Published online 21 May 2021 Nucleic Acids Research, 2021, Vol. 49, Web Server issue **W207–W215**
https://doi.org/10.1093/nar/gkab366

eVITTA: a web-based visualization and inference toolbox for transcriptome analysis

Xuanjin Cheng[1,2,3,†], **Junran Yan**[1,2,4,†], **Yongxing Liu**[1,2,3], **Jiahe Wang**[1,2,3] and **Stefan Taubert**[1,2,3,4,*]

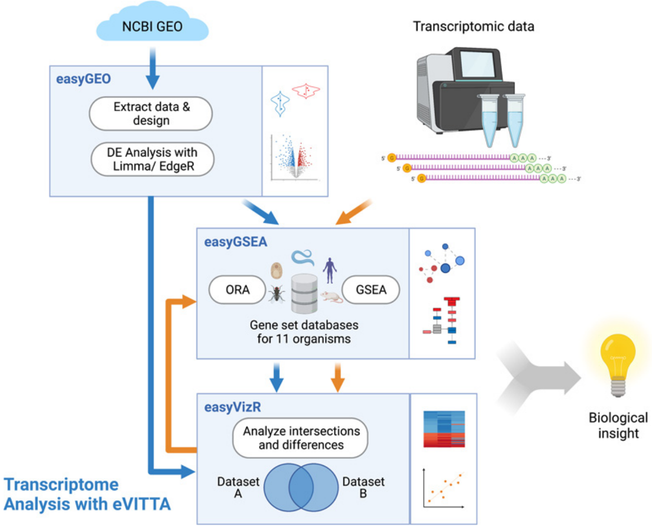
[1]Centre for Molecular Medicine and Therapeutics, The University of British Columbia, Vancouver, British Columbia, Canada, [2]British Columbia Children's Hospital Research Institute, The University of British Columbia, Vancouver, British Columbia, Canada, [3]Department of Medical Genetics, The University of British Columbia, Vancouver, British Columbia, Canada and [4]Graduate Program for Cell and Developmental Biology, The University of British Columbia, Vancouver, British Columbia, Canada

Received March 15, 2021; Revised April 12, 2021; Editorial Decision April 22, 2021; Accepted May 04, 2021

ABSTRACT

Transcriptome profiling is essential for gene regulation studies in development and disease. Current web-based tools enable functional characterization of transcriptome data, but most are restricted to applying gene-list-based methods to single datasets, inefficient in leveraging up-to-date and species-specific information, and limited in their visualization options. Additionally, there is no systematic way to explore data stored in the largest transcriptome repository, NCBI GEO. To fill these gaps, we have developed eVITTA (easy Visualization and Inference Toolbox for Transcriptome Analysis; https://tau.cmmt.ubc.ca/eVITTA/). eVITTA provides modules for analysis and exploration of studies published in NCBI GEO (easyGEO), detailed molecular- and systems-level functional profiling (easyGSEA), and customizable comparisons among experimental groups (easyVizR). We tested eVITTA on transcriptomes of SARS-CoV-2 infected human nasopharyngeal swab samples, and identified a downregulation of olfactory signal transducers, in line with the clinical presentation of anosmia in COVID-19 patients. We also analyzed transcriptomes of *Caenorhabditis elegans* worms with disrupted *S*-adenosylmethionine metabolism, confirming activation of innate immune responses and feedback induction of one-carbon cycle genes. Collectively, eVITTA streamlines complex computational workflows into an accessible interface, thus filling the gap of an end-to-end platform capable of capturing both broad and granular changes in human and model organism transcriptomes.

GRAPHICAL ABSTRACT


NCBI GEO
Transcriptomic data
easyGEO
Extract data & design
DE Analysis with Limma/ EdgeR
easyGSEA
ORA
GSEA
Gene set databases for 11 organisms
easyVizR
Analyze intersections and differences
Dataset A
Dataset B
Biological insight
Transcriptome Analysis with eVITTA

INTRODUCTION

Transcriptome profiling is an essential technique to study gene regulation in development and disease (1). The emergence of affordable high-throughput microarray and sequencing technologies has resulted in the rapid expansion of transcriptome experiments, which in turn greatly increased the demand for robust analytical tools for data interpretation. Effective transcriptome interpretation involves three key aspects: drawing inference from published studies, translating the data into meaningful biological knowledge, and comparing multiple conditions to each other to discern underlying regulatory changes.

First, knowledge from past studies is essential for hypothesis generation and data interpretation. The Gene Expression Omnibus (GEO) database, funded by the Na-

[*]To whom correspondence should be addressed. Tel: +1 604 875 3860; Fax: +1 604 875 3819; Email: taubert@cmmt.ubc.ca
[†]The authors wish it to be known that, in their opinion, the first two authors should be regarded as joint First Authors.

tional Center for Biotechnology Information (NCBI), is the largest public repository for transcriptome datasets (144 751 data series, 4.2 million samples on 23 February 2021) (2,3). Despite the treasure trove of data in GEO, no tool yet exists that can systematically extract and process such data for inferential use. Most web-based GEO data analysis tools are limited in functionality: some only provide access to microarray data without differential expression (DE) analysis (4), while others analyzing RNA-sequencing (RNA-seq) data rely on their own processed data repositories, which tend to update slowly and often exclude datasets due to unsupported species or experiment type (5).

Second, uncovering mechanistic insights from gene expression data is central to all types of transcriptomic studies. Functional enrichment analysis (aka functional profiling) is the primary technique for this purpose, and is commonly used to interpret gene lists derived from many omics platforms (6). To date, a variety of web-based enrichment analysis tools have been developed (7–14), but these are sometimes suboptimal for interpreting transcriptome data. Surveys have shown that most tools are outdated in their gene annotation (gene set, GS) databases, sometimes by several years, which can severely impact functional interpretation and follow-up experiments (15). When multiple GS databases are analyzed together, most tools list results separately in tables or simple graphs, which is ineffective in integrating the information (8–12). Some tools rely on literature-curated resources such as Gene Ontology (GO), which are sometimes not precise enough to capture the functions of genes in the biological system of interest (8–12). Approach-wise, gene-list-based overrepresentation analysis (ORA) remains predominant (7–13); alternative methods for transcriptomic studies, such as pre-ranked Gene Set Enrichment Analysis (GSEA) based on gene-set scoring (16), are not supported by most tools. The only tool supporting GSEA to our knowledge (14) requires users to supply a separately generated rank file, setting a hurdle for non-bioinformaticians. Furthermore, most existing tools provide limited options for visualizing transcriptome patterns.

Third, multiple dataset comparisons play a crucial role in interpreting multi-group experiments and in comparing new results to published data. Despite the growth in demand, no web-based tool exists yet to our knowledge that provides a complete pipeline for identifying and visualizing intersections and disjoints among multiple transcriptome profiles. Tools exist for filtering gene lists (17) or plotting certain types of visualizations such as Venn diagrams (18–20), UpSet plots (19,21), and heatmaps (22), but stringing these modules into an inference workflow remains tedious. The one tool that does provide a graphical workflow for intersection analysis, to our knowledge, is only tailored to pairwise comparisons, and does not provide interactive or customizable visualizations (23).

