doi:10.20944/preprints202201.0353.v1

*Type of the Paper (**Article**, Review, Communication, etc.)*

Special Issue "Plant Bioinformatics: Applications and Databases"

GXP: Analyze and plot plant omics data in web browsers

Constantin Eiteneuer[1,#], David Velasco[4,#], Joseph Atemia[2,#], Dan Wang[1], Andrea Schrader[5], Roland Pieruschka[1], Sven Fahrner[1], Rainer Schwacke[2], Vanessa Wahl[3], Ulrich Schurr[1], Björn Usadel[2] and Asis Hallab [2,*]

[1] Affiliation 1; IBG-2 Plant Sciences, Forschungszentrum Jülich, 52428 Jülich, Germany
[2] Affiliation 2; IBG-4 Bioinformatics, Forschungszentrum Jülich, 52428 Jülich, Germany
[3] Affiliation 3; Max Planck Institute for Molecular Plant Physiology, 14476 Potsdam, Germany
[4] Affiliation 4; Norges Teknisk-Naturvitenskapelige Universitet, 7034 Trondheim, Norway
[5] Affiliation 5; RWTH Aachen University, Institute for Biology I, 52062 Aachen, Germany
* Correspondence: AH a.hallab@fz-juelich.de
# These authors contributed equally

Abstract:

Next generation sequencing and metabolomics have become very cost and work efficient and are integrated into an ever growing number of life science research projects. Typically, well established software pipelines provide quantitative data informing about gene expression or concentrations of metabolites from the raw data. This data needs to be visualized and further analyzed in order to support scientific hypothesis building and identification of underlying biological patterns. Some tools exist, but require installation or manual programming. We developed "Gene Expression Plotter" (GXP), an RNA-Seq and Metabolomics data visualization and analysis tool entirely running in the user's web browsers, thus not needing any custom installation, manual programming or upload of confidential data to third party servers. GXP enables the user to generate interactive plots, visually summarize genetic or metabolic responses in scientific sketches (Mapman), carry out cluster and principal component analysis, and conduct overrepresentation analyses. GXP can be used to publish research data along with interactive plots and results of analyses carried out with it. GXP is freely available on GitHub: https://github.com/usadellab/GeneExpressionPlots

Keywords: RNA sequencing, metabolomics, data visualization, overrepresentation analysis, cluster analysis, principal component analysis, scientific plotting, Mapman, Mercator

1. Introduction

RNA sequencing is a popular and well established method to assess gene expression in varying conditions and organisms. Typical analysis pipelines apply quality filters and adapter trimming to raw reads [1–3] and quantify expression by aligning the reads back to a reference genome or transcriptome [4–7]. Usually, a subsequent step is to compare different biological and experimental conditions and infer the gene level response to the contrasting stimuli, i.e. find transcripts whose expression responds to the experimental stimulus and are significantly up- or downregulated, respectively. At the end of such a standard analysis, expression tables provide numeric expression values typically expressed as counts, counts per million (cpm) or transcripts per kilobase million (TPM) and contrast tables inform about differential gene expression in terms of significance (i.e. *p*-Values). Given the commodity nature of short read RNA sequencing, such RNA-seq experiments often sample many genotypes and/or conditions exposed to a variety of experimental conditions producing a large number of tabular result files. A similar data structure occurs in metabolomics which is well established in the plant field and detailed guidelines exist [8].

After these analyses, the expression of selected genes or metabolites of interest is plotted, highlighting the biological response varying in different samples. On a more general level, genotypes or experimental conditions are compared by correlating their measured gene expression or metabolite accumulation, which is visualized in heatmaps and hierarchical clustering trees. Alternatively, the similarity of the genotype-specific response measured for the different experimental conditions or genotypes is assessed and plotted using principal component analyses. Typically, the response is not only described by lists of up- or down-regulated individual genes and metabolites, but additionally by the biological processes and pathways in which these genes and metabolites are involved. Visualization of such contexts helps to summarize transcriptional responses, e.g. by plotting gene expression into metabolic cellular sketches [9–12]. The molecular and metabolic responses to a stimulus are often assessed by the identification of over-represented functional annotations of selected genes compared to a suitable background, e.g. the whole genome or the whole set of expressed genes.

In order to enable scientific meaningful interpretation of these comprehensive intermediate results, these need to be summarized and visualized in scientific plots. Multiple software applications provide such functionality [9,12,13],[14–17]. However, these tools, at least partially, require the installation of a software [9,10,14–17], they need to submit potentially confidential data to servers [12,13,16] or they provide summarized sets of public data [13,18], respectively.
These observations inspired us to develop a simple browser based suite that enables the user to (i) visualize their own quantitative data, (ii) assess the transcriptional response to experimental conditions, and (iii) identify over-represented gene functions. Our new tool “Gene-Expression-Plotter” (GXP) is freely available online (https://github.com/usadellab/GeneExpressionPlots/), it does not require the user to install any software or manually program plots, and, importantly, does keep all data local on the computer of the user - no data is sent to any server. While GXP can be used to analyze and visualize any quantitative data, it was mainly developed for transcriptomics and metabolomics data. Finally, the output of GXP can be used for publication. A user can save a current work-session and export all data, generated plots, and analysis results into a single file, a GXP database, that can be shared with other users or uploaded to public servers like GitHub. By loading a GXP Database into GXP, the data, plots, and analysis results are restored and accessible. An option exists for advanced users to publish a custom version of GXP that includes research data, plots, and analysis results on their own website as well.
Thus our new tool Gene Expression Plotter enables the end-user to visualize and analyze quantitative data, typically taken from RNA-Seq or metabolomics analyses, identify similarity between experimental conditions by cluster and principal component analysis, generate visual summaries of genetic or metabolic responses, identify overrepresented transcripts or metabolite characteristics, and even use GXP to publish the data along with plots and analysis results.

2. Results

2.1 Handling input and output data

With the aim of providing a single suite in which a user can visualize and analyze quantitative data and also share and publish the output, we programmed “Gene-Expression-Plotter” (GXP). GXP is available freely with a GPL license on GitHub: https://github.com/usadellab/GeneExpressionPlots. GXP is provided with a comprehensive online documentation and a manual (see submenu “Docs”). To use GXP, the user is first asked to import quantitative RNA-seq data in a tabular format. Each row should represent a single transcript measured as presented by its unique gene identifier in the first column. All following columns should contain the transcript counts for each genotype, replicate or treatment as specified by the column names in the first line (figure 1.A). GXP is made aware of the factors studied in the experiments, in terms of simple column names in the count table. In doing so the user can for example specify the time after an experimental treatment at which a sample was obtained. Such a factor is then used by GXP to position points on the x-axis in plots. We dub such a factor used to position points on the x-axis "x-axis factors". Note, that a x-axis factor joins biological replicates into a bin. Such x-axis bins are used to calculate y-axis error bars for points on the axis several joined replicates are represented with. Another type of factors can optionally be introduced and used to compare for example genotypes or different treatments, e.g. a mutant versus wild-type. We dub such factors that group several biological conditions (x-axis factor values) spread out over different points on the x-axis "group factors". A user can specify as many group factors, as were investigated in a respective research project. As said, GXP is made aware of these factors simply through the column names of the counts table input.

Optionally, GXP can use any extra information associated with specific transcripts. This is done by loading a separate table in which each row corresponds to a single transcript, and each column holds the additional information (figure 1.B). Examples of additional information include functional annotations of proteins, e.g. terms from the Mapman4 framework, [19,20], from the Gene Ontology (GO) project [21] or from KEGG pathways [22], and differential expression between contrasted conditions, e.g. heat shock treatment versus control conditions. Information about differential expression can be provided in the form of logarithmic fold change and/or adjusted p-values. The user can provide any information that group genes of interest or contain data about specific traits these genes have. These can be numerical values, e.g. chemical properties of metabolites, or categorical annotations about the protein function, like the terms used by Mapman4, GO or KEGG. These data are displayed in the form of a tabular column and can be used for example, in subsequent over-representation analyses.

