plants

MDPI

Technical Note

GXP: Analyze and Plot Plant Omics Data in Web Browsers

Constantin Eiteneuer [1,†], David Velasco [2,†], Joseph Atemia [3,†], Dan Wang [1], Rainer Schwacke [3], Vanessa Wahl [4], Andrea Schrader [5], Julia J. Reimer [5,6], Sven Fahrner [1], Roland Pieruschka [1], Ulrich Schurr [1], Björn Usadel [3] and Asis Hallab [3,*]

1 IBG-2 Plant Sciences, Forschungszentrum Jülich, 52428 Jülich, Germany; c.eiteneuer@fz-juelich.de (C.E.); d.wang@fz-juelich.de (D.W.); s.fahrner@fz-juelich.de (S.F.); r.pieruschka@fz-juelich.de (R.P.); u.schurr@fz-juelich.de (U.S.)
2 Faculty of Natural Sciences, Norges Teknisk-Naturvitenskapelige Universitet, 7034 Trondheim, Norway; davidve@stud.ntnu.no
3 IBG-4 Bioinformatics, Forschungszentrum Jülich, 52428 Jülich, Germany; j.atemia@fz-juelich.de (J.A.); r.schwacke@fz-juelich.de (R.S.); b.usadel@fz-juelich.de (B.U.)
4 Max Planck Institute for Molecular Plant Physiology, 14476 Potsdam, Germany; vwahl@mpimp-golm.mpg.de
5 Institute for Biology I, RWTH Aachen University, 52062 Aachen, Germany; schrader@bio1.rwth-aachen.de (A.S.); julia.reimer@hs-emden-leer.de (J.J.R.)
6 Faculty of Technology, University of Applied Science Emden/Leer, Molecular Biosciences, 26723 Emden, Germany
* Correspondence: a.hallab@fz-juelich.de
† These authors contributed equally to this work.

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, established software pipelines analyze raw data and produce quantitative data informing about gene expression or concentrations of metabolites. These results need to be visualized and further analyzed in order to support scientific hypothesis building and identification of underlying biological patterns. Some of these tools already exist, but require installation or manual programming. We developed "Gene Expression Plotter" (GXP), an RNAseq and Metabolomics data visualization and analysis tool entirely running in the user's web browser, thus not needing any custom installation, manual programming or uploading of confidential data to third party servers. Consequently, upon receiving the bioinformatic raw data analysis of RNAseq or other omics results, GXP immediately enables the user to interact with the data according to biological questions by performing knowledge-driven, in-depth data analyses and candidate identification via visualization and data exploration. Thereby, GXP can support and accelerate complex interdisciplinary omics projects and downstream analyses. GXP offers an easy way to publish data, plots, and analysis results either as a simple exported file or as a custom website. GXP is freely available on GitHub (see introduction)

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

Citation: Eiteneuer, C.; Velasco, D.; Atemia, J.; Wang, D.; Schwacke, R.; Wahl, V.; Schrader, A.; Reimer, J.J.; Fahrner, S.; Pieruschka, R.; et al. GXP: Analyze and Plot Plant Omics Data in Web Browsers. *Plants* **2022**, *11*, 745. https://doi.org/10.3390/plants11060745

Academic Editors: Ji Huang and Yufeng Wu

Received: 11 January 2022
Accepted: 1 March 2022
Published: 11 March 2022

1. Introduction

Modern life science research projects often produce quantitative data, for example to quantify gene expression or metabolite concentration in tissue samples. Many well-established tools exist that carry out the wet lab and bioinformatics procedures to produce such count data from the samples. Not rarely, these pipelines are carried out by third party laboratories. In the subsequent step, the life scientist, having ordered such RNAseq or other Omics experiments, needs to investigate these count data to form scientific hypotheses and identify underlying biological patterns. Typically, this involves plotting the quantitative data, carrying out Principal Component Analyses and correlation-based hierarchical

clustering to elucidate differences between experimental conditions. Genetic or metabolic responses to the tested experimental conditions and treatments often are summarized by identification of enriched traits within significantly up- or down-regulated genes or metabolites of interest. These steps often require manual programming, installation of software, or sending potentially confidential data to webservers for analysis. Gene Expression Plotter (GXP) minimizes these requirements by enabling the user to load count data, generate a variety of informative plots, and carry out typical clustering, principal component and overrepresentation analyses, without the need to write any code, install any software, or send the data to third party servers. Furthermore, GXP enables the user to save all loaded data along with the work done, including generated plots and carried out analysis. With this feature the user can not only save the current work to continue at a later time, but also share data, plots, and analysis results with others, simply by sending the exported GXP database file. Naturally, such a file can be published for example in the form of an article's supplement thus enabling readers to directly obtain the data, see plots and analysis results and even carry out their own subsequent investigations.

Gene Expression Plotter consumes two types of input tables, which can be prepared with standard spreadsheet programs, e.g., Microsoft Excel, or can be generated directly by the bioinformaticians producing the quantitative data (for details see results Section 2.1). In short, in the input quantifications table, each row represents quantitative values assessed for a single gene or metabolite, and columns correspond to the different samples for which these quantifications were assessed. In addition to the pure quantifications data table a free format information table can be provided. In it, too, rows correspond to the genes or metabolites, as they appear in the quantifications data table, and columns can provide any free text, categorical, or numeric information the user wants to load into GXP. Such free format information typically comprises knowledge about molecular gene function, for example in the form of Mapman4 Bin annotations [1–3], InterPro conserved protein domains [4], or Gene Ontology terms [5]. Additionally, in this table, the user can provide information, e.g., in the form of logarithmic fold change values, quantifying how much the expression of a particular gene changes when contrasting two selected experimental conditions, e.g., control versus stress treatment. Or, in case metabolites have already been quantified, information about their respective chemical properties and involved pathways can be provided. Both of this optional categorical and numerical data can later be used in GXP's enrichment analyses or can be displayed in the captions of generated plots, here comprising free text information, too. Thus, GXP has been developed to consume generic input data, making it a highly versatile tool, particularly in the context of overrepresentation analysis.

To produce, analyze, visualize, and publish quantitative omics data many tools exist. Among others, they provide means to plot the data, carry out clustering, and conduct principal component and overrepresentation analyses. A number of these tools specialize in RNAseq analysis [6–21], most of which consume the raw gene expression count data produced by standard gene expression quantifiers [22–25] and enable the user to identify differentially expressed genes [6–9,11–13,15,16,20,21,26,27] and review the results in form of comprehensive reports and/or plots [6–13,15,16,18–21,26–28]. Some [7–14,18–20] are implemented as an R / Shiny [29,30] application or use other forms of graphical user interfaces (GUI) [15,16,21,27]. These GUI tools allow the user (i) to either execute the tool installed locally on their computer and/or use it on a public web-server and (ii), by means of the GUI, eliminate the need to program plots and analyses manually, with few exceptions [6,28] which require some manual coding to make use of the extended provided functionality. By means of integration of curated published data some tools offer specific analyses, e.g., providing high-confidence insights into molecular gene function. One of these, GENAVi [12], enables the identification of differentially expressed genes (DEGs) in human or mouse RNAseq data by contrasting input with published data. OnestopRNAseq [16] is another example and offers several useful analyses by integration of curated public data from several model animal organisms. In addition, Plant Physiospace [31] enables the user to compare differential gene expression data with curated signatures to identify similar

genetic responses investigated in other already-published studies. Other tools focus on metabolomics [32–34] or, integrating RNAseq and Metabolomics data, the identification of genotype–phenotype relationships [35]. Some of the introduced tools [7,13,28] offer the elucidation of interactive plots where, e.g., hovering with the mouse over data points in a plot summarizing differential gene expression opens another plot illustrating the expression counts of the particular gene represented by the hovered-over data point. For a more detailed review of the above tools, see supplemental Text S1. In comparison with the above tools, GXP has a "downstream focus" on the visualization of quantitative omics data and subsequent clustering, principal component, and overrepresentation analysis. In this context, the key features of GXP are that (i) it does not require manual programming, nor installation of particular software, (ii) it can thus be used on a simple tablet or even smartphone, and (iii) GXP is versatile, in that it can consume any quantitative omics data stemming from RNAseq or metabolomics analyses. This genericity especially includes any additional arbitrary information on the quantified entities, that is, either genes or metabolites. As explained above, this particularly enables the user to carry out overrepresentation analysis on any, numerical or categorical, data the user provides (see results sections "2.1 Handling input and output data" and "2.5 Overrepresentation (enrichment) analysis" for more details). Furthermore, (iv) GXP is the first, immediately accessible, mature implementation of the popular Mapman tool that visually combines quantitative omics data analysis results with diagrams of metabolic pathways and other processes (see results section "2.4 MapMan web browser plots" for more details). Finally, (v) GXP ensures complete data safety. To explain this, consider that even though GXP is deployed on a webserver, once it has been loaded into the user's browser, it runs there completely independent of that server. This form of implementation is called a "single page application" (SPA) in which the webserver is little more than a file system delivering GXP to the user's browser. Among other things the implementation as a SPA implies that at no time is any data sent to any server. All analyses are carried out on the user's computer in the used web-browser and all data remains exactly there.

Gene Expression Plotter is freely available on GitHub (https://usadellab.github.io/GeneExpressionPlots/; accession date 11 January 2022). Its code has been released as open source (https://github.com/usadellab/GeneExpressionPlots; accession date 11 January 2022) under the GNU public license, version 3. All functions carrying out the above-described analyses are tested with automatic software unit tests to ensure correct calculations. As mentioned, GXP provides means to export all loaded data, generated plots, and carried out analyses results into a single file, dubbed the "GXP database". Such a GXP database can be used for publication, e.g., in the form of an article supplement. Additionally, GXP can easily be copied and this copy can be published including custom quantitative omics data, plots, and carried-out analyses. Such a copy can be made available, e.g., on GitHub free of charge or any other webserver. The GXP manual has detailed instructions on a simple procedure to set up such a custom copy of GXP and includes screenshots on seven easy steps that only require a GitHub account and a web browser.

