



## OPEN ACCESS

## EDITED BY

Stefano Tiozzo,  
Université Paris-Sorbonne, France

## REVIEWED BY

Maja Adamska,  
Australian National University, Australia  
Peter Ladurner,  
University of Innsbruck, Austria

## \*CORRESPONDENCE

Omri Wurtzel,  
✉ owurtzel@tauex.tau.ac.il

## SPECIALTY SECTION

This article was submitted to Evolutionary Developmental Biology, a section of the journal Frontiers in Cell and Developmental Biology

RECEIVED 22 January 2023

ACCEPTED 16 March 2023

PUBLISHED 23 March 2023

## CITATION

Hoffman M and Wurtzel O (2023),  
PLANAtools—An interactive gene  
expression repository for the planarian  
*Schmidtea mediterranea*.  
*Front. Cell Dev. Biol.* 11:1149537.  
doi: 10.3389/fcell.2023.1149537

## COPYRIGHT

© 2023 Hoffman and Wurtzel. This is an open-access article distributed under the terms of the [Creative Commons Attribution License \(CC BY\)](https://creativecommons.org/licenses/by/4.0/). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

# PLANAtools—An interactive gene expression repository for the planarian *Schmidtea mediterranea*

Michael Hoffman<sup>1</sup> and Omri Wurtzel<sup>1,2\*</sup>

<sup>1</sup>The George S. Wise Faculty of Life Sciences, School of Neurobiology, Biochemistry, and Biophysics, Tel Aviv University, Tel Aviv, Israel, <sup>2</sup>Sagol School of Neuroscience, Tel Aviv University, Tel Aviv, Israel

## KEYWORDS

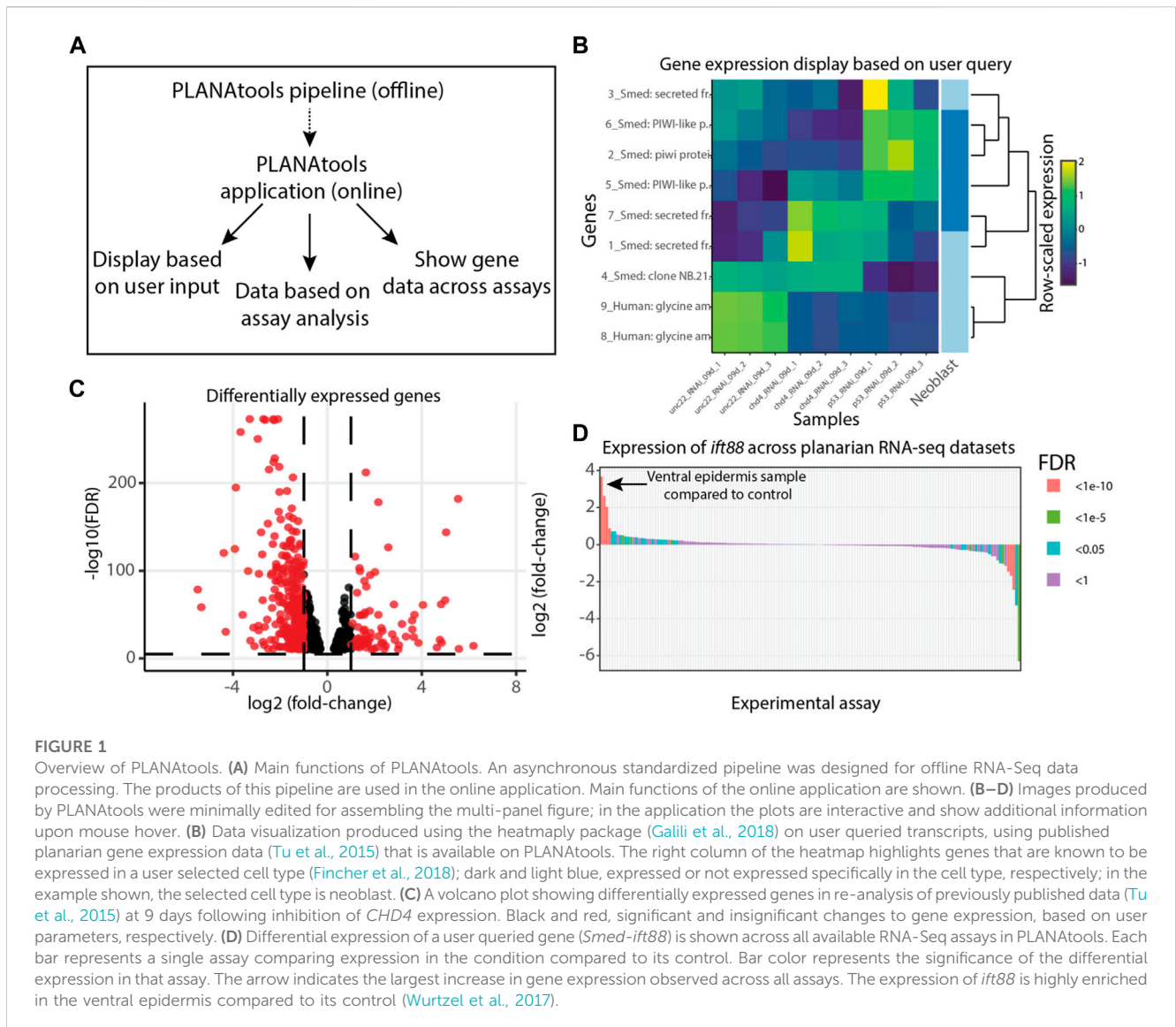
planarian, database, gene expression, regeneration, RNAseq, data visualization