2.1.1 Browsing and searching gene information

After having loaded the input data, the user can browse gene information in a searchable interface and make this information available to all who have access (figure 1.C). The presented information includes the respective count data extracted from the "counts table" (see section 2.1) and any further information about the respective transcripts or metabolites extracted from the "information table" (see section 2.1).

2.1.2 Saving work and exporting data

At any stage of using GXP, the user can export and save all imported data, plots, and analyses by using the dedicated "Export GXP Database" functionality (figure 1.D). The generated database can be used later to resume previous work or share plots and results. The exported GXP database contains a configuration file that can be used to change the order of "x-axis-factors", the unit of the expression values, and the various field separators used to load the tables into the application. At any stage all data is strictly kept on the user's computer and at no point is user data sent through the web.

2.2 Visualizing quantitative data

The user can generate interactive plots showing the expression of individually selected genes. Available visualizations are bar- (figure 2.A) and line-plots (figure 2.B). In these plots, "x-axis factors" (see section 2.1) define the position of points on the x-axis. If the user has provided "group factor" information (see section 2.1) that further groups replicates, e.g. genotype or different treatments, this information will be visualized by the color of plotted values (figure 2). Group factor values can either be plotted side by side (figure 2.A and 2.B) or as stacked lines (figure 2.C). The user can also visualize the expression of multiple genes/metabolites in the same interactive plot using all mentioned types: bar, single, or stacked lines (figure 2.D for an example with bars). In this case, the color distinguishes genes while the group factors are shown in their separate panels side by side. All plots are interactive. Hovering with the mouse over a specific point will display the underlying plotted data corresponding to that point, or bar, respectively. All plots a user generates can be saved and downloaded in high quality scalable vector graphics and used for publication purposes. Furthermore, exporting the GXP data and state optionally saves generated plots and analyses as well (see section 2.1).

2.3 Assessing similarity of gene expression between biological replicates

As explained in the introduction, to distinguish within and between group differences and enable drawing significant conclusions in terms of response when contrasting experimental conditions, it is mandatory to include biological replicates. The similarity of these replicates and thus also of gene expression/metabolite accumulation between experimental conditions can be assessed and visualized using correlation based hierarchical clustering and principal component analysis.

2.3.1 Hierarchical cluster analysis

Results of a hierarchical cluster analysis are visualized in a heatmap whose axis is accompanied with a tree dendrogram representing the hierarchical clusters, samples have been grouped into (figure 3.A). A transposed cluster analysis can also be carried out, in which all or a selected subset of transcripts/metabolites are grouped by similarity of their respective quantitative data. GXP computes and visualizes hierarchical cluster analysis on demand in the user’s browser. The user can either select correlation or euclidean distance between replicates’ gene expression vectors as a basis for the clustering analysis. The potentially demanding analysis is carried out in the background, using web-workers, and therefore does not block the user interface. While an analysis runs, a “loading” plot is immediately created, indicating ongoing calculation. Once complete, the plot color indicates either the correlation values or euclidean distances between respective replicates, depending on the original user's choice. A color scale is provided. The plot is interactive, hovering with the mouse over a heatmap cell will display the calculated likeness value assessed for the respective pair the cell corresponds to.

2.3.2 Principal component analysis

Results of principal component analysis (PCA) are typically visualized as a subclass of scatter-plots where the two axes represent the two principal components that explain most of the observed variance between samples. GXP carries out a PCA on user demand and runs the respective calculation in the user’s browser and in the background, thus not blocking the user interface. Once the calculation is finished, the loading icon indicating ongoing processing disappears and the respective scatter plot becomes visible (figure 3.B). Each gene expression vector of a biological replicate is color coded. Hovering with the mouse over a data point will display the name of the underlying biological replicate and the values of the two visualized principal components.

2.4 MapMan web browser plots

The Mapman frameworks (Mapman4 [20] and the older version Mapman v.3.6 [23]) comprise a manually curated vocabulary to describe the function of land plant proteins. Mercator and Mercator4 [19,20] are efficient and accurate genome scale annotation pipelines that assign the descriptions of Mapman v.3.6 and of Mapman4, respectively, to query proteins or transcripts. The desktop application MapMan [9,19,20,23]) has been developed to visualize annotations of the Mapman frameworks in the context of gene

doi:10.20944/preprints202201.0353.v1

expression data. Based on a proof of principle code [9,24] we also developed a simple MapMan web browser application. Like in the MapMan desktop application, a user can choose one of several basic metabolic cellular sketches, e.g. "Metabolism overview", "Photosynthesis", or "Cellular respiration" (figure 4). In these sketches small squares represent proteins or transcripts with functional annotations semantically corresponding to that region in the diagram. For example, all proteins with functions related to the Calvin cycle would appear as small boxes in the upper left area of the "Photosynthesis" sketch. The user chooses either a group of expression values for biological replicates belonging to the same experimental condition (x-axis factor; see section 2.1), or any arbitrary numerical information provided in the optional information table. This can be the logarithmic expression fold change as typically shown in pathway diagrams comparing two experimental conditions, e.g. control versus heat shock. Instead of using logarithmic fold change values as a measure of the intensity of a transcriptional response under different experimental conditions, the user can also choose adjusted *p*-values produced by differential gene expression analyses. Finally, the user can choose how the color gradient is dispersed over the selected numerical values. The choice is between a divergent scale from a fixed negative value to the positive counterpart (as in the MapMan desktop application which focuses on log fold change data), or a continuous scale ranging from zero, or the first quartile to the third quartile. The MapMan web browser plots are interactive, hovering with the mouse over a specific box displays the gene identifier, the Mapman4 protein description , and the numerical information assigned to the gene .

2.5 Over-representation (enrichment) analysis

To qualitatively describe a biological response, often, annotations about the biological process and the molecular function are analyzed. Those annotations found to be significantly over-represented among selected genes of interest characterize that group of genes. The selection criteria can be a significant up- or down-regulation contrasting two experimental conditions, e.g. control versus heat shock treatment. Typically, Fisher's exact test is used to determine whether the number of annotations within the selected genes significantly deviates from the number of annotations found within the background, i.e. the whole genome. GXP offers the user an easy way to carry out such overrepresentation analysis (ORAs). The user specifies a criterion shared by all selected genes and annotation terms for which overrepresentation should be tested. Alternatively, the user can select the genes of interest manually by entering gene identifiers. It is noteworthy that GXP is agnostic to the underlying data structure, consequently the user can use any information originally loaded with the "information table" (section 2.1) and in principle could do this for metabolite data as well, even though these analyses are less common for targeted metabolomics. Thus, a great variety of enrichment analyses can be carried out. Each single calculation of Fisher's exact test produces a corresponding single p-value, i.e. one *p*-value for each tested annotation. Currently, these *p*-values are corrected for multiple hypothesis testing using Bonferroni's method. All calculations are carried out in the background, so that user experience is not interrupted. The final result is a table in which for each tested annotation the corresponding adjusted *p*-value is shown (figure 5). These results can be exported along with the data and plots by clicking the "Export GXP Database" button in the "Data" sub-menu (see section 2.1).

2.6 Usage of GXP to publish data along with plots and analysis results

As previously mentioned, a user can save a GXP work-session by exporting all data, plots, and analysis results into a downloadable "GXP Database". Such a GXP Database file can be made publicly available, e.g. by uploading it to a web-server like GitHub . A link to this database can be included in scientific articles, and readers can download the published GXP Database, load it into GXP and explore the data, plots, and analysis results restored from the original work-session.

Another, more luxurious mode of publishing data, plots, and analysis results along with GXP exists. A user can deploy GXP together with an exported database to a dedicated web-server. GitHub-pages offers this option free of charge. This creates a link that can be cited in upcoming publications.

Thus, by using either of the above two methods, not only the raw expression counts and differential expression analysis result data could be provided, but also pre-generated supplemental interactive plots highlighting the scientific results could be discussed in publications. This makes the data free to be explored by third parties in their own context of interest, possibly reaching beyond the scope of the publication citing the installed custom version of GXP. The GXP manual ("Docs" menu) explains how to deploy custom versions of GXP in a few simple steps e.g. on GitHub.

