

levi package

February 19, 2024

1 Getting Start with Levi

Levi recognizes network files in the formats Medusa (DAT) Pavlopoulos *et al.* (2011), RedeR (DYN) Castro *et al.* (2012), Pajek (NET) Mrvar and Batagelj (2016) and STRING/STITCH Szklarczyk *et al.* (2019, 2016). The expression file must have a column with Gene Symbol data containing the gene nomenclature and at least one column with the expression values. It is also possible to compare two columns of expression values (Test / Control).

1.1 Cell Adhesion Network

We selected an example to demonstrate some of the possible results obtained using levi to prospect the projection of expression data over a biological network. The Cell Adhesion Network were built from the map hsa04514 (https://www.genome.jp/kegg-bin/show_pathway?hsa04514) obtained form the Kyoto Encyclopedia of Genes and Genes (KEGG). The set of genes from this map was used as entry for STRING database v11.0 (<https://string-db.org/>), and the following parameters were selected:

- active interaction sources: Experiments and Database
- Confidence score: 0.7 (high score)

We obtained a network with 133 genes and 700 interactions, according to the figure 1.

1.2 Expression Gene Expression

We used the expression data set GSE10072, deposited in the Gene Expression Omnibus (GEO) for the article “Gene expression signature of cigarette smoking and its role in lung adenocarcinoma development and survival” (<https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE10072>) Landi *et al.* (2008).The author used for this experiment the HG-U133A - Affymetrix Human Genome U133A Array platform. The 107 samples of lung adenocarcinoma and normal lung tissue were analyzed for smokers,former smokers and non-smokers. We performed the gene expression data normalization using MAS5 and we divided the data in six groups, according to the list below:

1. Normal Lung Tissue from Never Smoked - 15 samples
2. Normal Lung Tissue from Former Smoker - 18 samples
3. Normal Lung Tissue from Current Smoker - 16 samples
4. Lung Adenocarcinoma from Never Smoked - 16 samples
5. Lung Adenocarcinoma from Former Smoker - 18 samples
6. Lung Adenocarcinoma from Current Smoker - 24 samples

Each group consisted of the average expression of each gene for each tissue. These data were used as input to the levi.