Thus, our new tool, Gene Expression Plotter, enables the end-user to visualize and analyze quantitative data, typically taken from RNAseq 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

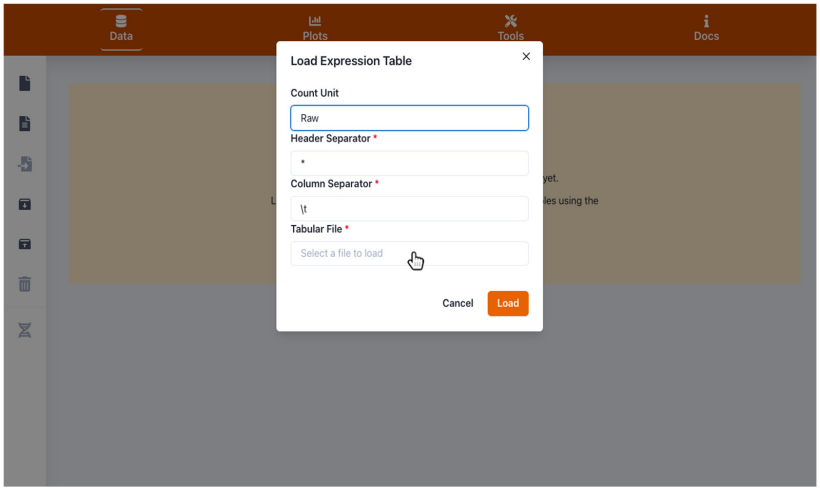
2.1. Handling Input and Output Data

With the aim of providing a single suite in which a user can visualize and analyze quantitative omics 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 (accession date 11 January 2022).

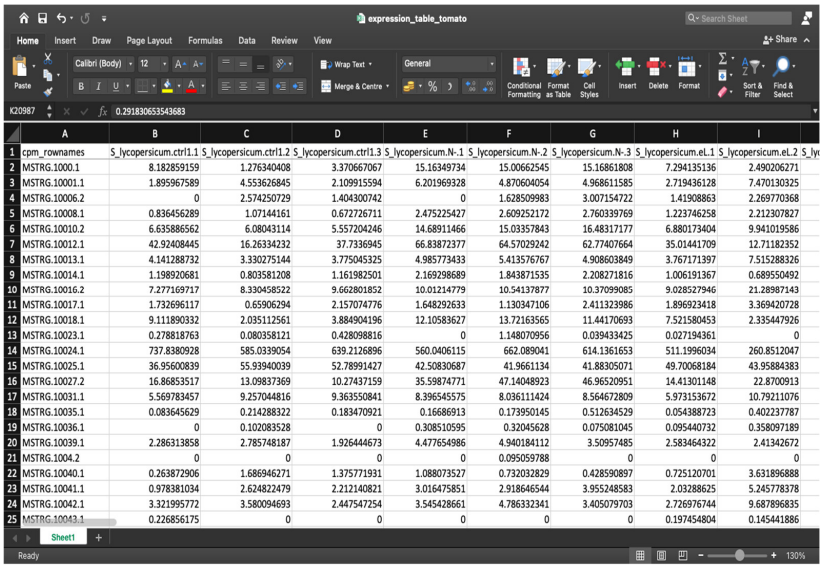
GXP is provided with a comprehensive online documentation and a manual (see menu "Docs"). To use GXP, the user is first asked to import quantitative RNA-seq or metabolomics data in a tabular format. Each row should represent, e.g., a single quantified transcript. The transcript identifiers are typically provided in the first column. All following columns should contain the transcript quantifications for each respective genotype, replicate or treatment as specified by the column names (Figure 1a,b). GXP is made aware of the statistical factors differentiating in the experiment the respective biological replicates. This is encoded directly in the respective column names in the input quantification table. Here, as one studied factor the user can for example specify the time after an experimental treatment at which a sample was obtained or the type of stress treatment a sample was exposed to. Such a factor is then used by GXP to position points on the x-axis in plots. We consequently dub such factors "x-axis factors". In the example data included in GXP, a column name positioning data points over the "ctrl1" x-axis tick (Figure 2) would be, for example, "S_lycopersicum.ctrl1.1" or "S_lycopersicum.ctrl1.2". Note that the x-axis factor is identified by its location between the first and second "." in the column name. The user can select another character, e.g., "*" instead of ".". Strictly speaking, GXP accepts a single x-axis factor that can have multiple values. In the example data these can be "ctrl1" (control) and the stress treatments "cold" (chilling temperature), "eL" (extended light), and "N-" (nitrogen deficiency). Thus, x-axis factors join (in typical RNAseq experiments three) biological replicates that were subjected to the same experimental treatment into a single bin. Such x-axis bins are subsequently used to calculate y-axis error bars for data points representing several joined replicates (see Section 2.2 and Figure 2). Another type of factor can optionally be introduced and is used to compare for example biological species, genotypes, or different treatments. A good example for such a type of factor can be the comparison of a wild (*Solanum pennellii*) versus a domesticated (*S. lycopersicum*) tomato species. We dub such factors that group several biological conditions "group factors". Note that these group factors are not used to generate tick labels on the x-axis, but rather imply two plots showing, e.g., gene expression in the wild type and domesticated species side-by-side (Figure 2a,b). A user can specify as many group factors as were investigated in a respective research project. Additionally, in the case of group factors, GXP is made aware of these simply through the column names of the counts table input. In the example data the column names "S_lycopersicum.ctrl1.1" and "S_pennellii.ctrl1.3" indicate the group factor "species" with values *S. lycopersicum* and *S. pennellii*, respectively. Note that the group factors are also identified by their position in the column names, appearing before the x-axis factor and separated by the "." character. As mentioned before, the user can specify any other character as separator, e.g., "*".

Optionally, GXP can use any extra information associated with specific quantified transcripts or metabolites. This is done by loading a separate table in which each row corresponds to a single quantified entity (transcript or metabolite), and in which each column holds additional generic information (Figure 1c,d). Examples of such additional generic information include ontological annotations informing about the molecular function of proteins, e.g., terms from the Mapman4 framework [3,36], from the Gene Ontology (GO) project [5], or from KEGG pathways [37], or differential expression between contrasted conditions, e.g., cold stress treatment versus control conditions. Information about differential expression can be provided in the form of logarithmic fold change of gene expression and/or adjusted *p*-values used to identify significant changes in gene expression. In fact, in this optional information table, the user can provide any information in the form of columns that contain either free text, numerical values, e.g., chemical properties, such as hydrophobicity, of metabolites, or categorical annotations about molecular protein function, similar to the terms obtained from Mapman4 ontology [36], the GO [5], the KEGG [37], or InterPro [4]. Note that GXP offers tools to help the user obtain and import molecular gene function annotations in the form of Mapman Bins for his data in the respective "Mapman functional annotations" section of the "Tools" menu. Importantly numeric and categorical information can be used to carry out subsequent overrepresentation analysis, while all

information is used in the gene browser to provide the user with a rich database about the studied transcripts or metabolites (Section 2.1.1 and Figure 1e).

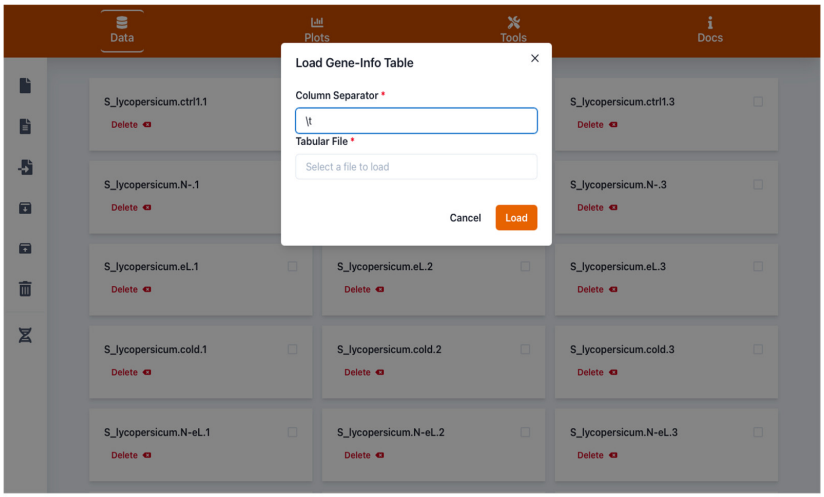


(**a**)