3. Discussion

The availability, efficiency, and relatively low cost of next generation sequencing and metabolomics technologies allow their application in a wide variety of plant science research projects. Quantification of results and contrasting these quantifications between different experimental conditions is implemented in many standard pipelines. However, the need for simple visualization, summarization, and further selected analyses revealing similarity between biological replicates and overrepresented molecular functions in sets of selected transcripts or metabolites is the key for a biological interpretation of these datasets. Consequently, platforms have been developed to provide the user with tools to generate scientific plots and carry out clustering, principal component, and over-representation analysis [14–17]. However, these tools require either a manual installation of software, or sending data via the web to dedicated servers. . Spreadsheet applications are often used to fill this gap, but the resulting plots are not interactive, and these applications do not allow clustering, principal component or overexpression analyses. Alternatively, tools like MeV [25] allow an even deeper analysis but require installation. For this reason also MeV has been ported to the web as WebMeV [25,26] which even allows to run complex workflows, but it requires a login and to send the data to a proprietary server for analysis and storage.

Gene Expression Plotter (GXP) does all calculations locally in the browser without any need to submit data to servers. Hence, calculations in GXP are made efficient by usage of the web worker technology that enables running background processes in parallel and GXP offers simple solutions to explore data and to generate high quality plots.

4. Materials and Methods

Gene-Expression-Plotter (GXP) was implemented in TypeScript (https://www.typescriptlang.org/) as a stand alone application (single page application) running in the web-browser. The ReactJS (https://reactjs.org/) and Chakra UI (https://chakra-ui.com/) libraries were used to build the user interface. The ViteJS library was used for tooling. GXP can be accessed on GitHub pages (https://usadellab.github.io/GeneExpressionPlots); everytime a new version is pushed to GitHub, the new code is compiled and automatically deployed to GitHub pages. GXP's source code is freely available on GitHub (https://github.com/usadellab/GeneExpressionPlots) under a GPL-3 license.

Example data has been generated manually based on random Gaussian distributions to highlight different aspects in the plots and analyses.

4.1 Input and output data

doi:10.20944/preprints202201.0353.v1

Expression or metabolite data, and additional information about transcripts or metabolites, can be loaded into GXP (see section 2.1). Alternatively, a previous work-session can be restored using the “Import GXP Database” function in the “Data” menu (see section 2.1). All data is stored in memory, no data is ever sent via the internet to any backend-server. State management has been implemented using the MobX library (https://mobx.js.org).

4.2 Gene expression plots

All introduced plots (see section 2.2) are implemented with the plotly.js v2.6.3 Javascript library (https://plotly.com/javascript/). Plot data and definitions are stored in memory and thus can be exported to and restored from GXP Databases (see section 2.1.2).

4.3 Hierarchical cluster analysis

Similarity between biological replicates is either assessed using correlation or euclidean distance between the respective gene expression vectors (see section 2.3.1). In this, correlation values $c_{i,k}$ are transformed to distance values $d_{i,k}$ as follows:

$$d_{i,k} = 1 - \mathrm{abs}(\, c_{i,k} \,)$$

Thus, complete anti-correlation as well as complete correlation are interpreted as maximum likeliness of gene expression vectors.

Euclidean distance measures are computed with the ml-distance v3.0.0 (https://github.com/mljs) Javascript library. And hierarchical clusters are identified using the ml-hclust v3.1.0 library (https://github.com/mljs). The heatmap and the respective dendrogram visualizing the results of hierarchical clustering are plotted with the visx v2.4.0 (https://github.com/airbnb/visx) library that incorporates the popular and well proven D3 library (https://d3js.org/) into React.js.

4.4 Principal component analysis (PCA)

In GXP, a PCA can be carried out to identify and visualize likeliness between gene expression or metabolite concentrations of biological replicates (see section 2.3.2). The principal component analysis is computed with the help of the ml-pca v4.0.2 library (https://github.com/mljs). The respective plot visualizing the first two principal components contributing most to the observed differences is created with plotly.js v2.6.3 (https://plotly.com/javascript/).

4.5 MapMan visualizations

GXP offers to summarize genetic expression or responses to contrasting experimental conditions in the form of MapMan plots [9] (see section 2.4). All available canvas sketches (version X4.3) upon which to draw boxes to represent transcripts’ values were downloaded from the “MapMan Store” (https://mapman.gabipd.org/mapmanstore) and included in the GXP package. Based on the proof of concept implementation [24] the visualization was programmed with the D3 library v7.1.1 (https://github.com/d3/d3).

4.6 Over-representation analysis

Gene Expression Plotter offers the user to define sets of transcripts (genes) of interest, either by stating the respective identifiers one by one or by defining a selection criterion (see section 2.5). Subsequently annotations assigned to the selected genes are tested for being over-presented in comparison to the background, which is the whole information table (see section 2.1). Each of these tests is carried out as Fisher’s exact test resulting in a single *p*-value indicating how likely the observed annotation numbers can be explained by the null hypothesis, i.e. variations of the background annotations. In Fisher’s exact test contingency tables are created and *p*-values calculated using the hypergeometrical probability distribution (HGD) [27]. The calculation of specific HGD *p*-values is carried out with the GNU scientific library [28] which was compiled to web-assembly (https://webassembly.org/) for usage in the web-browser with Javascript. This compilation was done with emscripten (https://emscripten.org/). To calculate the likelihood of the alternative hypothesis that the observed numbers of annotations are not just as is in the contingency table but potentially greater, i.e. more extreme, more contingency tables of more extreme distributions need to be created and tested. Resulting *p*-values are summed up until no more extreme contingency tables can be generated, i.e. the respective cells contain zero. This procedure has been implemented manually in Javascript.

Author Contributions: A. H. and B. U. conceptualized the project. A. H. headed the software-design of the Gene Expression Plotter (GXP) browser based application. J.A., C. E., D. V., and D. W. (in alphabetical order) programmed the application-software and carried out software tests. C.E. and D.V. designed the graphical interface. B.U. provided scientific guidance, especially of methods to be applied, and feedback particularly about the user interface. B.U, J. A. and A.H. wrote, reviewed, and edited the article. B.U. and R.S. implemented a proof of concept Javascript software implementation of the MapMan visualizations [24]. A.S. provided extensive review in the form of testing and editing of the GXP user manual, and delivered detailed user feedback. R.P., S.F. and U.S. provided project administration, integration and application of GXP into ongoing scientific research projects and user feedback. A.H., J.A. and B.U wrote the manuscript with the help of all authors.

Funding: This research was funded by the German Ministry of Education and Research BMBF, grant number BreedPath 031B0890B and the European Commission for the project EPPN2020 under grant agreement agreement No 731013, the project EOSC-Life under the grant agreement No 824087, and the project EMPHASIS-PREP under the grant agreement No 739514.

Data Availability Statement: All code and test data is available at https://github.com/usadellab/GeneExpressionPlots. Note that the test data has been generated based on random Gaussian distributions to enable exploring the various functions and plots of GXP (see Materials and Methods section) except for Figure 4A which used data from [29].

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Bolger AM, Lohse M, Usadel B. Trimmomatic: a flexible trimmer for Illumina sequence data. Bioinformatics. 2014;30: 2114–2120.

2. Andrews S. FastQC: A Quality Control Tool for High Throughput Sequence Data. 2010. Available: http://www.bioinformatics.babraham.ac.uk/projects/fastqc/

3. Ewels P, Magnusson M, Lundin S, Käller M. MultiQC: summarize analysis results for multiple tools and samples in a single report. Bioinformatics. 2016. pp. 3047–3048. doi:10.1093/bioinformatics/btw354

4. Bray NL, Pimentel H, Melsted P, Pachter L. Near-optimal probabilistic RNA-seq quantification. Nat Biotechnol. 2016;34: 525–527.

5. Patro R, Duggal G, Love MI, Irizarry RA, Kingsford C. Salmon provides fast and bias-aware quantification of transcript expression. Nat Methods. 2017;14: 417–419.