To address these challenges in transcriptome analysis and interpretation, we have developed eVITTA (easy Visualization and Inference Toolbox for Transcriptome Analysis; https://tau.cmmt.ubc.ca/eVITTA/). It consists of three modules that can work together or as standalones: easyGEO accesses, analyzes, and visualizes transcriptome data in NCBI GEO; easyGSEA visually delineates gene expression patterns by functional profiling; and easyVizR compares and contrasts multiple datasets via an integrated intersection analysis workflow. Above all, eVITTA's interactive and user-friendly interface makes transcriptome analysis accessible for wet and dry lab biologists alike. The multiple entry and exit points in the workflow also allow users to adapt one or more individual tool(s) into their own custom analysis pipeline (Figure 1, Table 1).

To test eVITTA and demonstrate its capabilities, we performed two independent evaluation studies on published gene expression datasets: (i) transcriptomes of SARS-CoV-2 infected human nasopharyngeal (NP) swab samples and (ii) transcriptomes of *Caenorhabditis elegans* worms deficient in *sams-1/MAT1A* or *sbp-1/SREBP*. We were able to recapitulate original findings and also discovered additional biological insights that were experimentally confirmed in studies following the original profiling experiments, demonstrating eVITTA's effectiveness in transcriptome interpretation.

MATERIALS AND METHODS

Implementation

The eVITTA web server runs on Ubuntu Linux (18.04.5 LTS) with 32 GB memory, 16-core CPUs and a 10TB hard drive with Apache (version 2.4.29, https://httpd.apache.org), R (version 4.0.2, https://www.r-project.org/), and R Shiny Server (version 1.5.14.948, https://rstudio.com/products/shiny/download-server/). eVITTA utilizes several third-party tools, including GEOquery (24), edgeR (25), limma (26), plotly (https://plotly.com/r/), Tidyverse (27), fgsea (28), gprofiler2 (29), pathview (30), visNetwork (https://datastorm-open.github.io/visNetwork/), VennDiagram (31), UpsetR (32), RRHO (33) and others (Supplementary Table S1). eVITTA has been tested successfully on several browsers, including Safari v13.1.2, Firefox v65.0, and Chrome v86.0.4240.111. Detailed methods in Supplementary Data Section 1.

easyGEO: An interface to access, analyze and visualize transcriptome data from NCBI GEO

Inputs: The unique GEO identifier of an NCBI GEO series, which begins with 'GSE'.

Data processing: Gene expression data matrix and design matrix are retrieved automatically. If the count table in an RNA-seq study is specified as raw, genes expressed at a level less than 1 count per million (CPM) reads in at least five of the samples, or the minimum number of biological repeats in each condition, whichever less, are excluded from further analysis; if the count table in an RNA-seq experiment is normalized, or if the dataset is based on microarrays, a threshold of 1 is applied likewise to exclude barely expressed genes. Next, for RNA-seq datasets, raw read counts are normalized using the trimmed mean of M-values (TMM) in edgeR (25) to adjust samples for differences in library size and Limma-voom (26) transformed using the default parameters; for microarray and normalized RNA-seq counts, Limma-voom (26) transformation is applied with the normalize = 'quantile' option. Then, a linear model using weighted least squares for each gene is fitted with Limma-lmFit (26); batch effect, if any, is processed as

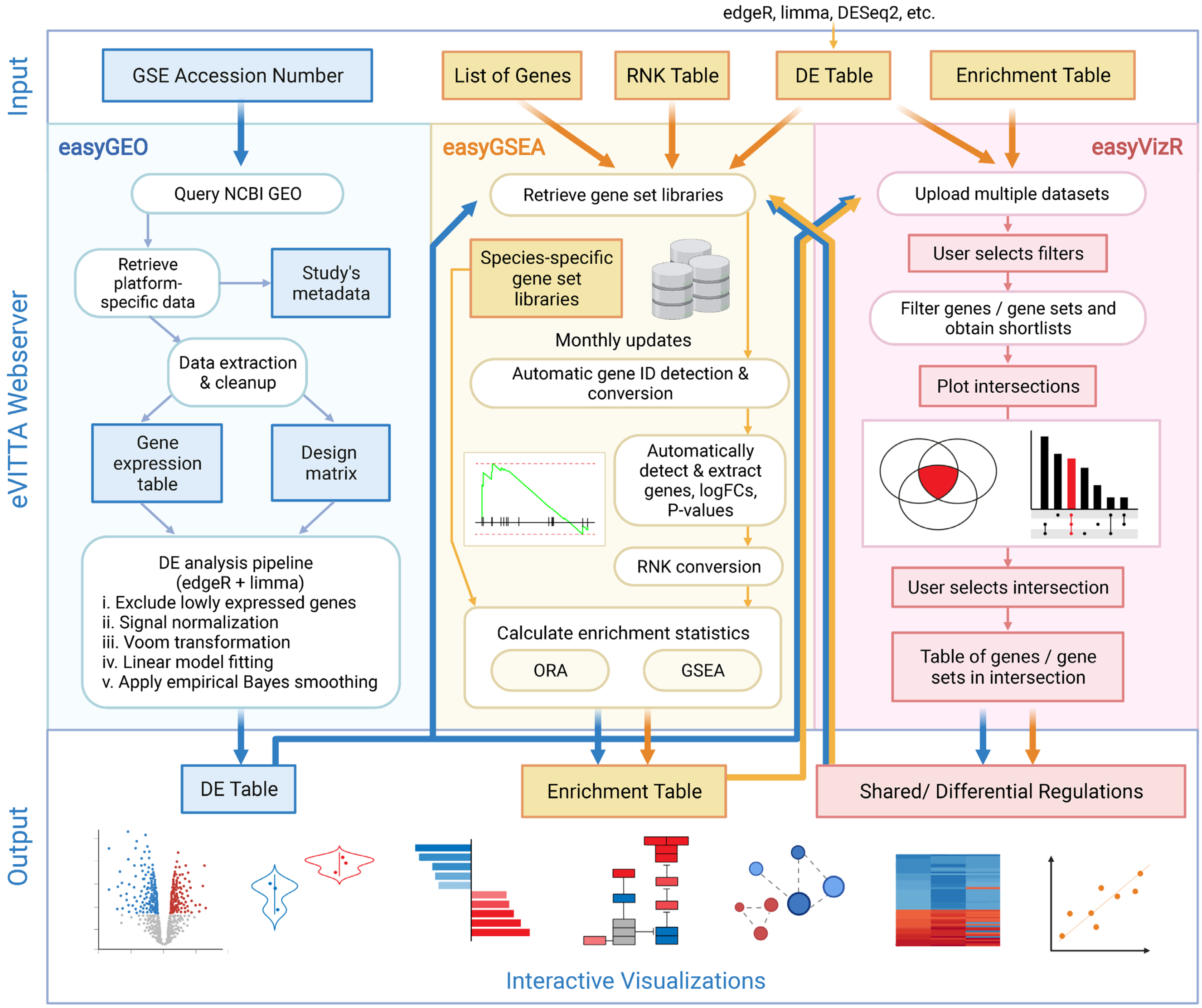

edgeR, limma, DESeq2, etc.
Input
GSE Accession Number
List of Genes
RNK Table
DE Table
Enrichment Table
eVITTA Webserver
easyGEO
Query NCBI GEO
Retrieve platform-specific data
Study's metadata
Data extraction & cleanup
Gene expression table
Design matrix
DE analysis pipeline (edgeR + limma)
i. Exclude lowly expressed genes
ii. Signal normalization
iii. Voom transformation
iv. Linear model fitting
v. Apply empirical Bayes smoothing
easyGSEA
Retrieve gene set libraries
Species-specific gene set libraries
Monthly updates
Automatic gene ID detection & conversion
Automatically detect & extract genes, logFCs, P-values
RNK conversion
Calculate enrichment statistics
ORA
GSEA
easyVizR
Upload multiple datasets
User selects filters
Filter genes / gene sets and obtain shortlists
Plot intersections
User selects intersection
Table of genes / gene sets in intersection
Output
DE Table
Enrichment Table
Shared/ Differential Regulations
Interactive Visualizations