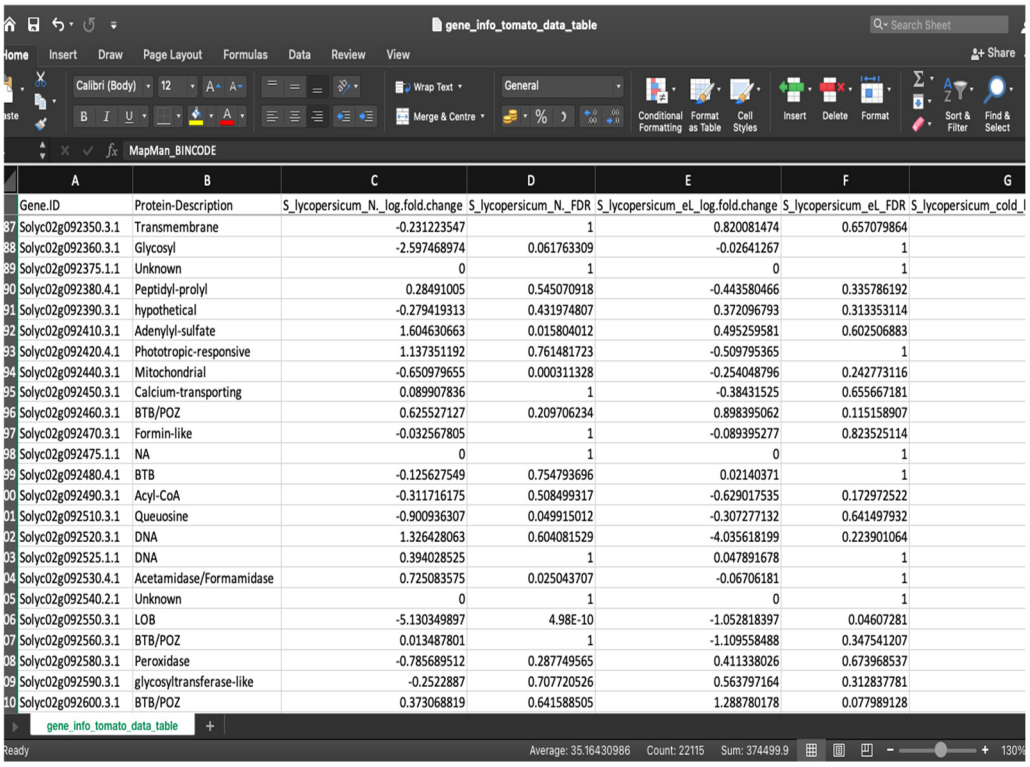
	A	B	C	D	E	F	G	H	I
1	cpm_rownames	S_lycopersicum.ctrl1.1	S_lycopersicum.ctrl1.2	S_lycopersicum.ctrl1.3	S_lycopersicum.N-.1	S_lycopersicum.N-.2	S_lycopersicum.N-.3	S_lycopersicum.eL.1	S_lycopersicum.eL.2
2	MSTRG.10000.1	8.182859159	1.276340408	3.370667067	15.16349734	15.00662545	15.16861808	7.294135136	2.490206271
3	MSTRG.10001.1	1.895967589	4.553626845	2.109915594	6.201969328	4.870604054	4.968611585	2.719436128	7.470130325
4	MSTRG.10006.2	0	2.574250729	1.404300742	0	1.628509983	3.007154722	1.41908863	2.269770368
5	MSTRG.10008.1	0.836456289	1.07144161	0.672726711	2.475225427	2.609252172	2.760339769	1.223746258	2.212307827
6	MSTRG.10010.2	6.635886562	6.08043114	5.557204246	14.68911466	15.03357843	16.48317177	6.880173404	9.941019586
7	MSTRG.10012.1	42.92408445	16.26334232	37.7336945	66.83872377	64.57029242	62.77407664	35.01441709	12.71182352
8	MSTRG.10013.1	4.141288732	3.330275144	3.775045325	4.985773433	5.413576767	4.908603849	3.767171397	7.515288326
9	MSTRG.10014.1	1.198920681	0.803581208	1.161982501	2.169298689	1.843871535	2.208271816	1.006191367	0.689550492
10	MSTRG.10016.2	7.277169717	8.330458522	9.662801852	10.01214779	10.54137877	10.37099085	9.028527946	21.28987143
11	MSTRG.10017.1	1.732696117	0.65906294	2.157074776	1.648292633	1.130347106	2.411323986	1.896923418	3.369420728
12	MSTRG.10018.1	9.111890332	2.035112561	3.884904196	12.10583627	13.72163565	11.44170693	7.521580453	2.335447926
13	MSTRG.10023.1	0.278818763	0.080358121	0.428098816	0	1.148070956	0.039433425	0.027194361	0
14	MSTRG.10024.1	737.8380928	585.0339054	639.2126896	560.0406115	662.089041	614.1361653	511.1996034	260.8512047
15	MSTRG.10025.1	36.95600839	55.93940039	52.78991427	42.50830687	41.9661134	41.88305071	49.70068184	43.95884383
16	MSTRG.10027.2	16.86853517	13.09837369	10.27437159	35.59874771	47.14048923	46.96520951	14.41301148	22.8700913
17	MSTRG.10031.1	5.569783457	9.257044816	9.363550841	8.396545575	8.036111424	8.564672809	5.973153672	10.79211076
18	MSTRG.10035.1	0.083645629	0.214288322	0.183470921	0.16686913	0.173950145	0.512634529	0.054388723	0.402237787
19	MSTRG.10036.1	0	0.102083528	0	0.308510595	0.32045628	0.075081045	0.095440732	0.358097189
20	MSTRG.10039.1	2.286313858	2.785748187	1.926444673	4.477654986	4.940184112	3.50957485	2.583464322	2.41342672
21	MSTRG.1004.2	0	0	0	0	0.095059788	0	0	0
22	MSTRG.10040.1	0.263872906	1.686946271	1.375771931	1.088073527	0.732032829	0.428590897	0.725120701	3.631896888
23	MSTRG.10041.1	0.978381034	2.624822479	2.212140821	3.016475851	2.918646544	3.955248583	2.03288625	5.245778378
24	MSTRG.10042.1	3.321995772	3.580094693	2.447547254	3.545428661	4.786332341	3.405079703	2.726976744	9.687896835
25	MSTRG.10043.1	0.226856175	0	0	0	0	0	0.197454804	0.145441886

(**b**)



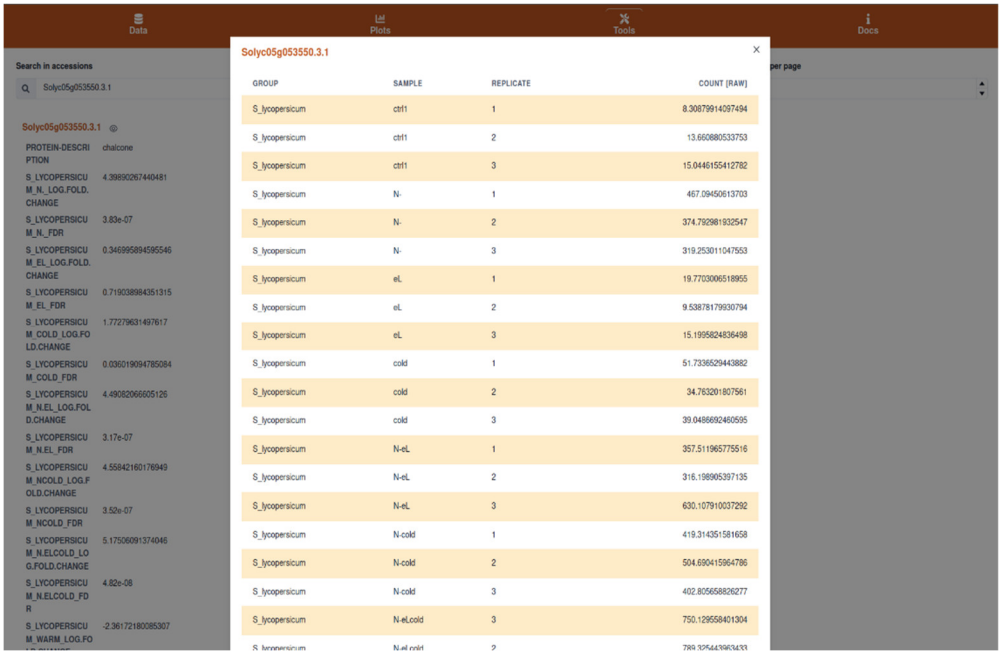
(**c**)

Figure 1. *Cont.*



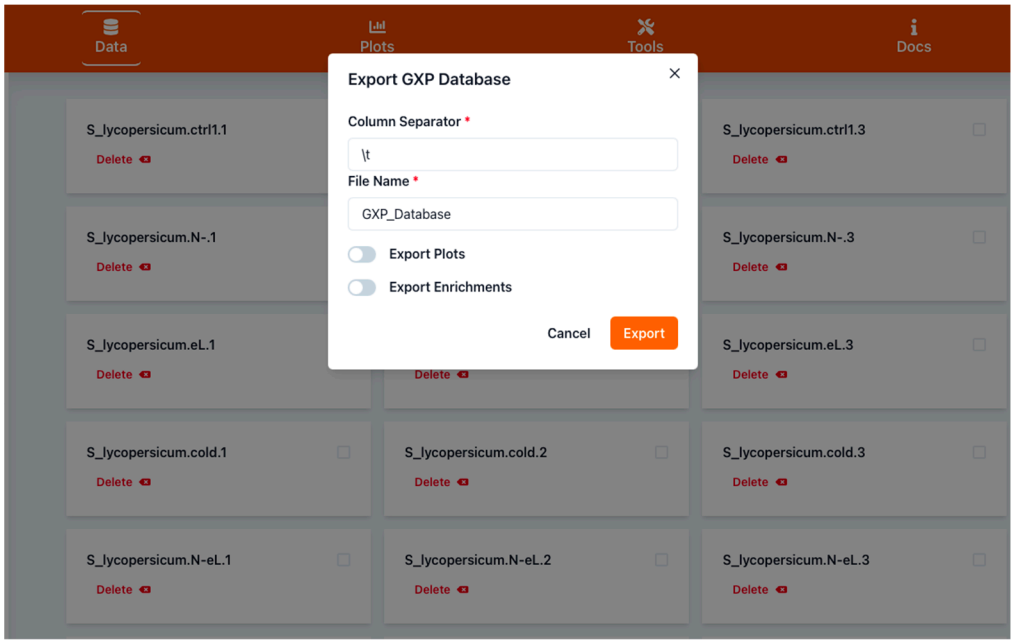
	A	B	C	D	E	F	G
	Gene.ID	Protein-Description	S_lycopersicum_N._log.fold.change	S_lycopersicum_N._FDR	S_lycopersicum_eL_log.fold.change	S_lycopersicum_eL_FDR	S_lycopersicum_cold_l
87	Solyc02g092350.3.1	Transmembrane	-0.231223547	1	0.820081474	0.657079864	
88	Solyc02g092360.3.1	Glycosyl	-2.597468974	0.061763309	-0.02641267	1	
89	Solyc02g092375.1.1	Unknown	0	1	0	1	
90	Solyc02g092380.4.1	Peptidyl-prolyl	0.28491005	0.545070918	-0.443580466	0.335786192	
91	Solyc02g092390.3.1	hypothetical	-0.279419313	0.431974807	0.372096793	0.313353114	
92	Solyc02g092410.3.1	Adenylyl-sulfate	1.604630663	0.015804012	0.495259581	0.602506883	
93	Solyc02g092420.4.1	Phototropic-responsive	1.137351192	0.761481723	-0.509795365	1	
94	Solyc02g092440.3.1	Mitochondrial	-0.650979655	0.000311328	-0.254048796	0.242773116	
95	Solyc02g092450.3.1	Calcium-transporting	0.089907836	1	-0.38431525	0.655667181	
96	Solyc02g092460.3.1	BTB/POZ	0.625527127	0.209706234	0.898395062	0.115158907	
97	Solyc02g092470.3.1	Formin-like	-0.032567805	1	-0.089395277	0.823525114	
98	Solyc02g092475.1.1	NA	0	1	0	1	
99	Solyc02g092480.4.1	BTB	-0.125627549	0.754793696	0.02140371	1	
00	Solyc02g092490.3.1	Acyl-CoA	-0.311716175	0.508499317	-0.629017535	0.172972522	
01	Solyc02g092510.3.1	Queuosine	-0.900936307	0.049915012	-0.307277132	0.641497932	
02	Solyc02g092520.3.1	DNA	1.326428063	0.604081529	-4.035618199	0.223901064	
03	Solyc02g092525.1.1	DNA	0.394028525	1	0.047891678	1	
04	Solyc02g092530.4.1	Acetamidase/Formamidase	0.725083575	0.025043707	-0.06706181	1	
05	Solyc02g092540.2.1	Unknown	0	1	0	1	
06	Solyc02g092550.3.1	LOB	-5.130349897	4.98E-10	-1.052818397	0.04607281	
07	Solyc02g092560.3.1	BTB/POZ	0.013487801	1	-1.109558488	0.347541207	
08	Solyc02g092580.3.1	Peroxidase	-0.785689512	0.287749565	0.411338026	0.673968537	
09	Solyc02g092590.3.1	glycosyltransferase-like	-0.2522887	0.707720526	0.563797164	0.312837781	
10	Solyc02g092600.3.1	BTB/POZ	0.373068819	0.641588505	1.288780178	0.077989128	

