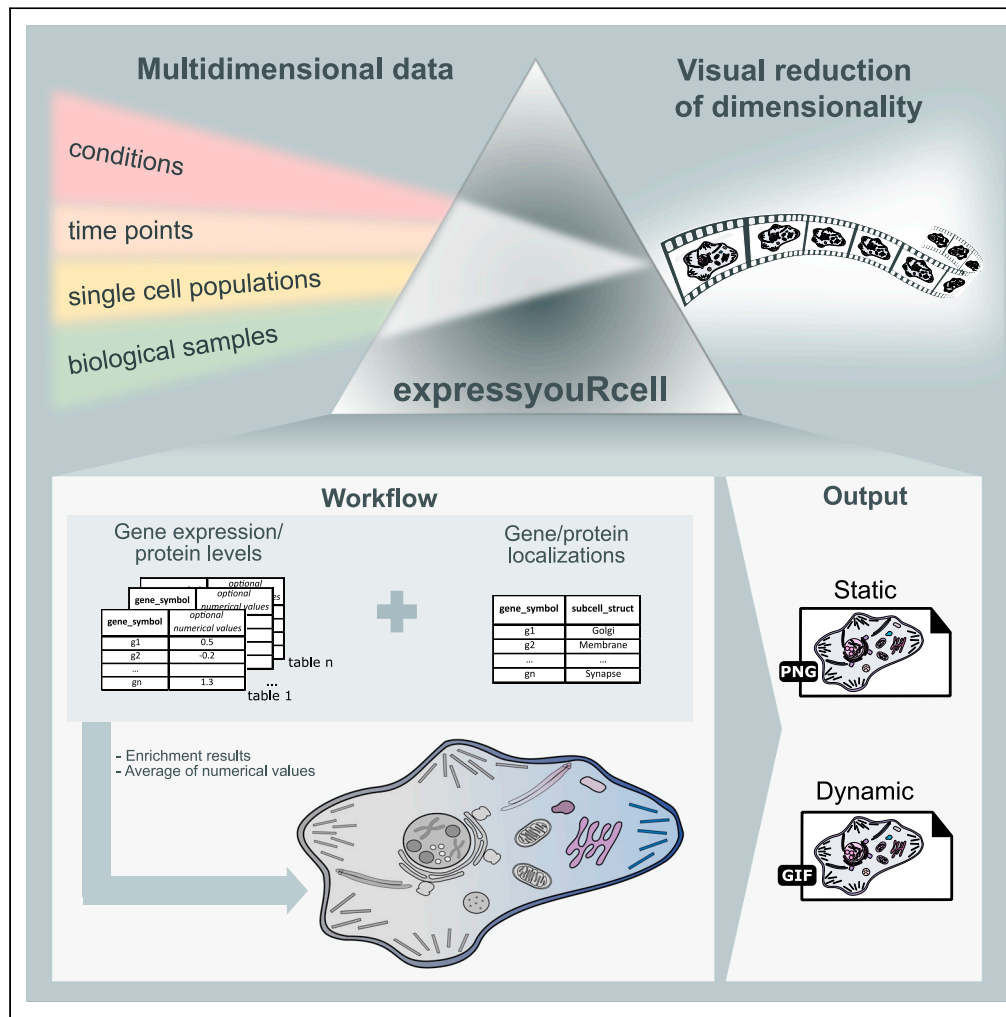**Article**

Visualizing gene expression changes in time, space, and single cells with *expressyouRcell*

Martina Paganin,
Toma Tebaldi,
Fabio Lauria,
Gabriella Viero

fabio.lauria@ibf.cnr.it (F.L.)
gabriella.viero@cnr.it (G.V.)

Highlights

Dynamic visualization of omics data in multiple cell maps

Improved communication of gene expression data with multiple variables

Reduction of data visualization complexity across time or single cells

Handling of heterogeneous data types (single cell, bulk RNA-seq, and proteomics)

Article

Visualizing gene expression changes in time, space, and single cells with *expressyouRcell*

Martina Paganin,[1] Toma Tebaldi,[2,3] Fabio Lauria,[1,*] and Gabriella Viero[1,4,*]

SUMMARY

The last decade has witnessed massive advancements in high-throughput techniques capable of producing increasingly complex gene expression datasets across time and space and at the resolution of single cells. Yet, the large volume of big data available and the complexity of experimental designs hamper an easy understanding and effective communication of the results.
We present *expressyouRcell*, an easy-to-use R package to map the multi-dimensional variations of transcript and protein levels in dynamic cell pictographs. *expressyouRcell* visualizes gene expression variations as pictographic representations of cell-type thematic maps. *expressyouRcell* visually reduces the complexity of displaying gene expression and protein level changes across multiple measurements (time points or single-cell trajectories) by generating dynamic representations of cellular pictographs.
We applied *expressyouRcell* to single cell, bulk RNA sequencing (RNA-seq), and proteomics datasets, demonstrating its flexibility and usability in the visualization of complex variations in gene expression. Our approach improves the standard quantitative interpretation and communication of relevant results.

INTRODUCTION

In the world of big data we are living in today, visualization tools and technologies are essential to analyze massive amounts of information, guide data-driven decisions, and allow a deeper understanding of the complexity of biological systems in physiological and diseased conditions.[1–4] During the last two decades, the generation of sequencing-based big data has witnessed an unprecedented explosion and a massive increase in the amount of raw information at our disposal.[5–11] As the complexity in the experimental designs—characterized by multiple variables, factors, or covariates (e.g., conditions, time points and tissues, cell types)—increases, the biological data mining process has become more challenging. Thus, the scientific community requires next-generation visualization tools and approaches to support the effective interpretation and intuitive presentation of results from complex experimental designs.

Typically, the biological meaning of single-cell or bulk RNA-seq and proteomics data is assessed through differential analyses and downstream pipelines, such as annotation enrichment analysis, coupled with network and clustering analysis. However, with complex experimental designs, these approaches are of difficult interpretation, and they do not always satisfactorily disentangle the biological meaning of the data nor can effectively and rapidly communicate biological results.

Humans are visual organisms, as Aristotle said in his Metaphysics: "All men naturally desire knowledge. An indication of this is our esteem for the senses; for apart from their use we esteem them for their own sake, and most of all the sense of sight.".[12] Today more than ever, the visualization of complex data in simple and information-rich graphical formats is of the utmost relevance to visually exploit complex abstract information. The overarching aim is to fulfill the double ambitious purpose of quantitative data mining and effective communication in a blink of an eye.

Dedicated graphical tools have been proposed as means to represent biological information through graphical approaches in the past.[13–17] Available web applications, such as GeneCards, UniProt, or Human Protein Atlas,[13,16,17] mainly consist of a user-friendly front end for exploring gene and protein expression data obtained from experiments performed in various tissues, cell types, and species. To highlight the localization of genes/proteins, the experimental results can be visualized using schematic representations

[1]Institute of Biophysics, CNR Unit Trento, Trento, Italy

[2]Yale Comprehensive Cancer Center, Yale University School of Medicine, New Haven, CT, USA

[3]Department CIBIO, University of Trento, Trento, Italy

[4]Lead contact

*Correspondence: fabio.lauria@ibf.cnr.it (F.L.), gabriella.viero@cnr.it (G.V.)

of cells or organisms.[13,16,17] However, these tools are not always versatile. In fact, they allow the users to just browse the information available in the database rather than customize the visualization of his/her own data. An R package and an accompanying web application have been developed to draw representations of expression datasets in pictographs of human and mouse tissues.[15] However, this tool requires a series of advanced computational steps to handle and visualize the user's own data.[15] Recently, an R-based application has been proposed to compute and represent results from enrichment analyses through pictographs[14] and a user-friendly interface, but it still lacks flexibility.

Overall, none of the described tools can reduce the biological complexity of results from sophisticated experimental designs. The development of efficient tools for next-generation data exploitation is an urgent need in biology today.

