UpSet plot can be limited to a user-selected set of biomedical entries (e.g. compare reported pathways, or biological processes or both). In addition, Flame offers the option to visualize the combined results as a *Functions* versus *Genes* association network (as described in the previous section), in which the weight of the edges corresponds to the number of tools supporting the corresponding gene–function associations.

2.5 Implementation

Flame is mainly written in R (Shiny framework). The enrichment analysis with aGOtool is executed through its corresponding API (https://agotool.org/api_orig). Enrichment

library. The volcano plot input along with all other enriched term plots (heatmaps, barcharts and scatter plots) are generated with the use of the *plotly* R library. A Manhattan plot is also offered, specifically for the *g: Profiler* results, through its respective package. Finally, UpSet plots are drawn with the use of *upsetjs* R library.

The SNP input and namespace/orthology conversions are handled through the *g: Profiler* package. Text-mining input is handled by the EXTRACT tagger API (<https://tagger.jensenlab.org>). The PPI network analysis is offered by the STRING API (<https://string-db.org/api>). Flame handles its GET API requests through R and its POST API requests through a dedicated node.js server.

2.6 Application programming interface

Flame comes with its own API and currently supports both GET and POST requests. Shorter requests can be done with a GET request using the following format:

```
https://bib.fleming.gr:8084/app/flame/?
url_genes=MCL1,TTR;APOE,ACE2;TLR4,HMOX1
```

In this example, gene lists are separated with semicolon (;) and genes with comma (.). Longer requests can be performed via Flame's POST request at:

<https://bib.fleming.gr/bib/api/flame>

using the following JSON object format:

```
{ "gene_list1": ["GNAS", "ABCG2", "WT1",
"CDK2", "FLT1", "HBA1", "CCN2", "MDM2"],
"gene_list2": ["MMP1", "PTGS2", "PON1",
"LDLR", "HBA1", "CYP1B1", "PTEN", "SNCA"],
"gene_list3": ["UGT1A1", "CDH1", "MDM2",
"EGFR", "FMR1", "VEGFA", "ERCC1"] }
```

3 Discussion

Flame (v2.0) expands on its previous capabilities in many aspects (Table 2); from offering users a plethora of different input options, to a significant organism-space expansion, the addition of multiple enrichment libraries/APIs and the combination of their results, as well as through enhancing its own API. We believe that the multiple input options along with STRING's organisms, make Flame a very competitive application which greatly enhances user experience. By also offering the streamlined execution of multiple functional enrichment tools and the aggregation of their results, we facilitate the knowledge extraction of analysis, and partially tackle the

Table 2. Main differences between Flame versions.

	Flame (v1.0)	Flame (v2.0)
Enrichment categories	<ol style="list-style-type: none"> 1) Gene ontology 2) Biological pathways 3) Proteins 4) Phenotypes 5) Tissues 6) Regulatory motifs 7) Literature enrichment 	<ol style="list-style-type: none"> 1) Gene ontology 2) Biological pathways 3) Proteins 4) Phenotypes 5) Tissues 6) Regulatory motifs 7) Literature enrichment 8) Drug associations 9) Protein families 10) Diseases
Enrichment tools	<ol style="list-style-type: none"> 1) <i>g: Profiler</i> 2) aGOTool (only for literature enrichment) 	<ol style="list-style-type: none"> 1) aGOTool 2) <i>g: Profiler</i> 3) WebGestalt 4) enrichR
Supported organisms	197	14 436
Input options	<ol style="list-style-type: none"> 1) Gene/identifier lists 	<ol style="list-style-type: none"> 1) Gene/identifier lists 2) SNP lists 3) NER and text-mining on a free text 4) Gene selection from an interactive volcano plot 5) Support of custom background lists
List comparison with the use of UpsetPlots	<ol style="list-style-type: none"> 1) Input gene lists 	<ol style="list-style-type: none"> 1) input gene lists 2) reported results from any of the four supported tools
Other types of analysis	<ol style="list-style-type: none"> 1) Protein–protein interaction networks 2) Gene orthology search 3) Database id conversion 	<ol style="list-style-type: none"> 1) Protein–protein interaction networks 2) Gene orthology search 3) Database id conversion 4) Mapping of SNPs to genes 5) Statistical testing when combining multiple <i>P</i>-values from various tools 6) Functional and literature enrichment for the generated STRING PPI network
Visualization	<ol style="list-style-type: none"> 1) Plots (Bar, UpSet, Scatter, Manhattan) 2) Heatmaps 3) 2D Networks with the use of VisNetwork 	<ol style="list-style-type: none"> 1) Plots (Bar, UpSet, Scatter, Manhattan) 2) Heatmaps 3) 2D Networks with the use of VisNetwork 4) Gene–function network visualization for entities from multiple sources 5) 3D multilayer graphs with Arena3D^{web}
API requests	<ol style="list-style-type: none"> 1) GET 	<ol style="list-style-type: none"> 1) GET 2) POST