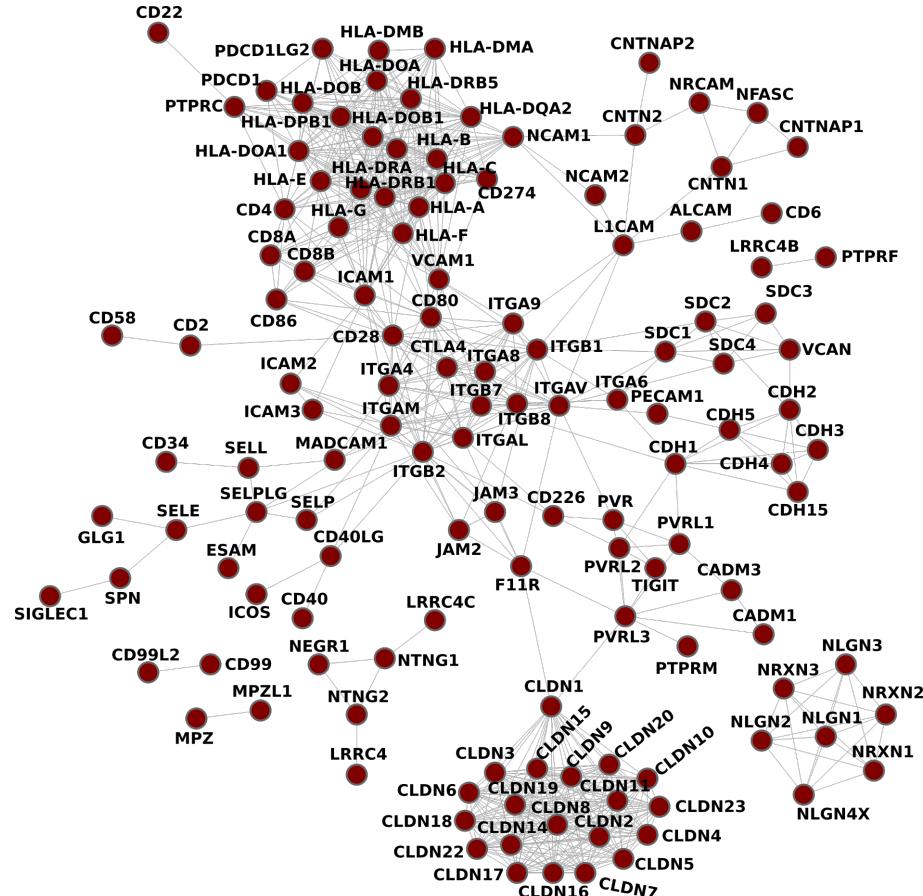


Figure 1: Cell Adhesion Network generated in STRING with biomolecular components.

2 How to Install levig

The levi package is available on Bioconductor open source software for bioinformatics:
<http://bioconductor.org/packages/release/bioc/html/levi.html>.

```
1 #Install levi on R
2 if (!requireNamespace("BiocManager", quietly = TRUE))
3   install.packages("BiocManager")
4
5 BiocManager::install("levi")
```

After the installation, levi can be loaded in the browser (default) or with the R engine itself. The lines of code for loading and running the modes of the levi package are available below.

```
1 library(levi)
2 #Run levi in the default browser
3 LEVIui(TRUE)
4
5
6 #Run the levi in R
7 LEVIui(FALSE)
8
9 #Run levi in script mode by command line
10 levi(expressionInput , fileTypeInput , networkCoordinatesInput ,
11 networkInteractionsInput , geneSymbolInput , readExpColumn ,
12 contrastValueInput , zoomValueInput , resolutionValueInput ,
13 smoothValueInput , expressionLog , contourLevi , setcolor)
```

3 Running levi in script mode to batch execution

We used two examples to show the execution of levi in the R console. The first example used the expression data of smokers that do not have adenocarcinoma compared to people who never smoked and do not have adenocarcinoma. The second example was the comparison between smokers that do not have adenocarcinoma and people who have never smoked and do not have adenocarcinoma. We executed levi and generated the landscapes using the following commands in the R console:

```
1 library(levi)
2 levi(expressionInput = "expression.dat", fileTypeInput = "stg",
3 networkCoordinatesInput = "string_network_coordinates.txt",
4 networkInteractionsInput = "string_interactions.dat",
5 geneSymbolInput = "ID", readExpColumn = readExpColumn("NormalCurrentSmoker-NormalNeverSmoker",
6 "NormalFormerSmoker-NormalNeverSmoker"),
7 contourLevi = FALSE)
```

In our examples, the smoothing, zoom, contrast, and resolution parameters were not shown. In this case, they assume the default values (value = 50).

Two landscapes were acquired in the first example, which corresponds to: i) normal lung from smokers (group 3) compared to normal lung from non-smokers (group 1); ii) Normal lung from former smokers (group 2) compared to normal lung from non-smokers. These landscapes had the predominance of the green color due the similarity among samples in these regions, which corresponds to similar expression levels of the genes. Spots in yellow indicated differences in samples, in which case samples had more expressed genes than the normal lungs samples (**Figure 2**).

