

# PaintOmics 4: new tools for the integrative analysis of multi-omics datasets supported by multiple pathway databases

Tianyuan Liu<sup>1</sup>, Pedro Salguero<sup>2</sup>, Marko Petek<sup>1b3</sup>, Carlos Martinez-Mira<sup>4</sup>, Leandro Balzano-Nogueira<sup>5</sup>, Živa Ramšak<sup>1b3</sup>, Lauren McIntyre<sup>6</sup>, Kristina Gruden<sup>3</sup>, Sonia Tarazona<sup>2</sup> and Ana Conesa<sup>1b7,8,\*</sup>

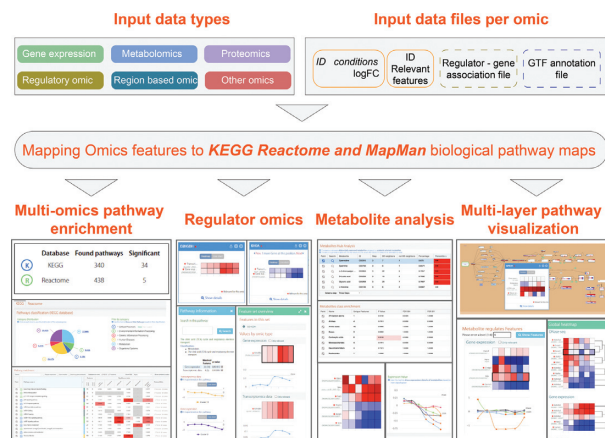
<sup>1</sup>Department of Mechanical Engineering, School of Engineering, Cardiff University, Cardiff, UK, <sup>2</sup>Department of Applied Statistics, Operations Research and Quality, Universitat Politècnica de València, Valencia, Spain, <sup>3</sup>Department of Biotechnology and Systems Biology, National Institute of Biology, Ljubljana, Slovenia, <sup>4</sup>Biobam Bioinformatics, Valencia, Spain, <sup>5</sup>Diabetes Institute, University of Florida, Gainesville, USA, <sup>6</sup>Department of Molecular Genetics and Microbiology, Genetics Institute, University of Florida, Gainesville, USA, <sup>7</sup>Institute for Integrative Systems Biology, Spanish National Research Council (CSIC), Paterna, Spain and <sup>8</sup>Department of Microbiology and Cell Science, University of Florida, Gainesville, FL, USA

Received March 12, 2022; Revised April 22, 2022; Editorial Decision April 25, 2022; Accepted April 25, 2022

## ABSTRACT

PaintOmics is a web server for the integrative analysis and visualisation of multi-omics datasets using biological pathway maps. PaintOmics 4 has several notable updates that improve and extend analyses. Three pathway databases are now supported: KEGG, Reactome and MapMan, providing more comprehensive pathway knowledge for animals and plants. New metabolite analysis methods fill gaps in traditional pathway-based enrichment methods. The metabolite hub analysis selects compounds with a high number of significant genes in their neighbouring network, suggesting regulation by gene expression changes. The metabolite class activity analysis tests the hypothesis that a metabolic class has a higher-than-expected proportion of significant elements, indicating that these compounds are regulated in the experiment. Finally, PaintOmics 4 includes a regulatory omics module to analyse the contribution of trans-regulatory layers (microRNA and transcription factors, RNA-binding proteins) to regulate pathways. We show the performance of PaintOmics 4 on both mouse and plant data to highlight how these new analysis features provide novel insights into regulatory biology. PaintOmics 4 is available at <https://paintomics.org/>.

## GRAPHICAL ABSTRACT



## INTRODUCTION

Multi-omics approaches have become popular in the study of a wide range of biological domains, with multi-omics datasets being now commonly obtained by individual investigators as well as large consortia (1–3). Moreover, the number and diversity of measured omics modalities have also increased, with former studies combining at most two or three omics platforms, and more recently genomics, transcriptomics, methylomics, chromatin accessibility, proteomics, metabolomics and/or lipidomics often combined in a single study (4–6). Multiple approaches have been proposed for the integrative analysis of these data (see (7,8) for reviews). Current methods can be broadly divided into three groups: detecting biomarkers, classifying samples, and in-

\*To whom correspondence should be addressed. Tel: +34 963544771; Email: ana.conesa@csic.es

ferring functional relationships between molecular layers (9). Regardless of the aim of these analyses, the subsequent biological interpretation of the analysis results is frequently difficult and time-consuming, as it requires human interaction and comprehension. Biological interpretation of multi-omics models is a grand challenge faced by investigators interrogating multi-omics data (9).

Three major strategies may be applied for the interpretation of multi-omics data. Overrepresentation and enrichment analyses are widely used in genomics and transcriptomics analyses (10). These methodologies have been adapted to the different omics data types (11–13), as well as to their integrative analysis (14–16). Enrichment methods are powerful tools to identify which biological processes are regulated in a given condition, but they are limited by the vocabulary, lack of comprehensive annotations and absence of mechanistic insight within and across omics layers.

An alternative approach is to use multi-partite networks, where nodes depict molecular entities and edges indicate regulatory or covariance information that can be extracted from the integrative statistical analysis (17). A variety of network (graphical) resources (e.g. Cytoscape (18), 3Omics (19), VisANT (20), OmicsAnalyst (21)) may be used to visualise relationships among biomolecules. Unfortunately, networks are frequently too large and interpretability is limited by the lack of context. A recent novel related method is multiSlide, which visualizes interconnected molecular features in heatmaps of multi-omics data sets (22).

A third option is to leverage existing biological knowledge represented in pathway maps to project multi-omics data and visualise them within a highly interpretable format. Examples are kaPPA-view 4 (23), MapMan4 (24) and PaintOmics 3 (14). Some tools may include several of these options. For instance, Cytoscape, OmicsAnalyst and Paintomics 3 also support enrichment analysis.

Pathway-based visualization methods also have limitations. First, they lack the flexibility to incorporate significant features suggested by the statistical analysis since absent in the available map. Moreover, interpretation is limited by the pathway boundaries decided by curators and by the amount and identity of the molecular features captured in the maps (25). Some of the measured features may not have a pathway location and some pathway components might not be measured. Additionally, some pathway elements might correspond to multiple measured features, such as a protein complex or a gene family. The limitations of pathway methods are particularly evident in untargeted and semi-targeted metabolomics data, since many measured compounds are unidentified, or only identified based on a large class of similar compounds (i.e. lipids) and therefore not present on current maps. Consequently, pathway-based enrichment methods behave poorly on metabolomics data and other strategies, such as metabolite-class enrichment (26) might be more appropriate. Here we present PaintOmics 4, a substantial expansion of the Paintomics 3 web server for pathway-based multi-omics data analysis (14,27), that addresses some of these current limitations in pathway definitions, metabolomics data integration and visualisation across molecular layers. An overview of the new functionalities is provided in Supplementary Figure S1.