To portray quantitative gene expression changes in space and time in a very intuitive and effective manner, we propose *expressyouRcell*, a user-friendly and flexible R package that generates both static and dynamic cellular pictographs based on a variety of experiments, from bulk or single-cell RNA-seq and proteomics. Leveraging the concept of choropleth maps, we used specific types of thematic maps for representing variations of variables, i.e., fluctuations in gene and transcripts expression levels. These values can be directly mapped to specific cellular compartments and visualized in pictographic representations of multiple cell types. *expressyouRcell* also generates movies of dynamic changes in the cellular pictographs at subcellular and organelle-level resolution. This function is uniquely suitable for visualizing fluctuations in gene expression levels across multiple time points or to represent outcomes from differential longitudinal analyses, such as single-cell developmental or differentiation trajectories. *expressyouRcell* unlocks an enormous simplification of data understanding and a conceptual shift as to how scientific data communication of biological results can be achieved.

RESULTS

Workflow

Here, we present *expressyouRcell*, an R package that generates static and dynamic cellular pictographs to represent, at subcellular and organelle-level resolution, the outcomes of gene expression data obtained from bulk and single-cell RNA-seq or proteomics analysis (Figure 1A). The input of *expressyouRcell* consists of a list of one or multiple tabular data structures with the biological information that the user desires to represent as cellular pictographs.

A table that associates genes ID with subcellular compartments is also required and is provided for model organisms or custom generated through the *map_gene_localization* function that is available in *expressyouRcell*. The list of data tables and the gene localization table represent the input of the *color_cell* function, which assigns colors to cellular compartments according to the statistical significance of enrichment analysis, or based on fold change averages for genes in each compartment. Then, the function outputs static cellular pictographs and the data structure required to generate animated cellular pictographs or movies (through the *animate* function).

expressyouRcell provides the users with custom options for obtaining multiple static cellular pictographs or dynamic pictographs in the form of movies and currently offers six thematic maps, based on different cellular types: i) a typical eukaryotic cell, ii) a neuron, iii) a fibroblast, iv) a microglial cell, v) a lymphocyte, and vi) a macrophage (Figure 1B). Each cell type comprises a set of organelles and macromolecular complexes (nucleus, Golgi apparatus, endoplasmic reticulum, cell membrane, cytosol, chromosome, vesicles, endosomes, cytoskeleton, lysosomes, mitochondria, ribosomes, nuclear bodies, nucleolus, extracellular matrix [Figure 1B]). The lymphocyte and macrophage pictographs include also pseudopodia and receptors. In addition, the neuron pictograph includes the myelin sheath and neurotransmitters, and its global structure is organized in three regions: cell body, axon, and synapse (Figure 1B, and Table S1).

Visualize longitudinal differential analysis

To demonstrate the flexibility and power of *expressyouRcell*, we used seven studies (Table 1), which produced particularly suitable datasets for applying our tool. The selected studies share complex experimental designs with multiple variables or covariates (e.g., conditions, time points and biological samples, single-cell populations) based on three experimental techniques (bulk/single-cell RNA-seq and mass spectrometry).[5–11]
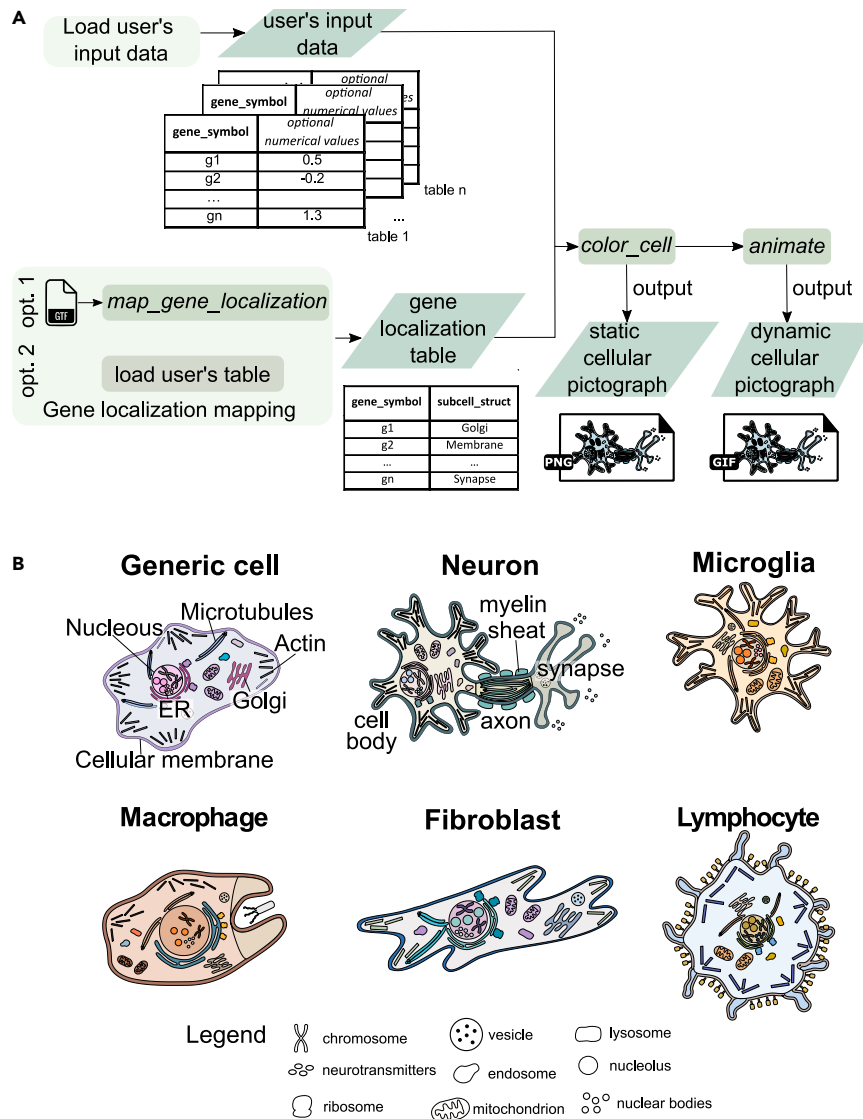
Figure 1. Workflow overview and cellular pictographs

(A) Flowchart representing the basic steps of *expressyouRcell*, the input requirements and the outputs. User's input data are loaded within the package through the *color_cell* function, which also needs the gene-localization table. This data structure can be either generated through the *map_gene_localization* function (option 1) or provided by the user (option 2). The *color_cell* function assigns colors to cellular compartments according to i) the statistical significance of enrichment analysis, ii) fold changes from differential analyses, or iii) gene expression or protein level abundances, averaged for genes in each compartment. Then, it outputs static cellular pictographs (in PNG format) and the data structure required to the *animate* function for the generation of animated cellular pictographs (in GIF or MPG format). User's input data are defined in light green boxes. Diamond boxes denote intermediate and final output data. Functions provided within the package are indicated in italic font.

(B) Set of available cellular pictographs, based on different cellular types: i) a generic eukaryotic cell, ii) a neuron, iii) a microglial cell, iv) a macrophage, v) a fibroblast, and vi) a lymphocyte. The chosen cellular pictograph has to be provided by the user to a dedicated parameter of the *color_cell* function. Default colors are assigned to organelles and subcellular compartments.