Figure 1. A transcriptome interpretation workflow using the eVITTA toolbox. DE: differential expression; GEO: Gene Expression Omnibus; GSE: Gene Series Expression; GSEA: Gene Set Enrichment Analysis; RNK: ranked gene list; ORA: overrepresentation analysis.

Table 1. Overview of eVITTA tools

Tool	Input	Data processing	Output data	Output visualizations
easyGEO	GEO accession number	DE analysis with edgeR and limma	Gene expression, study design, DE tables, comma-separated	Volcano, heatmap, box, violin
easyGSEA	RNK file, DE table, or gene list	ORA or GSEA	Enrichment table, comma-separated	Bar, bubble, keyword, Manhattan, volcano; GSEA plot, density, box, violin; pathway maps; enrichment network, dendrogram, cluster bar/bubble/table
easyVizR	GSEA or DE results	Intersection analysis based on user-filtered lists	Intersection table, comma-separated	Venn, UpSet plot; heatmap, 2-D & 3-D scatter; RRHO plot, rank-rank scatter; volcano, bar; leading-edge network; word cloud

a factor in the design matrix to exclude sequencing artifacts. Lastly, eBayes (26) empirical Bayes smoothing of standard errors is applied to assess for DE.

Outputs: (a) Gene expression and study design tables (importable into DE analyzers); (b) DE table (importable into easyGSEA and easyVizR); and (c) visualizations (volcano, heatmap, box and violin) that highlight most significantly altered genes and display expression changes of a single gene.

easyGSEA: functional profiling with integrative gene annotation databases

Inputs: GSEA module: a ranked gene list (RNK) file or a DE table, comma- or tab-delimited. ORA module: list of genes or proteins, delimited by newline, tab or whitespace.

Data processing: To start, users select species of interest and adjust choices of databases if needed (Supplementary Table S2); or, users upload GS libraries supplied in Gene Matrix Transposed format (*.gmt) for custom analysis. Gene identifiers, if specified as Other/Mixed, are automatically converted into HUGO symbols. In the GSEA module, rank tables are automatically calculated if input is a DE table. Next, ORA or GSEA is performed (by default, min GS size = 15, max GS size = 200, permutation = 1000) and visualizations are automatically generated, each customizable with its own plotting parameters.

Outputs: (a) converted rank and DE tables (GSEA module); (b) enrichment table (importable into easyVizR); (c) results summary with interactive bar, bubble, keyword, Manhattan and volcano plots; (d) individual GS's statistics, and its distribution in the genome background (GSEA module) delineated with enrichment, density, box and violin plots; (e) pathway maps (KEGG (34), Reactome (35) and Wikipathways (36)); and (f) enrichment network with clustering dendrogram.

easyVizR: a systematic workflow for comparing regulatory patterns in multiple datasets

Inputs: Comma-delimited data table(s) containing identifiers, differential expression metric (e.g. $\log_2$-transformed fold change, enrichment score), *P*-value (pval), and FDR or adjusted *P*-value (padj).

Data processing: For each selected dataset, filtered lists of genes or GSs are generated from user-selected filters (default: pval < 0.05). From filtered lists, users may select an intersection of interest by defining set relationships (Supplementary Figure S1). The selected intersection is highlighted in Venn and Upset plots, and terms in the intersection are used to generate interactive visualizations.

Outputs: (a) filtered gene lists (importable into easyGSEA or other ORA tools) and corresponding expression tables; (b) Venn and UpSet plots; (c) heatmap for terms in chosen intersection; (d) 2D and 3D scatter plots, rank-rank hypergeometric overlap (RRHO) plot and rank-rank scatter for correlation analysis (33); (e) volcano and bar plot for single datasets; (f) leading-edge network (for GSEA outputs); and (g) text enrichment word cloud for identifiers.

EVALUATION

SARS-CoV-2 infected human nasopharyngeal transcriptomes show deregulation of olfactory signal transducers

Since early 2020, the global spread of SARS-CoV-2 has led to concerted efforts to characterize its etiology in human patients. To test the analytical pipeline of eVITTA, we reanalyzed a published RNA-seq dataset (GSE152075 (37)) of nasopharyngeal (NP) swabs from 430 SARS-CoV-2-infected individuals and 54 uninfected controls (for detailed steps, see Supplementary Data Section 2.1). First, using easyGEO, we retrieved the count data and design matrix submitted by the authors, and performed DE analysis. In line with the original findings (37), we found that SARS-CoV-2 infection induced an interferon-driven antiviral response in the nasopharynx, upregulating genes encoding antiviral factors (e.g. *IFIT1/2/3/6*, *RSAD2*) and chemokines (e.g. *CXCL9/10/11*) (Figure 2A). Next, to test if eVITTA's analytical capacity using combinatorial GS databases improves the sensitivity of finding molecular patterns, we performed GSEA using the default selection of biological process and pathway databases in easyGSEA. We found that olfactory transduction GSs were downregulated (Figure 2B, Supplementary Figure S2A, Supplementary Table S3). At the gene level, key olfactory transducers, including G protein subunit alpha L (*GNAL*) (38) and cyclic nucleotide-gated channel subunit alpha 4 (*CNGA4*) (39), showed reduced expression (Figure 2C, D, Supplementary Figure S2B). The observation of deregulated olfactory signaling during SARS-CoV-2 infection agrees well with the clinical presentation of anosmia in COVID-19 cases (40) and with a recent report of transient olfactory dysfunction in mice infected with SARS-CoV-2 (41). Together, this demonstrates eVITTA's capacity to capture both broad and granular patterns in gene expression, which facilitates the identification of biological insights.