## MATERIALS AND METHODS

### Implementation of Reactome and MapMan pathway databases into PaintOmics 4

Next to KEGG, PaintOmics 4 adds Reactome and MapMan to the list of pathway databases supported by the application (Supplementary Figure S2) (24,28). The new pathway data was incorporated into the existing PaintOmics MongoDB where pathways are classified into categories and features are mapped to the lowest pathway level. All PaintOmics data structures were modified to include a ‘source database’ field. This resulted in 904 new pathways distributed into 29 categories from Reactome and 25 new pathways in two categories from MapMan added to PaintOmics 4. Queries to the integrated database now proceed in batches to accommodate the larger number of entities to be searched. PaintOmics represents multi-omics data on pathways by overlaying feature expression and intensity values on their box positions in the pathway map using a colour scheme. In order to display multi-omics data on Reactome pathways, node coordinates available in pathway XML files were used. Since MapMan BIN coordinates are approximate, all MapMan pathway images were manually inspected and when required, XML files were edited to ensure the correct painting of data. Currently, Paintomics 4 implements KEGG Release 102.0 April 1 2022, Reactome Version 76 21 March 2021 and GoMapMan 25 May 2018. Databases are updated annually.

### Regulatory Omics functionality

PaintOmics 3 provided the Regulatory Omics option designed to upload data on features such as microRNA-seq, acting as regulators of gene expression. PaintOmics 4 extends this functionality to accept any type of trans-acting element operating on genes, transcripts or proteins and includes filtering functions to extract meaningful regulatory relationships. In addition to microRNA-seq, transcription factors (TF) and splicing factors (SF), detected by RNA-seq, RNA-binding proteins identified by CLIP-seq, etc., can be analysed with this option. The Regulatory Omics option takes a trans-regulatory-feature data matrix with expression or activity values for regulators in the conditions of the study. The regulator-gene/protein mapping file is provided by the user, together with an optional list of significant differentially expressed regulators. PaintOmics 4 filtering options include thresholds for positive or negative correlation to select the expected regulatory relationships. Applying these criteria, regulatory features will be mapped to their targeted features and their corresponding pathways. A pathway enrichment score is calculated either based on the number of regulators mapping to each pathway or on the number of regulated genes present in the pathway. Enriched pathways for the Regulatory Omics modality represent biological processes that are significantly impacted by that regulatory layer.

### Novel metabolomics interpretation methods

*Metabolite hub analysis.* One of the goals of multi-omics studies that combine metabolomics and gene expression or

proteomics data is to associate changes in metabolite levels with the regulation of the enzymes that may contribute to metabolite turnover. PaintOmics 4 leverages pathway information to identify metabolites that have a high proportion of differentially expressed features in their close network. Two tests were implemented (Figure 1). The binomial test is used to evaluate for each differentially expressed metabolite (DEM) the hypothesis that, given the overall percentage of differentially expressed genes (DEG)  $p_0$  in the dataset, the proportion of DEG linked to the metabolite is significantly higher than  $p_0$ . The hub analysis evaluates genes directly connected in the network (step 1) or genes associated with the metabolite through up to 3 intermediate nodes (steps 2 to 4).  $P$ -values are corrected for multiple testing (29). Alternatively, the distribution in the overall metabolic network of the percentage of neighbouring DEG for metabolite nodes is computed and the percentile position of each measured metabolite in this distribution is calculated. Metabolites with a high percentile value have a higher proportion of connecting DEG than the majority of the metabolites in the database.

**Metabolite class activity.** To test the hypothesis of a metabolite class being regulated, PaintOmics 4 implements a metabolite class activity analysis tool, where a binomial test is used to assess the hypothesis of the proportion of significant compounds in a given measured metabolite class being higher than a user-defined threshold. In case the user does not define an activity threshold, PaintOmics 4 will use the average percentage of significant metabolites as threshold for the null hypothesis.  $P$ -values are corrected for multiple testing (29). These novel metabolomics analysis tools are provided as a separate tab in the main PaintOmics results panel that includes hyperlinks to facilitate navigation between metabolite data, neighbouring genes and metabolic pathways (Figure 1).

### Metagenes for nodes and pathways

PaintOmics displays omics data on pathways maps by colouring the node position of the omic feature according to its experimental value. When a node contains multiple features, e.g. MapMan BINs, the map topology may not be able to accommodate the amount of data. In order to address this problem, metagenes are computed for pathway nodes with more than four matching features (30), resulting in a compressed representation of omics data in complex nodes that fits available space on the map. Note that when one node contains features with different profiles, the analysis might return multiple metagenes for the node, one per profile type.

### Use case datasets

PaintOmics 4 functionalities were demonstrated using two different multi-omics datasets. The STATegra data that collects multi-omics data for a mouse B-cell differentiation process triggered by the expression of the Ikaros TF (4).

The second dataset corresponds to an *Arabidopsis* study that evaluated the root transcriptional and metabolic profile of a BRL3 overexpressing mutant in drought conditions.

BRL3 is a vascular-enriched member of the brassinosteroid family, which was speculated to confer drought tolerance (31).

