

# PaintOmics 3: a web resource for the pathway analysis and visualization of multi-omics data

Rafael Hernández-de-Diego<sup>1,†</sup>, Sonia Tarazona<sup>1,2,†</sup>, Carlos Martínez-Mira<sup>1</sup>, Leandro Balzano-Nogueira<sup>3,4</sup>, Pedro Furió-Tarí<sup>1</sup>, Georgios J. Pappas, Jr<sup>5</sup> and Ana Conesa<sup>1,3,4,\*</sup>

<sup>1</sup>Genomics of Gene Expression Lab, Centro de Investigación Príncipe Felipe, Valencia, Spain, <sup>2</sup>Department of Applied Statistics, Operations Research and Quality, Universitat Politècnica de València, Spain, <sup>3</sup>Microbiology and Cell Science Department, Institute for Food and Agricultural Sciences, University of Florida, Gainesville, FL, USA, <sup>4</sup>Genetics Institute, University of Florida, Gainesville, FL, USA and <sup>5</sup>Department of Cellular Biology, University of Brasilia, Biological Sciences Institute, Brasília, Brazil

Received February 22, 2018; Revised April 20, 2018; Editorial Decision May 09, 2018; Accepted May 16, 2018

## ABSTRACT

The increasing availability of multi-omic platforms poses new challenges to data analysis. Joint visualization of multi-omics data is instrumental in better understanding interconnections across molecular layers and in fully utilizing the multi-omic resources available to make biological discoveries. We present here PaintOmics 3, a web-based resource for the integrated visualization of multiple omic data types onto KEGG pathway diagrams. PaintOmics 3 combines server-end capabilities for data analysis with the potential of modern web resources for data visualization, providing researchers with a powerful framework for interactive exploration of their multi-omics information. Unlike other visualization tools, PaintOmics 3 covers a comprehensive pathway analysis workflow, including automatic feature name/identifier conversion, multi-layered feature matching, pathway enrichment, network analysis, interactive heatmaps, trend charts, and more. It accepts a wide variety of omic types, including transcriptomics, proteomics and metabolomics, as well as region-based approaches such as ATAC-seq or ChIP-seq data. The tool is freely available at [www.paintomics.org](http://www.paintomics.org)

## INTRODUCTION

Due to the increasing popularity of multi-omic experiments, there is a growing demand for tools that facilitate integrative analyses. The heterogeneity, high dimensionality and multiple interconnectivity of the multi-omics data are distinct aspects that all require special attention. While there

are many analytical strategies to facilitate multi-omic data interpretation (1,2), graphical representations of the combined dataset are both common and valuable techniques to simplify data and facilitate interpretation (3).

Several resources for integrative visualization are already available in the context of systems biology. Two well-known tools for graph analysis and visualization are Cytoscape (4) and Gephi (5). These applications provide numerous functions for exploring, manipulating and analyzing complex networks and are supplemented with many plug-ins that enable specialized analysis of molecular data (3). Similarly, the web-based workbench VisANT (6) includes several tools for drawing and analyzing large biological networks and the ability to combine multiple types of networks to systematically analyze correlations with the phenotype. Another interesting tool is 3Omics (7), a web application specifically designed for the analysis of human data, which supports transcriptomics, metabolomics and proteomics datasets. Using 3Omics, users can perform correlation analysis, co-expression profiling, phenotype mapping, pathway and GO enrichment analysis on each dataset, and visualize results graphically.

Other software solutions assist with the discovery process by integrating information from existing curated databases, such as those for metabolic and signaling pathways. Pathways are a fundamental part of interpreting omics data, as they provide the biological context for a given observation (8). One popular tool for pathway-based visualization is MapMan (9), which allows for large datasets, including multiple conditions or time-series experiments, to be displayed as pathway diagrams. Another example is KaPPa-View (10) a web-based tool for integrating transcript and metabolite data into pathway maps. Luo and Brouwer introduced Pathview (11), an R/Bioconductor package for data integration and visualization using KEGG pathways, which

\*To whom correspondence should be addressed. Tel: +34 963 289 680; Fax: +34 963 289 701; Email: [aconesa@cipf.es](mailto:aconesa@cipf.es)

†The authors wish it to be known that, in their opinion, the first two authors should be regarded as joint First Authors.

has recently been launched as a web tool (12). Pathview integrates a wide variety of biological data based on pathways analysis, provided the omics features are previously mapped to genes. Finally, Garcia-Alcalde *et al.* developed PaintOmics 2, a web-based tool for integrated visualization of transcriptomics and metabolomics data using KEGG pathways as a template (13).

Despite being useful, these tools do have some limitations in terms of effective data integration and visualization. As a general rule, network-based tools are useful for identifying the interconnections between multiple biological features, but the size and complexity of the network coupled with poor integration to existing knowledge (e.g. pathway diagrams) can hinder interpretation. On the other hand, pathway-based solutions reduce the size of the displayed data by grouping the information based on previous knowledge, but they do not allow new molecular connections to easily be inferred. Moreover, most of the tools do not allow researchers to easily integrate data from chromatin profiling experiments or regulatory elements such as microRNAs (miRNAs) or transcription factors (TFs).