S-adenosylmethionine (SAM) mediates innate immune response and lipogenesis in *C. elegans*

To test eVITTA's functional profiling capacity and computational reproducibility, we analyzed published *C. elegans* transcriptomes characterizing the response to *S*-adenosylmethionine (SAM) deficiency. The universal methyl donor SAM is produced by SAM synthase (*SAMS* in *C. elegans*; *MAT* in mammals) in the one-carbon cycle (1CC). We reanalyzed published microarray and RNA-seq data characterizing responses to *sams-1* RNA interference (RNAi), which were previously analyzed by ORA (42) and used to validate the *C. elegans* functional database and annotation tool WormCat (43). We first compared three independent transcriptome analyses of *sams-1* RNAi-treated worms, including two microarray datasets and one RNA-seq dataset (42,44) (Supplementary Table S4; detailed steps in Supplementary Data Section 2.2). Despite the difference of the RNA-seq study compared to the two microarrays in terms of experimental platform and upstream processing, enrichment analysis with eVITTA showed a substantial overlap (Figure 3A) and strong correlation in terms of significantly regulated GSs (R^2 = 0.95 and 0.87; Figure 3B, C), and also in terms of overall enrichment profiles

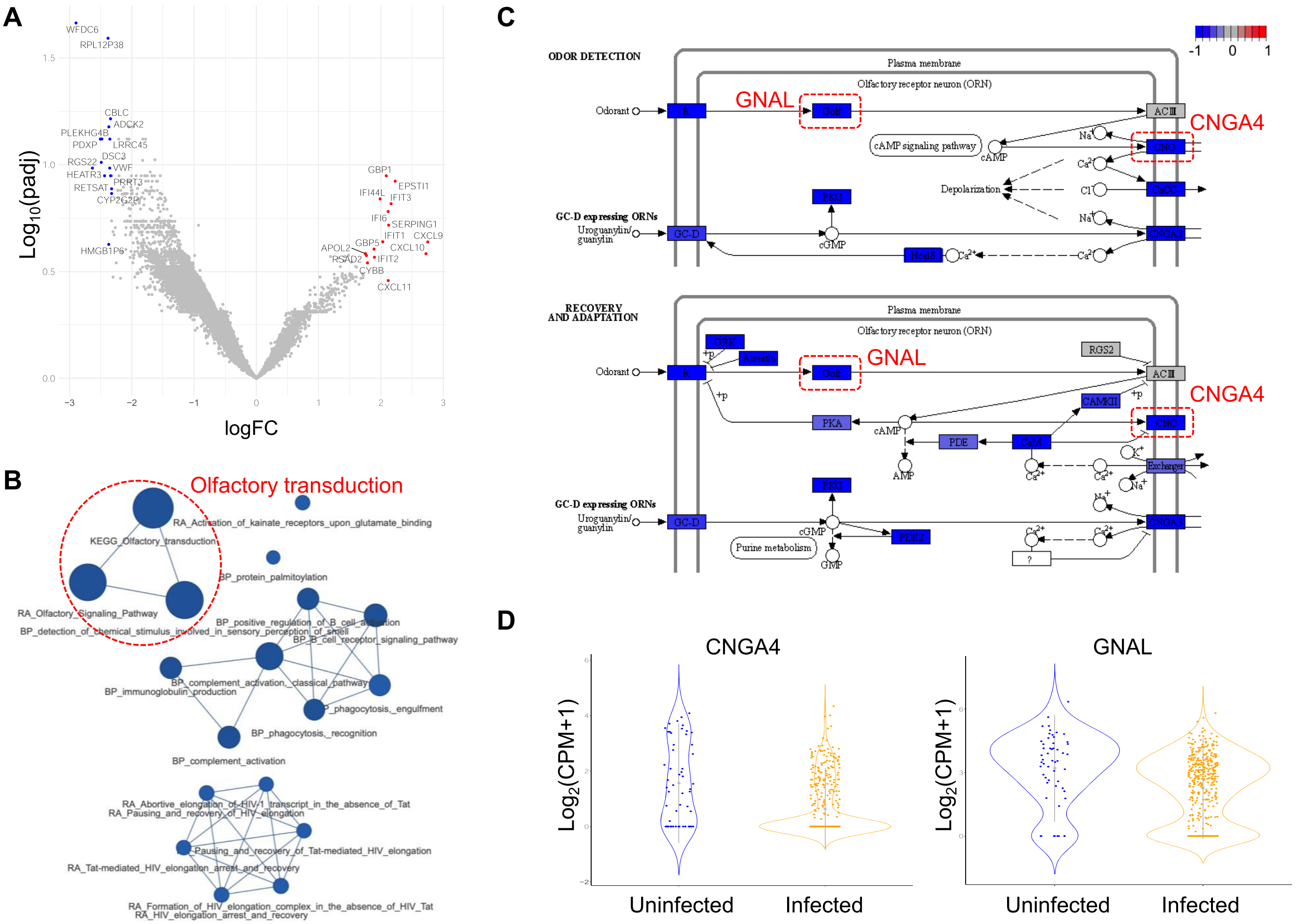


Figure 2. Evaluation study on SARS-CoV-2 infected human nasopharyngeal transcriptomes (GEO repository: GSE152075). (**A**) The volcano plot highlights the 15 most upregulated (red) and 15 most downregulated (blue) genes by logFC in SARS-CoV-2 infected NP samples relative to uninfected individuals. (**B**) The graph shows the enrichment network of GSs deregulated by SARS-CoV-2 infection, including significantly (pval < 0.001, padj < 0.3) downregulated olfactory transduction processes. Node denotes GS; node size reflects the number of leading-edge genes in each GS; blue, downregulation. Edge reflects significant overlap of leading-edge genes as defined by a Jaccard Coefficient larger than or equal to 0.25. Detailed statistics are provided in Supplementary Table S3. (**C**) The pathway map depicts gene expression changes of the KEGG olfactory transduction pathway in SARS-CoV-2 infected NP samples. Blue, downregulation. Notable genes highlighted in red. (**D**) Violin plots show reduced expression of two key olfactory transducers, *CNGA4* (logFC = -0.84, pval = 1.69E-01, padj = 4.88E-01) and *GNAL* (logFC = -1.65, pval = 1.30E−02, padj = 3.2E−01), in SARS-CoV-2 infected NP samples. CPM, counts per million; ES, enrichment score; GEO, Gene Expression Omnibus; GS, gene set; KEGG, Kyoto Encyclopaedia of Genes and Genomes; logFC, $\log_2$-transformed fold change; NP, nasopharyngeal; padj, adjusted *P*-value; pval, *P*-value.

(rho = 0.77 and 0.72; Figure 3D, E). This suggests that the eVITTA pipeline is robust enough to handle comparisons of studies with differences in upstream platforms and processing.

In *C. elegans*, SAM deficiency induces immune responses in the absence of pathogen infection, and a similar response occurs upon depletion of the SAM-regulated lipid synthesis regulator *sbp-1/SREBP* (42). We thus tested eVITTA's efficacy to capture convergent and divergent regulations following *sams-1* or *sbp-1* RNAi. Consistent with previous findings (42,43), we confirmed a strong immune signature in both *sams-1* and *sbp-1* deficiency (Figure 4A-B). Interestingly, eVITTA's comprehensive GS databases allowed us to discover specific changes in one branch of innate immunity, Toll-like receptor (TLR) signaling (Figure 4B). Prior studies have shown that, in *C. elegans*, TLR signaling is required for the innate immune response against Gram-negative bacteria (45,46); this may explain why the original study (44) found *sams-1* RNAi worms to be exquisitely susceptible to infection by *P. aeruginosa*, a Gram-negative bacterium.