## RESULTS

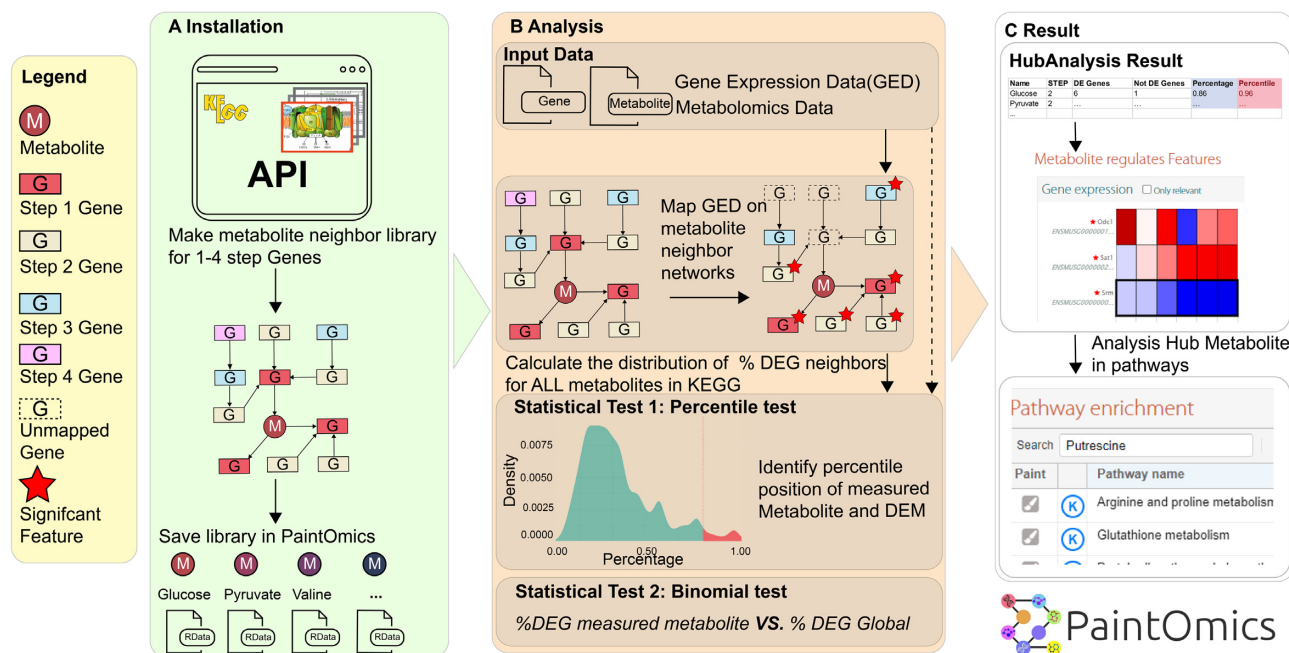
### STATegra data analysis with PaintOmics 4 reveals novel regulatory events during B-cell differentiation

We used the STATegra multi-omics dataset describing the differentiation of the murine B3 cell line from a proliferating pre-BI state to differentiated pre-BII (4) to demonstrate PaintOmics 4 functionalities. RNA-seq, micro-RNA-seq, DNase-seq, metabolomics and TF data were available. The dataset contains temporal data for 13 123 genes, 469 microRNAs, 10 272 DNaseq regions, 320 TFs and 60 metabolites, of which 5224, 172, 5099, 180 and 40 features, respectively, were found to have significant differences along the differentiation course (4). Five STATegra omics modalities were run in PaintOmics 4 selecting both KEGG and Reactome as pathway databases. KEGG disease and organismal pathways were excluded from the analysis. Data mapped to a total of 169 KEGG and 439 Reactome pathways, of which 14 and 11, respectively, were found significant by the Fisher combined  $P$ -value method (32) that jointly considers all omics modalities. The full list of significant pathways is provided in Supplementary Table S1.

**PaintOmics 4 indicates a multi-layered control of B-cell differentiation.** We first analysed the overall patterns of pathway changes across molecular layers using the pathway network analysis in PaintOmics 4 to focus attention on the analysis of our Regulatory Omics types, microRNA and TFs (Supplementary Figure S3). This tool revealed that pathways change during B-cell differentiation according to 2–3 patterns. For gene expression, most metabolic and genetic information pathways were downregulated, while signalling pathways showed both up and down regulation trends. microRNAs associated genes in these pathways tended to be upregulated at late-time points (pre-BII stage) in metabolic and genetic information processing pathways, while upregulation for signalling pathways took place at early time points (pre-BI stage). Finally, TF regulation showed the opposite behaviour, with TFs that bind metabolic and genetic information processing pathways being downregulated as differentiation progresses but upregulated for signalling pathway genes. These results indicate a highly coordinated control of biological pathways during B-cell differentiation, characterized by the transcriptional activation of signalling pathways and the downregulation of metabolic activities, with transcriptional (TF) and post-transcriptional (microRNA) mechanisms contributing to this program.

**KEGG and Reactome complemented each other in the analysis of STATegra multi-omics data.** The analysis of enriched KEGG and Reactome pathways indicated commonalities and differences between the two resources (Supplementary Table S1). Both databases reported enrichment of glucose, amino-acids and nucleotide metabolic processes, and of





**Figure 1.** New metabolomics analysis in PaintOmics 4. (A) Neighbouring genes for each metabolite at 1 to 4 network steps are identified. (B) The percentile and binomial tests are used to identify metabolites with a high density of DEGs in their proximal network. (C) Metabolites and Genes identified in the analysis are shown as heatmaps with links to associated pathways. DEM: differentially expressed metabolite; DEG: differentially expressed gene.

p53 signalling. However, most significant signalling pathways did not coincide, possibly due to different pathway definitions between KEGG and Reactome. The combined analysis revealed many of the known processes operating during the differentiation of the hematopoietic and immune cell lineages, e.g. Interleukin-2 family signalling, Interferon gamma signalling, RAF-independent MAPK1/3 signalling, for Reactome, and JAK-STAT signalling, FOXO signalling, and Hippo signalling for KEGG (33). Reactome but not KEGG identified the RET signalling pathway as enriched (combined  $P$ -value = 0.029). RET is a tyrosine kinase receptor essential for embryonic development (34) which has also been found to be expressed in hematopoietic tissues, suggesting a role in the development of the immune system. Specifically, RET induces the expression of chemokines and cytokines, and downregulates chemokine/cytokine receptors (35). In the B3 cell differentiation process, we found a strong upregulation of RET and other associated membrane receptors (Figure 2A and B). A concordant regulation by many microRNAs and transcription factors was also found (Figure 2C). Interestingly, a different component of the RET signalling pathway is the DOK protein, represented by three family members (1, 3 and 4) in our data. These genes were downregulated at the transition between pre-BI and pre-BII stages, together with their only significant TF (STAT4: ENSMUSG00000062939), meanwhile the associated microRNA (mir-188-3p) was strongly upregulated. STAT4 is a key TF of the immune cell lineage (36), and mir-188-3p has been reported to regulate cell proliferation (37). Whether these different regulatory relationships in the RET signalling pathway represent specific contributions to the differentiation of B-cells remains to be investigated. Our results showcase the power of PaintOmics 4 pathway-based

multi-omics analysis to present and dissect a diversity of multi-layered regulatory relationships.