(**d**)



(**e**)

Figure 1. *Cont.*



(**f**)

Figure 1. Screenshot images of Gene Expression Plotter (GXP) showing interfaces for data input. (**a**) shows the quantitative data table file import form, triggered by the first document button in the upper left panel. This form is used to load a quantification table such as the one shown in (**b**). (**c**) shows the transcript information table file import form, triggered by the second document icon in the left panel. This form is used to load an optional information table (see Section 1) like the one shown in (**d**). After successful import, the user can search for genes or their annotations in the "gene browser" (helix icon on the lower left panel). In the shown example, the user searched for "chalcone synthase", a polyketide synthase involved in flavonoid biosynthesis. In (**e**) the user now inspects this gene's expression quantifications (highlighted foreground) and additional information such as the logarithmic fold changes of gene expression assessed for various comparisons of control and stress treatments (lightly faded out background). Furthermore, as shown in (**f**), by using the GXP export function triggered by the fifth, upward arrow on the box icon in the left panel, GXP enables the user to save the current state, i.e., all imported data, generated plots, and analysis results for later continuation or to share it with other researchers.

Gene Expression Plotter is freely available on GitHub for direct use (https://usadellab.github.io/GeneExpressionPlots; accession date 11 January 2022). It is provided with example RNAseq data from a study on stress response contrasting wild with domesticated tomato species [38]. In collaboration with one of the original authors, this data has been used to directly compare and benchmark the results produced by GXP with the already published findings. This example data can be loaded by clicking on "Load example data" in the "data" menu.

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 1e). The presented information includes the respective quantitative data extracted from the "quantifications table" (see Section 2.1) and any further information about the respective transcripts or metabolites extracted from the "information table" (see Section 2.1). A user can, e.g., search for a gene of interest "chalcone synthase" and inspect all transcripts associated with this molecular function, their respective expression, and additional information such as conserved protein domains.

Figure 2. Screenshots of plots visualizing and comparing quantified gene expression between different treatments, experimental conditions, and genes: (**a**) bar plot, (**b**) individual lines plot and (**c**) stacked lines plot, are different modes of how Gene Expression Plotter (GXP) visualizes the expression profile of the example gene Solyc05g053550.3.1 (*CHALCONE SYNTHASE*). The three plots highlight how the expression of the example *CHALCONE SYNTHASE* responds to the experimental conditions. This *CHALCONE SYNTHASE's* expression is up-regulated in *S. lycopersicum* but conversely not in *S. pennellii* following stress treatments of nitrogen deficiency (N-) and in combination with chilling temperatures (cold) and elevated light intensity (eL). Plot (**d**) compares the genetic response of this *CHALCONE SYNTHASE* with another gene of interest Solyc08g075570.4.1 (*UREA PROTON SYMPORTER)*. In contrast to the expression of *CHALCONE SYNTHASE*, gene expression of the *UREA PROTON SYMPORTER* is relatively low in both *S. pennellii* and *S. lycopersicum*.

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 1f). The generated database can be used later to resume previous work or share plots and results. The exported GXP database contains a configuration file ("GXP_settings.json", see manual for more details) that can be used, e.g., 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 plots showing the expression of individually selected genes. Available visualizations are bar- (Figure 2a) and line-plots (Figure 2b). 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 biological replicates, e.g., species, genotype, or different treatments. This information will be visualized by the color or position of plotted values (Figure 2; domesticated tomato *S. lycopersicum* in red on the left side and wild *S. pennellii* in green on the right side). Quantifications differing between group factors, e.g., domesticated versus wild tomato species, can either be plotted in two graphs side-by-side (Figure 2a,b) or in a single graph as differentially colored stacked curves (Figure 2c). The user can also visualize the expression of multiple genes/metabolites in the same plot using all mentioned types: bar, single, or stacked curves (Figure 2d for an example with bars). In this case, as mentioned, the color distinguishes genes while the group factors are shown in their separate graphs side-by-side. All plots are interactive in that upon hovering over data points with the mouse, the user is presented with the respective values in a little overlay window. 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 Biological Replicates Based on Either Gene Expression or Quantified Metabolites

Typically, during the analysis and interpretation or RNAseq or metabolomics experiment results, one wants to distinguish within, i.e., biological background noise, from between-group differences to enable drawing significant conclusions in terms of an organism's response to contrasting experimental conditions. Only if the in-between-group variation is not silenced by the background noise, can the data be used to elucidate the original biological questions motivating the study. To assess background noise and in-between-group variation, typically several (at least three in RNAseq experiments) biological replicates sharing the same experimental condition are quantified. Subsequently, similarity of the quantified replicates should be recognizable over the background noise to indicate that the quantifications can be used for the investigation of the original biological question motivating the study. This similarity can be assessed and visualized using correlation or Euclidean distance based hierarchical clustering and principal component analysis.

2.3.1. Hierarchical Cluster Analysis

Results of a hierarchical cluster analysis, executed on optionally z-transformed values, are visualized in a heatmap whose axis is accompanied with a tree dendrogram representing the hierarchical clusters that the respective biological replicates have been grouped into (Figure 3a). 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'

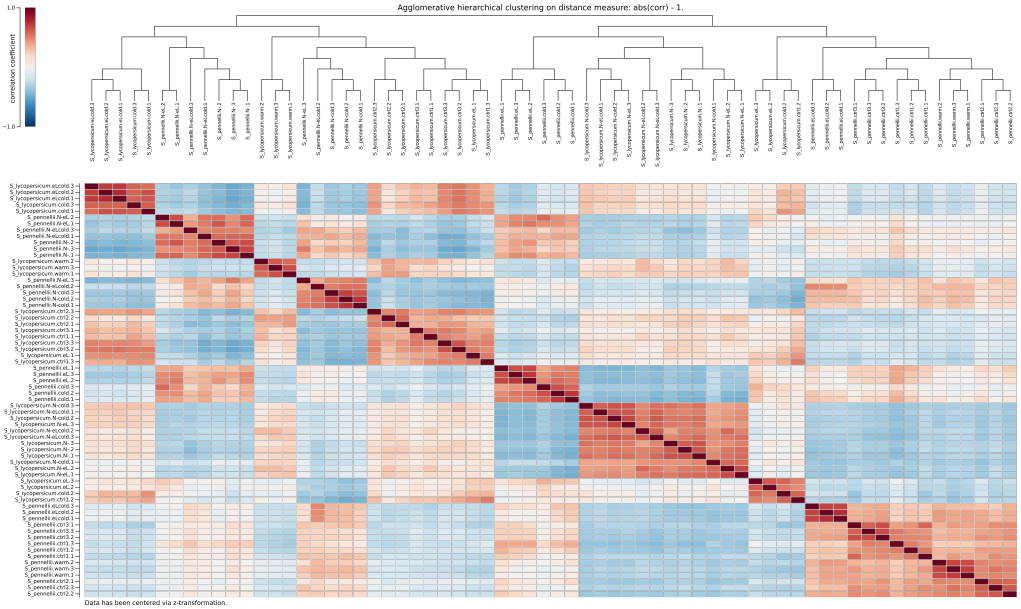
gene expression vectors as a basis for the clustering analysis (see Section 4.3 for more details). The potentially demanding analysis is carried out in the background, using so called "web-workers", and therefore does not block the user interface. While an analysis runs, a plot showing a loading icon 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 user's original choice. A color scale is provided, and the plot is interactive, in that hovering with the mouse over a heatmap cell will display the calculated likeness value assessed for the respective pair the cell's row and column corresponds to, respectively.

