

# LEVI: an R package for landscape expression visualization

José Rafael Pílan<sup>1</sup>, Isabelle M. da Silva<sup>1</sup>, Agnes A. Sekijima Takeda<sup>1</sup>,  
and José L. Rybarczyk-Filho<sup>1</sup>

<sup>1</sup> Department of Biophysics and Pharmacology, Institute of Bioscience of Botucatu, Universidade Estadual Paulista (UNESP)

DOI: 10.xxxxxx/draft

## Software

- [Review](#)
- [Repository](#)
- [Archive](#)

Editor: [✉](#)

Submitted: 26 February 2024

Published: unpublished

## License

Authors of papers retain copyright and release the work under a Creative Commons Attribution 4.0 International License ([CC BY 4.0](#)).

## Summary

The integration of biological data emerges as a powerful tool for extracting information capable of describing the biological system in greater detail. The use of transcriptomic data has become ubiquitous among researchers in addressing various biological questions. In the past decade, we have witnessed the integration of biological network analysis to enrich and complement these responses. This study proposes the synergistic integration of transcriptomic data and networks, adopting an approach analogous to the Heatmap technique. The Landscape Expression Visualization Interface (LEVI) is an open-source package developed for the R environment, aiming to provide an enhanced visualization of gene expression projection onto a biological network.

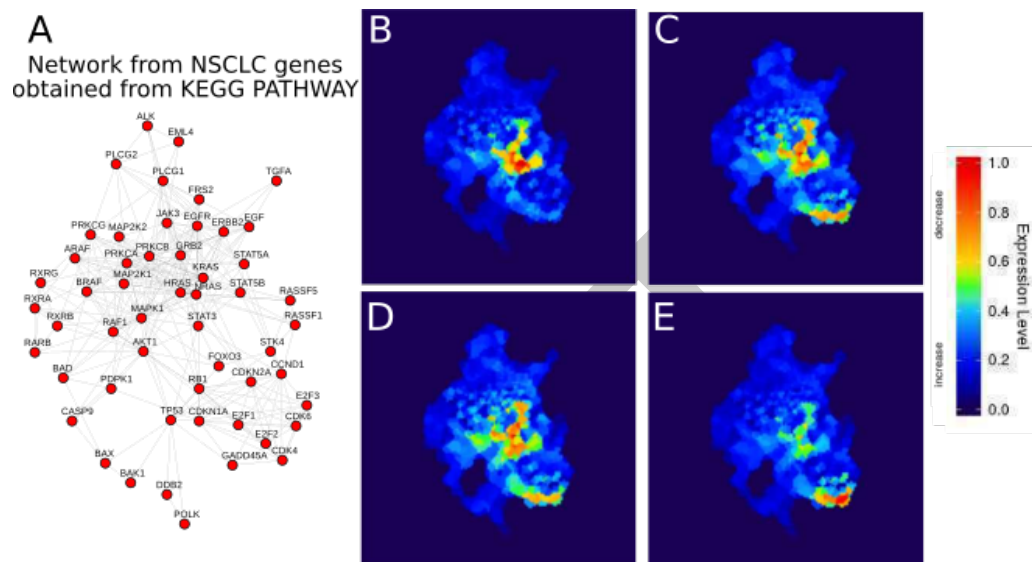
## Introduction

The high throughput technologies have been developed over the last decades have revolutionized the acquisition of life sciences knowledge, resulting in 'omics'-scale data (Gehlenborg et al., 2010). Consequently, developing methods and tools for data integration and visualization is essential to qualitatively and quantitatively enhance our understanding of biological systems (Kannan et al., 2016). Biological networks allow for the visualization of relationships among molecules within a pathway. When additional information (e.g. expression data) is integrated to the network it can help to understand which molecules play central roles in complex systems (Charitou et al., 2016). An effective approach for visual identification of patterns in biological data involves its integration into biological networks, resulting in 'colored networks' that serve as smoothed data maps (Camilo et al., 2013; Castro et al., 2009). We present the Landscape Expression Visualization Interface (LEVI), a package developed for the R environment that integrates gene expression levels or other features (e.g., clustering coefficient, connectivity, stress) with biological networks. LEVI displays a heatmap with surface curves (landscape) to highlight altered regions, facilitating data processing and evaluation within a single environment.