question “which tool should I use/trust?”. Given its accessibility and ease of use, we hope Flame will quickly become the go-to tool for advanced functional enrichment analysis and visualization.

Conflict of interest

The authors declare no conflict of interest.

Funding

F.A.B was supported by Fondation Santé and the Onassis Foundation. E.K. was supported by the Hellenic Foundation for Research and Innovation (H.F.R.I) under the ‘First Call for H.F.R.I Research Projects to support faculty members and researchers and the procurement of high-cost research equipment grant’, Grant ID: 1855-BOLOGNA. G.A.P. was supported by the project ‘The Greek Research Infrastructure for Personalized Medicine (pMedGR)’ (MIS 5002802), which is implemented under the Action ‘Reinforcement of the Research and Innovation Infrastructure’, funded by the Operational Program ‘Competitiveness, Entrepreneurship and Innovation’ (NSRF 2014-2020) and co-financed by Greece and the European Union (European Regional Development Fund). P.I.K. and P.G.B. were supported by the project “GENOMIC GENOMIC Analysis of Organisms of Agricultural and liveStock Interest in Sterea” [5045902]. P.I.K. and P.G.B. were also partially supported by the project ELIXIR-GR: The Greek Research Infrastructure for Data Management and Analysis in Life Sciences (MIS 5002780), which is implemented under the Action Reinforcement of the Research and Innovation Infrastructure funded by the Operational Programme Competitiveness, Entrepreneurship and Innovation (NSRF 2014-2020) and co-financed by Greece and the European Union (European Regional Development Fund).

Data availability

The software and data underlying this article are available at <https://github.com/PavlopoulosLab/Flame>.