2.3.2. Principal Component Analysis

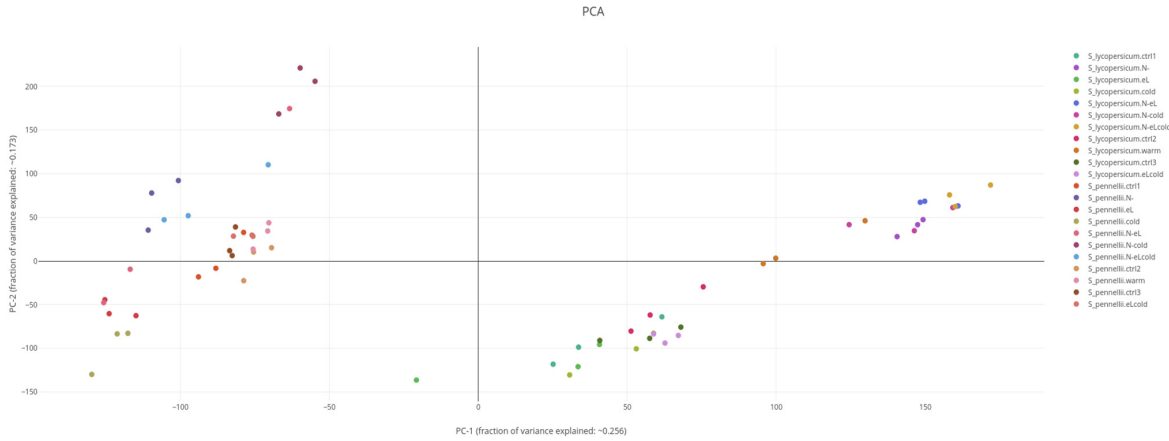
Results of principal component analysis (PCA) are typically visualized as a subclass of scatter-plots where the two axes represent the two principal components that explain most of the observed variance between samples. GXP carries out a PCA on user demand and runs the respective calculation in the user's browser 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 3b). Biological replicates are color coded so that replicates that have identical x-axis factors (Section 2.1), e.g., all replicates belonging to the wild tomato species *S. pennellii* that also have been exposed to the same cold stress conditions receive the same color. In the PCA scatter plot 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 interactively.

2.4. Mapman Web Browser Plots

The Mapman frameworks (Mapman4 [36] and the older version Mapman v.3.6 [39]) comprise a manually curated vocabulary (ontology) to describe the function of land plant proteins. Mercator and Mercator4 [3,36] 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 [1,3,36,39]) has been developed to visualize annotations of the Mapman frameworks in the context of gene expression data. Based on a proof-of-principle code [1,40] we also developed a simple MapMan web browser application. The same as in the MapMan desktop application, a user can choose one of several basic metabolic cellular sketches, e.g., "Metabolism overview", "Photosynthesis", or "Secondary metabolism" (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 (Figure 4b).



(**a**)



(**b**)

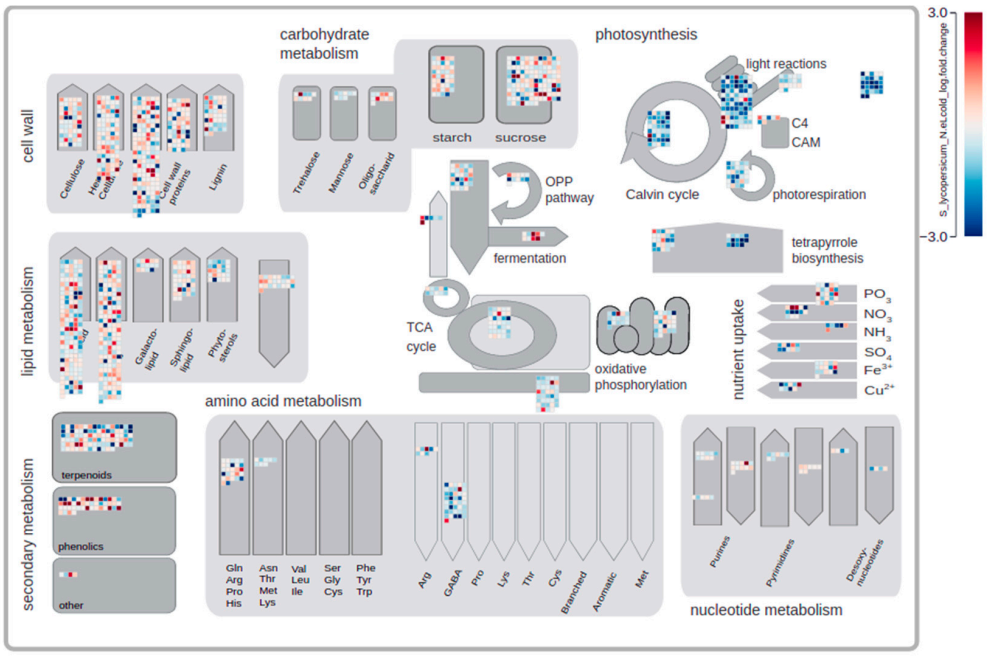
Figure 3. Screenshots of plots investigating likeliness of gene expression assessed in different biological replicates. Plot (**a**) shows the result of correlation-based hierarchical clustering in the form of a dendrogram and a correlation heatmap. In the top left corner, a scale color-codes the calculated correlation coefficients. In the top middle, the dendrogram represents the result of hierarchical clustering of all loaded biological replicates. In this example, the plot informs the user of their choice to z-transform the data before the calculation of correlation (lower left corner). Upon hovering with the mouse over single cells of the correlation matrix, the user is presented with the respective correlation value between the two biological replicates represented by the cell's row and column, respectively. This example shows how Gene Expression Plotter (GXP) helps the user to assess how well the applied experimental conditions and treatments are reflected in terms of quantified gene expression. Here, serving as a quality check, the statistical factors "species" and "stress treatments" mostly imply the grouping of biological replicates, highlighting that the experimental setup and bioinformatics analysis yielded data fit to carry out the original biological question of the study, namely to elucidate the genetic responses to the applied stress treatments and subsequently compare

these genetic responses between the two studied species of tomato. A plot highlighting similar patterns is shown in (**b**). Here, a principal component analysis (PCA) has been carried out on z-transformed data. The resulting scatter plot of the two most important principal components (PC) confirms that the color-coded biological replicates (legend in the top right corner) mostly group by the factors "species" and "stress treatment", i.e., are found in close proximity within the scatter plot. When hovering with the mouse over single data points, the user is presented with the exact PC values and the name of the respective biological replicate represented by the data point. Using the axes labels, the user is informed about how much of the observed variation is explained by the two respective principal components PC1 (here: approx 25.6 %) and PC2 (here: approx. 17.3 %). As in (**a**), the PCA and resulting scatter plot indicate that biological replicates group well together, implying that within this example study, the influence of treatment and genotype on gene expression is well distinguishable from the biological background noise.

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 cold stress. 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. Overrepresentation (Enrichment) Analysis

To qualitatively describe a biological response, often, annotations about biological processes and molecular functions are analyzed. Those annotations found to be significantly overrepresented among selected genes or metabolites of interest characterize that group of genes. The selection criteria can, e.g., be significant up- or down-regulation of gene expression contrasting two experimental conditions, e.g., control versus cold stress 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 or metabolome. GXP offers the user an easy way to carry out such overrepresentation analysis (ORAs). The user specifies a criterion shared by all selected quantified entities, i.e., either transcripts or metabolites, and additionally selects annotation terms for which overrepresentation should be tested. Alternatively, the user can select the transcripts or metabolites of interest manually by entering their respective 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). In principle, ORAs can be carried out for metabolite data, even though these analyses are less common for targeted metabolomics studies. Consequently, a great variety of enrichment analyses can be carried out. In this, 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 or the Benjamini-Hochberg 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" submenu (see Section 2.1).

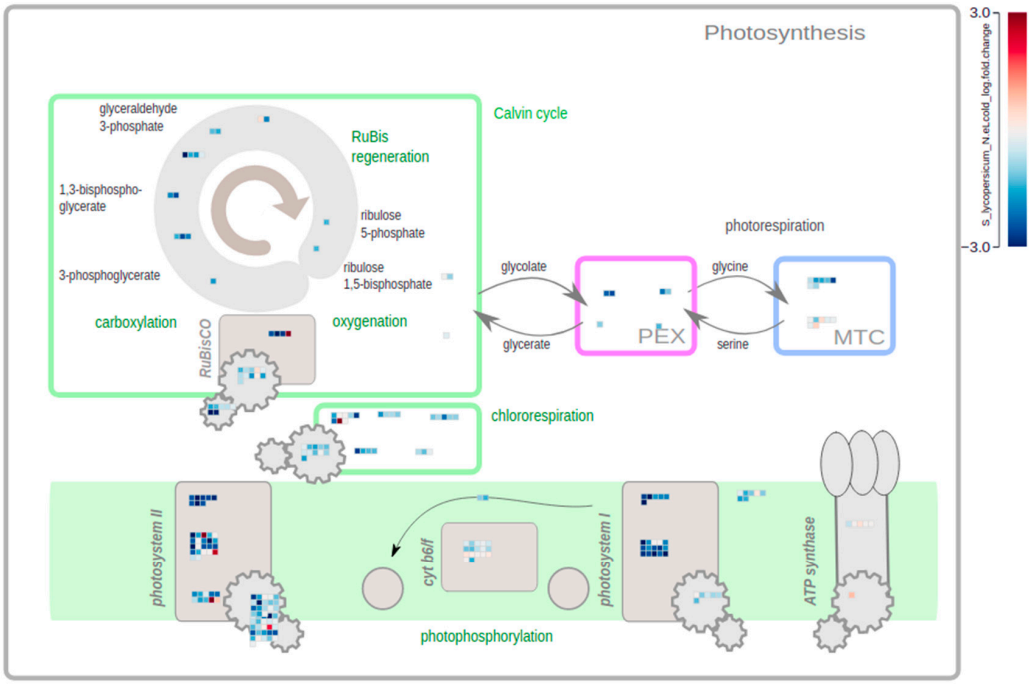


MIN	Q1	MEDIAN	MEAN	Q3	MAX
-12.12	-0.81	-0.10	-0.24	0.34	11.60

[5] Adjust Box Size

Distribution of S_lycopersicum_N.eLcold_log.fold.change

(**a**)