Here we introduce PaintOmics 3, a web-based application for integrative visualization of multiple biological datasets on KEGG pathway diagrams (Figure 1). As opposed to other visualization tools, PaintOmics 3 covers a comprehensive multi-omics pathway analysis workflow, including support for data from virtually any type of omics measurement, automatic feature name/identifier conversion, pathway enrichment and network analysis. PaintOmics 3 implements the latest technologies in web-based visualization, offers powerful exploratory tools of complex data and enables visual exploration of biological processes.

## MATERIALS AND METHODS

### The PaintOmics 3 architecture

The PaintOmics 3 platform is a Client-Server web-application developed entirely using open-source resources (Supplementary Table S1).

The server side is built using Python 2.7 and R languages and is in charge of processing remote user requests, managing access to biological information stored in a non-relational database implemented in MongoDB and performing a variety of statistical and bioinformatics analyses. The client side is responsible for the data presentation to the user web-browser and was entirely developed using the web technologies JavaScript and HTML5. Communication between the client applications and the server-end is handled by AJAX mechanisms where data are exchanged in JSON format.

### Input data

Input data for PaintOmics 3 are omic quantitative measurements per feature and experimental group, which must have already undergone pre-processing and normalization. Each omic type must be uploaded in a tab-delimited file format. When available, a second file can be provided with a list of relevant features, for instance a list of differentially

expressed (DE) genes (Supplementary Figure S1). Experimental conditions should be the same across omics. Ideally, the quantification file should contain one column per experimental condition, therefore replicated measurements should be averaged. To maximally benefit from the coloring rules implemented in the tool, quantification values should be computed as the log fold change between a case condition and a control/reference condition. The omic data types accepted by PaintOmics 3 can be broadly classified into four categories.

*Gene-based omics.* The biological features measured are or can be translated into genes. Some examples are mRNA-seq or microarrays, where measurements are made at the gene or transcript level and proteomics, where protein quantification can be imputed to the respective coding gene. PaintOmics 3 accepts Entrez Gene IDs as gene identifiers, but it also includes a module for Name/ID translation that allows users to input many other identifiers or naming domains. This module fetches the translation information from public databases such as Ensembl, PDB, NCBI RefSeq, and KEGG, whenever they are available, greatly facilitating data input.

*Metabolite-based data.* The biological features are metabolites. PaintOmics 3 includes some tools for resolving any existent ambiguity when matching user's metabolites to KEGG compound names. For each input metabolite, it generates a list of potentially related compounds that are ranked with a score based on the similarity of names, and users can manually review, and eventually change, metabolite assignments (Supplementary Figure S2).

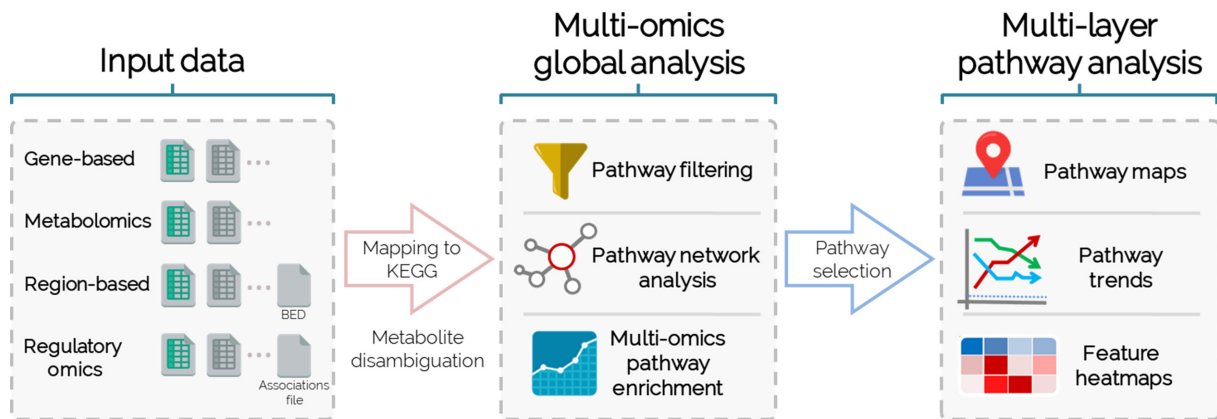
*Region-based omics.* The measured features are a set of genomic regions, such as those obtained by ChIP-seq, DNase-seq, ATAC-seq, Methyl-seq, etc. For omics in this category, the chromosome, start and end position, and a quantification value for the regions must be provided, along with a GTF file with the reference genome annotation. PaintOmics 3 maps each region to its proximal gene(s) with the RGMATCH tool (14), which takes into account the relative position of the region with respect to the specific areas of the gene (promoter region, first exon, intronic areas, etc.).