In the first case study, Chen and collaborators[6] used TRAP-seq to characterize the changes in the translatome of a specific population of neurons at 30, 60, and 120 min after induction of long-term potentiation (LTP). The canonical analysis and data exploration with volcano plots display the number and magnitude of differentially expressed genes (DEGs) identified in LTP compared to control condition at each time point

Table 1. List of datasets used to evaluate expressyouRcell features

Organism	Cells/tissues	# time points	Technique	Study title	Reference and identifiers
Mouse	Microglia	4	RNA-seq	Microglia activation in response to pathological tau	Wang et al., 2018[7] GEO ID: GSE123467
Mouse	Neurons	3	RNA-seq	RNA Expression in Excitatory Neurons during Hippocampal LTP	Chen et al., 2017[6] GEO ID: GSE79790
Mouse	B/SC/M/H/K	2	RNA-seq	Changes in GE during disease progression	Doktor et al., 2017[8] ArrayExpress ID: E-MTAB-3664
Mouse	Schwann cells	6	RNA-seq	Wound reprogramming and peripheral nerve regeneration	Clements et al., 2017[9] GEO ID: GSE103039
Mouse	Neurons	9	RNA-seq	Temporal regulation of alternative splicing during neural development	Weyn-Vanhentenryck et al., 2018[5] NCBI Short Read Archive ID: SRP055008
Mouse	Embryos	5	scRNA-seq	Organogenesis of mammalian embryo	Cao et al., 2019[11] https://oncoscape.v3.sttrcancer.org/atlas.gs.washington.edu.mouse.rna/downloads
Human	hiPSCs	7	Mass spectrometry	Protein levels in cells during differentiation from hiPSCs to hepatocyte-like cells	T. Hurrell et al., 2019[10]

B: Brain, SC: Spinal Cord, M: Muscle, H: Heart, K: Kidney.

and the corresponding Gene Ontology (GO) enrichment analysis (Figures 2A and 2B), demonstrating the complexity of data interpretation.

When we generated the neuron pictographs with *expressyouRcell,* we swiftly visualized: i) a gradual increase in the color intensity over time, in particular for downregulated genes (Figure 2C, red pictographs), and ii) changes in gene expression spreading across several cellular compartments (Figure 2C and Videos S1 and S2). With this visualization we can appreciate the spatial localization of gene expression changes and observe the emergence of downregulated genes at 30 min, with robust negative fold changes in mRNAs encoding for proteins localized in the endoplasmic reticulum. At later time points, genes with negative fold changes are connected to mitochondria, nucleoplasm, and the synaptic region of neurons.

We also visualized as heatmaps the results of gene set enrichment analyses of differentially expressed transcripts at all time points (Figure 2B). A significant enrichment of downregulated genes related to the cell membrane can immediately be observed, alongside an intensification at the last time point (120 min). Upregulated transcripts show a clear enrichment for chromatin and nuclear-related themes, in particular at 120 min after LTP induction. Next, we used *expressyouRcell* to display the results of gene set enrichment analysis through cellular pictographs of neurons for up- and downregulated genes at each time point (Figure 2D, and Videos S3 and S4). Coherently with the heatmap visualization, the representation obtained with our tool very effectively and intuitively shows that downregulated genes are enriched in the cell membrane at 120 min after LTP induction. From cellular pictographs generated for the enrichment analysis of upregulated genes, we can observe that nucleoplasm and chromosomes are highlighted at all the time points, with an even more evident enrichment at 120 min.

To further appreciate the reduction in data complexity and the power of graphical representations provided by *expressyouRcell*, we analyzed a second case study.[5] In terms of number of time points, this dataset is characterized by a higher complexity compared to the previous one. The experimental design consists of transcriptomes of mouse cortices at nine different time points during which the authors investigated temporal alterations in splicing events.[5] To test whether *expressyouRcell* can be exploited to visualize alternative splicing switches, we generated neuron pictographs (Figure 3A, Video S5) based on the outcome of alternative splicing analysis (Ψ, *psi*) for each developmental time point. We performed enrichment analysis of genes undergoing robust alternative splicing events. The false discovery rate (FDR) was used to derive the color shade for each organelle, obtaining representative plots of potential splicing switches over time. In line with the original study, synapse-related themes emerge at early time points, while
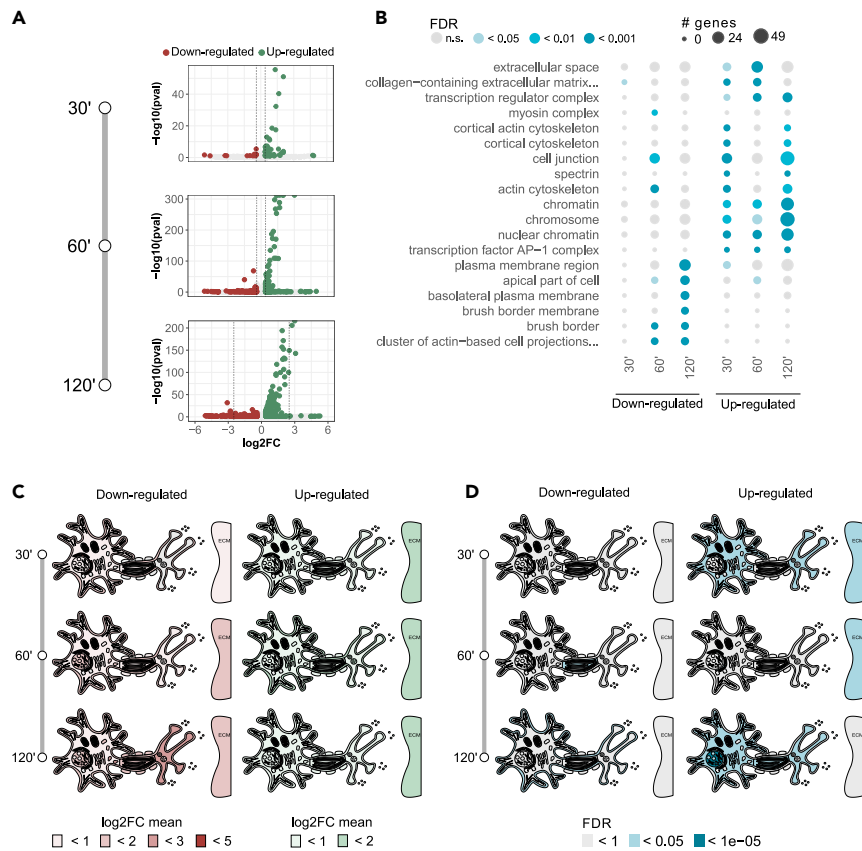
Figure 2. Standard approaches and cellular pictographs for visualizing results of differential analyses across multiple time points

(A) Volcano plot of DEGs between LTP and control condition at indicated time point from TRAP-seq of neurons. Fold changes are plotted against the –log(p value). The vertical dashed lines indicate the 0.4-fold change cutoff thresholds.

(B) Enrichment analysis of Gene Ontology (Cellular Component) terms at indicated time point of differentially expressed genes from TRAP-seq of neurons. Colors of dots indicate the significance of the enrichments. Dot sizes define the number of genes associated with each category.

(C) Neuron pictographs from TRAP-seq. Color shades of cellular compartments and organelles are based on logFC values from differential analysis. Up- and downregulated genes were defined with the threshold described in Chen et al., 2017[6] (i.e., abs($\log_2$FC) > 0.4 and FDR <0.1).

(D) Neuron pictographs from TRAP-seq. Color shades of cellular compartments and organelles are based on the significance of the enrichments. Up- and downregulated genes were defined with the threshold described in Chen et al., 2017[6] (i.e., abs($\log_2$FC) > 0.4 and FDR <0.1).

cytoskeleton-related genes are enriched at later stages. Furthermore, we specifically visualized the dynamic expression changes across development of *Mbnl2* at single-gene resolution. Coherently with what observed by the authors, we observed that the expression levels of *Mbnl2*, which localizes to the nuclear compartments, increase after day 7 (Figure S4). The localization of single genes was also inspected through pictographs of neurons visualizing at time point P4 the RNA-seq expression levels of *Rbfox3*, *Nefl*, and *Syp*, which codify for marker proteins of cell body, axon, and synapse[18–20](Figure 3B).