6. Langmead B, Trapnell C, Pop M, Salzberg SL. Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. Genome Biol. 2009;10: R25.

7. Kim D, Langmead B, Salzberg SL. HISAT: a fast spliced aligner with low memory requirements. Nat Methods. 2015;12: 357–360.

8. Alseekh S, Aharoni A, Brotman Y, Contrepois K, D'Auria J, Ewald J, et al. Mass spectrometry-based metabolomics: a guide for annotation, quantification and best reporting practices. Nat Methods. 2021;18: 747–756.

9. Bolger M, Schwacke R, Usadel B. MapMan Visualization of RNA-Seq Data Using Mercator4 Functional Annotations. Methods Mol Biol. 2021;2354: 195–212.

10. Usadel B, Poree F, Nagel A, Lohse M, Czedik-Eysenberg A, Stitt M. A guide to using MapMan to visualize and compare Omics data in plants: a case study in the crop species, Maize. Plant Cell Environ. 2009;32: 1211–1229.

11. Hernández-de-Diego R, Tarazona S, Martínez-Mira C, Balzano-Nogueira L, Furió-Tarí P, Pappas GJ Jr, et al. PaintOmics 3: a web resource for the pathway analysis and visualization of multi-omics data. Nucleic Acids Res. 2018;46: W503–W509.

12. Naithani S, Gupta P, Preece J, D'Eustachio P, Elser JL, Garg P, et al. Plant Reactome: a knowledgebase and resource for comparative pathway analysis. Nucleic Acids Res. 2020;48: D1093–D1103.

13. Waese J, Fan J, Pasha A, Yu H, Fucile G, Shi R, et al. ePlant: Visualizing and Exploring Multiple Levels of Data for Hypothesis Generation in Plant Biology. Plant Cell. 2017;29: 1806–1821.

14. Lohse M, Bolger AM, Nagel A, Fernie AR, Lunn JE, Stitt M, et al. RobiNA: a user-friendly, integrated software solution for RNA-Seq-based transcriptomics. Nucleic Acids Res. 2012;40: W622–7.

15. Pimentel H, Bray NL, Puente S, Melsted P, Pachter L. Differential analysis of RNA-seq incorporating quantification uncertainty. Nat Methods. 2017;14: 687–690.

16. Julkowska MM, Saade S, Agarwal G, Gao G, Pailles Y, Morton M, et al. MVApp—Multivariate Analysis Application for Streamlined Data Analysis and Curation. Plant Physiology. 2019. pp. 1261–1276. doi:10.1104/pp.19.00235

17. Su S, Law CW, Ah-Cann C, Asselin-Labat M-L, Blewitt ME, Ritchie ME. Glimma: interactive graphics for gene expression analysis. Bioinformatics. 2017;33: 2050–2052.

18. Winter D, Vinegar B, Nahal H, Ammar R, Wilson GV, Provart NJ. An "Electronic Fluorescent Pictograph" browser for exploring and analyzing large-scale biological data sets. PLoS One. 2007;2: e718.

19. Lohse M, Nagel A, Herter T, May P, Schroda M, Zrenner R, et al. Mercator: a fast and simple web server for genome scale functional annotation of plant sequence data. Plant Cell Environ. 2014;37: 1250–1258.

20. Schwacke R, Ponce-Soto GY, Krause K, Bolger AM, Arsova B, Hallab A, et al. MapMan4: A Refined Protein Classification and Annotation Framework Applicable to Multi-Omics Data Analysis. Mol Plant. 2019;12: 879–892.

21. Ashburner M, Ball CA, Blake JA, Botstein D, Butler H, Cherry JM, et al. Gene ontology: tool for the unification of biology. The Gene Ontology Consortium. Nat Genet. 2000;25: 25–29.

22. Goto MKA. KEGG: Kyoto Encyclopedia of Genes and Genomes. Nucleic Acids Res. 2000.

23. Thimm O, Bläsing O, Gibon Y, Nagel A, Meyer S, Krüger P, et al. MAPMAN: a user-driven tool to display genomics data sets onto diagrams of metabolic pathways and other biological processes. Plant J. 2004;37: 914–939.

24. Usadel B. MapManJS - pure web implementations of MapMan. 2018. Available: https://github.com/usadellab/MapManJS

25. Howe E, Holton K, Nair S, Schlauch D, Sinha R, Quackenbush J. MeV: MultiExperiment Viewer. Biomedical Informatics for Cancer Research. 2010. pp. 267–277. doi:10.1007/978-1-4419-5714-6_15

26. Howe E, Holton K, Nair S, Schlauch D, Sinha R, Quackenbush J. WebMeV: MultiExperiment Viewer. [cited 1 Mar 2022]. Available: https://webmev.tm4.org/#/about

27. Fisher RA. On the Interpretation of χ 2 from Contingency Tables, and the Calculation of P. Journal of the Royal Statistical Society. 1922. p. 87. doi:10.2307/2340521

28. The Gnu Scientific Library Team. Gnu Scientific Library 2.0. Samurai Media Limited; 2015.

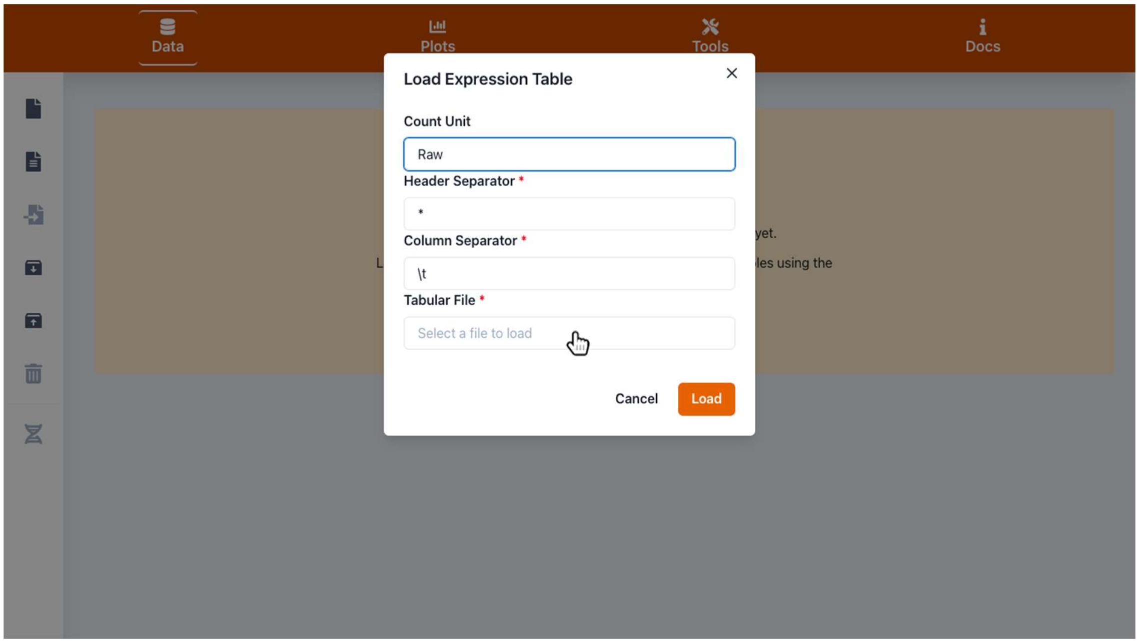
doi:10.20944/preprints202201.0353.v1

29. Reimer JJ, Thiele B, Biermann RT, Junker-Frohn LV, Wiese-Klinkenberg A, Usadel B, et al. Tomato leaves under stress: a comparison of stress response to mild abiotic stress between a cultivated and a wild tomato species. Plant Mol Biol. 2021;107: 177–206.