*PaintOmics 4 novel metabolomics analysis tools highlight metabolite roles in B-cell differentiation.* Pathway enrichment analysis based on metabolomics data did not detect any significant pathways, possibly due to the limited number of metabolites in this dataset. However, the Metabolite Class Activity analysis identified amino acids as significant and, marginally, also carboxylic acids (Figure 3A). Accordingly, most amino acids had higher values at early time points (Figure 3B), which is consistent with the high proliferative state of the pre-BI stage where protein synthesis is highly active (38). Moreover, the metabolite hub analysis of the STATegra data highlighted a number of compounds as having a high proportion of DEG in their proximal network, among them three polyamines: spermidine, putrescine, and spermine (Figure 3C). These metabolites have higher levels at the pre-BI stage and decrease as cells differentiate towards pre-BII (Figure 3D). Neighbouring genes for these metabolites included Srm (Spermidine synthase), Sms (Spermine synthase) and Amd1 (*S*-adenosylmethionine decarboxylase proenzyme 1) which were downregulated during differentiation (Figure 3E). Ikaros triggers pre-B-cell differentiation through repression of the c-Myc transcription factor (39), which is known to regulate the expression of polyamine synthesis genes such as Srm, Sms and Amd1 genes (40). Therefore, repression of c-Myc is consistent with the observed downregulation of these genes and of the three polyamines in the STATegra dataset. Moreover, spermidine, putrescine, and spermine have an established role in cell proliferation (41) and are likely to play a key role in T-cell and B-cell differentiation (42). The PaintOmics 4 analysis highlights the polyamine

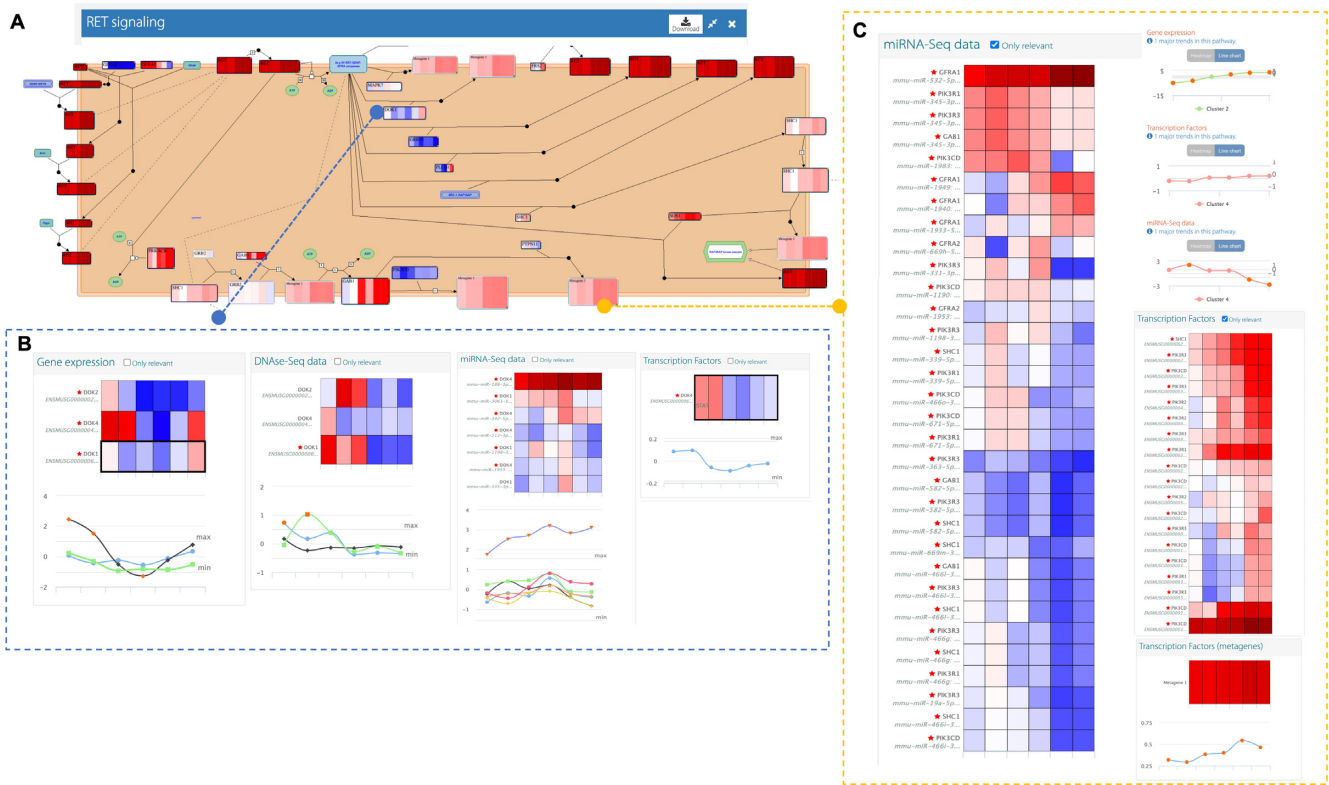


Figure 2. PaintOmics multi-omics analysis of RET signalling pathway. (A) RET signalling pathway. (B) Multi-omics data for metagene 1 containing receptor proteins. (C) Multi-omics data for DOK gene family.