## Introduction

Planarians are highly regenerative flatworms. They can regrow any missing organ by activating injury and regeneration-specific gene expression programs (Reddien, 2018). The planarian genome encodes over 20,000 genes (Rozanski et al., 2019), and therefore understanding the genetic basis for regeneration in this seemingly simple organism is a complex task. The use of RNA-sequencing (RNA-Seq) for gene expression analysis has transformed the study of planarian gene function, by facilitating the collection of thousands of data points in a single experiment. The accessibility of RNA-Seq technologies is demonstrated by the rapid growth in the number of RNA-Seq libraries from planarians, with over 3,450 planarian RNA BioSamples deposited to the publicly available Sequence Read Archive (Leinonen et al., 2011). The availability of transcriptional profiling has driven the discovery of key factors in planarian regeneration and stem cell (neoblast) biology. This includes the discovery of transcription factors associated with lineage selection and maturation (Forsthoefel et al., 2012; Tu et al., 2015), injury response genes (Wenemoser et al., 2012; Wurtzel et al., 2015), and regulators of neoblast differentiation (Lapan and Reddien, 2012; Zhu et al., 2015).

RNA-Seq data allows, in principle, highly reproducible research using common bioinformatic processing tools that are available to the scientific community. Compared to other research modalities, the format of high-throughput sequencing data is standardized, machine readable, and found in accessible repositories (Cock et al., 2010; Leinonen et al., 2011). However, despite these properties, using or comparing RNA-Seq data across projects and research groups presents several challenges. Data processing and analysis differ between scientists and teams, and continuously evolve based on the available tools, protocols, and technology. Moreover, RNA-Seq data is commonly found in raw fastq format, requiring computational proficiency and resources for even basic analysis. Even processed RNA-Seq data, often available as supplementary information in manuscripts, can be difficult to use. For example, different transcriptome assemblies are used in the planarian community for RNA-Seq data mapping, limiting comparisons of available data. Major planarian computational resources, such as PlanMine and Planosphere, have become instrumental for the research community by providing powerful tools for studying gene function and genome browsing using a gene or transcript focused interface (Rozanski et al., 2019; Nowotarski et al., 2021). Yet, a major strength of RNA-Seq data is the ability to examine changes in expression of multiple transcripts simultaneously.

Here, we curated published planarian RNA-Seq data and implemented a computational pipeline, based on broadly used RNA-Seq analysis tools, for analyzing gene expression. We developed a web-application, PLANAtools, that facilitates browsing, mining, downloading,



and visualizing planarian RNA-Seq data interactively, requiring minimal computational skills and resources from the user.