*Regulatory omics.* This category refers to miRNAs, transcription factors (TF) or other factors, which act as regulators of gene expression and are mapped accordingly to their respective genes. To do this mapping PaintOmics 3 requires an additional tabulated file containing the associations between miRNA/TF and their target genes. The system processes the input quantification data and assigns the expression values to the known list of target genes for each regulator.

PaintOmics 3 can process multi-omic datasets with any combination of omic platforms.

### Pathway analysis options

Once data are submitted, PaintOmics 3 maps all features to the KEGG database and returns a pathway analysis interface where users can easily keep or hide pathways for down-



**Figure 1.** PaintOmics 3 workflow diagram. PaintOmics 3 takes as input tab delimited files containing processed data from different omic types. After mapping to KEGG database and resolving metabolite ambiguities, a global analysis interface allows filtering, network and enrichment analysis of pathways. Selected pathways can be further analyzed in pathway maps displaying omic data trends and feature-level heatmaps of multi-omic measurements.

stream analysis (Supplementary Figure S3), explore pathway networks at different omic layers, visualize single and joint pathway enrichment results, and access specific pathways for interactive browsing.

**Pathway enrichment.** For each omic dataset, provided that a list of relevant features is included, PaintOmics 3 performs a single Pathway Enrichment Analysis (PEA) using the Fisher's Exact test. Both  $P$ -value and false discovery rate (FDR) adjusted  $P$ -values are returned for each dataset. A joint pathway enrichment  $P$ -value considering all available omics data is also computed by applying either the Fisher combined probability test (15) or the Stouffer's method (16). The latter introduces a weight for each individual  $P$ -value, thereby controlling the contribution of each omic type to the computation of the joint enrichment. This is interesting as not all platforms have the ability to measure molecular features with the same comprehensiveness. By default, PaintOmics 3 assigns a weight to each omic that is the percentage of features mapped to pathways in that omic.

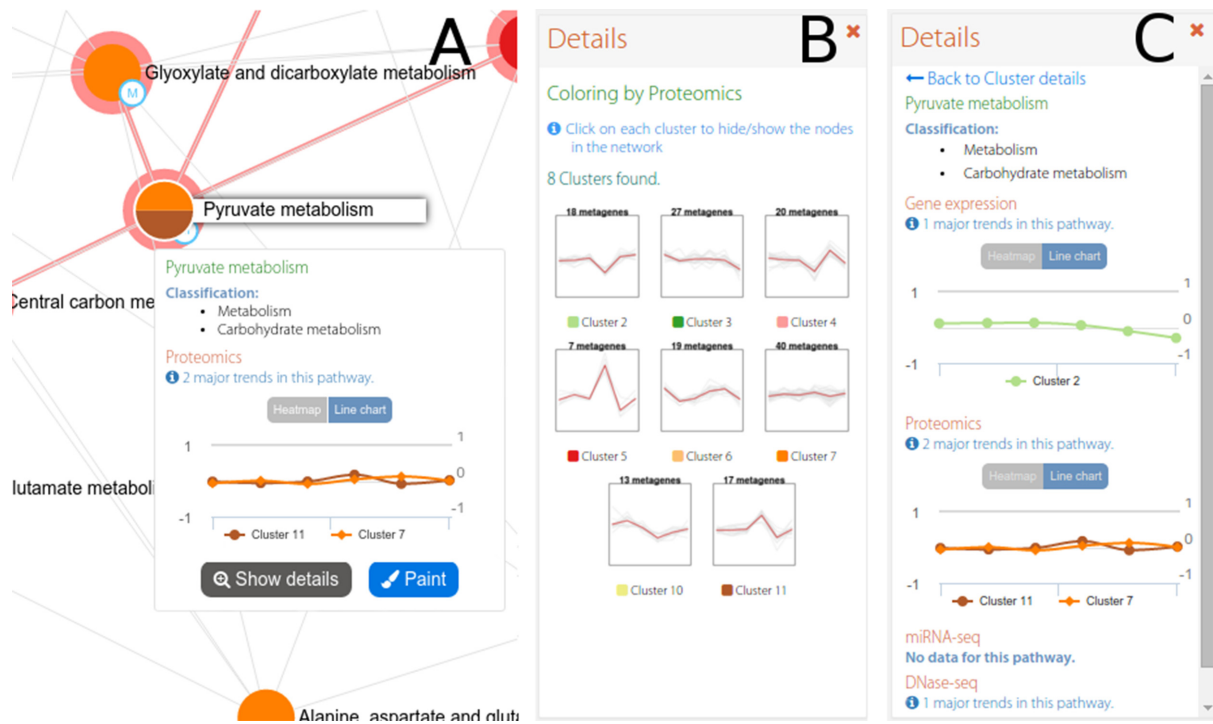
**The multi-omic pathway interaction network.** PEA results, which are presented as a list of pathways, often fall short in revealing relationships between cellular processes or in offering comprehensive representations of biological systems. PaintOmics 3 includes a tool to create *pathway networks* based on the multi-omics data. In this network, nodes represent pathways and edges indicate shared features among them or KEGG database connections. Each pathway in the network is summarized by one (in some cases several) representative profile(s) obtained by dimension reduction techniques, that recapitulate the major behavior of the pathway along each of the conditions of the study (17). These pathway profiles are then used to obtain clusters of pathways with similar trends, and the pathway network is colored according to this clustering (Figure 2). In this way, pathways with the same pattern of change can be easily grouped and, if also connected by edges, their molecular relationships revealed. Pathway profiles can be obtained for any of the available omics data and hence networks can be built from each molecular layer perspective. The pathway network tool also

includes several options for node selection based either on static (KEGG database) or dynamic (experimental) data.