References

- Aleksander SA, Balhoff J, Carbon S *et al*. The gene ontology knowledgebase in 2023. *GENETICS* 2023;224. [10.1093/genetics/iyad031](https://doi.org/10.1093/genetics/iyad031).
- Amberger JS, Bocchini CA, Schiettecatte F *et al*. OMIM.org: online Mendelian inheritance in man (OMIM[®]), an online catalog of human genes and genetic disorders. *Nucleic Acids Res* 2015;43:D789–98.
- Arita M, Karsch-Mizrachi I, Cochrane G *et al*. The international nucleotide sequence database collaboration. *Nucleic Acids Res* 2021;49:D121–4.
- Baltoumas FA, Zafeiropoulou S, Karatzas E *et al*. Biomolecule and entity interaction databases in systems biology: a comprehensive review. *Biomolecules* 2021a;11:1245.
- Baltoumas FA, Zafeiropoulou S, Karatzas E *et al*. OnTheFly^{2.0}: a text-mining web application for automated biomedical entity recognition, document annotation, network and functional enrichment analysis. *NAR Genom Bioinform* 2021b;3:lqab090.
- Bindea G, Mlecnik B, Hackl H *et al*. ClueGO: a cytoscape plug-in to decipher functionally grouped gene ontology and pathway annotation networks. *Bioinformatics* 2009;25:1091–3.
- Blum M, Chang H-Y, Chuguransky S *et al*. The InterPro protein families and domains database: 20 years on. *Nucleic Acids Res* 2021;49:D344–54.
- Bu D, Luo H, Huo P *et al*. KOBAS-i: intelligent prioritization and exploratory visualization of biological functions for gene enrichment analysis. *Nucleic Acids Res* 2021;49:W317–25.
- Chen J, Bardes EE, Aronow BJ *et al*. ToppGene suite for gene list enrichment analysis and candidate gene prioritization. *Nucleic Acids Res* 2009;37:W305–11.
- Chicco D, Jurman G. A brief survey of tools for genomic regions enrichment analysis. *Front Bioinform* 2022;2:968327.
- Cinar O, Viechtbauer W. The poolr package for combining independent and dependent p values. *J Stat Soft* 2022;101(1):1–42.
- Cunningham F, Allen JE, Allen J *et al*. Ensembl 2022. *Nucleic Acids Res* 2022;50:D988–95.
- Davis P, Zarowiecki M, Arnaboldi V *et al*. WormBase in 2022—data, processes, and tools for analyzing *Caenorhabditis elegans*. *Genetics* 2022;220:iyac003.
- Doncheva NT, Morris JH, Holze H *et al*. Cytoscape stringApp 2.0: analysis and visualization of heterogeneous biological networks. *J Proteome Res* 2023;22:637–46.
- Eden E, Navon R, Steinfeld I *et al*. GOrilla: a tool for discovery and visualization of enriched GO terms in ranked gene lists. *BMC Bioinformatics* 2009;10:48.
- Evangelista JE, Xie Z, Marino GB *et al*. Enrichr-KG: bridging enrichment analysis across multiple libraries. *Nucleic Acids Res* 2023;51:W168–79.
- Fabregat A, Jupe S, Matthews L *et al*. The reactome pathway knowledgebase. *Nucleic Acids Res* 2018;46:D649–55.
- Finn RD, Coghill P, Eberhardt RY *et al*. The Pfam protein families database: towards a more sustainable future. *Nucleic Acids Res* 2016;44:D279–85.
- Gerstner N, Kehl T, Lenhof K *et al*. GeneTrail 3: advanced high-throughput enrichment analysis. *Nucleic Acids Res* 2020;48:W515–20.
- Giurgiu M, Reinhard J, Brauner B *et al*. CORUM: the comprehensive resource of mammalian protein complexes—2019. *Nucleic Acids Res* 2019;47:D559–63.
- Gremse M, Chang A, Schomburg I *et al*. The BRENDA tissue ontology (BTO): the first all-integrating ontology of all organisms for enzyme sources. *Nucleic Acids Res* 2011;39:D507–13.
- Hu Y, Comjean A, Attrill H *et al*. PANGEA: a new gene set enrichment tool for Drosophila and common research organisms. *Nucleic Acids Res* 2023;51:W419–26.
- Huang DW, Sherman BT, Lempicki RA *et al*. Bioinformatics enrichment tools: paths toward the comprehensive functional analysis of large gene lists. *Nucleic Acids Res* 2009;37:1–13.
- Huang H-Y, Lin Y-C-D, Li J *et al*. miRTarBase 2020: updates to the experimentally validated microRNA–target interaction database. *Nucleic Acids Res* 2020;48:D148–54.
- Jourquin J, Duncan D, Shi Z *et al*. GLAD4U: deriving and prioritizing gene lists from PubMed literature. *BMC Genomics* 2012;13:S20.
- Kanehisa M, Goto S. KEGG: Kyoto encyclopedia of genes and genomes. *Nucleic Acids Res* 2000;28:27–30.
- Karatzas E, Baltoumas FA, Panayiotou NA *et al*. Arena3Dweb: interactive 3D visualization of multilayered networks. *Nucleic Acids Res* 2021;49:W36–45.
- Karlsson M, Zhang C, Méar L *et al*. A single-cell type transcriptomics map of human tissues. *Sci Adv* 2021;7:eabh2169.
- Kokoli M, Karatzas E, Baltoumas FA *et al*. Arena3Dweb: interactive 3D visualization of multilayered networks supporting multiple directional information channels, clustering analysis and application integration. *NAR Genom Bioinform* 2023;5:lqad053.
- Koutrouli M, Karatzas E, Paez-Espino D *et al*. A guide to conquer the biological network era using graph theory. *Front Bioeng Biotechnol* 2020;8:34.
- Kuleshov MV, Diaz JEL, Flamholz ZN *et al*. modEnrichr: a suite of gene set enrichment analysis tools for model organisms. *Nucleic Acids Res* 2019;47:W183–90.