Although *sams-1* and *sbp-1* RNAi affected the transcriptome in similar ways, a small set of GSs were upregulated in *sams-1* RNAi but downregulated in *sbp-1* RNAi (Figure 4C-E; Supplementary Table S5). Most of these GSs pertain to lipid metabolism, recapitulating published findings that lipogenesis is elevated by *sams-1* deficiency but suppressed by *sbp-1* deficiency (42,47). Interestingly, the 1CC also follows this pattern, not only confirming a known negative feedback loop from *sbp-1* to the 1CC (42,47), but also indicating that SAM deficiency alone causes compensatory induction of the 1CC, in line with a recent study (48). Overall, this exemplifies the utility of eVITTA in revealing both

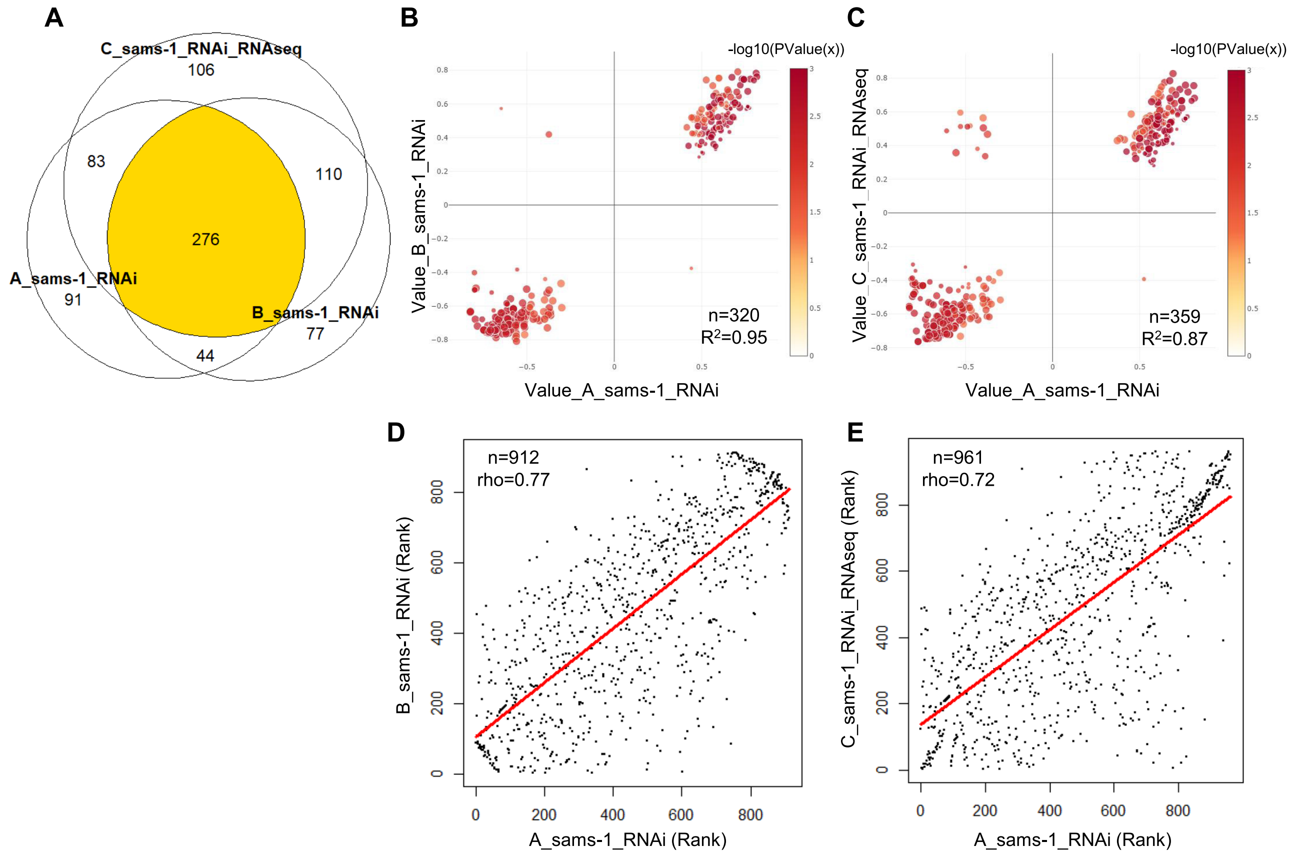


Figure 3. Comparison of three different transcriptome profiles of *C. elegans* worms with perturbations in *S*-adenosylmethionine synthesis (*sams-1*/*MAT1A* RNAi). The figure compares GSEA profiles of three independent experiments of *sams-1* RNAi treated worms: dataset A – microarray from GEO accession GSE70692; dataset B – microarray from GSE70693; and dataset C – RNA-seq from GSE121508. (**A**) The Venn diagram shows (in yellow) GSs that are significantly changed (pval < 0.05, padj < 0.25) in all three *sams-1* RNAi datasets. Each circle represents GSs that are significantly changed in the corresponding dataset. easyVizR parameters: for datasets A–C, pval< 0.05, padj < 0.25, intersection = true. (**B**) The 2D scatter plot shows GSs that are significantly regulated (pval< 0.05, padj < 0.25) in both datasets A and B. $n = 320$; correlation $R^2 = 0.95$. X-axis: ES in dataset A; Y-axis: ES in dataset B. (**C**) The 2D scatter plot shows GSs that are significantly regulated (pval< 0.05, padj < 0.25) in both datasets A and C. $n = 359$; correlation $R^2 = 0.87$. X-axis: ES in dataset A; Y-axis: ES in dataset C. (**D**) The rank scatter plot shows the Spearman correlation between unfiltered datasets A and B. X and Y axes: ranks of $-\log_{10}$-transformed *P*-values signed by ES ($-\log_{10}$(pval)*sign(ES)) in datasets A and B, respectively. $n = 912$; rho = 0.77; pval< 2.2e-16. (**E**) The rank scatter plot shows the Spearman correlation between unfiltered datasets A and C. X and Y axes: ranks of $-\log_{10}$-transformed *P*-values signed by ES ($-\log_{10}$(pval)*sign(ES)) in datasets A and C, respectively. $n = 961$; rho = 0.72; pval< 2.2e-16. ES, enrichment score, also displayed as 'Value'; GEO, Gene Expression Omnibus; GS, gene set; GSEA, pre-ranked gene set enrichment analysis; intersection, selected parameters in easyVizR '3.2 Intersection of Interest'; padj, adjusted *P*-value, also displayed as 'FDR'; pval, *P*-value, also displayed as 'PValue'; RA, Reactome Pathways; rho, Spearman's rank coefficient; RNA-seq, RNA sequencing.

convergent and differential patterns in multiple datasets at high resolution.

DISCUSSION

Assembling a dedicated analytical pipeline to interpret transcriptomes is a complex task with many challenges. eVITTA addresses these challenges by automating the query and analysis of NCBI GEO transcriptome data with a standardized pipeline (easyGEO), performing functional profiling with 100+ monthly-updated, species-specific GS libraries (easyGSEA), and providing a workflow for systematic comparison of expression patterns in multiple datasets (easyVizR). As illustrated in the evaluation studies, eVITTA's workflow and interactive visualizations enable efficient discovery of both broad and subtle changes in expression, which other tools were unable to fully capture.

Although we developed eVITTA for transcriptome analysis and interpretation, its tools can also be applied to other omics studies. For instance, easyGSEA can functionally characterize lists of genes or proteins generated from any omics platform, and easyVizR can handle any differential expression data with statistical significance (https://tau.cmmt.ubc.ca/eVITTA/#userguide).