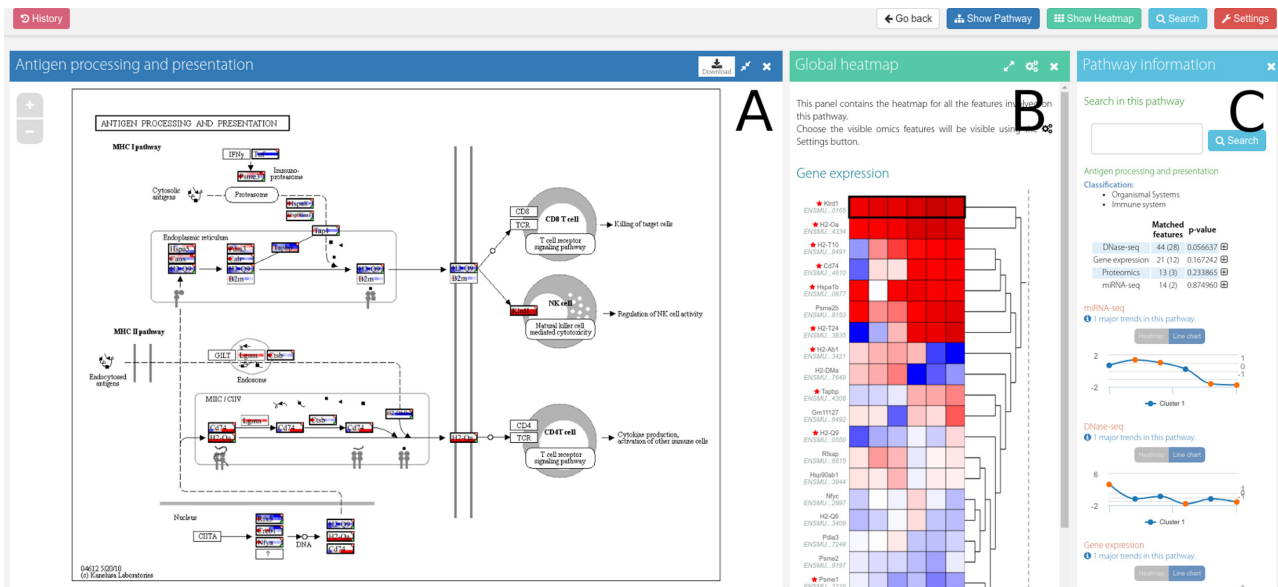
**Multi-omic visualization of single pathways.** One of the core features of PaintOmics 3 is the visualization of user input data onto individual KEGG pathways. Figure 3 illustrates an example of the typical workspace for pathway exploration. The main panel contains the pathway diagram colored according to the input data. Users can easily navigate through this panel and visualize the different values associated with each biological feature mapped on the KEGG map (Figure 3A and Supplementary Figure S4). Feature (genes or metabolites) boxes are divided in as many sections as columns in the input files, and in up to three rows to display different omics measurements. Boxes are colored based on a blue (low) to red (high) gradient scale calculated on each omics range of values after discarding extreme observations. Relevant features are highlighted by a thicker border and a special mark at the top right corner. Two additional panels are available at the pathway workspace. The *Pathway Information* panel includes summary information for all the omic data in the pathway, including the major trends for each type of measurement (Figure 3C). The *Global Heatmap* panel displays the pathway features values in the form of consecutive heatmaps, one per omic dataset (Figure 3B). More details about the functionalities of this module can be found in Supplementary Data.

#### Availability and requirements

PaintOmics 3 is free to use and is distributed under the GNU General Public License Version 3. A public copy of the application is hosted at the CIPF facilities (<http://bioinfo.cipf.es/paintomics>) and sources are available at GitHub (<https://github.com/fikipollo/paintomics3>). The documentation and guides for users and administrators are available at the free web platform Read the Docs, and also stored at the GitHub repository. PaintOmics 3 server-side application has been extensively tested on Ubuntu and Debian Linux servers.



**Figure 2.** The interactive pathway network in PaintOmics 3. The interactive network panel (A) is complemented by a secondary panel showing the trends for all pathway clusters in a given omic (B), or the trends for each omic in the chosen pathway (C).



**Figure 3.** Workspace for pathway exploration in PaintOmics 3. The layout for pathway exploration is divided into three panels. The main panel (A) contains the interactive pathway diagram, the *Global Heatmap* panel (B) displays multi-omics data in the form of heatmaps, and the *Pathway Information* panel (C) contains search and summarizing functions.