## Results

### Data retrieval and processing

A list of deposited *Schmidtea mediterranea* bulk RNA-Seq metadata was retrieved from the NCBI Sequence Read Archive (Leinonen et al., 2011). The list was curated manually and each library was associated with its controls and replicates. Each set of replicates and their controls was processed as a single biological experiment. Raw RNA-Seq libraries were retrieved from the SRA using the fasterq-dump tool (Leinonen et al., 2011) or by using bowtie2 --sra-acc parameter (Langmead and Salzberg, 2012). Raw RNA-Seq data was mapped to the planarian dd\_v6 transcriptome assembly (Rozanski et al., 2019) using bowtie2 with parameter --fast. Paired-end RNA-Seq data was mapped as single-end by considering

the first read in a read pair. The mapped data was transformed to Binary Alignment Map (BAM) format using SAMtools (Li et al., 2009). For each biological experiment (i.e., replicates and controls), a read count table was produced using featureCounts v2 with parameters [-M -s 0] (Liao et al., 2014). Data normalization and differential gene expression calling were then performed using DESeq2 using the rlog and DESeqDataSetFromMatrix functions (Love et al., 2014). Differential gene expression analysis was not performed on conditions where no biological replicates were available. Processed data tables were then indexed using the R data.table package and stored as binary files using the saveRDS function.

### Application implementation

PLANAtools is accessible from <https://wurtzellab.org/planatools>. It was developed in R using the Shiny framework and packaged to a docker image. The current release includes

gene expression analyses of 168 assays from over 40 manuscripts. Data browsing was implemented to accomplish four main uses (Figure 1): 1) Showing gene expression in published datasets as a heatmap or volcano plot based on transcripts that are selected by the users. Transcript selection is performed using free text search of transcript IDs or their putative annotation, as previously described (Dagan et al., 2022). 2) Showing differentially expressed genes of an assay as a heatmap, a volcano plot, or a table in the datasets that are available and analyzed with PLANAtools. 3) Presenting gene expression across datasets by comparing the expression of a single gene in control and treatment samples across assays. 4) Finally, to intersect a list of queried genes with a list of differentially expressed genes in an assay. Resulting visualizations are interactive online, and downloadable as vector images for further processing. The processed gene expression tables and differential gene expression analyses are also available for download.

## Usage

PLANAtools is a hub for gene expression analyses, and facilitates data visualization, browsing and acquisition. In addition, PLANAtools provides links to the original publication, to the raw deposited data in SRA (Leinonen et al., 2011), and to major organism-specific gene resources, such as FlyBase, PlanMine, and WormBase (Rozanski et al., 2019; Thurmond et al., 2019; Harris et al., 2020). The functionality of the website is showcased in a video tutorial that is found on the website demonstrating the major features of the application.

## Summary

PLANAtools provides non-experts access to planarian gene expression data, gene annotations, and external data sources. In contrast to most planarian resources, it is not gene- or transcript-centric. Instead, it allows an overview of gene expression using interactive visualizations and direct downloads of the original data. Therefore, it bridges the gaps in the ability to access, use, and re-use

planarian gene expression resources. Periodic updates to PLANAtools are planned, pursuant to the availability of newly deposited planarian gene expression datasets.

## Data availability statement

Publicly available datasets were analyzed in this study. This data can be found here: <https://wurtzellab.shinyapps.io/planatools/>.

## Author contributions

MH and OW designed and implemented the research. MH and OW wrote the manuscript.

## Funding

This work is supported by the European Research Council Horizon 2020 research and innovation programme (No. 853640) and the Israel Science Foundation (grant 2039/18). OW is a Zuckerman Faculty Scholar.

## Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

## Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

## References

- Cock, P. J. A., Fields, C. J., Goto, N., Heuer, M. L., and Rice, P. M. (2010). The Sanger FASTQ file format for sequences with quality scores, and the Solexa/Illumina FASTQ variants. *Nucleic Acids Res.* 38, 1767–1771. doi:10.1093/nar/gkp1137
- Dagan, Y., Yesharim, Y., Bonneau, A. R., Frankovits, T., Schwartz, S., Reddien, P. W., et al. (2022). m6A is required for resolving progenitor identity during planarian stem cell differentiation. *EMBO J.* 41, e109895. doi:10.15252/embj.2021109895
- Fincher, C. T., Wurtzel, O., De Hoog, T., Kravarik, K. M., and Reddien, P. W. (2018). Cell type transcriptome atlas for the planarian *Schmidtea mediterranea*. *Science* 360, eaaq1736. doi:10.1126/science.aaq1736
- Forsthoefel, D. J., James, N. P., Escobar, D. J., Stary, J. M., Vieira, A. P., Waters, F. A., et al. (2012). An RNAi screen reveals intestinal regulators of branching morphogenesis, differentiation, and stem cell proliferation in planarians. *Dev. Cell* 23, 691–704. doi:10.1016/j.devcel.2012.09.008
- Galili, T., O'Callaghan, A., Sidi, J., and Sievert, C. (2018). Heatmaply: an R package for creating interactive cluster heatmaps for online publishing. *Bioinformatics* 34, 1600–1602. doi:10.1093/bioinformatics/btx657
- Harris, T. W., Arnaboldi, V., Cain, S., Chan, J., Chen, W. J., Cho, J., et al. (2020). WormBase: A modern model organism information resource. *Nucleic Acids Res.* 48, D762–D767. doi:10.1093/nar/gkz920
- Langmead, B., and Salzberg, S. L. (2012). Fast gapped-read alignment with Bowtie 2. *Nat. Methods* 9, 357–359. doi:10.1038/nmeth.1923
- Lapan, S. W., and Reddien, P. W. (2012). Transcriptome analysis of the planarian eye identifies ovo as a specific regulator of eye regeneration. *Cell Rep.* 2, 294–307. doi:10.1016/j.celrep.2012.06.018
- Leinonen, R., Sugawara, H., and Shumway, M. (2011). The sequence read archive. *Nucleic Acids Res.* 39, D19–D21. doi:10.1093/nar/gkq1019
- Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., et al. (2009). The sequence alignment/map format and SAMtools. *Bioinformatics* 25, 2078–2079. doi:10.1093/bioinformatics/btp352
- Liao, Y., Smyth, G. K., and Shi, W. (2014). featureCounts: an efficient general purpose program for assigning sequence reads to genomic features. *Bioinformatics* 30, 923–930. doi:10.1093/bioinformatics/btt656

- 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. doi:10.1186/s13059-014-0550-8
- Nowotarski, S. H., Davies, E. L., Robb, S. M. C., Matentzoglou, N., Doddihal, V., et al. (2021). Planarian anatomy ontology: A resource to connect data within and across experimental platforms. *Development* 148, dev196097. doi:10.1242/dev.196097
- Reddien, P. W. (2018). The cellular and molecular basis for planarian regeneration. *Cell* 175, 327–345. doi:10.1016/j.cell.2018.09.021
- Rozanski, A., Moon, H., Brandl, H., Martin-Duran, J. M., Grohme, M. A., Huttner, K., et al. (2019). PlanMine 3.0-improvements to a mineable resource of flatworm biology and biodiversity. *Nucleic Acids Res.* 47, D812–D820. doi:10.1093/nar/gky1070
- Thurmond, J., Goodman, J. L., Strelets, V. B., Attrill, H., Gramates, L. S., Marygold, S. J., et al. (2019). FlyBase 2.0: The next generation. *Nucleic Acids Res.* 47, D759–D765. doi:10.1093/nar/gky1003
- Tu, K. C., Cheng, L. C., T K Vu, H., Lange, J. J., McKinney, S. A., Seidel, C. W., et al. (2015). Egr-5 is a post-mitotic regulator of planarian epidermal differentiation. *eLife* 4, e10501. doi:10.7554/eLife.10501
- Wenemoser, D., Lapan, S. W., Wilkinson, A. W., Bell, G. W., and Reddien, P. W. (2012). A molecular wound response program associated with regeneration initiation in planarians. *Genes Dev.* 26, 988–1002. doi:10.1101/gad.187377.112
- Wurtzel, O., Cote, L. E., Poirier, A., Satija, R., Regev, A., and Reddien, P. W. (2015). A generic and cell-type-specific wound response precedes regeneration in planarians. *Dev. Cell* 35, 632–645. doi:10.1016/j.devcel.2015.11.004
- Wurtzel, O., Oderberg, I. M., and Reddien, P. W. (2017). Planarian epidermal stem cells respond to positional cues to promote cell-type diversity. *Dev. Cell* 40, 491–504.e5. doi:10.1016/j.devcel.2017.02.008
- Zhu, S. J., Hallows, S. E., Currie, K. W., Xu, C., and Pearson, B. J. (2015). A mex3 homolog is required for differentiation during planarian stem cell lineage development. *eLife* 4, e07025. doi:10.7554/eLife.07025