To display the versatility of our tool, we applied *expressyouRcell* to a proteomics dataset.[10] In their study, the authors analyzed protein levels during differentiation of human induced pluripotent stem cells (hiPSCs) into hepatocytes-like cells (HLCs) at eight time points. For each time point we generated a pictograph of a generic eukaryotic cell, based on protein levels (Figure 4A). We can observe that the protein levels increase and decrease in waves during time and across various cellular localizations. In the large majority of the time points, high protein levels are detected in almost all cell compartments and organelles, while at days 3, 7 and 25 proteins increase within specific organelles. Coherently with what observed by Hurrel and colleagues,[10] our cell pictographs show that the endoplasmic reticulum lumen
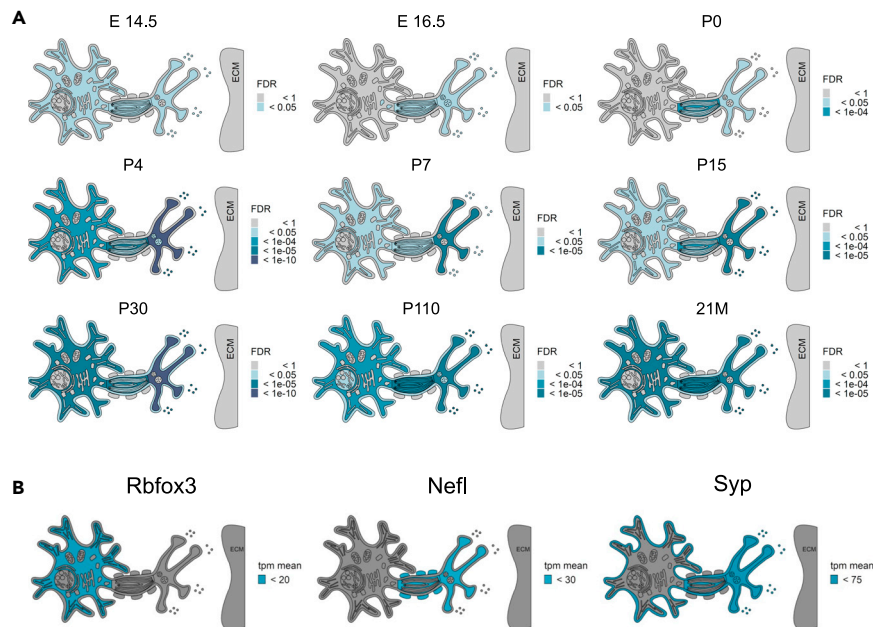
Figure 3. Neuron cell pictographs from RNA-seq of cortical neurons

(A) Neuron cell pictographs from alternative splicing score of cortical neurons. Color shades of cellular compartments and organelles are based on the enrichment of genes with high alternative splicing score ($\Psi > 0.9$), from Weyn-Vanhentenryck et al., 2018.[5]

(B) Neuron cell pictographs from RNA-seq of cortical neurons. Color shades of cellular compartments and organelles are based on gene expression levels (measured in TPM) at P4 of *Rbfox*, *Nefl*, and *Syp* genes, respectively. Counts of RNA-seq data from this study were processed through https://www.refine.bio.

displays low protein levels at early time points and a switch at day 7, leading to higher protein abundance at later stages. Moreover, the authors in the original study reported that the level of proteins associated with the cell cycle progression decreases during the differentiation process. In accordance with this, we observed that pictographs generated by *expressyouRcell* display darker shades (i.e., higher protein abundance) for the nuclear compartment at day 1–5 and lighter colors at later time points. Dynamic cellular pictographs, generated with the *animate* function, clearly show protein level fluctuations across time points and cellular compartments (Video S6). Furthermore, to exploit *expressyouRcell* functionality in displaying variations at single-gene resolution, we visualized mean protein levels of COX5B, SRSF6, MAP7D2, and RPS6[10] (Figure 4B). Consistently with their biological functions, COX5B and RPS6 mainly localize in mitochondria and ribosomes, respectively. For these genes, additional organelles and cytoplasmic regions are colored due to multiple compartment localizations of the proteins, while in the case of SRSF6 and MAP7D2 unique nuclear and microtubule localization can be observed, respectively.

Finally, we evaluated *expressyouRcell* functionalities using scRNA-seq data[11] obtained in mouse embryos from E9.5 to E13.5 days. By looking at thousands of cells one at a time, single-cell approaches capture the diversity of cell populations with unprecedented resolution but also raise the complexity of data display and interpretation. The dataset consists in 2,058,652 total cells. To investigate cell differentiation over embryogenesis, the authors identified distinct pseudo time trajectories with Monocle3.[11] Among these, we selected cells from the enteric neuron sub-trajectory 1. UMAP embedding of 31,189 cells from this trajectory shows that enteric neurons, colored by stage, follow a gene expression transition along embryogenesis, as captured by the first UMAP component (Figure 5A).

To display the activation of specific cellular compartments along this differentiation trajectory, we applied *expressyouRcell* functionalities and generated pictographs of neuronal cells (Figure 5B and Video S7) based on the stage-specific marker genes expressed at each time point. We observed a significant enrichment of genes localized in the nuclear compartments at the earlier embryo stage E9.5. During the following stages, other organelles are also significantly enriched (e.g., mitochondria and ribosomes), with an
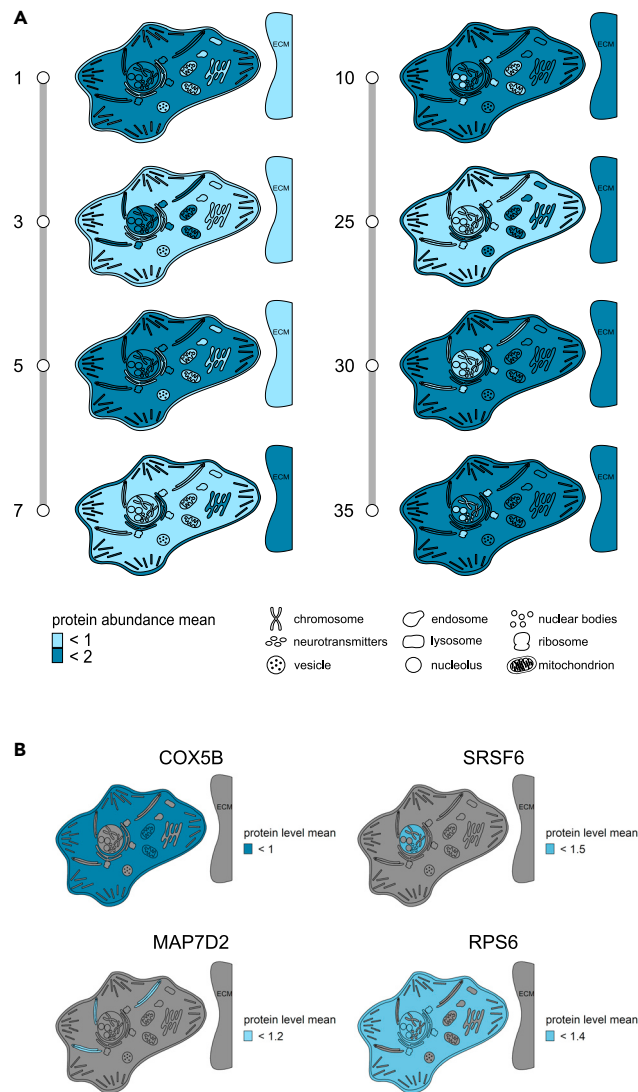
Figure 4. Mammalian cell pictographs from proteomic of hiPSCs

(A) Color shades of cellular compartments and organelles are based on protein abundance levels from Hurrel et al., 2019.[10]

(B) Color shades of cellular compartments and organelles are based on mean protein levels at day 1 of COX5B, SRSF6, MAP7D2, and RPS6, respectively.

increased activation of the axon and synapse compartments. At the last time point (stage E13.5) the enrichment is evident in the entire cell structure. The representation generated by our tool is coherent with the heatmap visualization (Figure 5C), which shows that genes are enriched in the nuclear compartment theme at day E9.5, while other compartments start to be enriched at day E10.5. To display differences with single-cell resolution, we applied *expressyouRcell* to an enteric neuron and a visceral neuron, showing compartment-specific expression signatures (Figure 5D).