**USE CASE**

To illustrate how to use PaintOmics 3, we selected a multi-omic dataset from Cacchiarelli *et al.* where they analyze changes in gene expression and chromatin state that occur during human reprogramming of immortalized fibroblasts (18). This study includes transcriptomics (RNA-seq and small RNA-seq), methylation (RRBS-seq) and

region-based histone modification (H3K4me3 ChIP-seq) data taken at different time points after reprogramming. Details on data preprocessing and input for PaintOmics 3 can be found in the Supplementary Data file.

In order to analyze the Cacchiarelli’s multi-omic dataset with PaintOmics 3, we started by excluding *disease* and some *organismal systems* pathways that were considered ir-

Pathway enrichment									
Search <input type="text"/>		<input type="checkbox"/> Regular expression	<input type="checkbox"/> Case sensitive	Show FDR: <input type="text" value="None"/>	Show combined p-values: <input type="text" value="Fisher"/>	<a href="#">Configure</a>	<a href="#">Download as XLS</a>		
Paint	Pathway name	Features			Significance tests				
		Un... ge...	Un... me...	Gene expression	Methyl-Seq data	miRNA-Seq data	ChIP-seq... data H3K4...	Combined p-value (Fisher)	
<input checked="" type="checkbox"/>	Neuroactive ligand-receptor interaction	219	0	5.2161e-16	0.21271	0.53530	5.0137e-23	3.5844e-34	
<input checked="" type="checkbox"/>	Cell adhesion molecules (CAMs)	124	0	2.0059e-7	0.68555	0.26537	2.6022e-6	4.7232e-10	
<input checked="" type="checkbox"/>	Cytokine-cytokine receptor interaction	164	0	3.8441e-6	0.04258	0.56106	0.04442	1.3777e-8	
<input checked="" type="checkbox"/>	Calcium signaling pathway	161	0	9.2863e-4	0.95149	0.44550	2.5301e-7	2.3155e-7	
<input checked="" type="checkbox"/>	Rap1 signaling pathway	191	0	1.5737e-4	0.87912	0.53208	3.2736e-5	3.6630e-6	

**Figure 4.** Part of the results for the PaintOmics pathway enrichment analysis for Cacchiarelli's data (complete list in Supplementary Figure S5). The enriched pathways are ordered by the combined *P*-value. Upper positions correspond to the most significant pathways. A color scale is used to highlight the level of enrichment for each pathway where the higher the intensity of red, the higher the significance. Gray cells indicate that the corresponding omic data type is not present in the pathway.

relevant for the study. This resulted in 199 pathways being mapped by multi-omic data, of which 25 were found to be significant by the combined enrichment test (Figure 4 and Supplementary Figure S5). Most enriched pathways were regulated at the gene expression and H3K4me3 levels with substantial overlap, and this is in agreement with its role as an active promoter mark associated with H3K4me3. At the Methyl-seq and miRNA-seq layers, there were much fewer enriched pathways, suggesting that these regulatory levels were not as coordinated or ubiquitous.

Significant pathways include hormone and drug detoxification, but mainly signaling, communication and cell lineage. To understand the dynamics and relationships between these processes, we analyzed the multi-omic pathway network (Figure 5). The network is dominated by signaling and lineage pathways that are activated between 10 (cluster 5) and 24 days (cluster 6) after reprogramming, including Ca<sup>2+</sup> signaling and Wnt signaling (Figure 5A), pathways which were also identified by Cacchiarelli *et al.* as transient early up-regulation. These pathways are part of H3K4me3 cluster 1 (Figure 5B), which represents an increase of active promoter histone marks in time. Previous reports indicated that these epigenetic changes enable direct reprogramming to pluripotency (19). On the contrary, cluster 4, containing extracellular matrix component and focal adhesion pathways, shows a down-regulation pattern, which was also observed in the original publication and might reflect the need of loose inter-cellular structures during the reprogramming process.

A significant pathway detected in PaintOmics 3 was 'Signaling pathways regulating pluripotency of stem cells' (Figure 6). This is a complex pathway in which both naïve and primed stem cells signaling are represented. Major variation trends computed for this pathway by PaintOmics 3 (Figure 6, bottom panels) indicate a general gene expression up-regulation from 10 days after reprogramming that is halted at the last hIPSC stage. H2K4me3 is strong from day 24 while the DNA methylation is consistently down-regulated as reprogramming progresses. For genes in this pathway, miRNA regulation becomes active at 24 days. However, regulatory patterns may vary for different genes. Some examples are shown on the right side of the figure. Most activated genes, including the three key pluripotency genes, POU5F,

SOX2 and NANOG, are over-expressed and have positive H3K4me3 marks at their promoters (detailed data shown for SOX2). This is in agreement with the activity of several epigenetic regulators like SMARCD1 and SETDB1, which are histone modifiers that increase the DNA accessibility at promoters of these genes (20). Other genes in the pathway are down-regulated. PaintOmics 3 is able to reveal the potential mechanisms of these down-regulated genes as well. For example, expression of AKT3 and SMAD3 remain low together with absence of H3K4me3 marks while PIK3R1 is down-regulated with a strong DNA methylation pattern. Interestingly, down-regulation of PIK3R1 has been shown to promote a stem-like phenotype in renal cancer cells through the AKT/GSK3 $\beta$ /CTNNB1 pathway, due to AKT phosphorylation (21). Although the AKT genes appear down-regulated in Cacchiarelli's data, their active involvement in hIPSC has been reported (22). We speculate that the observed lower expression of the AKT isoforms could be compensated by their post-translational activation through the PIK3R1 pathway. Finally, other regulatory patterns appear in this pathway. For example, the INHBA gene details box reveals down-regulation is correlated with an strong induction of associated miR-509-3.