Like all similar web servers, eVITTA has some limitations. easyGEO cannot handle datasets where count data are missing; it also relies on user-supplied count data, which may be processed using different methods and thus cannot be used for between-study comparisons. In addition, it does not yet support datasets deposited in ArrayExpress or the European Nucleotide Archive (ENA). Future iterations of eVITTA may include access to these resources

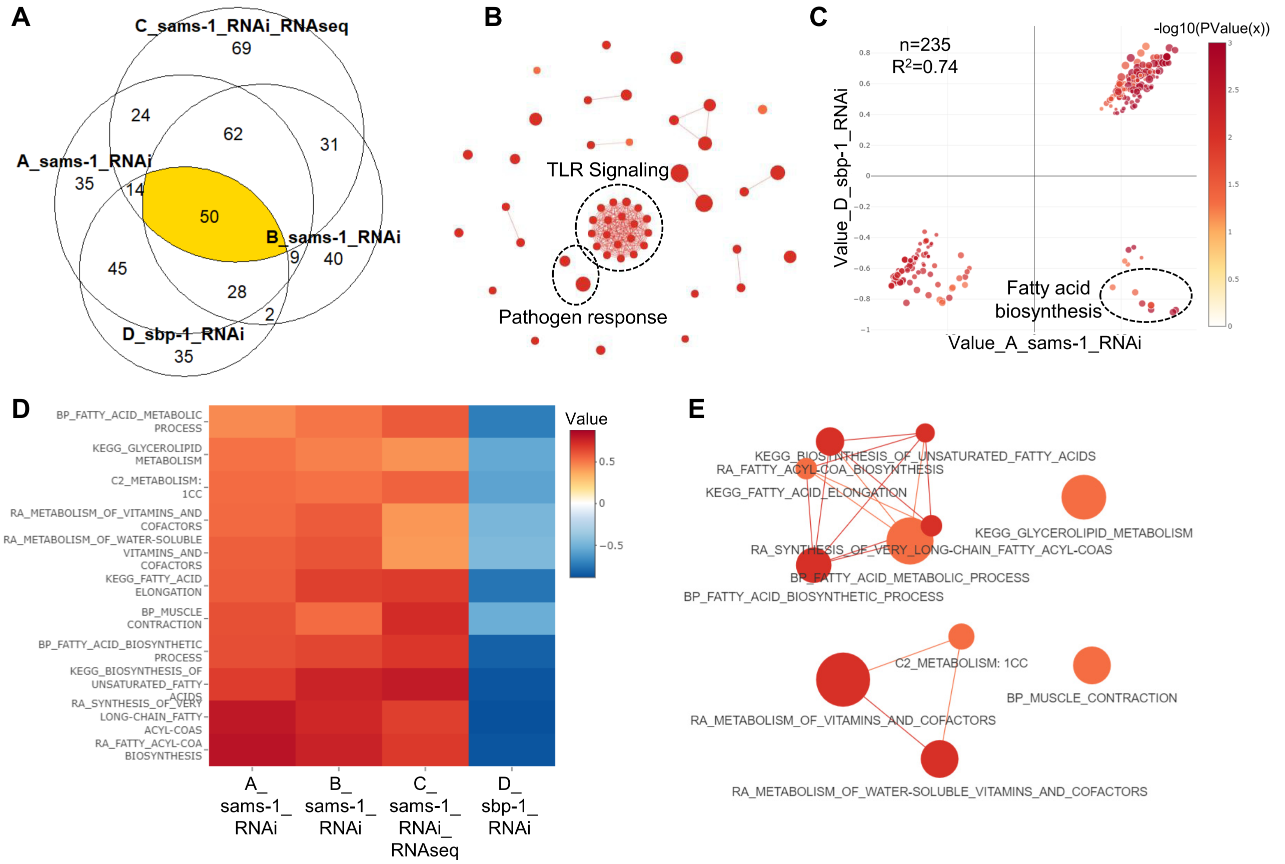


Figure 4. Evaluation study on *C. elegans* worms with perturbations in *S*-adenosylmethionine synthesis (*sams-1/MAT1A* RNAi) and lipogenesis (*sbp-1/SREBP* RNAi). The figure compares GSEA profiles of *sams-1* RNAi treated worms (datasets A–C in Figure 3) versus dataset D, microarray profile of *sbp-1* RNAi treated worms (GSE70692). (**A**) The Venn diagram shows (in yellow) GSs that are significantly upregulated (pval$<$ 0.05, padj $<$ 0.25, ES $>$ 0) in all four datasets. Each circle represents GSs that are significantly upregulated in the corresponding dataset. easyVizR parameters: for datasets A–D, pval$<$ 0.05, padj $<$ 0.25, sign = +, intersection = true. (**B**) The graph shows the enrichment network of 50 shared significantly upregulated GSs (from A). Node denotes GS; node size reflects the number of leading-edge genes in each GS. Edge reflects significant overlap of leading-edge genes in dataset A, as defined by a Jaccard coefficient larger than or equal to 0.25. (**C**) The 2-dimensional scatter plot shows the ESs of GSs that are significantly regulated (pval$<$ 0.05, padj $<$ 0.25) in both datasets A and D. $n = 235$; correlation $R^2 = 0.74$. A set of GSs that are regulated in opposite directions is found in quadrant four (bottom right); these mostly pertain to fatty acid biosynthesis. (**D**) The heatmap shows the 11 categories that are positively regulated in *sams-1* RNAi (pval$<$ 0.05, padj $<$ 0.25, ES $>$ 0 in datasets A, B and C) and negatively regulated in *sbp-1* RNAi (pval$<$ 0.05, padj $<$ 0.25, ES $<$ 0 in dataset D). Cell colors represent ES value, ordered by dataset A. Detailed statistics are provided in Supplementary Table S5. (**E**) The graph shows the enrichment network of 11 GSs from Figure 4D. Edges reflect significant overlap of leading-edge genes in dataset A; other parameters are as in B. 1CC, one-carbon cycle; GO, gene ontology; BP, biological process; C2, WormCat Category 2; ES, enrichment score, also displayed as 'Value'; GEO, Gene Expression Omnibus; GS, gene set; GSEA, pre-ranked gene set enrichment analysis; Intersection, selected parameters in easyVizR '3.2 Intersection of Interest'; KEGG, Kyoto Encyclopaedia of Genes and Genomes; padj, adjusted *P*-value, also displayed as 'FDR'; pval, P-value, also displayed as 'PValue';RA, Reactome Pathways; RNA-seq, RNA sequencing; TLR, Toll-like receptor.

and offer customizable DE analysis with limma (26), edgeR (25) and DESeq2 (49). The GSEA method in easyGSEA assumes genes in a GS change in one direction (either predominantly up- or down-regulated); methods to evaluate GSs regardless of the direction are to be incorporated (50). Future iterations of eVITTA may also adopt more effective weighing techniques in prioritizing GSs with high phenotype relevance, especially in the context of other omics data such as ChIP-seq and genomic mutation data (51). In easyVizR, most modules rely on comparisons between filtered gene lists; more options for unbiased comparisons, such as Spearman's correlation heatmap, may be included in the future. Future iterations of eVITTA may also address challenges in comparing transcriptomes across species (51,52). Lastly, future releases of eVITTA may provide a more seamless user experience by providing direct links between its three modules. Overall, eVITTA aims to both improve existing pipelines for omics data analysis and to make transcriptome interpretation more accessible to the wider research community.