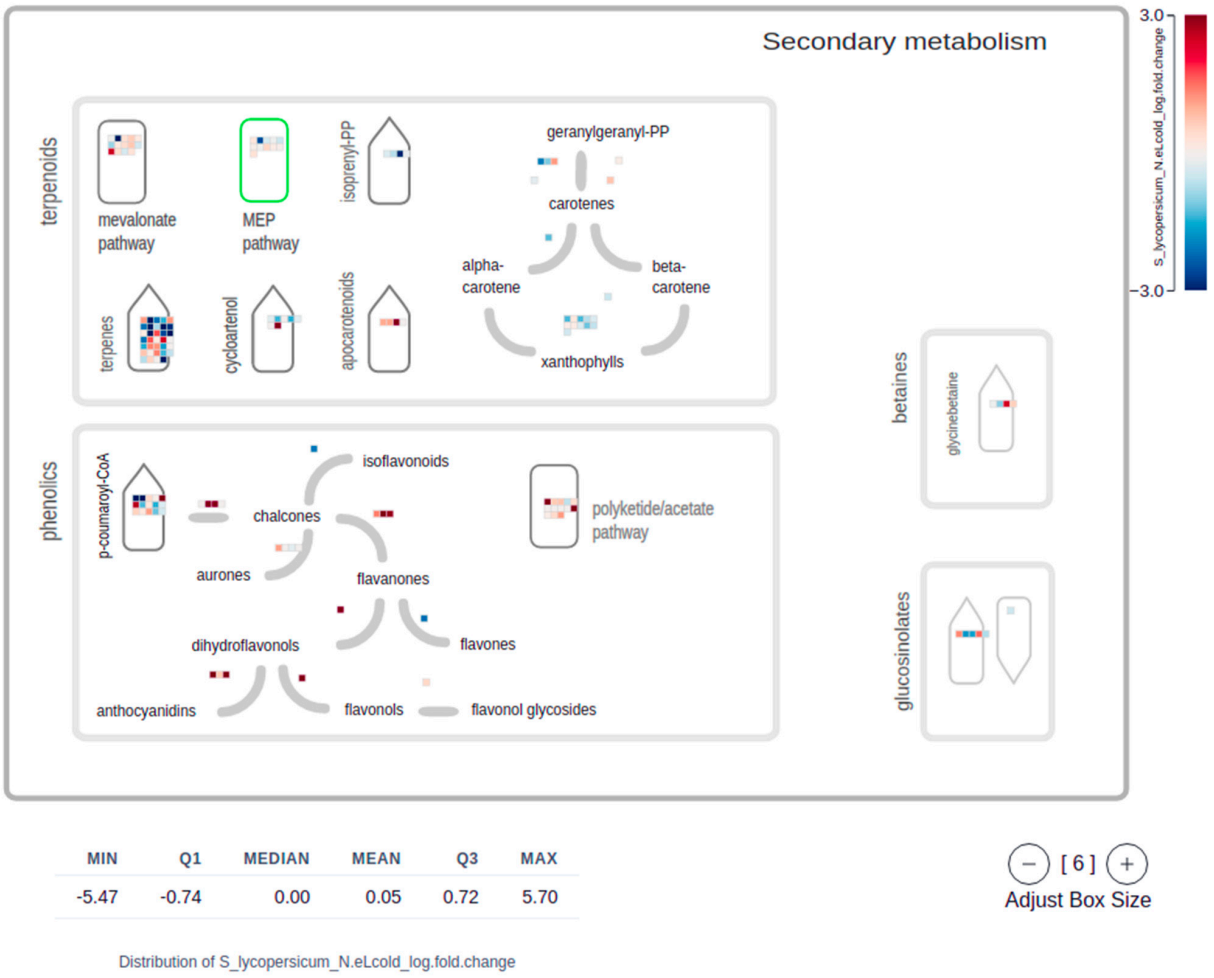
MIN	Q1	MEDIAN	MEAN	Q3	MAX
-8.22	-1.87	-1.13	-1.13	-0.29	6.70

[6] Adjust Box Size

Distribution of S_lycopersicum_N.eLcold_log.fold.change

(**b**)

Figure 4. *Cont.*



(**c**)

Figure 4. Screenshots of Mapman plots [1,2] used to visualize the genetic response to experimental stimuli in the form of metabolic sketches. Genes are mapped to areas in the sketches according to their molecular function. This gene function is directly extracted from the respective Mapman Bins [36] the genes are assigned to [3]. Each gene is represented by a single-colored box, where the color represents a numeric value, in this example the logarithmic fold change of gene expression (log-FC) between control and stress treatment. A legend in the top-right corner informs about the color-scale used to represent these numeric values. At the bottom of each Mapman plot, a summary statistic informs the user about the distribution of the respective numerical values, here the log-FC, shown in the plot. An interactive control in the bottom-right corner allows the user to adjust the sizes of the boxes, each representing one gene. Plot (**a**) shows a metabolic overview sketch and highlights how in the example data the expression of genes associated with photosynthesis is down regulated in *S. lycopersicum* following stress treatments (blue boxes in the respective top-right corner matrices). This down-regulation particularly affects genes of the light reaction, calvin cycle, and photorespiration pathways. Plot (**b**) sheds more light on this genetic response and zooms into the effect of stress treatments on the expression of genes associated with Photosystem I and II. Another detailed representation of the observed genetic response to stress treatment is shown in plot (**c**), elucidating how the expression of genes involved in terpene and carotene synthesis is down-regulated in *S. lycopersicum*.

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 such as GitHub or by

providing it in the form of a supplement to a scientific article. Readers can thus 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 is included in the GXP manual. A user can deploy a copy of GXP together with an exported database to a dedicated webserver. GitHub-pages offers this option free of charge. This creates a link that can be cited in upcoming publications, does not require any maintenance work, and will be functioning as long as GitHub is maintained. A comprehensive (seven steps) step-by-step guide to set up such a tailored copy of GXP with specific user data, plots, and analysis results has been included in the online manual.

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 pregenerated supplemental 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.

Figure 5. Screenshot of the result of an enrichment analysis (EA) carried out on the example data. This analysis is available in the "Tools" menu (screwdriver and wrench icon in the top panel). In the lightly faded-out background, the carried-out enrichment analysis can be seen. If more such analyses were done by clicking on the round plus icon in the bottom-right corner, they would also appear in this list. Clicking on the respective analysis opens the table shown in the highlighted foreground overlay. In it, the user is presented with significant results, while the button in the bottom-right corner "Show all Entries" enables the inspection of all, not only the annotations significantly tested for overrepresentation. In the shown example, the EA identified molecular gene functions overrepresented among genes whose expression is down-regulated in *S. lycopersicum* in response to the applied stress treatments, a combination of nitrogen deficiency (N-), chilling temperatures (cold), and elevated light intensity (eL). In this case, the results support the observation made for the example data earlier in Figure 4a,b, i.e. the response to stress treatments in the form of down-regulation of genes associated with (i) photosynthesis and (ii) terpene and carotene biosynthesis. Among the down-regulated genes, the molecular functions (i) "Chlorophyll a-b binding protein" in the "LHC-II complex" (Mapman Bin 1.1.1.1.1) and (ii) "UmamiT solute transporter" (Mapman Bin 24.2.1.5), a "sesquiterpene synthase", and "diterpene synthase" (Mapman Bin 9.1.4.2) are significantly overrepresented (all adjusted p-values < 0.006). Thus the Mapman plots and enrichment analyses truly help to elucidate the genetic response in *S. lycopersicum* to the stress treatments applied in the example study.

3. Discussion

The availability, efficiency, and relatively low cost of next-generation sequencing and metabolomics technologies allows their application in a wide variety of plant science research projects. Quantification of gene expression or metabolites 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 key for biological interpretation of these datasets. We revised platforms and software solutions that have been developed to provide the user with tools to quantify RNAseq raw data and contrast this quantified data in differential expression analysis. The revision includes tools that generate scientific plots, carry out clustering, principal component, and overrepresentation analysis [1,3,6–28,31,32,35,41]. However, the presented tools require either some programming expertise, or manual installation of software, or send potentially confidential data via the web to dedicated servers. Spreadsheet applications are often used to partially fill this gap, but the resulting plots are not interactive, and spreadsheet programs do not easily allow clustering, principal component, or overexpression analyses. In this context, we introduced Gene Expression Plotter (GXP) that provides the user with the means to load, analyze, and visualize quantitative and qualitative Omics data in the browser without the need for programming expertise or software installation. Additionally, GXP does all calculations locally in the browser without any need to submit data to servers. We incorporated into GXP the first mature and no-installation-required version of the popular Mapman tool [1,2]. This enables the user to summarize, in high quality plots, the gene functions, particularly up- and down-regulation, in a genetic response to experimental treatments, e.g., contrasting control and cold stress treatments. Hence, GXP offers simple solutions to explore and analyze Omics data and to generate publication grade plots. We furthermore explained how GXP can be used to publish Omics data along with plots and analysis results either by simply providing the community with a ready to use GXP database file, e.g., in the form of an article supplement, or by setting up an online copy of GXP already including the mentioned data, plots, and results. In this, the latter method can be done directly on GitHub free of charge by following seven simple steps that only require a web browser and a GitHub account.

To help the reader and user explore the value of using GXP and to benchmark GXP's functions, we included real research data from a published study on stress response in two tomato species [38]. We verified with the aid of one of the authors of this original study that (i) GXP reproduces and visualizes the already published findings, (ii) aids in the exploration of Omics data and promotes the formation of scientific hypotheses (Figures 1–5), and (iii) thus helps to elucidate, e.g., the genetic response to experimental stimuli, i.e., the original biological question motivating the study.