## Implementation and main functions

LEVI was developed in R programming language, incorporating the Shiny package to create a graphical user interface (GUI). Normalized gene expression data can be adjusted to logarithmic scale for better visualization of the expression levels projected onto the network. The network upload requires files in formats Medusa (DAT), RedeR (DYN), Pajek (NET) or STRING / STITCH. The GUI mode allows the user the landscape visualization from a single gene expression data or a comparison of data expression. Moreover, it allows the interaction with the final image, selecting areas in the landscape to identify the genes in a chosen region. LEVI

settings contrast, resolution, smoothing, zoom and color palettes are also available. The script mode enables the batch execution to create landscapes for different comparisons between sample types for a single pathway/network. Both modes export the landscapes as images in TIFF, BMP, JPEG and PNG formats.



**Figure 1:** Landscape analysis to Non-Small Cell Lung Cancer signaling pathway. (A) Biological network for Non-Small Cell Lung Cancer. (B) Landscape for normal lung and never smoked. (C) Landscape for tumor and current smoker. (D) Landscape for tumor and former smoker. (E) Landscape for tumor and never smoked.

## Case Study

To illustrate the LEVI application, we built a Non-Small Cell Lung Cancer (NSCLC) network, which was obtained by prospection of the genes from the Non-Small Cell Lung Cancer signaling pathway (map05223) from the Kyoto Encyclopedia of Genes and Genomes(KEGG) database (Kanehisa et al., 2019). These genes were used as input for the STRING v11 (Szklarczyk et al., 2018) to built the NSCLC network using the following parameters: active interaction sources - Experiments and database, confidence score 0.40 (considered medium confidence) Figure (A). The files “network-coordinates” and “as simple tabular text output” were selected and used as input for LEVI package. Then, we prepared the expression file. In our example, we selected a gene expression dataset of lung adenocarcinoma and lung tissue from current, former and never smokers (GSE10072) (Landi et al., 2008) available in Gene Expression Omnibus (GEO) (Barrett et al., 2013). Before the upload in LEVI, the data was previously normalized with MAS5 and the probes were mapped with gene symbol using the R packages affy and hgu133a2.db, respectively.

The data processing in LEVI allowed the visualization of the expression data onto the biological network to evidence difference among samples (Figure 1 (B-E)). The comparison among images indicated a difference of colors in one area that corresponds to the E2F2 and CDK4 genes, which were more expressed in the samples of lung cancer patients than in the control samples. This result was corroborated by previous studies, which also suggested these genes as potential biomarkers for prognosis in detection of NSCLC (Du et al., 2022; Wu et al., 2011).

## Conclusion

The landscape visualization, provided by LEVI — a tool available on the Bioconductor.org platform — results from the integration of expression data and the network. This allows for a more comprehensive qualitative analysis of the biological problem, due to the improved visualization of both upregulated and downregulated genes, as well as their interactions with the surrounding network.

## Funding

This study was supported by CNPq 473789/2013-2, 134469/2016-0, 134467/2016-7 and financed in part by the Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - Brasil (CAPES) - Finance Code 001.