- Li W, O'Neill KR, Haft DH *et al.* RefSeq: expanding the prokaryotic genome annotation pipeline reach with protein family model curation. *Nucleic Acids Res* 2021;49:D1020–8.
- Liberzon A, Subramanian A, Pinchback R *et al.* Molecular signatures database (MSigDB) 3.0. *Bioinformatics* 2011;27:1739–40.
- Maere S, Heymans K, Kuiper M *et al.* BiNGO: a cytoscape plugin to assess overrepresentation of gene ontology categories in biological networks. *Bioinformatics* 2005;21:3448–9.
- Maier H, Döhr S, Grote K *et al.* LitMiner and WikiGene: identifying problem-related key players of gene regulation using publication abstracts. *Nucleic Acids Res* 2005;33:W779–82.
- Maleki F, Ovens K, Hogan DJ *et al.* Gene set analysis: challenges, opportunities, and future research. *Front Genet* 2020;11:654.
- Matys V, Fricke E, Geffers R *et al.* TRANSFAC: transcriptional regulation, from patterns to profiles. *Nucleic Acids Res* 2003;31:374–8.
- Mendez D, Gaulton A, Bento AP *et al.* ChEMBL: towards direct deposition of bioassay data. *Nucleic Acids Res* 2019;47:D930–40.
- Mi H, Ebert D, Muruganujan A *et al.* PANTHER version 16: a revised family classification, tree-based classification tool, enhancer regions and extensive API. *Nucleic Acids Res* 2021;49:D394–403.
- Mosteller F, Fisher RA. Questions and answers. *Am Stat* 1948;2:16–7.
- Otasek D, Morris JH, Bouças J *et al.* Cytoscape automation: empowering workflow-based network analysis. *Genome Biol* 2019;20:185.
- Pafilis E, Buttigieg PL, Ferrell B *et al.* EXTRACT: interactive extraction of environment metadata and term suggestion for metagenomic sample annotation. *Database (Oxford)* 2016;2016:baw005.
- Piñero J, Ramírez-Anguita JM, Saüch-Pitarch J *et al.* The DisGeNET knowledge platform for disease genomics: 2019 update. *Nucleic Acids Res* 2020;48:D845–55.
- Pletscher-Frankild S, Pallegà A, Tsafou K *et al.* DISEASES: text mining and data integration of disease–gene associations. *Methods* 2015;74:83–9.
- Ringwald M, Richardson JE, Baldarelli RM *et al.* Mouse genome informatics (MGI): latest news from MGD and GXD. *Mamm Genome* 2022;33:4–18.
- Robinson PN, Köhler S, Bauer S *et al.* The human phenotype ontology: a tool for annotating and analyzing human hereditary disease. *Am J Hum Genet* 2008;83:610–5.
- Saito R, Smoot ME, Ono K *et al.* A travel guide to cytoscape plugins. *Nat Methods* 2012;9:1069–76.
- Sayers EW,avanaugh M, Clark K *et al.* GenBank. *Nucleic Acids Res* 2022;50:D161–4.
- Schölz C, Lyon D, Refsgaard JC *et al.* Avoiding abundance bias in the functional annotation of post-translationally modified proteins. *Nat Methods* 2015;12:1003–4.
- Schriml LM, Mitraka E, Munro J *et al.* Human disease ontology 2018 update: classification, content and workflow expansion. *Nucleic Acids Res* 2019;47:D955–62.
- Sherman BT, Hao M, Qiu J *et al.* DAVID: a web server for functional enrichment analysis and functional annotation of gene lists (2021 update). *Nucleic Acids Res* 2022;50:W216–21.