DATA AVAILABILITY

eVITTA is free and open to all users and there is no login requirement (https://tau.cmmt.ubc.ca/eVITTA/). Its source code is available in the GitHub repository (https://github.com/easygsea/eVITTA.git).

SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online.

ACKNOWLEDGEMENTS

We thank Dr T.F. Chan (The Chinese University of Hong Kong, Shatin), Dr O. Fornes (The University of British Columbia, Vancouver, BC), Dr D. Huang (Tianjin Medical University, Tianjin), Dr H.S. Kwan (The Chinese University of Hong Kong, Shatin), Dr M.J. Li (Tianjin Medical University, Tianjin), Dr S. Mostafavi (The University of British Columbia, Vancouver, BC), Dr W. Nong (The Chinese University of Hong Kong, Shatin), Dr A.K. Walker (University of Massachusetts Medical School, Worcester, MA), and Taubert lab members for critical comments on the manuscript. Figure 1 and Graphical Abstract created with BioRender.com, Toronto, Canada.

FUNDING

Canadian Institutes of Health Research (CIHR) [PJT-153199, PJT-165988 to S.T.]; Natural Sciences and Engineering Research Council of Canada (NSERC) [RGPIN-2018-05133 to S.T.]; Cancer Research Society (CRS) [22727 to S.T.]; J.Y. was supported partly by University of British Columbia Edward Squires Memorial Fellowship; Y.L. was partly supported by BioTalent Canada. Funding for open access charge: Canadian Institutes of Health Research (CIHR) [PJT-153199].

Conflict of interest statement. None declared.

REFERENCES

1. Wang,Z., Gerstein,M. and Snyder,M. (2009) RNA-Seq: a revolutionary tool for transcriptomics. *Nat. Rev. Genet.*, **10**, 57–63.
2. Edgar,R., Domrachev,M. and Lash,A.E. (2002) Gene Expression Omnibus: NCBI gene expression and hybridization array data repository. *Nucleic Acids Res.*, **30**, 207–210.
3. Clough,E. and Barrett,T. (2016) The Gene Expression Omnibus database. *Methods Mol. Biol. Clifton NJ*, **1418**, 93–110.
4. Dumas,J., Gargano,M.A. and Dancik,G.M. (2016) shinyGEO: a web-based application for analyzing gene expression omnibus datasets. *Bioinforma. Oxf. Engl.*, **32**, 3679–3681.
5. Mahi,N.A., Najafabadi,M.F., Pilarczyk,M., Kouril,M. and Medvedovic,M. (2019) GREIN: an interactive web platform for re-analyzing GEO RNA-seq data. *Sci. Rep.*, **9**, 7580.
6. Creixell,P., Reimand,J., Haider,S., Wu,G., Shibata,T., Vazquez,M., Mustonen,V., Gonzalez-Perez,A., Pearson,J., Sander,C. *et al.* (2015) Pathway and network analysis of cancer genomes. *Nat. Methods*, **12**, 615–621.
7. Zhou,Y., Zhou,B., Pache,L., Chang,M., Khodabakhshi,A.H., Tanaseichuk,O., Benner,C. and Chanda,S.K. (2019) Metascape provides a biologist-oriented resource for the analysis of systems-level datasets. *Nat. Commun.*, **10**, 1523.
8. Chen,E.Y., Tan,C.M., Kou,Y., Duan,Q., Wang,Z., Meirelles,G.V., Clark,N.R. and Ma'ayan,A. (2013) Enrichr: interactive and collaborative HTML5 gene list enrichment analysis tool. *BMC Bioinformatics*, **14**, 128.
9. Huang,D.W., Sherman,B.T., Tan,Q., Collins,J.R., Alvord,W.G., Roayaei,J., Stephens,R., Baseler,M.W., Lane,H.C. and Lempicki,R.A. (2007) The DAVID gene functional classification tool: a novel biological module-centric algorithm to functionally analyze large gene lists. *Genome Biol.*, **8**, R183.
10. Raudvere,U., Kolberg,L., Kuzmin,I., Arak,T., Adler,P., Peterson,H. and Vilo,J. (2019) g:Profiler: a web server for functional enrichment analysis and conversions of gene lists (2019 update). *Nucleic Acids Res.*, **47**, W191–W198.
11. Alonso,R., Salavert,F., Garcia-Garcia,F., Carbonell-Caballero,J., Bleda,M., Garcia-Alonso,L., Sanchis-Juan,A., Perez-Gil,D., Marin-Garcia,P., Sanchez,R. *et al.* (2015) Babelomics 5.0: functional interpretation for new generations of genomic data. *Nucleic Acids Res.*, **43**, W117–W121.
12. Xie,C., Mao,X., Huang,J., Ding,Y., Wu,J., Dong,S., Kong,L., Gao,G., Li,C.-Y. and Wei,L. (2011) KOBAS 2.0: a web server for annotation and identification of enriched pathways and diseases. *Nucleic Acids Res.*, **39**, W316–W322.
13. Mi,H., Ebert,D., Muruganujan,A., Mills,C., Albou,L.-P., Mushayamaha,T. and Thomas,P.D. (2020) PANTHER version 16: a revised family classification, tree-based classification tool, enhancer regions and extensive API. *Nucleic Acids Res.*, **49**, D394–D403.
14. Liao,Y., Wang,J., Jaehnig,E.J., Shi,Z. and Zhang,B. (2019) WebGestalt 2019: gene set analysis toolkit with revamped UIs and APIs. *Nucleic Acids Res.*, **47**, W199–W205.
15. Wadi,L., Meyer,M., Weiser,J., Stein,L.D. and Reimand,J. (2016) Impact of outdated gene annotations on pathway enrichment analysis. *Nat. Methods*, **13**, 705–706.
16. 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.
17. Listopad,S.A. and Norden-Krichmar,T.M. (2019) A-Lister: a tool for analysis of differentially expressed omics entities across multiple pairwise comparisons. *BMC Bioinformatics*, **20**, 595.
18. Lam,F., Lalansingh,C.M., Babaran,H.E., Wang,Z., Prokopec,S.D., Fox,N.S. and Boutros,P.C. (2016) VennDiagramWeb: a web application for the generation of highly customizable Venn and Euler diagrams. *BMC Bioinformatics*, **17**, 401.
19. Khan,A. and Mathelier,A. (2017) Intervene: a tool for intersection and visualization of multiple gene or genomic region sets. *BMC Bioinformatics*, **18**, 287.
20. Heberle,H., Meirelles,G.V., da Silva,F.R., Telles,G.P. and Minghim,R. (2015) InteractiVenn: a web-based tool for the analysis of sets through Venn diagrams. *BMC Bioinformatics*, **16**, 169.
21. Lex,A., Gehlenborg,N., Strobelt,H., Vuillemot,R. and Pfister,H. (2014) UpSet: visualization of Intersecting Sets. *IEEE Trans. Vis. Comput. Graph.*, **20**, 1983–1992.
22. Rue-Albrecht,K., Marini,F., Soneson,C. and Lun,A.T.L. (2018) iSEE: Interactive SummarizedExperiment Explorer. *F1000Research*, **7**, 741.
23. Seo,M., Yoon,J. and Park,T. (2015) GRACOMICS: software for graphical comparison of multiple results with omics data. *BMC Genomics*, **16**, 256.
24. Davis,S. and Meltzer,P.S. (2007) GEOquery: a bridge between the Gene Expression Omnibus (GEO) and BioConductor. *Bioinformatics*, **23**, 1846–1847.
25. Robinson,M.D., McCarthy,D.J. and Smyth,G.K. (2010) edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics*, **26**, 139–140.
26. Ritchie,M.E., Phipson,B., Wu,D., Hu,Y., Law,C.W., Shi,W. and Smyth,G.K. (2015) limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic Acids Res.*, **43**, e47.
27. Wickham,H., Averick,M., Bryan,J., Chang,W., McGowan,L.D., François,R., Grolemund,G., Hayes,A., Henry,L., Hester,J. *et al.* (2019) Welcome to the Tidyverse. *J. Open Source Softw.*, **4**, 1686.
28. Korotkevich,G., Sukhov,V. and Sergushichev,A. (2019) Fast gene set enrichment analysis. bioRxiv doi: https://doi.org/10.1101/060012, 01 February 2021, preprint: not peer reviewed.
29. Kolberg,L., Raudvere,U., Kuzmin,I., Vilo,J. and Peterson,H. (2020) gprofiler2 – an R package for gene list functional enrichment analysis and namespace conversion toolset g:Profiler. *F1000Research*, **9**, ELIXIR–709.
30. Luo,W. and Brouwer,C. (2013) Pathview: an R/Bioconductor package for pathway-based data integration and visualization. *Bioinformatics*, **29**, 1830–1831.
31. Chen,H. and Boutros,P.C. (2011) VennDiagram: a package for the generation of highly-customizable Venn and Euler diagrams in R. *BMC Bioinformatics*, **12**, 35.
32. Conway,J.R., Lex,A. and Gehlenborg,N. (2017) UpSetR: an R package for the visualization of intersecting sets and their properties. *Bioinforma. Oxf. Engl.*, **33**, 2938–2940.
33. Plaisier,S.B., Taschereau,R., Wong,J.A. and Graeber,T.G. (2010) Rank–rank hypergeometric overlap: identification of statistically significant overlap between gene-expression signatures. *Nucleic Acids Res.*, **38**, e169.