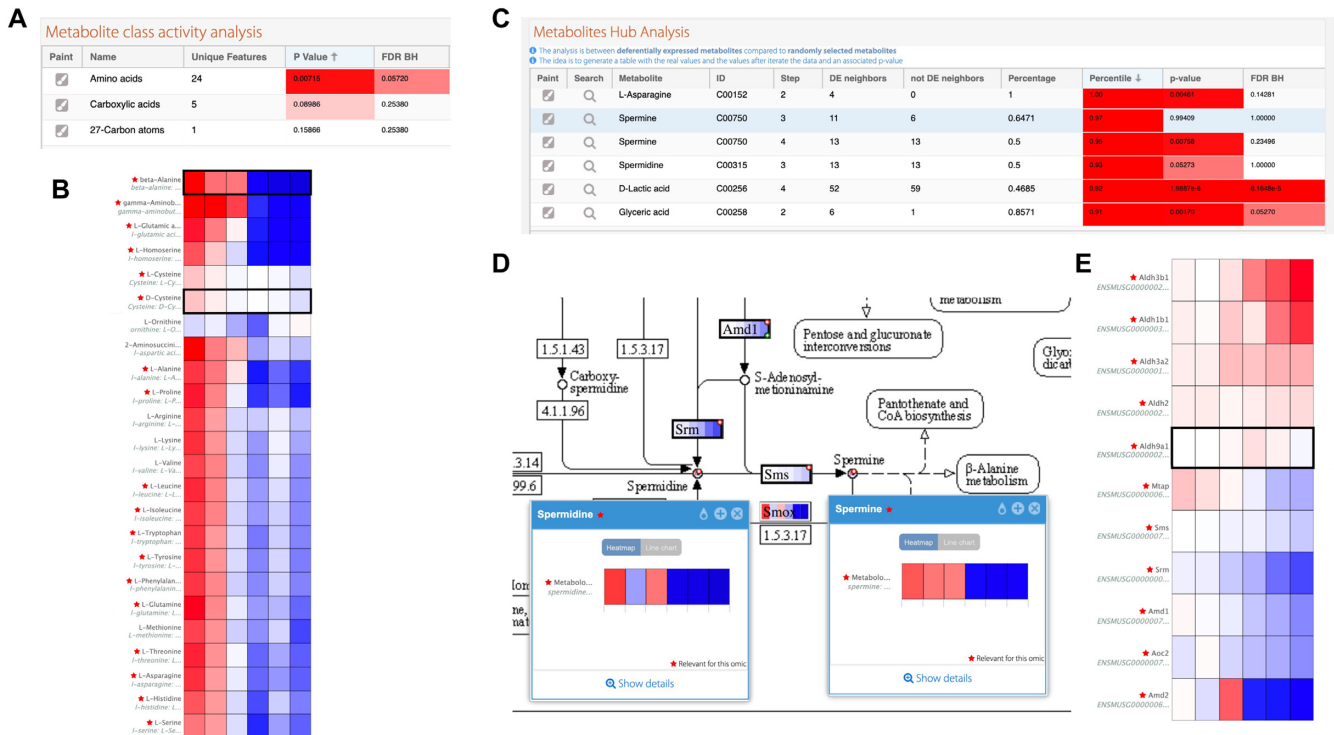


Figure 3. Metabolite class and hub analyses of STATegra data. (A) PaintOmics 4 panel with Metabolite Class Activity analysis results. (B) Heatmap of amino-acid values during B3 cell differentiation. (C) PaintOmics 4 panel with Metabolite Hub Analysis results. (D) Neighbouring network for spermine and spermidine in KEGG. (E) Expression values for spermine neighbouring differentially expressed genes.

metabolite/gene regulatory hub during the murine B3 cell differentiation course.

#### PaintOmics 4 analysis of *Arabidopsis* drought response leverages MapMan and KEGG pathways for novel pathway insights

We used the *Arabidopsis* BRL3ox study (31) to showcase the utility of PaintOmics 4 for the interpretation of multi-omics data from plants. This study evaluated the response to drought conditions of a mutant overexpressing BRL3, a plant brassinosteroid receptor. Roots RNA-seq and metabolomics data after 5 days of drought treatment, together with a list of differentially expressed features, were available.

We run PaintOmics 4 on the BRL3ox data using the KEGG and MapMan databases. While KEGG is a general pathway database, MapMan is tailored to plants and contains a more detailed representation of plant-specific pathways. A total of 18 and 8 enriched pathways were found for KEGG and MapMan databases, respectively (combined adjusted  $P$ -value < 0.05) (Figure 4A, Supplementary Table S2). KEGG results indicated enrichment of multiple metabolic pathways (e.g. Phenylpropanoid biosynthesis, Biosynthesis of secondary metabolites, Brassinosteroid metabolism, among others), as well as signalling pathways, including MAPK signalling pathway, ABC transporters and the general Plant hormone signal transduction. MapMan, however, returned an enrichment picture that highlighted the role of specific hormones (synthesis of jasmonic acid, GABA, abscisic acid), secondary metabolites (flavonoids, chorismate, polyamines) and the synthesis of lignin in the BRL3ox response to drought, complementing the biological interpretation provided by KEGG. Many of these processes were discussed in the original publication, supporting the robustness of PaintOmics 4 analysis. Here, we discussed two novel results not identified in the previous study.

The Synthesis of Chorismate pathway was found enriched by the PaintOmics 4 MapMan analysis. Visual inspection of this pathway indicated a general downregulation in the BRL3ox mutant under drought conditions (Figure 4B). This pathway catalyses the formation of chorismate, the last step in the shikimate pathway and a branch-point metabolite used for the synthesis of aromatic amino acids, p-aminobenzoic acid, folate, and other cyclic metabolites such as ubiquinone (43). Under abiotic stress conditions, plants activate the synthesis of aromatic compounds through the shikimate pathway, improving salt stress tolerance but not causing oxidative or drought stress (44). However, some aromatic compounds, such as m-tyrosine, inhibit the growth of many plant species by slowing down root development, and high tryptophan levels have been reported to inhibit root growth (45). Interestingly, both tyrosine and tryptophan were down regulated in BRL3ox plants compared to the WT plants under drought conditions (31). We speculate that BRL3ox overexpression results in downregulation of chorismate pathway and aromatic compound synthesis under drought, improving drought tolerance without growth arrest.

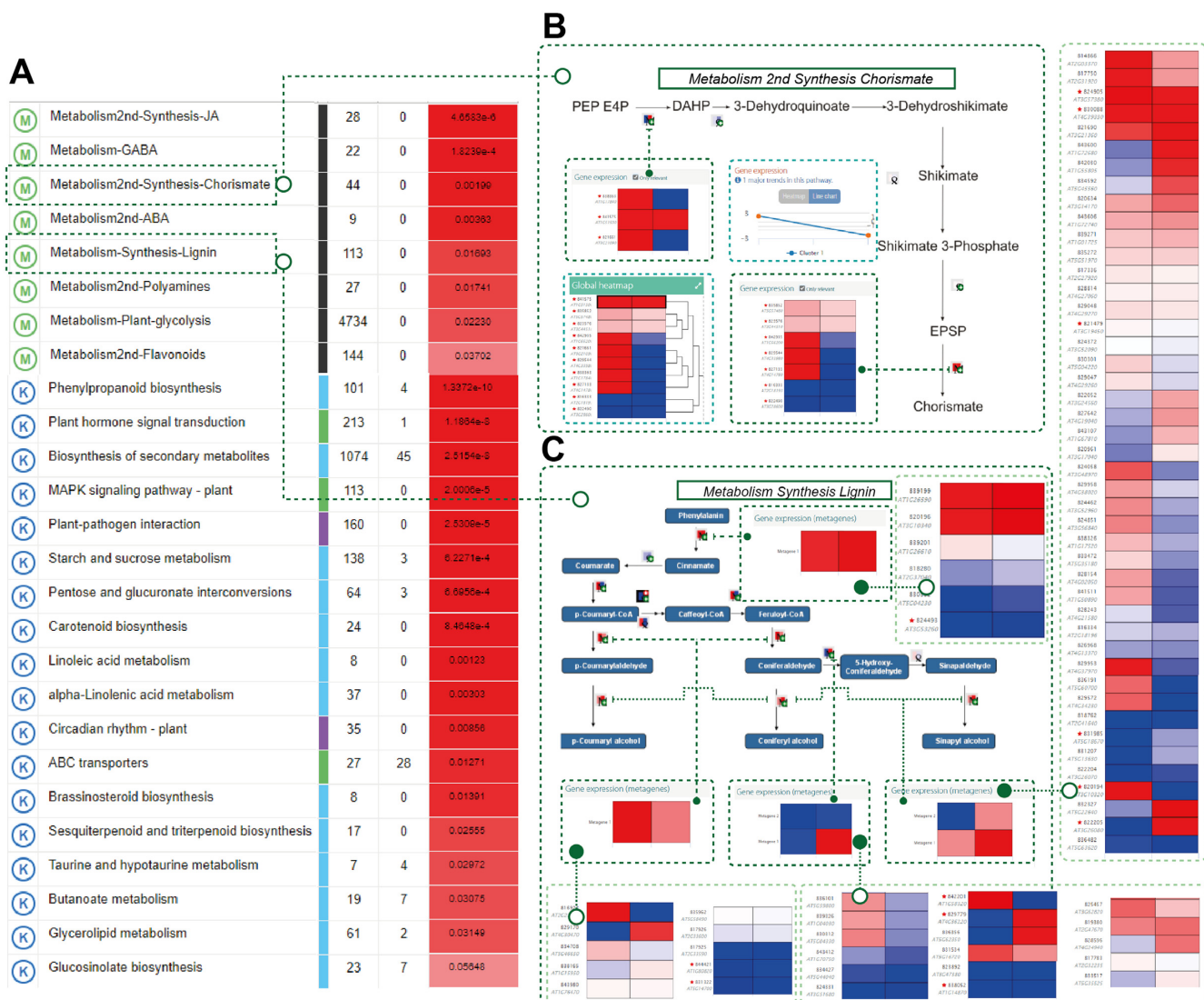
Another significant MapMan pathway was synthesis of lignin, which shows upregulation in treated BRL3ox plants compared to the WT (Figure 4C). This pathway is represented in the MapMan database by 113 different genes distributed into 12 reactions, which implies that multiple steps in the pathways are associated with a large number of genes. For example, the last steps of the pathways represent the conversion from aldehyde to alcohol of coumaryl, coniferyl and synapil, catalysed by cinnamyl alcohol dehydrogenase protein family, with 44 associated genes, which would be difficult to include in the map. The PaintOmics metagene function calculated two major upregulating trends for this node, providing a more interpretable representation for the reaction (Figure 4C). Lignin is known to play an important role in improving plants' drought resistance through water transport and mechanical support (46). The PaintOmics 4 analysis suggests that the coordinated upregulation of gene families catalysing monolignol synthesis is part of the BRL3ox mechanism of drought resistance.