## DISCUSSION

We introduce here a novel application for integrative visualization of many different types of omics data, PaintOmics 3, a one-click web tool which enables comprehensive exploration of multi-omics data via pathway models. Using this tool, researchers can easily move through different levels of regulation within biological systems while leveraging the combined benefits of network graphs for studying global pathway relations and multi-omics painted pathway maps to understand molecular interactions. The network provides an overall view of regulated processes while maintaining a manageable size. This is achieved by using pathways as nodes and by implementing the filtering options offered by the application. By representing pathways as omic profiles and coloring the network accordingly, PaintOmics 3 provides a helpful method of understanding functional relationships across different molecular layers. The single pathway colored map gives the user a full view of the path-

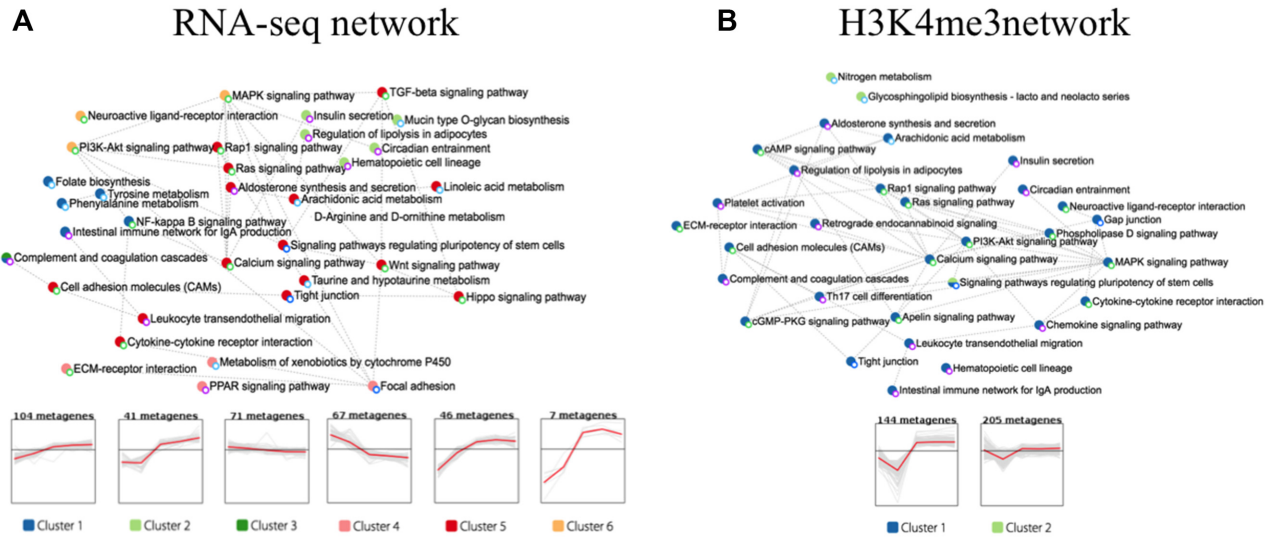


Figure 5. Pathway networks and cluster profiles of representative temporal patterns. Network A is colored according to gene expression data. Network B is colored according to H3K4me3 ChIP-seq data.