GXP is open-source software, runs entirely in the web browser and all code is automatically unit-tested, and thus is ensured to carry out correct calculations. Additionally, all code has been written adhering to current cutting-edge coding standards. Importantly, GXP is versatile in terms of its input. GXP consumes data about quantified entities, typically transcripts or chemical compounds (metabolites). Optionally, the user can supply further arbitrary either free text, categorical, and/or numerical information about the quantified entities and use, especially the latter two types of input data in GXP's plots and analyses. This generic approach to quantitative data visualization, exploration, and analysis, along with the option to easily setup a copy of GXP with specific user data, plots, and analysis results, specifically qualifies the GXP code base for reuse and extension in the typical open-source community approach. We indeed believe that GXP can become a platform for which over time, more and more functions, plots, and analyzes can be provided by third party developers.

4. Materials and Methods

Gene-Expression-Plotter (GXP) was implemented in TypeScript (version 4.5.2; https://www.typescriptlang.org/; access date 11 January 2022) as a standalone application (single page application; SPA) executed in the web-browser. This form of implementation implies that even though the SPA is obtained from a webserver, after opening it in the browser no data is ever sent to any webserver for analysis. All calculations, plots, and analysis are carried out right on the user's computer in the browser itself; thus, data confidentiality is guaranteed. For implementation, the ReactJS (version 17.0.1; https://reactjs.org/; access date 11 January 2022) and Chakra UI (version 1.7.2; https://chakra-ui.com/; access date 11 January 2022) libraries were used to build the user interface. The ViteJS library (version 1.3.6; https://vitejs.dev/; access date 11 January 2022) was used for tooling. GXP can be accessed on GitHub pages (https://usadellab.github.io/GeneExpressionPlots; access date 11 January 2022); every time a new version is pushed to GitHub, the new code is compiled and automatically deployed to GitHub pages using GitHub actions. GXP's source code is freely available on GitHub (https://github.com/usadellab/GeneExpressionPlots; access date 11 January 2022) under a GPL-3 license.

4.1. Input and Output Data

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. Memory state management has been implemented using the MobX library (version 6.3.7; https://mobx.js.org; access date 11 January 2022).

4.2. Gene Expression Plots

All introduced plots (see Section 2.2) were implemented with the plotly.js Javascript library (version 2.6.3; https://plotly.com/javascript/; access date 11 January 2022). 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:

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

with "abs" returning the absolute value of its real number argument.

Thus, complete anticorrelation as well as complete correlation are interpreted as maximum likeliness of numeric quantification vectors.

Euclidean distance measures are computed with the ml-distance Javascript library (version 3.0.0; https://github.com/mljs; access date 11 January 2022). Hierarchical clusters are identified using the ml-hclust library (version 3.1.0; https://github.com/mljs; access date 11 January 2022). The heatmap and the respective dendrogram visualizing the results of hierarchical clustering are plotted with the visx (version 2.4.0; https://github.com/airbnb/visx; access date 11 January 2022) library that incorporates the popular and well proven D3 library (version 7.1.1; https://d3js.org/; access date 11 January 2022into 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 library (version 4.0.2; https://github.com/mljs; access date 11 January 2022). In this calculation, all data points are

considered. The respective plot visualizing the first two principal components contributing most to the observed differences is created with plotly.js (version 2.6.3; https://plotly.com/javascript/; access date 11 January 2022).

4.5. MapMan Visualizations

GXP offers to summarize genetic expression or responses to contrasting experimental conditions in the form of Mapman plots [1] (see Section 2.4). All available canvas sketches (version X4.3) upon which to draw boxes to represent transcripts' quantification values were downloaded from the respective "MapMan Store" online repository (version X4.3; https://mapman.gabipd.org/mapmanstore; access date 11 January 2022) and included in the GXP package. Based on the proof-of-concept implementation [40], the visualization was programmed with the D3 library (version 7.1.1; https://github.com/d3/d3; access date 11 January 2022).

4.6. Overrepresentation Analysis

Gene Expression Plotter offers the user the ability to define sets of transcripts (genes) or sets of metabolites 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, i.e., the genome or the metabolome (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 hypergeometric probability distribution (HGD) [42]. The calculation of specific HGD *p*-values is carried out with the GNU scientific library (version 2.6; git://git.savannah.gnu.org/gsl.git; accessed on 11 January 2022) [43] which was compiled to web-assembly (version 1.0; https://webassembly.org/; access date 11 January 2022) for usage in the web-browser with Javascript. This compilation was done with emscripten (version 2.0.25; https://emscripten.org/; access date 11 January 2022). 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 are created and tested, respectively. 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 in Javascript and correctness of the calculations is confirmed by dedicated automatic unit software tests.

4.7. Example Dataset

To demonstrate GXP's qualities, published data sets from two tomato species were used [38]. In brief, two tomato species (*S. lycopersicum* and *S. pennellii*) were grown in rockwool blocks and watered with water for 16 days. Afterwards, the seedlings were fertilized with half-strength Hoagland solution for 14 days, followed by full-strength Hoagland solution (5 mM KNO3, 5 mM Ca(NO3)2, 2 mM MgSO4, 1 mM KH2PO4, 90 µM FeEDTA, plus micronutrients) for a further 11 days. A total of 6 weeks after germination, plants were stressed by nitrogen deficiency (N-), chilling temperatures (cold), warmer temperature regime (warm), or elevated light intensity (eL) and combinations thereof (Ncold, N-eL, eLcold and N-eLcold). After 1 week of stress treatment, leaflets of the fourth leaf (counted from the tip) were sampled, immediately frozen in liquid nitrogen and stored at −80 °C. Total RNA was extracted and treated with DNase followed by mRNA enrichment, and subsequently analyzed using an Illumina-platform (HiSeq) sequencing 2 × 75 bp paired-end reads.

Raw reads were trimmed using Trimmomatic [44]. An artificial transcriptome was built using default settings of StringTie [45,46] and back mapped to the genome of *S. lycopersicum*

(version ITAG 4). Data analysis was performed using R (version 3.5.2) [29]. Read abundances were analyzed using R-packages limma [47], edgeR [48], and tximport [49].

4.8. Automated Software Tests Ensure Correctness of Implemented Analyses

The code written to carry out the z-transformation, correlation, clustering, principal component, Fisher's exact test, overrepresentation, and *p*-value adjustment calculations provided within GXP are all verified for correctness using automated software, so called unit tests. These tests use data obtained from real life science projects and ensure that the respective functions behave correctly even in "edge cases" where data is unexpectedly abnorm, e.g., empty. These tests can be run automatically and thus ensure correctness of calculations even if future extensions are programmed. All tests are located in the cypress/tests/integration directory in GXP's GitHub code repository.

Supplementary Materials: The following supporting information can be downloaded at: https://www.mdpi.com/article/10.3390/plants11060745/s1, Text S1: Review of comparable tools to visualize and analyze RNAseq or Metabolomics quantitative data, Text S2: Table of used software packages and frameworks.

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. 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. and R.S. implemented a proof-of-concept Javascript software implementation of the MapMan visualizations. A.S. provided an extensive review in the form of testing and editing the GXP user manual, and delivered detailed user feedback. V.W. iteratively used and tested GXP in a RNAseq project, contributed to feature design, and provided extensive user feedback. J.J.R. provided preprocessed biological data. 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., C.E. and B.U. wrote the manuscript with the help of all authors. All authors have read and agreed to the published version of the manuscript.

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

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: All code, example and test data is available at https://github.com/usadellab/GeneExpressionPlots (accession date 11 January 2022).

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

References

1. Bolger, M.; Schwacke, R.; Usadel, B. MapMan Visualization of RNA-Seq Data Using Mercator4 Functional Annotations. *Methods Mol. Biol.* **2021**, *2354*, 195–212. [PubMed]
2. 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. [CrossRef] [PubMed]
3. Lohse, M.; Nagel, A.; Herter, T.; May, P.; Schroda, M.; Zrenner, R.; Tohge, T.; Fernie, A.R.; Stitt, M.; Usadel, B. Mercator: A fast and simple web server for genome scale functional annotation of plant sequence data. *Plant Cell Environ.* **2014**, *37*, 1250–1258. [CrossRef] [PubMed]
4. The InterPro Consortium; Mulder, N.J.; Apweiler, R.; Attwood, T.; Bairoch, A.; Bateman, A.; Binns, D.; Biswas, M.; Bradley, P.; Bork, P.; et al. InterPro: An integrated documentation resource for protein families, domains and functional sites. *Brief. Bioinform.* **2002**, *3*, 225–235. [CrossRef]
5. Ashburner, M.; Ball, C.A.; Blake, J.A.; Botstein, D.; Butler, H.; Cherry, J.M.; Davis, A.P.; Dolinski, K.; Dwight, S.S.; Eppig, J.T.; et al. Gene ontology: Tool for the unification of biology. The Gene Ontology Consortium. *Nat Genet.* **2000**, *25*, 25–29. [CrossRef]
6. Pimentel, H.; Bray, N.L.; Puente, S.; Melsted, P.; Pachter, L. Differential analysis of RNA-seq incorporating quantification uncertainty. *Nat. Methods* **2017**, *14*, 687–690. [CrossRef]