## References

- Barrett, T., Wilhite, S. E., Ledoux, P., Evangelista, C., Kim, I. F., Tomashevsky, M., Marshall, K. A., Phillippy, K. H., Sherman, P. M., Holko, M., Yefanov, A., Lee, H., Zhang, N., Robertson, C. L., Serova, N., Davis, S., & Soboleva, A. (2013). NCBI GEO: Archive for functional genomics data sets—update. *Nucleic Acids Research*, 41, D991–5. <https://doi.org/10.1093/nar/gks1193>
- Camilo, E., Bovolenta, L. A., Acencio, M. L., Rybarczyk-Filho, J. L., Castro, M. A. A., Moreira, J. C. F., & Lemke, N. (2013). GALANT: a Cytoscape plugin for visualizing data as functional landscapes projected onto biological networks. *Bioinformatics*, 29(19), 2505–2506. <https://doi.org/10.1093/bioinformatics/btt377>
- Castro, M. A. A., Filho, J. L. R., Dalmolin, R. J. S., Sinigaglia, M., Moreira, J. C. F., Mombach, J. C. M., & Almeida, R. M. C. de. (2009). ViaComplex: software for landscape analysis of gene expression networks in genomic context. *Bioinformatics*, 25(11), 1468–1469. <https://doi.org/10.1093/bioinformatics/btp246>
- Charitou, T., Bryan, K., & Lynn, D. J. (2016). Using biological networks to integrate, visualize and analyze genomics data. *Genetics Selection Evolution*, 48(1), 27. <https://doi.org/10.1186/s12711-016-0205-1>
- Du, K., Sun, S., Jiang, T., Liu, T., Zuo, X., Xia, X., Liu, X., Wang, Y., & Bu, Y. (2022). E2F2 promotes lung adenocarcinoma progression through b-myb- and FOXM1-facilitated core transcription regulatory circuitry. *International Journal of Biological Sciences*, 18, 4151–4170. <https://doi.org/10.7150/ijbs.72386>
- Gehlenborg, N., O'Donoghue, S. I., Baliga, N. S., Goesmann, A., Hibbs, M. A., Kitano, H., Kohlbacher, O., Neuweger, H., Schneider, R., Tenenbaum, D., & Gavin, A.-C. (2010). Visualization of omics data for systems biology. *Nature Methods*, 7(3), S56–S68. <https://doi.org/10.1038/nmeth.1436>
- Kanehisa, M., Sato, Y., Furumichi, M., Morishima, K., & Tanabe, M. (2019). New approach for understanding genome variations in KEGG. *Nucleic Acids Research*, 47, D590–D595. <https://doi.org/10.1093/nar/gky962>
- Kannan, L., Ramos, M., Re, A., El-Hachem, N., Safikhani, Z., Gendoo, D. M. A., Davis, S., Gomez-Cabrero, D., Castelo, R., Hansen, K. D., Carey, V. J., Morgan, M., Culhane, A. C., Haibe-Kains, B., & Waldron, L. (2016). Public data and open source tools for multi-assay genomic investigation of disease. *Briefings in Bioinformatics*, 17, 603–615. <https://doi.org/10.1093/bib/bbv080>

- 107 Landi, M. T., Dracheva, T., Rotunno, M., Figueroa, J. D., Liu, H., Dasgupta, A., Mann, F. E.,  
108 Fukuoka, J., Hames, M., Bergen, A. W., Murphy, S. E., Yang, P., Pesatori, A. C., Consonni,  
109 D., Bertazzi, P. A., Wacholder, S., Shih, J. H., Caporaso, N. E., & Jen, J. (2008). Gene  
110 expression signature of cigarette smoking and its role in lung adenocarcinoma development  
111 and survival. *PloS One*, 3, e1651. <https://doi.org/10.1371/journal.pone.0001651>
- 112 Szklarczyk, D., Gable, A. L., Lyon, D., Junge, A., Wyder, S., Huerta-Cepas, J., Simonovic,  
113 M., Doncheva, N. T., Morris, J. H., Bork, P., Jensen, L. J., & Mering, C. von. (2018).  
114 STRING v11: protein–protein association networks with increased coverage, supporting  
115 functional discovery in genome-wide experimental datasets. *Nucleic Acids Research*, 47(D1),  
116 D607–D613. <https://doi.org/10.1093/nar/gky1131>
- 117 Wu, A., Wu, B., Guo, J., Luo, W., Wu, D., Yang, H., Zhen, Y., Yu, X., Wang, H., Zhou,  
118 Y., Liu, Z., Fang, W., & Yang, Z. (2011). Elevated expression of CDK4 in lung cancer.  
119 *Journal of Translational Medicine*, 9, 38. <https://doi.org/10.1186/1479-5876-9-38>

DRAFT