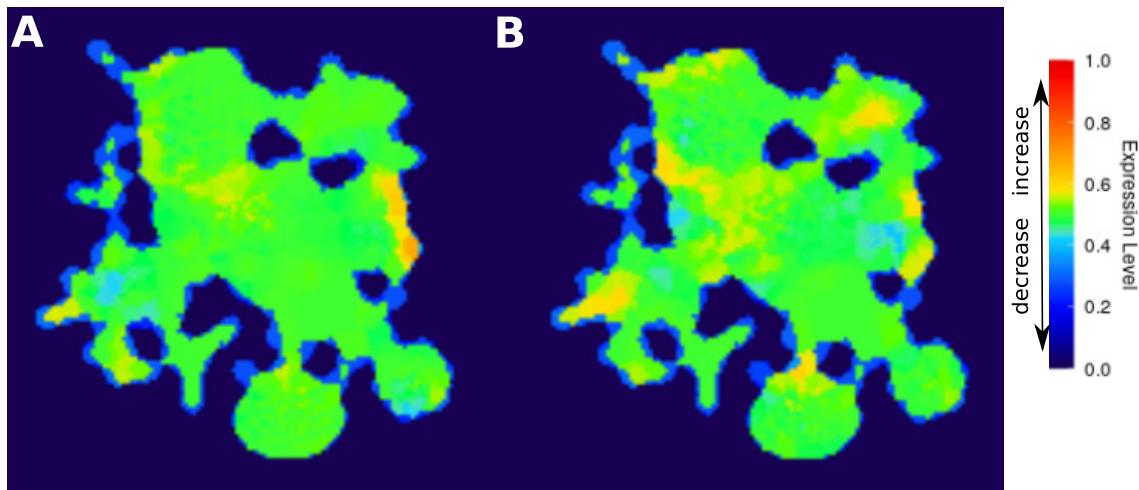


Figure 2: Landscapes generated by levi. (A) Normal lung from smokers compared to Normal lung from non-smokers. (B) Normal lung from former smokers compared to Normal lung from non-smokers.

The second example comprised three comparisons: i) Lung Adenocarcinoma from Never Smokers (group 4) compared to Normal lung from Never smokers (group 1); ii) Lung Adenocarcinoma from Former Smokers (group 5) compared to Normal lung from Never smokers. iii) Lung Adeno-carcinoma from Current Smoker (group 6) compared to Normal lung from Never smokers.

```

1 library(levi)
2
3 levi(expressionInput = "expression.dat", fileTypeInput = "stg",
4 networkCoordinatesInput = "string_network_coordinates.txt",
5 networkInteractionsInput = "string_interactions.dat",
6 geneSymbolInput = "ID", readExpColumn = readExpColumn(
7   "TumorCurrentSmoker-NormalNeverSmoker",
8   "TumorFormerSmoker-NormalNeverSmoker",
9   "TumorNeverSmoker-NormalNeverSmoker"),
9 contourLevi = FALSE)

```

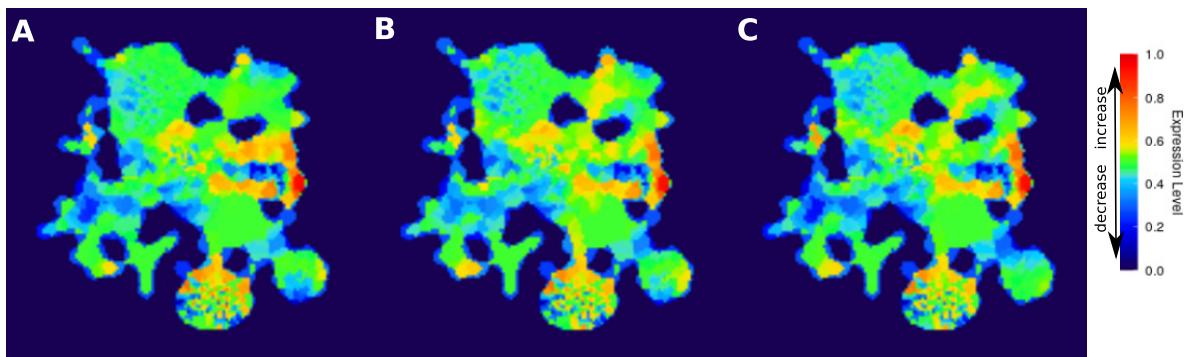


Figure 3: Landscapes generated by levi. (A) Lung Adenocarcinoma from Never Smoked compared to Normal lung from Never smokers. (B) Lung Adenocarcinoma from Former Smoker compared to Normal lung from Never smokers. (C) Lung Adenocarcinoma from Current Smoker compared to Normal lung from Never smokers.

The batch mode allowed the generation of all the tree images using a single set of lines of code. The **figure 3** showed more colored landscapes than the previous one, suggesting more differences among the comparisons. The green areas showed no differences between the groups gene expression levels. The blue areas indicated low expression levels of genes in the tumor samples than in the normal tissue. The yellow and red areas showed high gene expression levels in the tumors than in normal tissue.

4 Running levi using the Graphical User Interface (GUI)

The GUI mode was developed using Shiny package (Figure 4). This mode can be used in the R environment (browser=FALSE) or in the default web browser of the operating system (browser=TRUE) to generate the landscapes and adjust the settings without writing lines of code in R. The figure shows the features from levi GUI.