## DISCUSSION

While multi-omics studies have increased in number, scope, sample size and diversity of measured omics modalities, and a wide range of integrative statistical data analysis tools have been proposed, the biological interpretation of these data is still a major challenge. PaintOmics 4 addresses this problem by projecting processed values of multi-omics features onto pathway maps. However, the success of a pathway analysis strategy depends on the amount and identity of the features captured in the database and on their distribution among pathway definitions. Different biological pathway tools focus on different types of organisms, cellular processes, and types of reactions, thereby offering different but complementary views of the biology. Our use cases showed that, by implementing different databases in PaintOmics 4, complementary information about the experiment is gained, and interpretation of the data is improved. To the best of our knowledge, PaintOmics 4 is the only tool to combine these three pathway databases under one analysis.

Another distinctive aspect of PaintOmics 4 is its versatility in representing virtually any omics modality on pathways. This implies that omics layers that act as direct or indirect regulators of pathway activities, such as epigenetic marks or microRNAs, can be interpreted from the perspective of the pathways that they impact. PaintOmics 4 not only provides enrichment analysis for these layers, but further improves interpretability by showing the relationship between the regulatory feature and regulated gene on the pathway maps. Importantly, and in contrast to other methods that do accept regulatory omics data for pathway views (21) PaintOmics 4 displays multiple regulatory information simultaneously on the pathway representation both globally and for each pathway node, allowing for different levels of granularity in the analysis. This unique functionality implies that both *cis* and *trans* regulatory relationships can be directly linked to the mechanistic representation of the biological process captured by the pathway, thereby facilitating the understanding of the multi-layer component of the multi-omics study.





**Figure 4.** PaintOmics 4 analysis of Arabidopsis BRL3ox data. (A) List of enriched pathways. (B) MapMan synthesis chorismate pathway. The pathway shows a downregulation trend at 5 days. (C) MapMan synthesis lignin pathway. The pathway contains a large number of BINs with many associated genes. The figure shows both metagene representations and the heatmaps of the corresponding genes linked by green connectors.

Analysis and interpretation of metabolomics experiments are particularly challenging because usually pathways only accurately represent a fraction of the measured metabolites (e.g. in lipidomics, many different compounds might be represented by the same entity in the pathway) and/or, many metabolites present in the metabolic network are not measured in the metabolomics experiment. Moreover, *a priori* hypotheses of metabolite relevance for the study may dictate the type of metabolomics assay to be run (our STATEgra and BLR3ox use cases are examples of this). In such a case, a large fraction of the measured compounds may show significant changes, and this jeopardises any enrichment analysis strategy. To still be able to evaluate if the targeted metabolite types are affected in the experiment, we introduced the Metabolite Class Activity analysis, where we test the hypothesis of the measured metabolite class having a high proportion of compounds with significant changes. Another type of question that multi-omics studies involv-

ing metabolomics may pose is the link between the metabolite change and the regulation of the expression of the genes, proteins or other compounds that could modify their metabolite levels. This is relevant, for example, when looking for metabolite biomarkers or targets of metabolic control. This question can be addressed by a gene-metabolite bipartite network analysis (47) or by a flux balance analysis (FBA) strategy (48). However, gene-metabolite correlation networks usually lack the pathway context and FBA returns information on fluxes rather than compounds, requires a complex mathematical formulation, and still has limited adaptations to multi-omics data (49). In PaintOmics 4, we propose a simple approach based on the analysis of the local metabolite network to identify the proportion of differentially expressed features. Applied to our STATEgra dataset, this method identified a number of metabolites displaying hub-like properties (spermidine, spermine, citric acid, etc.) and complemented the limited metabolomics en-

richment analysis results. As the hub analysis panel also provides links to the involved features and pathways, the user has all the information on hand to navigate and interpret the data.

In summary, PaintOmics 4 is a web server for the multi-layered biological interpretation of multi-omics data that includes a wealth of resources for a comprehensive and interactive analysis. Future developments will address the growing applications of the multi-omics paradigm to assist precision medicine and single-cell analyses.

## DATA AVAILABILITY

PaintOmics 4 web server is available at <https://paintomics.org/>. PaintOmics 4 code and documentation are available at <https://github.com/ConesaLab/paintomics4>.

## SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online.

## ACKNOWLEDGEMENTS

We thank Ana Caño and Fidel Lozano for help accessing the BRL3ox data.