Figures

Figure 1 link

a)



Load Expression Table

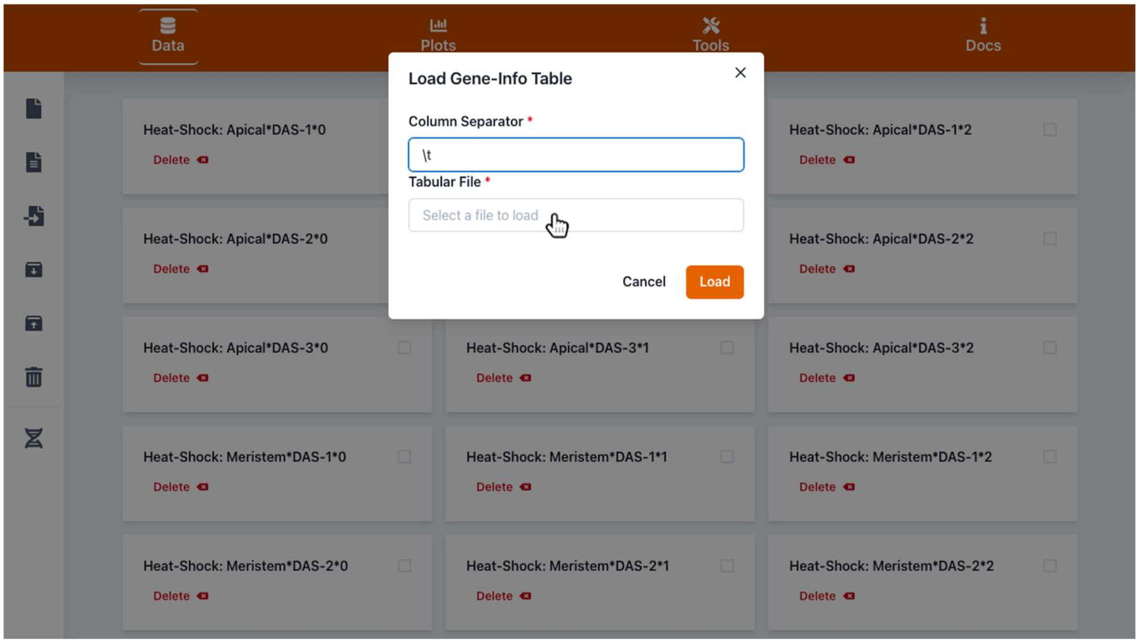
Count Unit
Raw

Header Separator *
*

Column Separator *
\t

Tabular File *
Select a file to load

b)



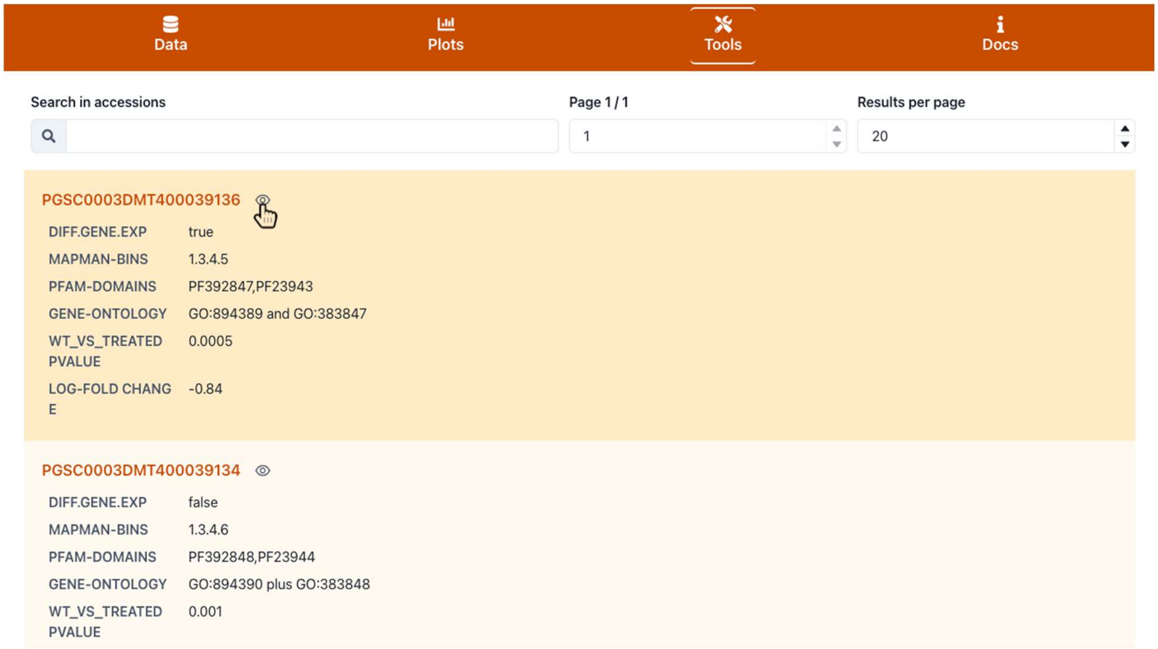
Load Gene-Info Table

Column Separator *
\t

Tabular File *
Select a file to load

Heat-Shock: Apical*DAS-1*0 Delete
Heat-Shock: Apical*DAS-1*2 Delete
Heat-Shock: Apical*DAS-2*0 Delete
Heat-Shock: Apical*DAS-2*2 Delete
Heat-Shock: Apical*DAS-3*0 Delete
Heat-Shock: Apical*DAS-3*1 Delete
Heat-Shock: Apical*DAS-3*2 Delete
Heat-Shock: Meristem*DAS-1*0 Delete
Heat-Shock: Meristem*DAS-1*1 Delete
Heat-Shock: Meristem*DAS-1*2 Delete
Heat-Shock: Meristem*DAS-2*0 Delete
Heat-Shock: Meristem*DAS-2*1 Delete
Heat-Shock: Meristem*DAS-2*2 Delete

c)



Search in accessions

Page 1 / 1
1

Results per page
20

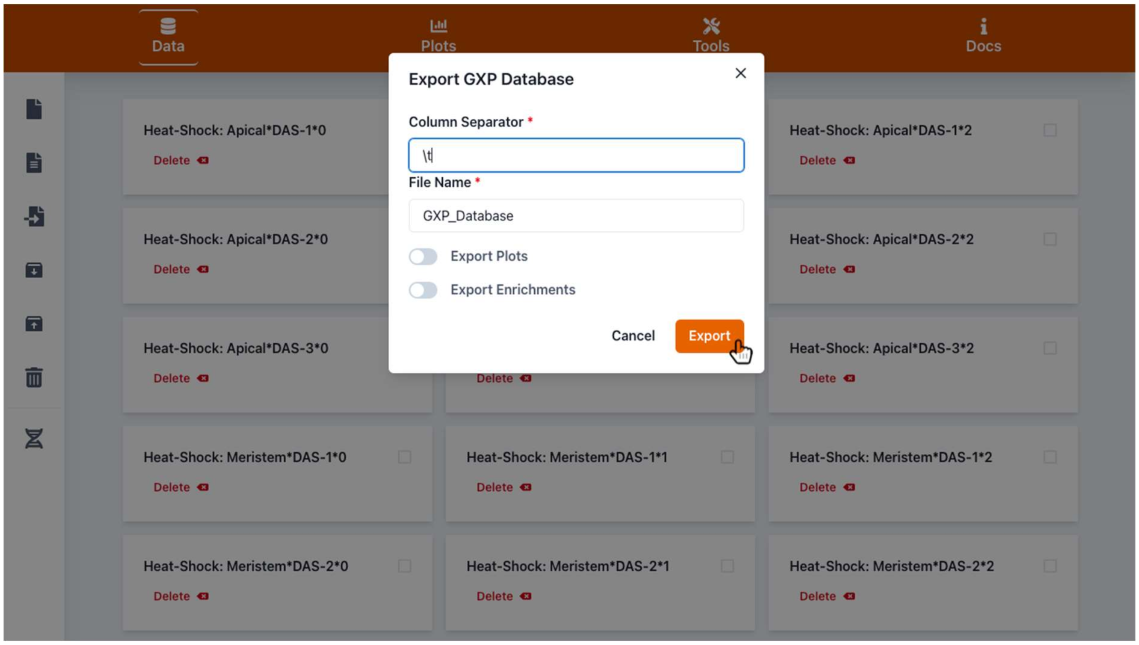
PGSC0003DMT400039136

DIFF.GENE.EXP	true
MAPMAN-BINS	1.3.4.5
PFAM-DOMAINS	PF392847,PF23943
GENE-ONTOLOGY	GO:894389 and GO:383847
WT_VS_TREATED PVALUE	0.0005
LOG-FOLD CHANG E	-0.84