Overall, these and additional examples with applications of *expressyouRcell* (Figures S1–S3, and, Videos S1-S19) demonstrate that our tool is flexible and powerful allowing simple and intuitive data mining and visualization for the next generation of biological data exploitation.

Comparison with other tools and performance evaluation

To further demonstrate the innovativeness of *expressyouRcell*, we compared its functionalities with currently available tools for data visualisation.[13–17] To this aim, we took into account several features,
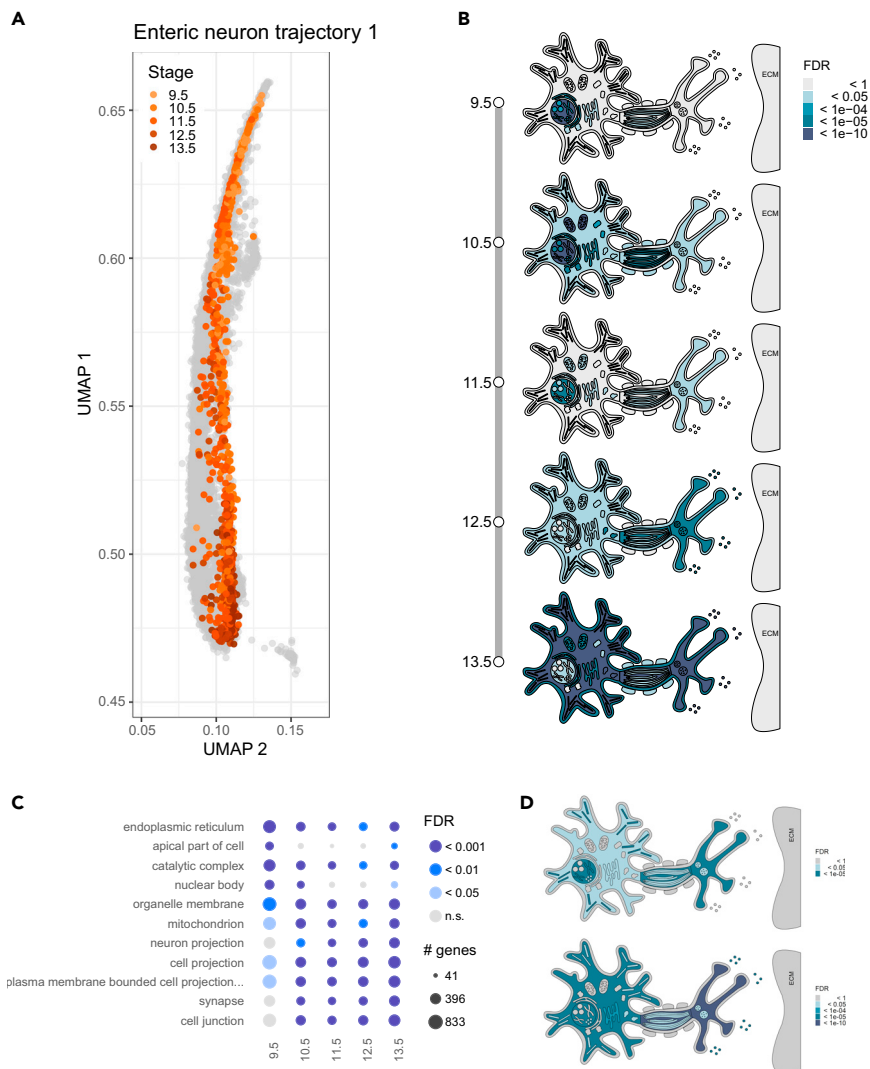
Figure 5. Comparisons between different visualization methods on data from scRNA-seq of enteric neurons

(A) UMAP visualization of cells in the main neural crest 1 trajectory from Cao et al., 2019.[11] Cells in the sub-trajectory enteric neuron 1 are colored by their developmental stage.

(B) Cell pictographs from scRNA-seq of neurons. Color shades of cellular compartments and organelles are based on the significance of the enrichments. Differentially expressed genes were identified across the five developmental stages as described by Cao et al., 2019.[11]

(C) Enrichment analysis of Gene Ontology (Cellular Component) terms in all at indicated time points. The heatmap is colored according to the significance of the enrichments, and the number of genes associated with each category is reported.

(D) Cell pictographs from scRNA-seq of an enteric neuron and a visceral neuron. Color shades of cellular compartments and organelles are based on the significance of the enrichments computed on genes expressed in one single cell of each cluster.

such as automatic handling of user's input data and management of multiple analyses (Table 2). We also valued a flexible management of heterogeneous biological data, ranging from gene symbols for basic enrichment analysis to various numerical values, such as the outcome of differential analyses or gene expression and protein levels variations across multiple time points or single-cell trajectories. The opportunity of generating different output representations (static or dynamic pictographs) into several cellular types was also considered.

Some currently available tools (i.e., Uniprot, Human Protein Atlas, and Jensen/GeneCards[13,16,17]) mainly consist of web interfaces, which rely on pictographic illustration of cells to display the location of protein

Table 2. List of functionalities of tools for the visualization of biological data

		Visualization tool			Web browser		
		expressyou Rcell	subcellulaRVis	gganatogram	Uniprot	Human Protein Atlas	Jensen/ GeneCard
Available analysis	Enrichment results	Yes	yes	no	no	no	no
	Average of numerical values	Yes	no	no	no	no	no
Automatic handling of users' data		Yes	yes	no	no	no	no
Number of genes per pictograph	One gene	Yes	no	no	yes	yes	yes
	Multiple genes	Yes	yes	no	no	no	no
Format	Static pictographs	Yes	no	no	yes	yes	yes
	Dynamic pictographs	Yes	no	no	no	no	no
Multiple cell pictograph		Yes	yes	no	yes	yes	no
Features achieved		100%	50%	0%	38%	38%	25%

Green ticks: achieved function, red crosses: not achieved function.

at subcellular resolution. As such, these tools are not suitable for visualizing users-supplied biological data neither from complex experimental designs nor from biological results of differential or enrichment analyses performed *a priori*.

Recently, two R packages (i.e., *gganatogram* and *subcellulaRVis*) have been developed to visualize user-supplied biological data. The R-based app *gganatogram*[15] offers pictographs of various organisms for visualizing biological data at organ and tissue levels. However, this tool is not specialized for visualizing biological data at the subcellular level as there is only one available cellular pictograph. As such, this tool cannot be used to represent biological data of cellular types, characterized by specific subcellular compartments, such as axons, synapses, or myelin sheaths in highly polarized cells as neurons. From a more technical perspective, the tool does not directly handle user's input data, requiring a series of advanced computational steps to integrate and visualize them. The R-based *subcellulaRVis* web application computes enrichment analysis and visualizes these results through a generic cellular pictograph or a more specific visualization of the endosomal compartments.[14] Despite its user-friendly web page, this tool currently lacks flexibility in managing heterogeneous data types, making it well suited just for performing enrichment analysis. In addition to the functionalities provided by *subcellulaRVis*,[14] our approach allows the user to visualize gene/transcript expression level and to use the results of differential analyses as input (i.e., fold changes and fold enrichments). As such, *expressyouRcell* emerges for the possibility of visualizing comparisons between multiple experimental conditions.

Thanks to its unique possibility of associating numerical values to weight input genes, *expressyouRcell* stands out for its high flexibility in handling multiple types of input data and expands to a great extent the range of available analyses (Tables 2, and S2).