1. Tabs:

- **File:** General options for uploading the data files;
- **Settings:** Options to build the **landscape**.

2. **Network input type:** Selection of the biological network input format. The options available are: Medusa(dat), RedeR(dyn), Pajek(NET), STRING/STITCH (stg). If the user chooses a STRING/STITCH format, the levi GUI will open a new box for the user upload the coordinates file from STRING/STITCH (see Figure 5);

3. **Upload the network file:** Input button to select the biological network file;

4. **Upload the expression file:** Input button to select gene expression levels file;

5. **Expression values in log scale:** select when there is low variation between expression data;

6. Selected fields:

- Two Samples: Comparison between two samples;
- One Sample: Single analysis.

7. **Selected ID field:** Select the gene ID from gene expression levels file. The gene ID in the biological network file must be the same in the gene expression levels file. **Select test field**-Select the case/test sample. **Select control field** - Select the control sample. If levi detects the case/test sample and control sample as equal, then levi will apply the "single sample" analysis;

8. **Chart Colors:** Color palette to build the **landscape**. There are two options: the **Multicolor** which has 20 combined color levels and the **Two colors**. The selection of **Two colors** displays as available options, the following pairs of colors: purple_pink, green_blue, blue_yellow, pink_green, orange_purple, green_marine;

9. **Chart with contour:** Enable or disable the contour lines in the **landscape**;

10. **Run:** Button to execute the application;

11. **Landscape:** Image display area. The user can select specific areas of the image to inspect the gene name and position;

12. **Expression area:** Total expression value of the selected area in the image;

13. Download options:

- **File format to landscape:** Select the output format of the **landscape**. The available image formats are: TIFF, BMP, JPEG and PNG. We recommend the use of this option in the browser;
- **Download Plot:** Button to save the file.

14. **Gene names visualization:** Visualization of gene name of the selected area in the **landscape**;

15. Settings:

- **Contrast:** Contrast value in the **landscape**. The variable range is 1 to 100. The default value is 50;

- **Resolution:** Image size of the **landscape**. The variable range is 1 to 100. The default value is 50. If this parameter is higher, then the total time required will be longer;
- **Smoothing:** Smoothing of the **landscape**. The variable range is 1 to 100. The default value is 50. If this parameter is higher, then the total time required will be longer;
- **Zoom:** Zoom value for the **landscape**. The variable range is 1 to 100. The default value is 50.