PGSC0003DMT400039134

DIFF.GENE.EXP	false
MAPMAN-BINS	1.3.4.6
PFAM-DOMAINS	PF392848,PF23944
GENE-ONTOLOGY	GO:894390 plus GO:383848
WT_VS_TREATED PVALUE	0.001

d)



Export GXP Database

Column Separator *
\t

File Name *
GXP_Database

○ Export Plots
○ Export Enrichments

Cancel
Export

Heat-Shock: Apical*DAS-1*0 Delete
Heat-Shock: Apical*DAS-1*2 Delete
Heat-Shock: Apical*DAS-2*0 Delete
Heat-Shock: Apical*DAS-2*2 Delete
Heat-Shock: Apical*DAS-3*0 Delete
Delete
Heat-Shock: Apical*DAS-3*2 Delete
Heat-Shock: Meristem*DAS-1*0 Delete
Heat-Shock: Meristem*DAS-1*1 Delete
Heat-Shock: Meristem*DAS-1*2 Delete
Heat-Shock: Meristem*DAS-2*0 Delete
Heat-Shock: Meristem*DAS-2*1 Delete
Heat-Shock: Meristem*DAS-2*2 Delete

Figure 1. Screenshot images of the Gene expression tool data page interfaces for: (a) RNA-seq expression table file import, (b) transcript information table file import, (c) the gene browser that allows the imported input data to be viewed and (d) data export. These interfaces can be accessed on the side menu of the data page.

Figure 2 link

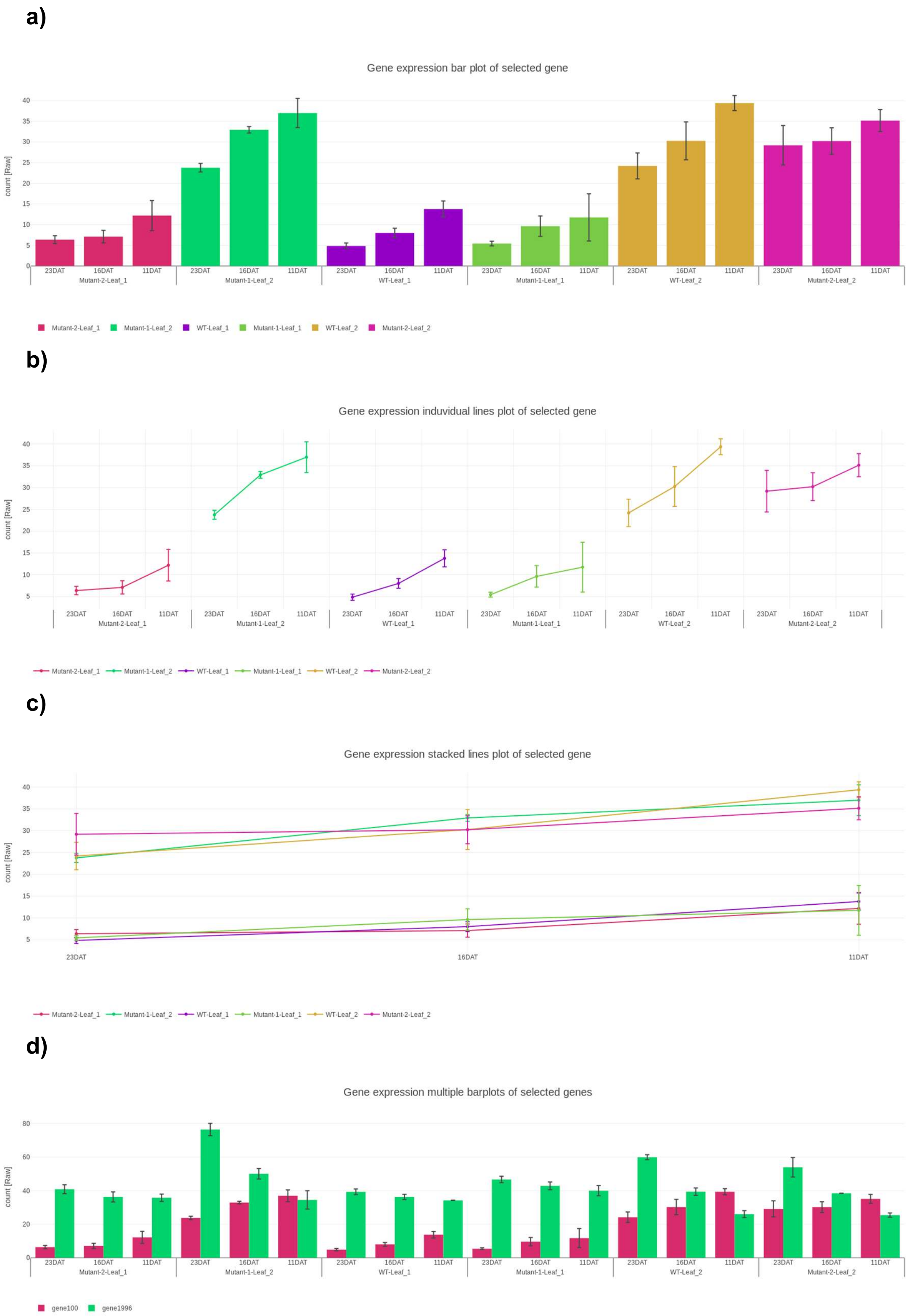


Figure 2. Interactive gene expression visualization plots describing expression levels (counts) (y axis) of selected genes in response to treatments of mutant and wild plant leaves (x axis). (a-d) The plots display mean values of replicates used in the experiment with standard deviation error bars. (a) Bar plots and (b) individual line plots display the same information. The stacked line plot (c) shows the same information

grouped per experimental sample type. (d) Multiple gene bar-plots. Note the gradual increase in expression of gene100 with respect to time (DAT) (a, b, c).