DISCUSSION

Differential gene expression outcomes (i.e., the absolute number of DEGs/transcripts, their magnitude and significance values) are usually displayed as bar plots, volcano plots, or heatmaps. As the experimental methodologies to peer inside the complexity of biological processes evolve, the amount of biological information to disentangle explodes, and standard visualization approaches are not always sufficient and powerful enough to fulfill this task. Hence, new visualization approaches become essential for supporting the interpretation of outcomes from data analysis pipelines and for visually conveying meaningful information. For instance, volcano plots are used to display significant changes in gene expression at either RNA or protein levels, but they are not suitable to represent the spatial localization of fluctuations across cellular compartments. Hence, the canonical representation of differential expression through volcano plots can be integrated with intuitive and informative representations of the localization and intensity of changes provided by *expressyouRcell* pictographs. Likewise, cell differentiation or transition trajectories captured with single-cell approaches can be visualized by pictographs enlightening specific cellular components activated in specific cell states or fate decisions.

Compared with standard methods, cellular pictographs generated through *expressyouRcell* offer an intuitive approach to data exploitation and a reduction of biological complexity. In particular, this method gives us the unique advantage of an immediate detection and intuitive illustration of the most affected cellular components, together with the intensity of the variations.

A handful of tools with alternative representation methods[13–17] have already been proposed, but at the time of this writing, there are no comparable alternatives to *expressyouRcell*. In fact, our tool addresses the data visualization problem with a different and original approach compared to existing ones.[13–17] It gives the unique possibility of creating dynamic illustrations, which can effectively support researchers in tracking the fluctuations in gene expression levels across multiple time points and reduces the complexity of biological information. Importantly, changes are dynamic by definition, and animated pictographs constitute a more appropriate and appealing solution for scientific communication than static illustrations. As of now, no other tool but *expressyouRcell* has been proposed to supply this need.

Some of the currently available tools mainly consist of web interfaces designed to query biological databases in a human interpretable manner, which comes particularly handy to retrieve information on single genes and proteins.[13,16,17] While these resources remain useful and provide us with a wealth of biological information, their purpose and functionality differ from our approach. For example, *gganatogram* is an R-based application, which mainly produces anatomical pictographs for different organisms.[15] It offers nicely detailed pictographs of various whole-body organisms for visualizing biological data at organs and tissue level. The set of available pictographs also includes one unspecific cellular map with multiple organelles. Even if visualizing biological data at the subcellular level is a great advantage, this option limits the users in representing specific cellular types with particular subcellular compartments. *subcellulaRVis* is a web R-based application, which performs and visualizes results from enrichment analysis. This tool provides the users with two graphical options: a generic cellular pictograph and a more specific visualization of endosomes,[14] which is particularly useful to represent detailed information only for this compartment. The user-friendly web page interface offers the possibility of performing enrichment analysis also for less-experienced users and allows us to generate easy-to-interpret graphical results. Unfortunately, this tool does not currently manage multiple data types, making it well suited for performing only enrichment analysis.

In conclusion, *expressyouRcell* meets users' needs for dealing with a variety of different data types and for customized visualizations. Our tool provides the scientific community with a fresh and original approach for representing gene expression changes and for effective, fast, and dynamic communication of results through various cellular pictographs. This visualization can support and complement the standard and already widely adopted methods with the additional knowledge on both the intensity and spatial localization of gene expression variations and opens a completely new scenario as to how biological data exploitation can move toward the future.

Limitations of the study

Currently, our tool provides a variety of options for dealing with data from multiple organisms and visualized on six different cellular maps. Despite the wide range of options provided by *expressyouRcell*, we are aware that they might be insufficient for supporting the visualization of specific organisms and cell types not included in the package. However, these issues may be addressed with updated versions of our tool, aiming at satisfying users' specific needs. Also, as the main purpose of *expressyouRcell* is to provide a non-complex and intuitive representation of broad biological changes, more detailed information, e.g., at pathways level, should be supported by additional and integrative analyses. Nevertheless, we do believe that our tool provides a valuable approach to improve the communication and interpretation of the most relevant biological results.

STAR★METHODS

Detailed methods are provided in the online version of this paper and include the following:

- KEY RESOURCES TABLE
- RESOURCE AVAILABILITY
  - Lead contact
  - Materials availability

SUPPLEMENTAL INFORMATION

Supplemental information can be found online at https://doi.org/10.1016/j.isci.2023.106853.

AUTHOR CONTRIBUTIONS

M.P. developed the R package, organized all figures, and wrote the draft of the paper. T.T. and F.L. advised on data interpretation and R package development. G.V. designed the work and obtained the funding. All authors wrote the paper and critically reviewed the manuscript.

DECLARATION OF INTERESTS

GV has been scientific advisor of IMMAGINA Biotechnology s.r.l. until November 2022.

REFERENCES

1. Hong, M., Tao, S., Zhang, L., Diao, L.T., Huang, X., Huang, S., Xie, S.J., Xiao, Z.D., and Zhang, H. (2020). RNA sequencing: new technologies and applications in cancer research. J. Hematol. Oncol. *13*, 166–216. https://doi.org/10.1186/s13045-020-01005-x.

2. Kerren, A., Kucher, K., Li, Y.F., and Schreiber, F. (2017). BioVis Explorer: A visual guide for biological data visualization techniques. PLoS One *12*, 1–14. e0187341. https://doi.org/10.1371/journal.pone.0187341.

3. Morash, M., Mitchell, H., Beltran, H., Elemento, O., and Pathak, J. (2018). The role of next-generation sequencing in precision medicine: a review of outcomes in oncology. J. Personalized Med. *8*, 30. https://doi.org/10.3390/JPM8030030.

4. Serratì, S., de Summa, S., Pilato, B., Petriella, D., Lacalamita, R., Tommasi, S., and Pinto, R. (2016). Next-generation sequencing: advances and applications in cancer diagnosis. OncoTargets Ther. *9*, 7355–7365. https://doi.org/10.2147/OTT.S99807.

5. Weyn-Vanhentenryck, S.M., Feng, H., Ustianenko, D., Duffié, R., Yan, Q., Jacko, M., Martinez, J.C., Goodwin, M., Zhang, X., Hengst, U., et al. (2018). Precise temporal regulation of alternative splicing during neural development. Nat. Commun. *9*, 2189. https://doi.org/10.1038/s41467-018-04559-0.

6. Chen, P.B., Kawaguchi, R., Blum, C., Achiro, J.M., Coppola, G., O'Dell, T.J., and Martin, K.C. (2017). Mapping gene expression in excitatory neurons during hippocampal late-phase long-term potentiation. Front. Mol. Neurosci. *10*, 39. https://doi.org/10.3389/FNMOL.2017.00039/BIBTEX.

7. Wang, H., Li, Y., Ryder, J.W., Hole, J.T., Ebert, P.J., Airey, D.C., Qian, H.R., Logsdon, B., Fisher, A., Ahmed, Z., et al. (2018). Genome-wide RNAseq study of the molecular mechanisms underlying microglia activation in response to pathological tau perturbation in the rTg4510 tau transgenic animal model. Mol. Neurodegener. *13*, 19–65. https://doi.org/10.1186/S13024-018-0296-Y/FIGURES/7.

8. Doktor, T.K., Hua, Y., Andersen, H.S., Brøner, S., Liu, Y.H., Wieckowska, A., Dembic, M., Bruun, G.H., Krainer, A.R., and Andresen, B.S. (2017). RNA-sequencing of a mouse-model of spinal muscular atrophy reveals tissue-wide changes in splicing of U12-dependent introns. Nucleic Acids Res. *45*, 395–416. https://doi.org/10.1093/NAR/GKW731.

9. Clements, M.P., Byrne, E., Camarillo Guerrero, L.F., Cattin, A.L., Zakka, L., Ashraf, A., Burden, J.J., Khadayate, S., Lloyd, A.C., Marguerat, S., and Parrinello, S. (2017). The wound microenvironment reprograms Schwann cells to invasive mesenchymal-like cells to drive peripheral nerve regeneration. Neuron *96*, 98–114.e7. https://doi.org/10.1016/J.NEURON.2017.09.008.

10. Hurrell, T., Segeritz, C.P., Vallier, L., et al. (2019). A proteomic time course through the differentiation of human induced pluripotent stem cells into hepatocyte-like cells. Sci. Rep. *9*, 3270–3311. https://doi.org/10.1038/s41598-019-39400-1.