## FUNDING

Spanish Ministry of Science and Innovation [BIO2015-1658-R and PID2020-119537RB-100 to A.C. and S.T.]; National Institutes of Health [R03 CA222444 to A.C. and L.M.M.]; Slovene Research Agency Program [P4-0165]; European Union's Horizon 2020 [GA 2020 862-858]; COST Action CA15110 STSM travel grant supported by European Cooperation in Science and Technology; Generalitat Valenciana [ACIF2019/081 to P.S.]. Spanish Ministry of Science and Innovation. Funding for open access charge: Spanish Ministry of Science and Innovation [PID2020-119537RB-100].

*Conflict of interest statement.* None declared.

## REFERENCES

- National Institutes of Health (2022) *NHLBI trans-omics for precision medicine*. <https://www.nhlbi.nih.gov/science/trans-omics-precision-medicine-topmed-program>.
- Cancemi,P., Buttacavoli,M., Di Cara,G., Albanese,N.N., Bivona,S., Pucci-Minafra,I. and Feo I,S. (2018) A multiomics analysis of S100 protein family in breast cancer. *Oncotarget*, **9**, 29064–29081.
- Stare,T., Ramšak,Ž., Križnik,M. and Gruden,K. (2019) Multiomics analysis of tolerant interaction of potato with potato virus Y. *Sci. Data*, **6**, 250.
- Gomez-Cabrero,D., Tarazona,S., Ferreirós-Vidal,I., Ramirez,R.N., Company,C., Schmidt,A., Reijmers,T., von Saint Paul,V., Marabita,F., Rodríguez-Ubrea,J. *et al.* (2019) STATegra, A comprehensive multi-omics dataset of B-cell differentiation in mouse. *Sci. Data*, **6**, 256.
- Ciriello,G., Miller,M., Aksoy,B., Senbabaoglu,Y., Schultz,N. and Sander,C. (2013) Emerging landscape of oncogenic signatures across human cancers. *Nat. Genet.*, **45**, 1127–1133.
- Stunnenberg,H. and Hirst,M. (2016) The international human epigenome consortium: A blueprint for scientific collaboration and discovery. *Cell*, **167**, 1897.
- Bersanelli,M., Mosca,E., Remondini,D., Giampieri,E., Sala,C., Castellani,G. and Milanese,L. (2016) Methods for the integration of multi-omics data: mathematical aspects. *BMC Bioinformatics*, **17**, S15.
- Meng,C., Zeleznik,O., Thallinger,G., Kuster,B., Gholami,A. and Culhane,A. (2016) Dimension reduction techniques for the integrative analysis of multi-omics data. *Brief. Bioinform.*, **17**, 628–641.
- Tarazona,S., Arzalluz-Luque,A. and Conesa,A. (2021) Undisclosed, unmet and neglected challenges in multi-omics studies. *Nat. Comput. Sci.*, **1**, 395–402.
- Subramanian,A., Tamayo,P., Mootha,V., Mukherjee,S., Ebert,B.L., Gillette,M.A., Paulovich,A., Pomeroy,S.L., Golub,T.R., Lander,E.S. *et al.* (2005) Gene set enrichment analysis: A knowledge-based approach for interpreting genome-wide expression profiles. *Proc. Nat. Acad. Sci. U.S.A.*, **102**, 15545–15550.
- Hong,M., Pawitan,Y., Magnusson,P. and Prince,J. (2009) Strategies and issues in the detection of pathway enrichment in genome-wide association studies. *Hum. Genet.*, **126**, 289–301.
- Garcia-Garcia,F., Panadero,J., Dopazo,J. and Montaner,D. (2016) Integrated gene set analysis for microRNA studies. *Bioinformatics*, **32**, 2809–2816.
- Maksimovic,J., Oshlack,A. and Phipson,B. (2021) Gene set enrichment analysis for genome-wide DNA methylation data. *Genome Biol.*, **22**, 173–173.
- Diego,R., Tarazona,S., Martínez-Mira,C., Balzano-Nogueira,L., Furió-Tarí,P., Pappas,J. and Conesa,A. (2018) PaintOmics 3: a web resource for the pathway analysis and visualization of multi-omics data. *Nucleic Acids Res.*, **46**, W503–W509.
- Canzler,S. and Hackermüller,J. (2020) multiGSEA: a GSEA-based pathway enrichment analysis for multi-omics data. *BMC Bioinformatics*, **21**, 561–561.
- Paczkowska,M., Barenboim,J., Sintupisut,N., Fox,N.S., Zhu,H., Abd-Rabbo,D., Mee,M.W., Boutros,P.C., Reimand,J., PCAWG Drivers and Functional Interpretation Working Group *et al.* (PCAWG Drivers and Functional Interpretation Working Group2020) Integrative pathway enrichment analysis of multivariate omics data. *Nat. Commun.*, **11**, 735.
- Pavlopoulos,G., Kontou,P., Pavlopoulou,A., Bouyioukos,C., Markou,E. and Bagos,P. (2018) Bipartite graphs in systems biology and medicine: a survey of methods and applications. *GigaScience*, **7**, giy014.
- Shannon,P., Markiel,A., Ozier,O., Baliga I,N.S., Wang,J.T., Ramage,D., Amin,N., Schwikowski I,B. and Ideker,T. (2003) Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Res.*, **13**, 2498–2504.
- Kuo,T., Tian,T. and Tseng,Y. (2013) 3Omics: a web-based systems biology tool for analysis, integration and visualization of human transcriptomic, proteomic and metabolomic data. *BMC Syst. Biol.*, **7**, 64.
- Hu,Z. (2014) Using VisANT to analyze networks. *Curr. Prot. Bioinformatics*, **45**, 8.8.1–8.8.39.
- Zhou,G., Ewald,J. and Xia,J. (2021) OmicsAnalyst: a comprehensive web-based platform for visual analytics of multi-omics data. *Nucleic Acids Res.*, **49**, W476–W482.
- Ghosh,S., Datta,A. and Choi,H. (2021) multiSLIDE is a web server for exploring connected elements of biological pathways in multi-omics data. *Nat. Commun.*, **12**, 2279.
- Sakurai,N., Ara,T., Ogata,Y., Sano,R., Ohno,T., Sugiyama,K., Hiruta,A., Yamazaki,K., Yano,K. and Aoki,K. (2011) KaPPA-View4: a metabolic pathway database for representation and analysis of correlation networks of gene co-expression and metabolite co-accumulation and omics data. *Nucleic Acids Res.*, **39**, D677–D684.
- Schwacke,R., Ponce-Soto,G., Krause,K., Bolger,A.M., Arsova I,B., Hallab I,A., Gruden,K., Stitt,M., Bolger I,M.E., Usadel,B. *et al.* (2019) MapMan4: A refined protein classification and annotation framework applicable to multi-omics data analysis. *Mol. Plant.*, **12**, 879–892.
- Conesa,A. and Hernández,R. (2014) Omics data integration in systems biology. **64**, 441–459.
- Molenaar,M., Jeucken,A., Wassenaar,T., Lest,C., Brouwers,J. and Helms,J. (2019) LION/web: a web-based ontology enrichment tool for lipidomic data analysis. *GigaScience*, **8**, giz061.
- García-Alcalde,F., García-Lopez,F., Dopazo,J. and Conesa,A. (2011) Paintomics: a web based tool for the joint visualization of transcriptomics and metabolomics data. *Bioinformatics*, **27**, 137–139.
- Haw,R., Hermjakob,H., D'Eustachio,P. and Stein,L. (2011) Reactome pathway analysis to enrich biological discovery in proteomics datasets. *Proteomics*, **11**, 3598–3613.