Figure 3 link

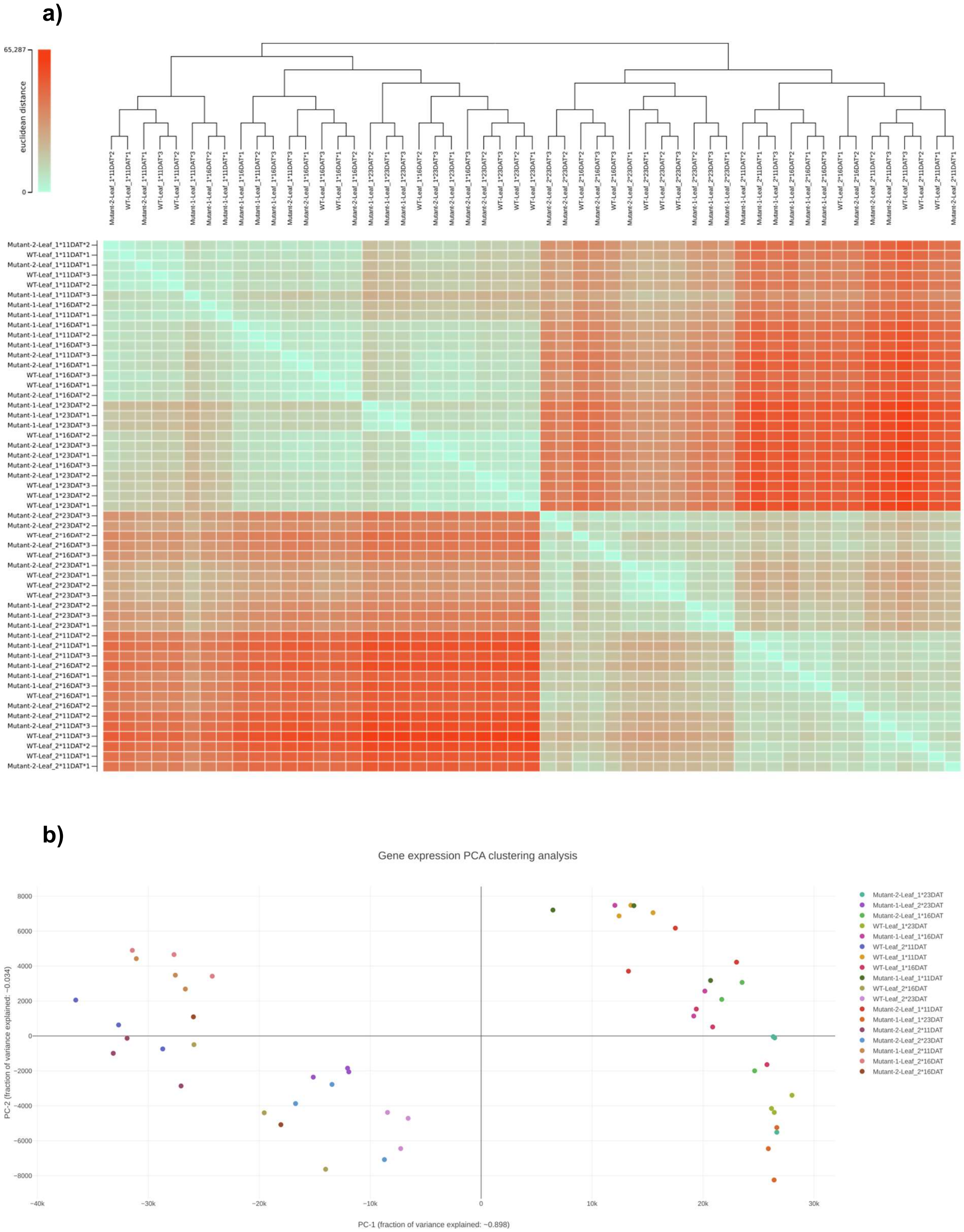
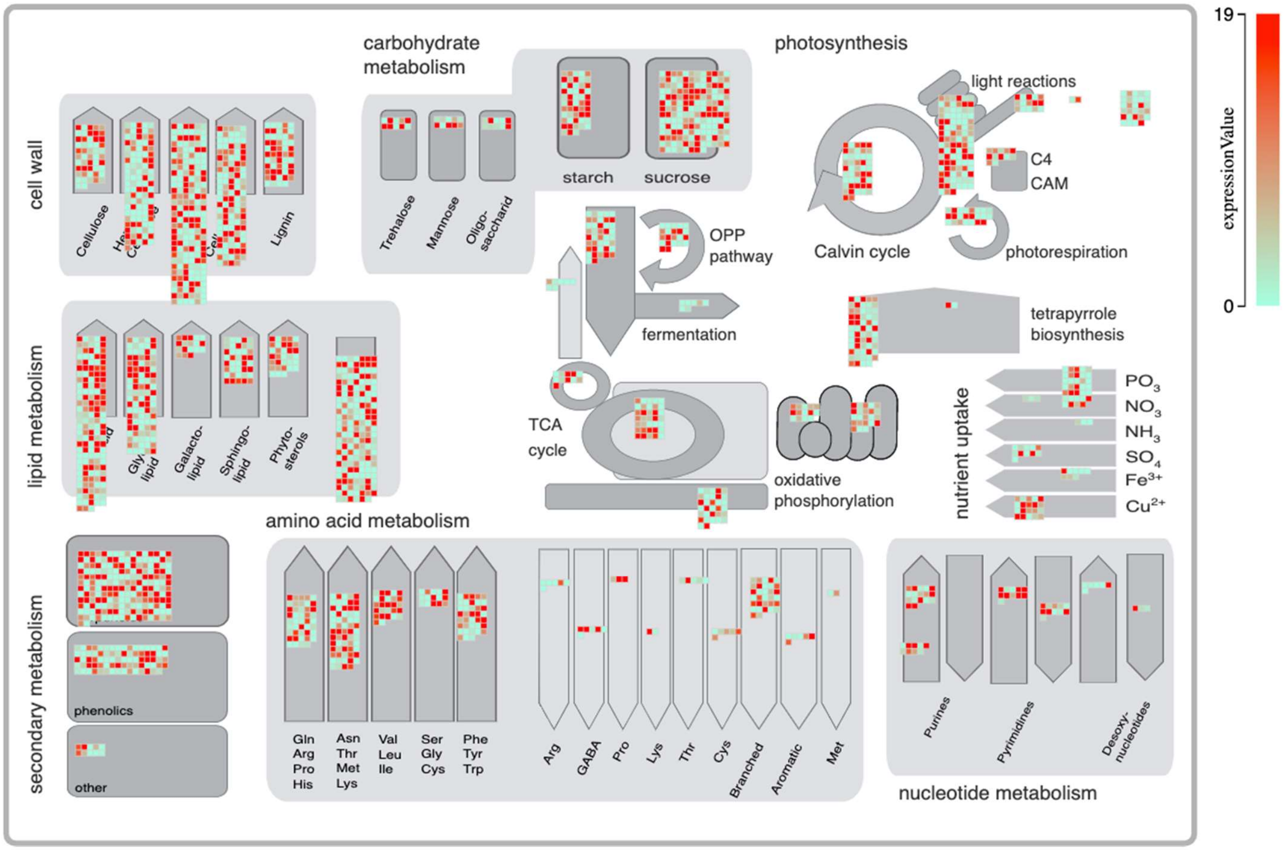


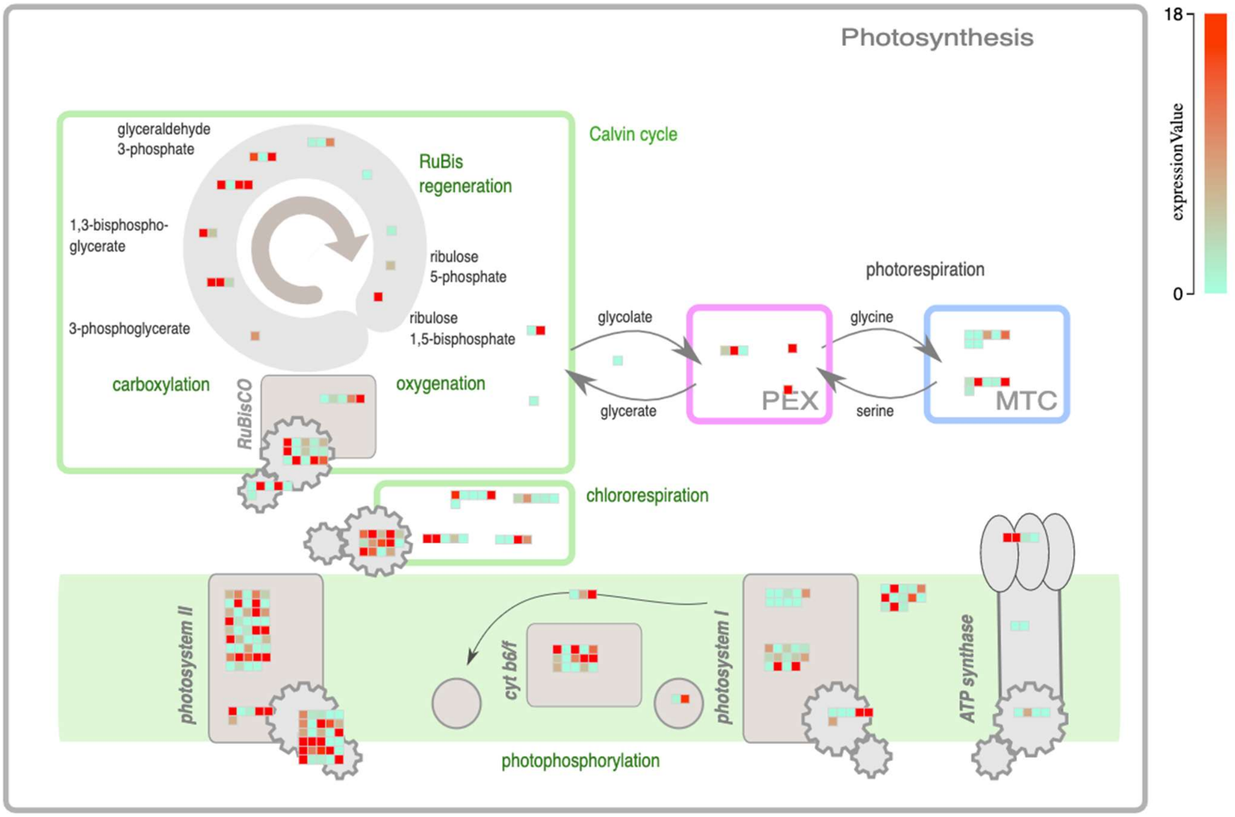
Figure 3. Hierarchical cluster analysis heatmap based on euclidean distance of different replicates (a). Clustering pattern of the individual samples based on the main principal components representing the highest variation observed revealed three distinct groups (b).