34. Kanehisa,M., Furumichi,M., Tanabe,M., Sato,Y. and Morishima,K. (2017) KEGG: new perspectives on genomes, pathways, diseases and drugs. *Nucleic. Acids. Res.*, **45**, D353–D361.
35. 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.
36. Martens,M., Ammar,A., Riutta,A., Waagmeester,A., Slenter,D.N., Hanspers,K.A., Miller,R., Digles,D., Lopes,E.N., Ehrhart,F. *et al.* (2020) WikiPathways: connecting communities. *Nucleic Acids Res.*, **49**, D613–D621.
37. Lieberman,N.A.P., Peddu,V., Xie,H., Shrestha,L., Huang,M.-L., Mears,M.C., Cajimat,M.N., Bente,D.A., Shi,P.-Y., Bovier,F. *et al.* (2020) In vivo antiviral host transcriptional response to SARS-CoV-2 by viral load, sex, and age. *PLoS Biol.*, **18**, e3000849.
38. Jones,D.T. and Reed,R.R. (1989) Golf: an olfactory neuron specific-G protein involved in odorant signal transduction. *Science*, **244**, 790–795.
39. Trudeau,M.C. and Zagotta,W.N. (2003) Calcium/calmodulin modulation of olfactory and rod cyclic nucleotide-gated ion channels. *J. Biol. Chem.*, **278**, 18705–18708.
40. Mastrangelo,A., Bonato,M. and Cinque,P. (2021) Smell and taste disorders in COVID-19: from pathogenesis to clinical features and outcomes. *Neurosci. Lett.*, **748**, 135694.
41. Ye,Q., Zhou,J., Yang,G., Li,R.-T., He,Q., Zhang,Y., Wu,S.-J., Chen,Q., Shi,J.-H., Zhang,R.-R. *et al.* (2020) SARS-CoV-2 infection causes transient olfactory dysfunction in mice. bioRxiv doi: https://doi.org/10.1101/2020.11.10.376673, 10 November 2020, preprint: not peer reviewed.
42. Ding,W., Smulan,L.J., Hou,N.S., Taubert,S., Watts,J.L. and Walker,A.K. (2015) s-Adenosylmethionine levels govern innate immunity through distinct methylation-dependent pathways. *Cell Metab.*, **22**, 633–645.
43. Holdorf,A.D., Higgins,D.P., Hart,A.C., Boag,P.R., Pazour,G.J., Walhout,A.J.M. and Walker,A.K. (2020) WormCat: an online tool for annotation and visualization of *Caenorhabditis elegans* genome-scale data. *Genetics*, **214**, 279–294.
44. Ding,W., Higgins,D.P., Yadav,D.K., Godbole,A.A., Pukkila-Worley,R. and Walker,A.K. (2018) Stress-responsive and metabolic gene regulation are altered in low S-adenosylmethionine. *PLoS Genet.*, **14**, e1007812.
45. Tenor,J.L. and Aballay,A. (2008) A conserved Toll-like receptor is required for *Caenorhabditis elegans* innate immunity. *EMBO Rep.*, **9**, 103–109.
46. Brandt,J.P. and Ringstad,N. (2015) Toll-like receptor signaling promotes development and function of sensory neurons required for a *C. elegans* pathogen-avoidance behavior. *Curr. Biol. CB*, **25**, 2228–2237.
47. Walker,A.K., Jacobs,R.L., Watts,J.L., Rottiers,V., Jiang,K., Finnegan,D.M., Shioda,T., Hansen,M., Yang,F., Niebergall,L.J. *et al.* (2011) A conserved SREBP-1/phosphatidylcholine feedback circuit regulates lipogenesis in metazoans. *Cell*, **147**, 840–852.
48. Giese,G.E., Walker,M.D., Ponomarova,O., Zhang,H., Li,X., Minevich,G. and Walhout,A.J. (2020) *Caenorhabditis elegans* methionine/S-adenosylmethionine cycle activity is sensed and adjusted by a nuclear hormone receptor. *eLife*, **9**, e60259.
49. Love,M.I., Huber,W. and Anders,S. (2014) Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.*, **15**, 550.
50. Geistlinger,L., Csaba,G., Santarelli,M., Ramos,M., Schiffer,L., Turaga,N., Law,C., Davis,S., Carey,V., Morgan,M. *et al.* (2020) Toward a gold standard for benchmarking gene set enrichment analysis. *Brief. Bioinform.*, **22**, 545–556.
51. Paczkowska,M., Barenboim,J., Sintupisut,N., Fox,N.S., Zhu,H., Abd-Rabbo,D., Mee,M.W., Boutros,P.C. and Reimand,J. (2020) Integrative pathway enrichment analysis of multivariate omics data. *Nat. Commun.*, **11**, 735.
52. Fukushima,K. and Pollock,D.D. (2020) Amalgamated cross-species transcriptomes reveal organ-specific propensity in gene expression evolution. *Nat. Commun.*, **11**, 4459.