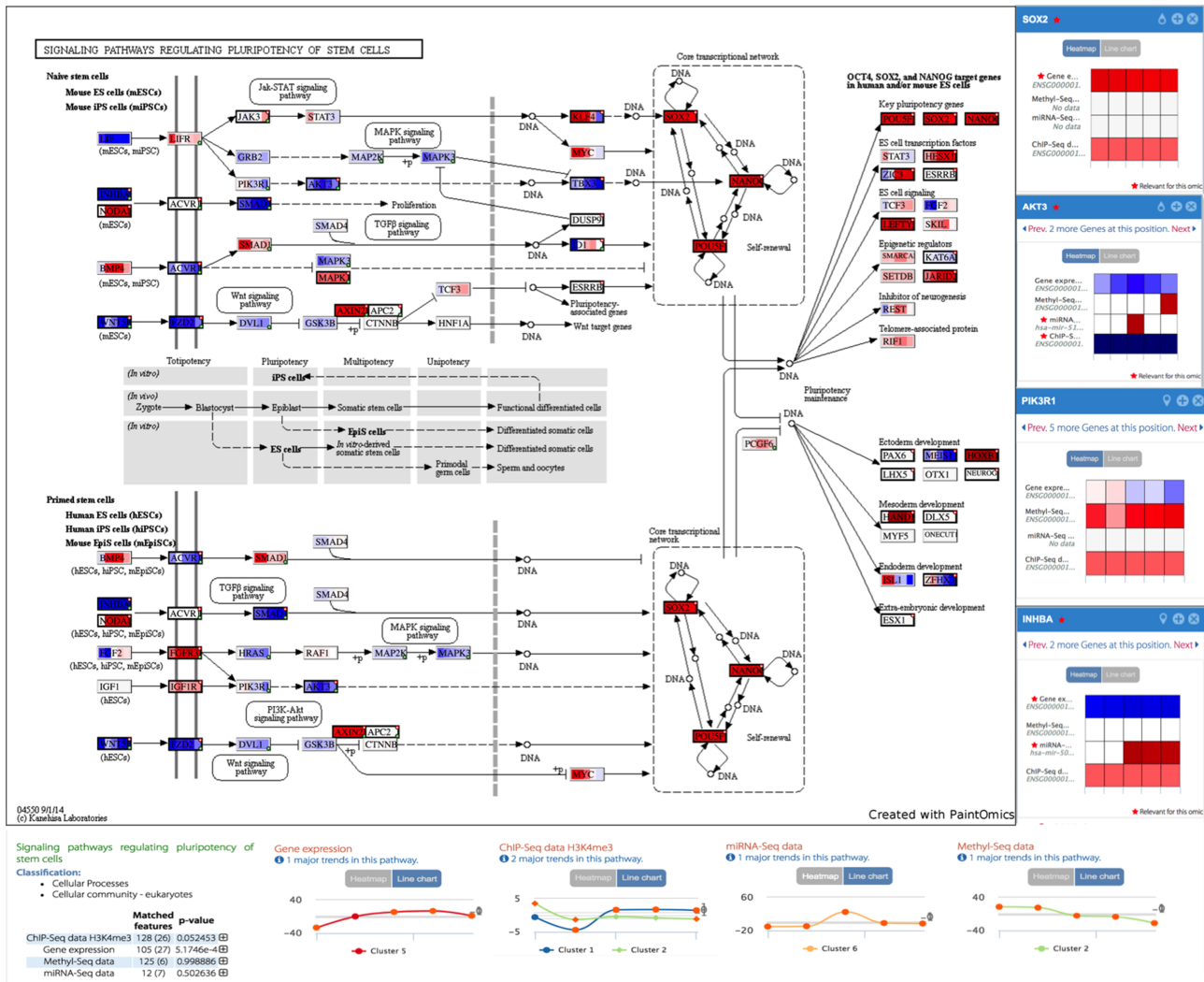


Figure 6. Interactive KEGG diagram for Signaling pathways regulating pluripotency of stem cells. Data obtained from Cacchiarelli's multi-omics study

way multi-omics data. The details panels available for each pathway node facilitate the analysis of complex nodes by providing separate multi-omics values for each of gene.

PaintOmics 3 employs the unique strategy of incorporating trans- and cis-regulatory factors into the pathway view. *cis*-Acting elements will be typically linked to gene nodes thanks to the *Region-based omics* input and PaintOmics 3 will calculate the corresponding associations directly from bed-like files. This kind of chromatin-based information is normally represented using genome browsers, which completely ignore pathway information and hence cannot study coordinated chromatin changes for cellular processes. Trans-acting regulatory elements (i.e. miRNAs, TFs) can be linked through the *Regulatory omics* input as well. In this case, an association file must be provided by the user. Although this might appear as a limitation, it is in fact an effective solution to provide the tool with full flexibility to visualize any kind of trans-regulatory data the user might have, including PAR-CLiP or regulatory lncRNA data.

Finally, PaintOmics 3 comes with a wide range of customization and auxiliary functions that allow data and job storage, id conversion, re-coloring, rescaling, filtering, etc. to create in each case the most informative representation of the complex multi-omics dataset. The current version of PaintOmics 3 has been available since 2013 and it has been used by >3000 users world-wide. The variety of organisms supported by PaintOmics 3 outperforms other domain specific tools such as MapMan, KaPPa-View (available for plants) or 3Omics (only available for human data), making PaintOmics 3 a highly versatile and broad representation tool of multi-omics information. Future versions of PaintOmics are planned to include pathways from other databases such as MapMan or Reactome, and novel pathways derived from text mining approaches. Future directions also include tool tuning for precision genomics analysis by allowing sample-specific data.

## SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online.

## FUNDING

European Union Seventh Framework Programme [FP7/2007–2013] under the grant agreement [306000-STATegra]; Marie Curie International Research Staff Exchange Scheme under grant agreement [612583-DEANN]; Spanish MINECO [BIO2012-40244]. INB Grant [PT17/0009/0015 - ISCIII-SGEFI / ERDF]. Funding for open access charge: MINECO [BIO2012-40244].

*Conflict of interest statement.* None declared.