11. Cao, J., Spielmann, M., Qiu, X., Huang, X., Ibrahim, D.M., Hill, A.J., Zhang, F., Mundlos, S., Christiansen, L., Steemers, F.J., et al. (2019). The single-cell transcriptional landscape of mammalian organogenesis.

Nature *566*, 496–502. https://doi.org/10.1038/s41586-019-0969-x.

12. Aristotle, Metaphysics 1.980a, Aristotle. Aristotle in 23 Volumes, Vols.17, 18, translated by Hugh Tredennick. Cambridge, MA, Harvard University Press; London, William Heinemann Ltd. 1933, 1989.

13. Thul, P.J., and Lindskog, C. (2018). The human protein atlas: a spatial map of the human proteome. Protein Sci. *27*, 233–244. https://doi.org/10.1002/PRO.3307.

14. Watson, J., Smith, M., Francavilla, C., and Schwartz, J.-M. (2022). SubcellulaRVis: a web-based tool to simplify and visualise subcellular compartment enrichment. Nucleic Acids Res. *50*, 718–725. https://doi.org/10.1093/nar/gkac336.

15. Maag, J.L.V., and Nakaya, H.I. (2018). gganatogram: an R package for modular visualisation of anatograms and tissues based on ggplot2. F1000Research, 1576. https://doi.org/10.12688/f1000research.16409.2.

16. UniProt Consortium (2019). UniProt: a worldwide hub of protein knowledge. Nucleic Acids Res. *47*, D506–D515. https://doi.org/10.1093/NAR/GKY1049.

17. Safran, M., Rosen, N., Twik, M., BarShir, R., Stein, T.I., Dahary, D., Fishilevich, S., and Lancet, D. (2021). The GeneCards suite. Pract. Guid. to Life Sci. Databases, 27–56. https://doi.org/10.1007/978-981-16-5812-9_2.

18. Kim, K.K., Adelstein, R.S., and Kawamoto, S. (2009). Identification of neuronal nuclei (NeuN) as fox-3, a new member of the fox-1 gene family of splicing factors. J. Biol. Chem. *284*, 31052–31061. https://doi.org/10.1074/JBC.M109.052969.

19. Yuan, A., Sershen, H., Basavarajappa, B., Basavarajappa, B.S., Kumar, A., Hashim, A., Berg, M., Lee, J.H., Sato, Y., Rao, M.V., et al. (2015). Neurofilament subunits are integral components of synapses and modulate neurotransmission and behavior in vivo. Mol. Psychiatr. *20*, 986–994. https://doi.org/10.1038/mp.2015.45.

20. Rehm, H., Wiedenmann, B., and Betz, H. (1986). Molecular characterization of synaptophysin, a major calcium-binding protein of the synaptic vesicle membrane. EMBO J. *5*, 535–541. https://doi.org/10.1002/J.1460-2075.1986.TB04243.X.

21. Wickham, H. (2016). ggplot2: Elegant Graphics for Data Analysis (Springer-Verlag).

22. Wu, T., Hu, E., Xu, S., Chen, M., Guo, P., Dai, Z., Feng, T., Zhou, L., Tang, W., Zhan, L., et al. (2021). clusterProfiler 4.0: a universal enrichment tool for interpreting omics data. Innovar *2*. https://doi.org/10.1016/j.xinn.2021.100141.

CellPress
OPEN ACCESS

STAR★METHODS

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Deposited data		
TRAP-seq - RNA Expression in Excitatory Neurons during Hippocampal LTP	Chen et al., 2017	https://www.frontiersin.org/articles/file/downloadfile/244243_ supplementary-materials_datasheets_1_xlsx/octet-stream/ Data%20Sheet%201.XLSX/1/244243 GEO ID: GSE79790
RNA-seq - temporal regulation of alternative splicing during neural development	Weyn-Vanhentenryck et al., 2018	https://www.refine.bio/experiments/SRP055008/developing-mouse-cortex-rna-seq NCBI Short Read Archive ID: SRP055008
RNA-seq - microglia activation in response to pathological tau	Wang et al., 2018	https://static-content.springer.com/esm/art%3A10.1186%2Fs13024-018-0296-y/MediaObjects/13024_2018_296_MOESM2_ESM.xlsx GEO ID: GSE123467
RNA-seq Wound reprogramming and peripheral nerve regeneration	Clements et al., 2017	https://www.cell.com/cms/10.1016/j.neuron.2017.09.008/attachment/229d65b5-db13-442e-9721-fe2763746e2f/mmc7.xlsx GEO ID: GSE103039
RNA-seq - Changes in GE during disease progression	Doktor et al., 2017	https://oup.silverchair-cdn.com/oup/backfile/Content_public/Journal/nar/45/1/10.1093_nar_gkw731/2/gkw731_Supp.zip?Expires=1654590951&Signature=e-4zy41FxwlXyOfa7onnHXKu6eaZe53WDLzeZx5xJTZWevpUxc48XW9BX64YZFbbcIhgJKzzjmqiY8-syWuxofEakd8-9UFJTY1cigUDhqqG2-mMa~4SOWuBUZIcJefNM0RWKeWzvxFrc-nsPeKqeskf1-zwSVsJMTmqKoemxoFORsbo8dNdVlMKbvYfApwpYHCj1o-Pm8rXdsXOg-dFi11f43BpHa0ELUW0aiRmdibACuK5F4tEFiLW0GbrEx2xEmcHDKuP5FiLoMWFA6QQ2AjjkrgsOsre7RodBRaYGuxSI60gEqunVWKoHbKuoZlCDYpXzzD4zJeot9EY7plSGg__&Key-Pair-Id=APKAIE5G5CRDK6RD3PGA ArrayExpress ID: E-MTAB-3664
Proteomics - protein levels in cells during differentiation from hiPSCs to hepatocyte-like cells	Hurrell et al., 2019	https://www.nature.com/articles/s41598-019-39400-1#Fig.2 Raw dataset ID not available
scRNA-seq – transcriptional dynamics during mouse organogenesis	Cao et al., 2019	https://oncoscape.v3.sttrcancer.org/atlas.gs.washington.edu.mouse.rna/downloads
Mouse reference genome 97	Ensembl	http://jul2019.archive.ensembl.org/
Software and algorithms		
R code for expressyouRcell package	This manuscript	https://github.com/LabTranslationalArchitectomics/expressyouRcell

RESOURCE AVAILABILITY

Lead contact
Further information and requests for resources and reagents should be directed to and will be fulfilled by the Lead Contact, Gabriella Viero (gabriella.viero@cnr.it).

Materials availability
This study did not generate new unique reagents.

Data and code availability
- This paper analyses existing, publicly available data. These accession numbers for the datasets are listed in the key resources table.

- The authors also declare that all original code has been deposited at https://github.com/LabTranslationalArchitectomics/expressyouRcell and is publicly available as of the date of publication. DOIs are listed in the key resources table.

- Any additional information required to reanalyse the data reported in this paper is available from the lead contact upon request.

EXPERIMENTAL MODEL AND STUDY PARTICIPANT DETAILS

This work considers already publicly available datasets and these accession numbers are listed in the key resources table.

METHOD DETAILS

User's input data

expressyouRcell is optimized for representing multiple gene expression datasets and accepts data in tabular format as input, for example lists of tables. Each data table must contain at least a column with gene symbols, which is sufficient to perform gene set enrichment analysis and visualize these results through cellular pictographs.

Additionally, the column with gene symbols can be coupled with i) one numerical column reporting the expression level of the gene, represented as read counts, count per million of reads (CPM) or reads per kilobase of gene per million (RPKM), protein MS intensities or any other type of expression measurement; or ii) two columns reporting expression changes (e.g. log fold-changes or Z-scores) and p values from upstream differential analyses.

To improve the flexibility in handling different data structures, *expressyouRcell* provides the user with multiple options for coloring subcellular localizations (see details in section visualizing gene expression data to the cellular pictograph for details).