7. Su, W.; Sun, J.; Shimizu, K.; Kadota, K. TCC-GUI: A Shiny-based application for differential expression analysis of RNA-Seq count data. *BMC Res. Notes* **2019**, *12*, 133. [CrossRef]
8. Choi, K.; Ratner, N. iGEAK: An interactive gene expression analysis kit for seamless workflow using the R/shiny platform. *BMC Genom.* **2019**, *20*, 177. [CrossRef]
9. Sundararajan, Z.; Knoll, R.; Hombach, P.; Becker, M.; Schultze, J.L.; Ulas, T. Shiny-Seq: Advanced guided transcriptome analysis. *BMC Res. Notes* **2019**, *12*, 432. [CrossRef]
10. Marini, F.; Binder, H. pcaExplorer: An R/Bioconductor package for interacting with RNA-seq principal components. *BMC Bioinform.* **2019**, *20*, 331. [CrossRef]
11. Wang, S.; Zhang, Y.; Hu, C.; Zhang, N.; Gribskov, M.; Yang, H. Shiny-DEG: A Web Application to Analyze and Visualize Differentially Expressed Genes in RNA-seq. *Interdiscip Sci.* **2020**, *12*, 349–354. [CrossRef]
12. Reyes, A.L.P.; Silva, T.C.; Coetzee, S.G.; Plummer, J.T.; Davis, B.D.; Chen, S.; Hazelett, D.J.; Lawrenson, K.; Berman, B.P.; Gayther, S.A.; et al. GENAVi: A shiny web application for gene expression normalization, analysis and visualization. *BMC Genom.* **2019**, *20*, 745. [CrossRef]
13. Haering, M.; Habermann, B.H. RNfuzzyApp: An R shiny RNA-seq data analysis app for visualisation, differential expression analysis, time-series clustering and enrichment analysis. *F1000Research* **2021**, *10*, 654. [CrossRef]
14. Kim, S.C.; Yu, D.; Cho, S.B. COEX-Seq: Convert a Variety of Measurements of Gene Expression in RNA-Seq. *Genom. Inform.* **2018**, *16*, e36. [CrossRef]
15. Zhang, C.; Fan, C.; Gan, J.; Zhu, P.; Kong, L.; Li, C. iSeq: Web-Based RNA-seq Data Analysis and Visualization. *Methods Mol. Biol.* **2018**, *1754*, 167–181.
16. Li, R.; Hu, K.; Liu, H.; Green, M.R.; Zhu, L.J. OneStopRNAseq: A Web Application for Comprehensive and Efficient Analyses of RNA-Seq Data. *Genes* **2020**, *11*, 1165. [CrossRef]
17. Hoek, A.; Maibach, K.; Özmen, E.; Vazquez-Armendariz, A.I.; Mengel, J.P.; Hain, T.; Herold, S.; Goesmann, A. WASP: A versatile, web-accessible single cell RNA-Seq processing platform. *BMC Genom.* **2021**, *22*, 195. [CrossRef]
18. Harshbarger, J.; Kratz, A.; Carninci, P. DEIVA: A web application for interactive visual analysis of differential gene expression profiles. *BMC Genom.* **2017**, *18*, 47. [CrossRef]
19. Nelson, J.W.; Sklenar, J.; Barnes, A.P.; Minnier, J. The START App: A web-based RNAseq analysis and visualization resource. *Bioinformatics* **2017**, *33*, 447–449. [CrossRef]
20. Li, Y.; Andrade, J. DEApp: An interactive web interface for differential expression analysis of next generation sequence data. *Source Code Biol. Med.* **2017**, *12*, 2. [CrossRef]
21. Russo, F.; Angelini, C. RNASeqGUI: A GUI for analysing RNA-Seq data. *Bioinformatics* **2014**, *30*, 2514–2516. [CrossRef]
22. Bray, N.L.; Pimentel, H.; Melsted, P.; Pachter, L. Near-Optimal probabilistic RNA-seq quantification. *Nat Biotechnol.* **2016**, *34*, 525–527. [CrossRef]
23. Langmead, B.; Trapnell, C.; Pop, M.; Salzberg, S.L. Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biol.* **2009**, *10*, R25. [CrossRef]
24. Kim, D.; Langmead, B.; Salzberg, S.L. HISAT: A fast spliced aligner with low memory requirements. *Nat. Methods* **2015**, *12*, 357–360. [CrossRef]
25. Patro, R.; Duggal, G.; Love, M.I.; Irizarry, R.A.; Kingsford, C. Salmon provides fast and bias-aware quantification of transcript expression. *Nat. Methods* **2017**, *14*, 417–419. [CrossRef]
26. Howe, E.; Holton, K.; Nair, S.; Schlauch, D.; Sinha, R.; Quackenbush, J. MeV: MultiExperiment Viewer. In *Biomed. Informatics for Cancer Research*; Springer: Berlin/Heidelberg, Germany, 2010; pp. 267–277. [CrossRef]
27. Howe, E.; Holton, K.; Nair, S.; Schlauch, D.; Sinha, R.; Quackenbush, J. WebMeV: MultiExperiment Viewer. Available online: https://webmev.tm4.org/#/about (accessed on 11 January 2022).
28. Su, S.; Law, C.W.; Ah-Cann, C.; Asselin-Labat, M.-L.; Blewitt, M.E.; Ritchie, M.E. Glimma: Interactive graphics for gene expression analysis. *Bioinformatics* **2017**, *33*, 2050–2052. [CrossRef]
29. R Core Team. R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing. 2021. Available online: https://www.R-project.org/ (accessed on 11 January 2022).
30. R Studio Inc. Easy Web Applications in R. 2013. Available online: https://www.rstudio.com/shiny/ (accessed on 11 January 2022).
31. Hadizadeh Esfahani, A.; Maß, J.; Hallab, A.; Schuldt, B.M.; Nevarez, D.; Usadel, B.; Ott, M.C.; Buer, B.; Schuppert, A. Plant PhysioSpace: A robust tool to compare stress response across plant species. *Plant Physiol.* **2021**, *187*, 1795–1811. [CrossRef]
32. Hernández-De-Diego, R.; Tarazona, S.; Martínez-Mira, C.; Balzano-Nogueira, L.; Furió-Tarí, P.; Pappas, G.J., Jr.; Conesa, A. PaintOmics 3: A web resource for the pathway analysis and visualization of multi-omics data. *Nucleic Acids Res.* **2018**, *46*, W503–W509. [CrossRef] [PubMed]
33. Naithani, S.; Gupta, P.; Preece, J.; D'Eustachio, P.; Elser, J.L.; Garg, P.; Dikeman, D.A.; Kiff, J.; Cook, J.; Olson, A.; et al. Plant Reactome: A knowledgebase and resource for comparative pathway analysis. *Nucleic Acids Res.* **2020**, *48*, D1093–D1103. [CrossRef] [PubMed]
34. Waese, J.; Fan, J.; Pasha, A.; Yu, H.; Fucile, G.; Shi, R.; Cumming, M.; Kelley, L.A.; Sternberg, M.J.; Krishnakumar, V.; et al. ePlant: Visualizing and Exploring Multiple Levels of Data for Hypothesis Generation in Plant Biology. *Plant Cell.* **2017**, *29*, 1806–1821. [CrossRef] [PubMed]

35. Julkowska, M.M.; Saade, S.; Agarwal, G.; Gao, G.; Pailles, Y.; Morton, M.; Awlia, M.; Tester, M. MVApp—Multivariate Analysis Application for Streamlined Data Analysis and Curation. *Plant Physiol.* **2019**, *180*, 1261–1276. [CrossRef]
36. Schwacke, R.; Soto, G.Y.P.; Krause, K.; Bolger, A.M.; Arsova, B.; Hallab, A.; Gruden, K.; Stitt, M.; Bolger, M.; Usadel, B. MapMan4: A Refined Protein Classification and Annotation Framework Applicable to Multi-Omics Data Analysis. *Mol. Plant* **2019**, *12*, 879–892. [CrossRef]
37. Goto, M.K.A. KEGG: Kyoto Encyclopedia of Genes and Genomes. *Nucleic Acids Res.* **2000**, *28*, 27–30.
38. Reimer, J.J.; Thiele, B.; Biermann, R.T.; Junker-Frohn, L.V.; Wiese-Klinkenberg, A.; Usadel, B.; Wormit, A. 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. [CrossRef]
39. Thimm, O.; Bläsing, O.; Gibon, Y.; Nagel, A.; Meyer, S.; Krüger, P.; Selbig, J.; Müller, L.A.; Rhee, S.Y.; Stitt, M. 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. [CrossRef]
40. Usadel, B. MapManJS—Pure Web Implementations of MapMan. 2018. Available online: https://github.com/usadellab/MapManJS (accessed on 11 January 2022).
41. Lohse, M.; Bolger, A.M.; Nagel, A.; Fernie, A.R.; Lunn, J.E.; Stitt, M.; Usadel, B. RobiNA: A user-friendly, integrated software solution for RNA-Seq-based transcriptomics. *Nucleic Acids Res.* **2012**, *40*, W622–W627. [CrossRef]
42. Fisher, R.A. On the Interpretation of χ^2 from Contingency Tables, and the Calculation of P. *J. R. Stat. Soc.* **1922**, *85*, 87. [CrossRef]
43. The Gnu Scientific Library Team. *Gnu Scientific Library 2.0*; Samurai Media Limited: Surrey, UK, 2015.
44. Bolger, A.M.; Lohse, M.; Usadel, B. Trimmomatic: A flexible trimmer for Illumina sequence data. *Bioinformatics* **2014**, *30*, 2114–2120. [CrossRef]
45. Pertea, M.; Pertea, G.M.; Antonescu, C.M.; Chang, T.-C.; Mendell, J.T.; Salzberg, S.L. StringTie enables improved reconstruction of a transcriptome from RNA-seq reads. *Nat. Biotechnol.* **2015**, *33*, 290–295. [CrossRef]
46. Pertea, M.; Kim, D.; Pertea, G.M.; Leek, J.T.; Salzberg, S.L. Transcript-level expression analysis of RNA-seq experiments with HISAT, StringTie and Ballgown. *Nat. Protoc.* **2016**, *11*, 1650–1667. [CrossRef]
47. Ritchie, M.E.; Phipson, B.; Wu, D.; Hu, Y.; Law, C.W.; Shi, W.; Smyth, G.K. limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic Acids Res.* **2015**, *43*, e47. [CrossRef]
48. Robinson, M.D.; McCarthy, D.J.; Smyth, G.K. edgeR: A Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics* **2010**, *26*, 139–140. [CrossRef]
49. Soneson, C.; Love, M.I.; Robinson, M.D. Differential analyses for RNA-seq: Transcript-level estimates improve gene-level inferences. *F1000Research* **2015**, *4*, 1521. [CrossRef]