Figure 4 link

a)



b)



c)

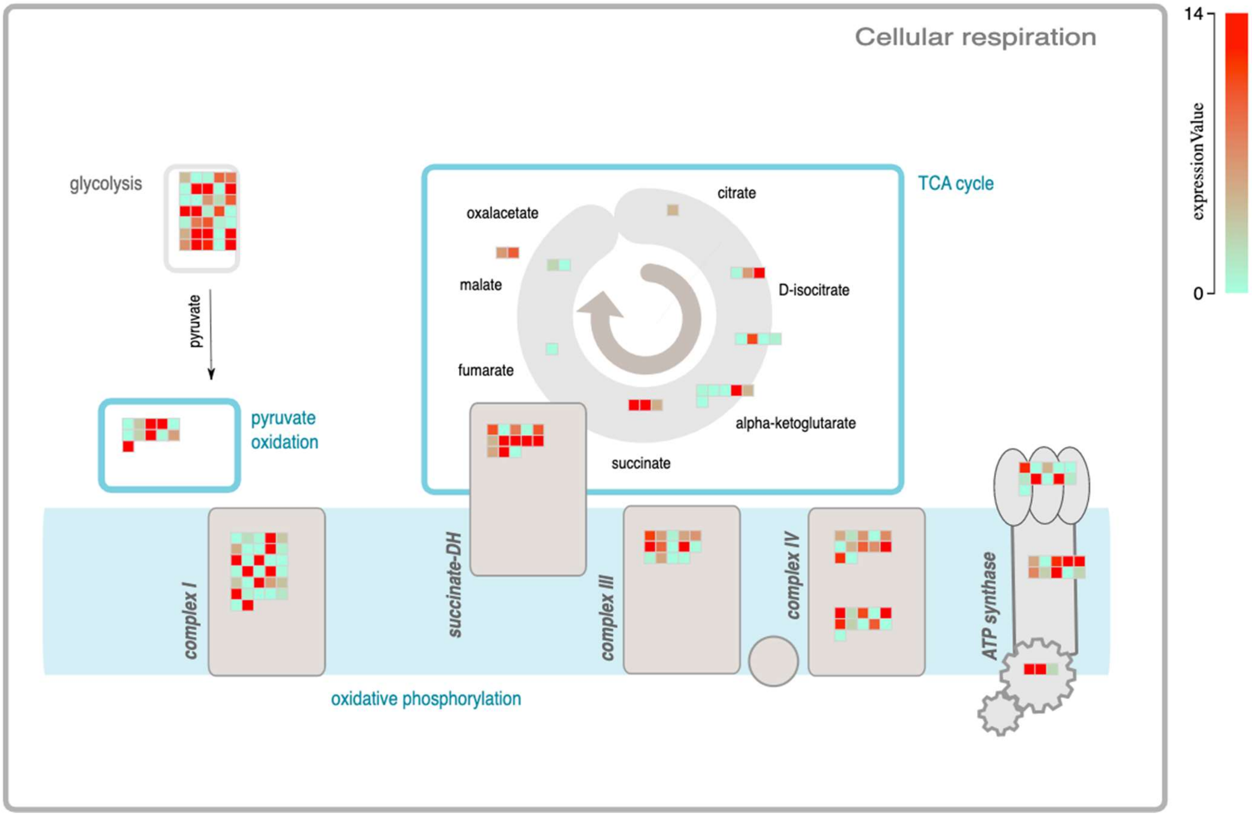
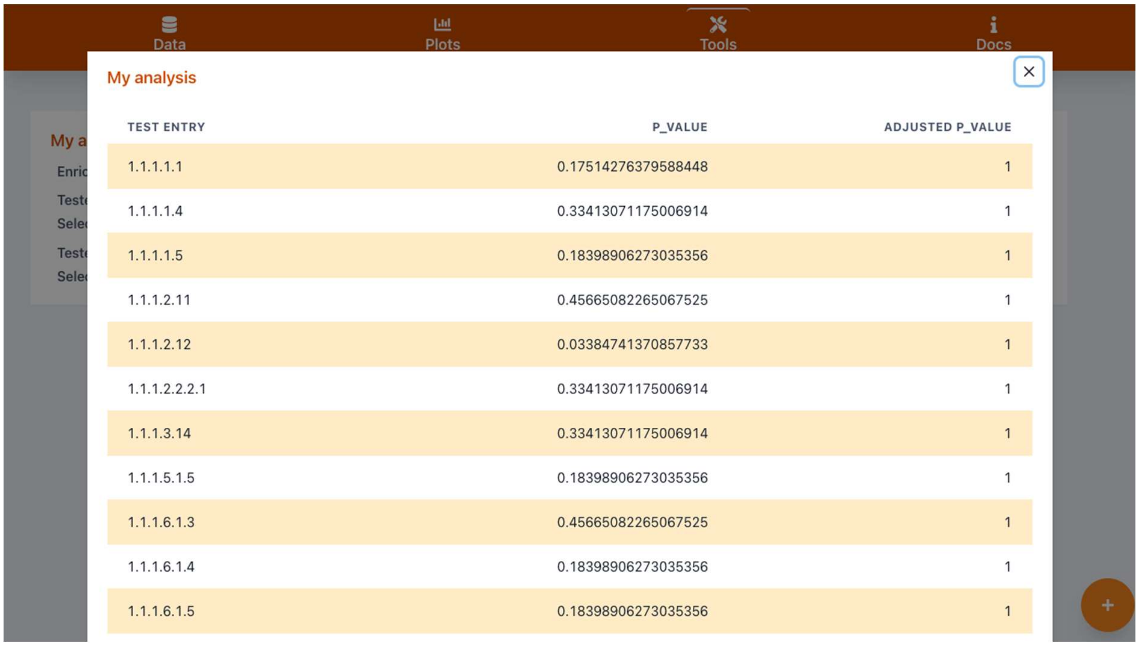


Figure 4. Images imported from the gene expression tool MapMan function. Each functional protein

annotation category (bin) is represented by a box where the expression value for each gene is translated into a color scale. The example images are a) Metabolism overview , b) Photosynthesis and c) Cellular respiration.

Figure 5 link



My analysis

TEST ENTRY	P_VALUE	ADJUSTED P_VALUE
1.1.1.1.1	0.17514276379588448	1
1.1.1.1.4	0.33413071175006914	1
1.1.1.1.5	0.18398906273035356	1
1.1.1.2.11	0.45665082265067525	1
1.1.1.2.12	0.03384741370857733	1
1.1.1.2.2.2.1	0.33413071175006914	1
1.1.1.3.14	0.33413071175006914	1
1.1.1.5.1.5	0.18398906273035356	1
1.1.1.6.1.3	0.45665082265067525	1
1.1.1.6.1.4	0.18398906273035356	1
1.1.1.6.1.5	0.18398906273035356	1

Figure 5. The image shows an interactive and downloadable tab-delimited text file format of an enrichment analysis performed on gene expression data. The tests were done for protein categories of the Mapman4 framework (here shown as code of the bins).