Gene localization mapping

To define the color and the shade of cellular pictographs, each gene must be mapped to a specific subcellular compartment. The localization of genes can be provided by the user as a custom table with gene symbols and associated cellular localizations. Alternatively, the gene-to-localization mapping is assessed by *expressyouRcell* through the dedicated *map_gene_localization* function. This function requires as input a gene annotation file, provided in GTF format, and performs gene ontology enrichment analysis on the sets of input gene symbols. Single or multiple terms of the cellular component ontology, either cellular compartment or macromolecular complex, will be assigned to each gene.

The subcellular compartments drawn in the pictographs are mapped to terms from the cellular component ontology. Since the logical data structure of the gene ontology is organized as an hierarchical tree graph, we selected a subset of higher-ranked terms as descriptors for the pictographic organelles and compartments.

Drawing cellular pictographs

expressyouRcell currently offers specific pictographs of six cell types: i) a typical eukaryotic cell, ii) a neuron, iii) a fibroblast, iv) a microglial cell, v) a lymphocyte, and vi) a macrophage. The desired cellular type has to be specified by the user in the *color_cell* function parameters.

The spatial coordinates of regions/organelles are extracted from multiple vector graphic files, one for each subcellular region or organelle, through the *rsvg_svg* function of the *rsvg* R package. The resulting cellular structures are then stored in RData objects, provided by *expressyouRcell*, and used by the function *geom_polygon* from the *ggplot* (Wickham, 2016) package to generate the final pictographs.

A selection of subcellular regions and organelles have been chosen and drawn to create the cell pictographs (Table S1). The selection was based on the highest ranked nodes in the gene ontology cellular component graph and includes the main cellular regions and organelles. Importantly, the subset of cellular component terms used for the cellular pictograph is cell-type specific.

Visualizing gene expression data to the cellular pictograph

The *color_cell* function provides two main approaches for assigning colors to the subcellular localizations and for visualizing gene expression data pictographs.

The first option is based on the outcome of gene set enrichment analysis and on its representation. This approach takes into consideration the identity and localization of genes, but not their expression level. Starting from the complete list of gene symbols, *expressyouRcell* performs the gene set enrichment analysis for each cellular component and assesses the statistical significance of the enrichment by the Fisher's test and the color shade of each compartment is defined by its associated p value. An optional gene classification can be specified by the user through distinct categorical variables (e.g., "down-regulated" and "up-regulated"). In this case, a cellular pictograph for each class is generated following the same procedure described above.

The second option allows the visualization of either gene expression levels or the result of differential analyses previously performed by the user. For each gene, this approach requires a numerical value reporting i) its gene or transcript expression level (reads count, TPM, CPM or RPKM) or protein levels, or ii) the outcome of differential analyses among multiple samples (e.g., fold changes, fold enrichments). Library-normalized counts (CPM or RPKM) are recommended to avoid biased and misleading results. If raw reads counts are provided, *expressyouRcell* can automatically handle this normalization step. Moreover, row-wise scaling of input values can be computed, resulting in gene-specific Z-scores.

The name of the column that defines the color of the pictograph must be specified by the user. To provide further flexibility in data visualization, *expressyouRcell* can generate cellular pictographs on the entire gene set or on subsets defined by the user. In the first case the color of each cellular compartment is defined by the mean (Equation 1), or median, of all gene-specific values associated with each localization

$$\forall l \in \{l_1, .., l_k\} \frac{\sum_{i=1}^{n} g_l^i}{n} \qquad \text{Equation 1}$$

where $l \in \{l1, ..., lk\}$ is a cellular localization, n is the total number of genes associated with the localization l and g_i^l is the gene-specific value (gene expression level or fold changes/enrichments). The absolute value of fold changes can be computed before calculating the mean or median value per each organelle, to avoid blank organelles arising from the integration of positive and negative values assigned to multiple genes in the same cellular compartment. Combined p values can also be used to assign colors to cellular compartments. In this case, more statistically sound methods (i.e. Fisher's method) should be applied beforehand to combine multiple p values assigned to the same cellular localization.

In the second case, *expressyouRcell* generates multiple pictographs, one for each subset of genes defined by the user according to an optional classification of genes (e.g., "down-regulated" and "up-regulated"). With this option, a cellular pictograph for each class is generated and the color of each cellular compartment is calculated as in Equation 2, for each gene subset.

$$\forall c \in \{c_1, .., c_m\}, \forall l \in \{l_1, .., l_k\} \frac{\sum_{i=1}^{n} g_{c,l}^i}{n} \qquad \text{Equation 2}$$

where $c \in \{c1, ..., ck\}$ defines a class of genes, $l \in \{l1, ..., lk\}$ is a cellular localization, n is the number of genes in class c associated with the localization l, and $g_{c,l}^i$ corresponds to their gene-specific value (gene expression level or fold changes/enrichments).

If no categorical variables are specified by the user, *expressyouRcell* classifies the genes according to a combination of cut-off values (e.g., *fold-change* and p *value* thresholds).

Animated pictographs

expressyouRcell generates dynamic representations of data shown in the cellular pictographs. This is particularly useful when the user's input data consists of multiple sets of time-course gene expression data. The generation of animated pictographs depends on information obtained by the *color_cell* function.

In this step, multiple data structures obtained by the *color_cell* function are required as input to the *animate* function, which creates a short Video or an animated picture. Additional mandatory inputs consist of the name of data tables for generating the animation, the transition duration (in seconds) and the number of frames/transitions. A bar timeline at the top of the animation plot allows a better tracking of the changes over time. Hence, the user can use as additional input a vector of labels for a customized timeline.

The working flow of this function begins with the generation of temporary frames with intermediate color shades for each transition. The complete set of frame pictures is merged into a single animated picture (GIF) or short Video in *mp4* format. The *gifski* and *av* packages are used to produce the animated GIF picture and the movie, respectively.

Output

expressyouRcell generates both static and dynamic representations of cellular pictographs that are associated with two main types of data structures.

The *color_cell* function generates static cellular pictographs and returns a list with multiple data structures. The first is a data table with six columns, reporting, for each subcellular component: i) its name, ii) the numeric value computed during the color assignment step, iii) a numeric identifier for grouping the cellular localizations by color, iv) its associated color shade, the v) the identifier of each dataset, and, if present, and vi) the variable used for grouping the genes by classification.

The second data structure is a data table summarizing the information on ranges used to categorize each subcellular localization (e.g., start, end, color, and labels).

The third and fourth data structures are lists of graphical objects of class *ggplot*, and datasets used to plot the resulting cellular pictographs, respectively. When multiple cell type names are input in the *color_cell* function, the output is composed by a graphical representation displaying the different cell pictographs at once, allowing the user to compare the visualization on different cell types.

The *animate* function generates dynamic cellular pictographs and saves them either as Video (in *mp4* format) or as animated pictures (in *gif* format). The saving step is performed directly within the function.

QUANTIFICATION AND STATISTICAL ANALYSIS

Color shades of pictographs based on the significance of the enrichments were defined by p values from one side Fisher's test.

Tool availability

expressyouRcell is a package written in the R programming language which depends on multiple R packages, such as *ggplot*[21] for data visualization, *clusterProfiler*[22] for gene ontology enrichment analysis, *gifski* and *av* for the GIF animation and video realization, respectively. *expressyouRcell* source code can be downloaded from the following GitHub repository: https://github.com/LabTranslationalArchitectomics/expressyouRcell, accompanied with installation instructions and functions documentation. *expressyouRcell* is an Open-Source software package (distributed under the MIT license), which is compatible with Linux, Mac, or Windows PCs. Future releases may be expanded by additional cellular pictographs to support the visualization of data on a wider range of cellular types. Also, an interactive web application could also be deployed through the Shiny framework to further expand the usability and reachability of our tool.

The package includes the R implementation of expressyouRcell, data used in this article, extensive documentation and a stable release