## REFERENCES

- Cavill, R., Jennen, D., Kleinjans, J. and Briedé, J.J. (2015) Transcriptomic and metabolomic data integration. *Brief. Bioinform.*, **17**, 891–901.
- Meng, C., Zeleznik, O.A., Thallinger, G.G., Kuster, B., Gholami, A.M. and Culhane, A.C. (2016) Dimension reduction techniques for the integrative analysis of multi-omics data. *Brief. Bioinform.*, **17**, 628–641.
- Pavlopoulos, G.A., Wegener, A.L. and Schneider, R. (2008) A survey of visualization tools for biological network analysis. *BioData Min.*, **1**, 12.
- Shannon, P., Markiel, A., Ozier, O., Baliga, N.S., Wang, J.T., Ramage, D., Amin, N., Schwikowski, B. and Ideker, T. (2003) Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Res.*, **13**, 2498–2504.
- Bastian, M., Heymann, S., Jacomy, M. *et al.* (2009) Gephi: an open source software for exploring and manipulating networks. *ICWSM*, **8**, 361–362.
- Hu, Z., Chang, Y.C., Wang, Y., Huang, C.L., Liu, Y., Tian, F., Granger, B. and Delisi, C. (2013) VisANT 4.0: Integrative network platform to connect genes, drugs, diseases and therapies. *Nucleic Acids Res.*, **41**, W225–W231.
- Kuo, T.C., Tian, T.F. and Tseng, Y.J. (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.
- Villaveces, J.M., Koti, P. and Habermann, B.H. (2015) Tools for visualization and analysis of molecular networks, pathways, and -omics data. *Adv. Appl. Bioinform. Chem.*, **8**, 11–22.
- Thimm, O., Blasing, O., Gibon, Y., Nagel, A., Meyer, S., Kruger, P., Selbig, J., Müller, L.A., Rhee, S.Y. and Stitt, M. (2004) MAPMAN: a user-driven tool to display genomics data sets onto diagrams of metabolic pathways and other biological processes. *Plant J.*, **37**, 914–939.
- Tokimatsu, T., Sakurai, N., Suzuki, H., Ohta, H., Nishitani, K., Koyama, T., Umezawa, T., Misawa, N., Saito, K. and Shibata, D. (2005) KaPPa-view: a web-based analysis tool for integration of transcript and metabolite data on plant metabolic pathway maps. *Plant Physiol.*, **138**, 1289–1300.
- Luo, W. and Brouwer, C. (2013) Pathview: an R/Bioconductor package for pathway-based data integration and visualization. *Bioinformatics*, **29**, 1830–1831.
- Luo, W., Pant, G., Bhavnani, Y.K., Blanchard, S.G. Jr and Brouwer, C. (2017) Pathview Web: user friendly pathway visualization and data integration. *Nucleic Acids Res.*, **45**, W501–W508.
- 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.
- Furió-Tari, P., Conesa, A. and Tarazona, S. (2016) RGmatch: matching genomic regions to proximal genes in omics data integration. *BMC Bioinformatics*, **17**, 427.
- Fisher, R.A. (1992) Statistical methods for research workers. In *Breakthroughs in Statistics*. Springer Series in Statistics. Springer, NY, pp. 66–70.
- Stouffer, S.A., Suchman, E.A., DeVinney, L.C., Star, S.A. and Williams, R.M. Jr (1949) *The American Soldier: Adjustment during Army Life. (Studies in Social Psychology in World War II)*. Princeton Univ. Press., Oxford, Vol. 1.
- Ponzoni, I., Nueda, M.J., 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.
- Cacchiarelli, D., Trapnell, C., Ziller, M.J., Soumillon, M., Cesana, M., Karnik, R., Donaghey, J., Smith, Z.D., Ratanasirintrawoot, S., Zhang, X. *et al.* (2015) Integrative analyses of human reprogramming reveal dynamic nature of induced pluripotency. *Cell*, **162**, 412–424.
- Ziller, M.J., Edri, R., Yaffe, Y., Donaghey, J., Pop, R., Mallard, W., Issner, R., Gifford, C.A., Goren, A., Xing, J. *et al.* (2015) Dissecting neural differentiation regulatory networks through epigenetic footprinting. *Nature*, **518**, 355.
- Boyer, L.A., Mathur, D. and Jaenisch, R. (2006) Molecular control of pluripotency. *Curr. Opin. Genet. Dev.*, **16**, 455–462.
- Lin, Y., Yang, Z., Xu, A., Dong, P., Huang, Y., Liu, H., Li, F., Wang, H., Xu, Q., Wang, Y. *et al.* (2015) PIK3R1 negatively regulates the epithelial-mesenchymal transition and stem-like phenotype of renal cancer cells through the AKT/GSK3 $\beta$ /CTNBN1 signaling pathway. *Scientific Rep.*, **5**, 8997.
- Tang, Y., Jiang, Z., Luo, Y., Zhao, X., Wang, L., Norris, C. and Tian, X.C. (2014) Differential effects of Akt isoforms on somatic cell reprogramming. *J. Cell Sci.*, **127**, 3998–4008.