- Sherry ST, Ward MH, Kholodov M *et al.* dbSNP: the NCBI database of genetic variation. *Nucleic Acids Res* 2001;29:308–11.
- Slenter DN, Kutmon M, Hanspers K *et al.* WikiPathways: a multifaceted pathway database bridging metabolomics to other omics research. *Nucleic Acids Res* 2018;46:D661–7.
- Szklarczyk D, Kirsch R, Koutrouli M *et al.* The STRING database in 2023: protein–protein association networks and functional enrichment analyses for any sequenced genome of interest. *Nucleic Acids Res* 2023;51:D638–46.
- Thanati F, Karatzas E, Baltoumas FA *et al.* FLAME: a web tool for functional and literature enrichment analysis of multiple gene lists. *Biology (Basel)* 2021;10:665.
- Tian T, Liu Y, Yan H *et al.* agriGO v2.0: a GO analysis toolkit for the agricultural community, 2017 update. *Nucleic Acids Res* 2017;45:W122–9.
- UniProt Consortium. UniProt: the universal protein knowledgebase in 2021. *Nucleic Acids Res* 2021;49:D480–9.
- van den Berg BHJ, Thanthiriwatte C, Manda P *et al.* Comparing gene annotation enrichment tools for functional modeling of agricultural microarray data. *BMC Bioinformatics* 2009;10:S9.
- Wang J, Vasaiakar S, Shi Z *et al.* WebGestalt 2017: a more comprehensive, powerful, flexible and interactive gene set enrichment analysis toolkit. *Nucleic Acids Res* 2017;45:W130–7.
- Wijesooriya K, Jadaan SA, Perera KL *et al.* Urgent need for consistent standards in functional enrichment analysis. *PLoS Comput Biol* 2022;18:e1009935.
- Wishart DS, Feunang YD, Guo AC *et al.* DrugBank 5.0: a major update to the DrugBank database for 2018. *Nucleic Acids Res* 2018;46:D1074–82.
- Xie Z, Bailey A, Kuleshov MV *et al.* Gene set knowledge discovery with enrichr. *Curr Protoc* 2021;1:e90.
- Ye J, Zhang Y, Cui H *et al.* WEGO 2.0: a web tool for analyzing and plotting GO annotations, 2018 update. *Nucleic Acids Res* 2018;46:W71–5.
- Yu G, Wang L-G, Han Y *et al.* clusterProfiler: an R package for comparing biological themes among gene clusters. *OMICS* 2012;16:284–7.
- Yu G, Wang L-G, Yan G-R *et al.* DOSE: an R/bioconductor package for disease ontology semantic and enrichment analysis. *Bioinformatics* 2015;31:608–9.
- Zafeiropoulos H, Paragkamian S, Ninidakis S *et al.* PREGO: A literature and data-mining resource to associate microorganisms, biological processes, and environment types. *Microorganisms* 2022;10:293. [10.3390/microorganisms10020293](https://doi.org/10.3390/microorganisms10020293).
- Zambon AC, Gaj S, Ho I *et al.* GO-Elite: a flexible solution for pathway and ontology over-representation. *Bioinformatics* 2012;28:2209–10.
- Zhang D, Hu Q, Liu X *et al.* AllEnricher: a comprehensive gene set function enrichment tool for both model and non-model species. *BMC Bioinformatics* 2020;21:106.
- Zhou Y, Zhou B, Pache L *et al.* Metascape provides a biologist-oriented resource for the analysis of systems-level datasets. *Nat Commun* 2019;10:1523.