29. Benjamini, Y. and Hochberg, Y. (1995) Controlling the false discovery rate: a practical and powerful approach to multiple testing. *J. R. Stat. Soc. Ser. B*, **57**, 289–300.
30. Ponzoni, I., Nueda, M., Tarazona, S., Götz, S., Montaner, D., Dussaut, J.S., Dopazo, J. and Conesa, A. (2014) Pathway network inference from gene expression data. *BMC Syst. Biol.*, **8**, S7.
31. Fàbregas, N., Lozano-Elena, F., Blasco-Escámez, D., Tohge, T., Martínez-Andújar, C., Albacete, A., Osorio, S., Bustamante, M., Riechmann, J.L., Nomura, T. *et al.* (2018) Overexpression of the vascular brassinosteroid receptor BRL3 confers drought resistance without penalizing plant growth. *Nat. Commun.*, **9**, 4680.
32. Mosteller, F. and Fisher, R. (1948) Questions and answers. *Am. Stat.*, **2**, 30–31.
33. Ferreirós-Vidal, I., Carroll, T., Zhang, T., Lagani, V., Ramirez, R.N., Ing-Simmons, E., Gómez-Valadés, A.G., Cooper, L., Liang, Z., Papoutsoglou, G. *et al.* (2019) Feedforward regulation of Myc coordinates lineage-specific with housekeeping gene expression during B cell progenitor cell differentiation. *PLoS Biol.*, **17**, e2006506.
34. Pachnis, V., Mankoo, B. and Costantini, F. (1994) Expression of the c-RET proto-oncogene during mouse embryogenesis. *Development (Cambridge, England)*, **119**, 1005–1017.
35. Rusmini, M., Griseri, P., Lantieri, F., Matera, I., Hudspeth, K.L., Roberto, A., Mikulak, J., Avanzini, S., Rossi, V., Mattioli, G. *et al.* (2013) Induction of RET dependent and independent pro-inflammatory programs in human peripheral blood mononuclear cells from Hirschsprung patients. *PLoS one*, **8**, e59066.
36. Darnell James, E., Kerr Ian, M. and Stark George, R. (1994) Jak-STAT pathways and transcriptional activation in response to IFNs and other extracellular signaling proteins. *Science*, **264**, 1415–1421.
37. Pichler, M., Stiegelbauer, V., Vychytilova-Faltejskova, P., Ivan, C., Ling, H., Winter, E., Zhang, X., Goblirsch, M., Wulf-Goldenberg, A., Ohtsuka, M. *et al.* (2017) Genome-Wide miRNA analysis identifies miR-188-3p as a novel prognostic marker and molecular factor involved in colorectal carcinogenesis. *Clin. Cancer Res.*, **23**, 1323.
38. Ferreiros-Vidal, I., Carroll, T., Taylor, B., Terry, A., Liang, Z., Bruno, L., Dharmalingam, G., Khadayate, S., Cobb, B.S., Smale, S.T. *et al.* (2013) Genome-wide identification of Ikaros targets elucidates its contribution to mouse B-cell lineage specification and pre-B-cell differentiation. *Blood*, **121**, 1769–1782.
39. Ma, S., Pathak, S., Mandal, M., Trinh, L., Clark, M.R. and Lu, R. (2010) Ikaros and aiolos inhibit pre-b-cell proliferation by directly suppressing c-myc expression. *Mol. Cell. Biol.*, **30**, 4149–4158.
40. Guo, Y., Ye, Q., Deng, P., Cao, Y., He, D., Zhou, Z., Wang, C., Zaytseva, Y.Y., Schwartz, C.E., Lee, E.Y. *et al.* (2020) Spermene synthase and MYC cooperate to maintain colorectal cancer cell survival by repressing Bim expression. *Nat. Commun.*, **11**, 3243.
41. Li, J., Meng, Y., Wu, X. and Sun, Y. (2020) Polyamines and related signaling pathways in cancer. *Cancer Cell Int.*, **20**, 539.
42. Hesterberg, R., Cleveland, J. and Epling-Burnette, P. (2018) Role of polyamines in immune cell functions. *Med. Sci. (Basel, Switzerland)*, **6**, 22.
43. Fagan, R. and Palfey, B. (2010) *7.03 - Flavin-Dependent Enzymes*. Elsevier Oxford, pp. 37–113.
44. Oliva, M., Guy, A., Galili, G., Dor, E., Schweitzer, R., Amir, R. and Hacham, Y. (2020) Enhanced production of aromatic amino acids in tobacco plants leads to increased phenylpropanoid metabolites and tolerance to stresses. *Front Plant Sci*, **11**, 604349.
45. Staswick, P. (2009) The tryptophan conjugates of jasmonic and indole-3-acetic acids are endogenous auxin inhibitors. *Plant Physiol.*, **150**, 1310–1321.
46. Lima, T., Carvalho, E., Martins, F., Oliveira, R., Miranda, R., Müller, C., Pereira, L., Bittencourt, P., Sobczak, J., Gomes-Filho, E. *et al.* (2018) Lignin composition is related to xylem embolism resistance and leaf life span in trees in a tropical semiarid climate. *New Phytologist.*, **219**, 1252–1262.
47. Mor, D., Huertas, F., Morse, A., Kaletsky, R., Murphy, C., Kalia, V., Miller, G.W., Moskalenko, O., Conesa, A. and Mor, D.E. (2022) GAIT-GM integrative cross-omics analyses reveal cholinergic defects in a *C. elegans* model of Parkinson's disease. *Sci. Rep.*, **12**, 3268.
48. Oberhardt, M., Chavali, A. and Papin, J. (2009) Flux balance analysis: interrogating genome-scale metabolic networks. *Methods Mol. Biol.*, **500**, 61–80.
49. Jensen, P. and Papin, J. (2011) Functional integration of a metabolic network model and expression data without arbitrary thresholding. *Bioinformatics*, **27**, 541–547.