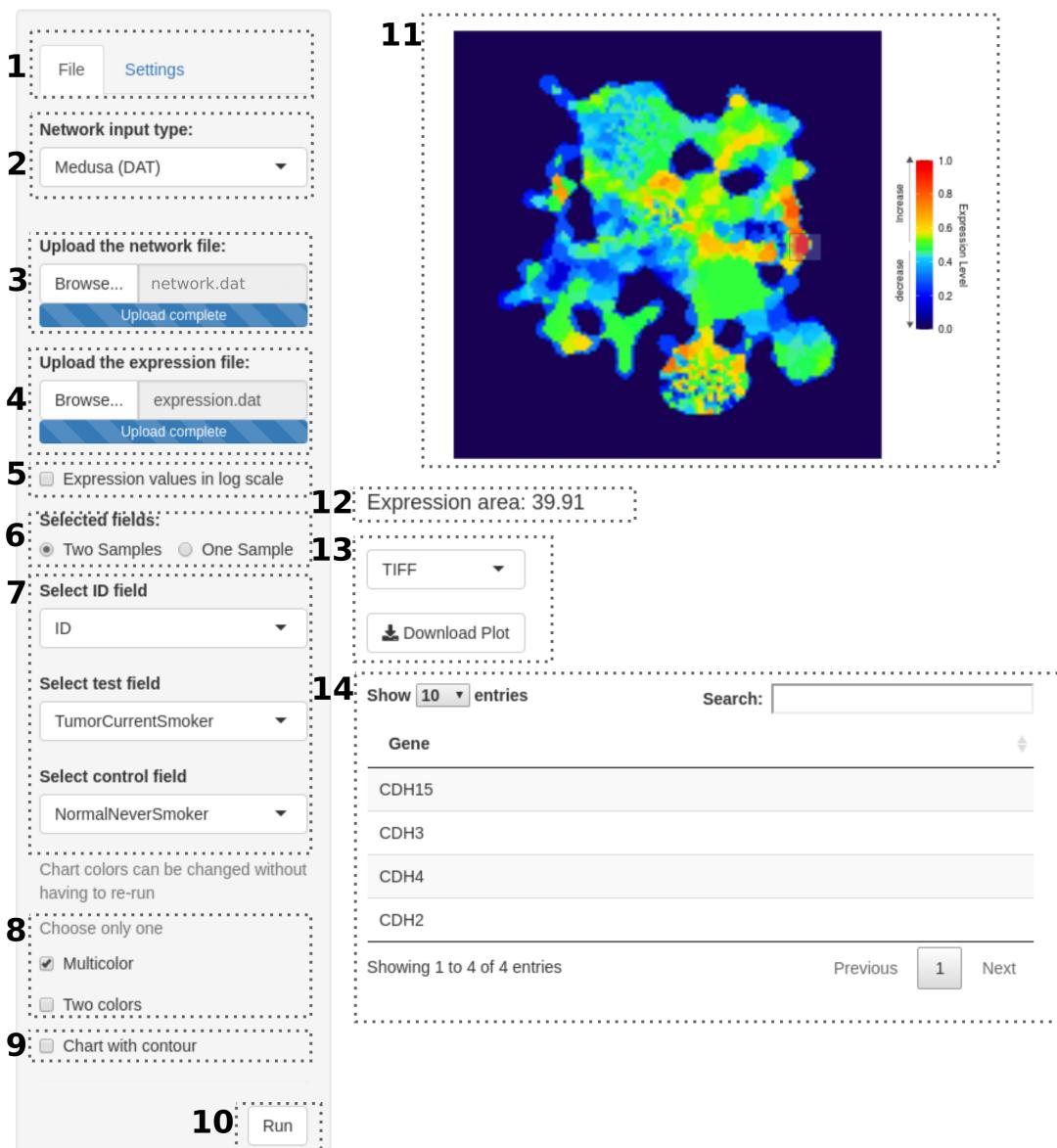


Figure 4: The graphical user interface of levi. The main tab contains the features to generate the landscapes (1-10). The landscape, displayed with its respective color key (11) can be selected to inspect an area (12) and saved as figure (13). The genes from the area are arranged in a table (14).

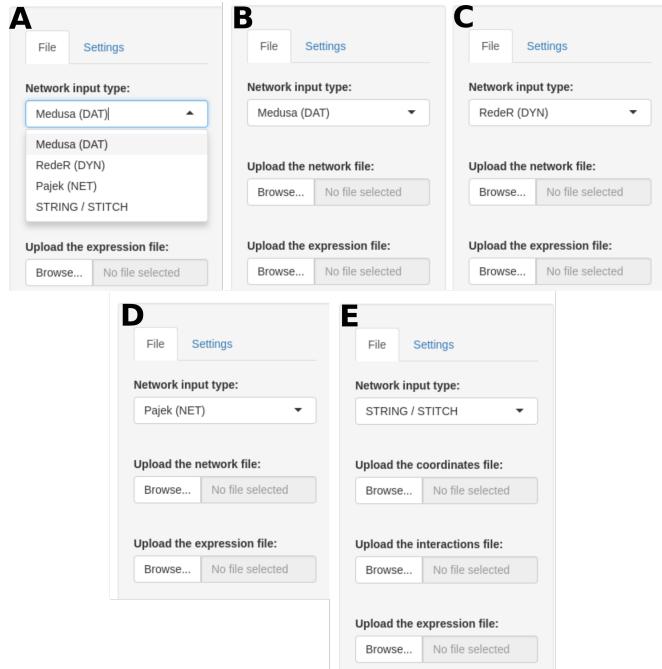


Figure 5: Biological Network inputs (A) Levi supports several extensions of biological networks files (Medusa, RedeR, Pajek, STRING/STITCH). (B) Medusa option required a single network file in format *.dat. (C) RedeR option required a single network file in format *.dyn. (D) Pajek option required a single network file in format *.net. (E) STRING/STITCH option required two files: The first file contains a coordinates for each node (protein/gene) and a second file contains the interactions of the network

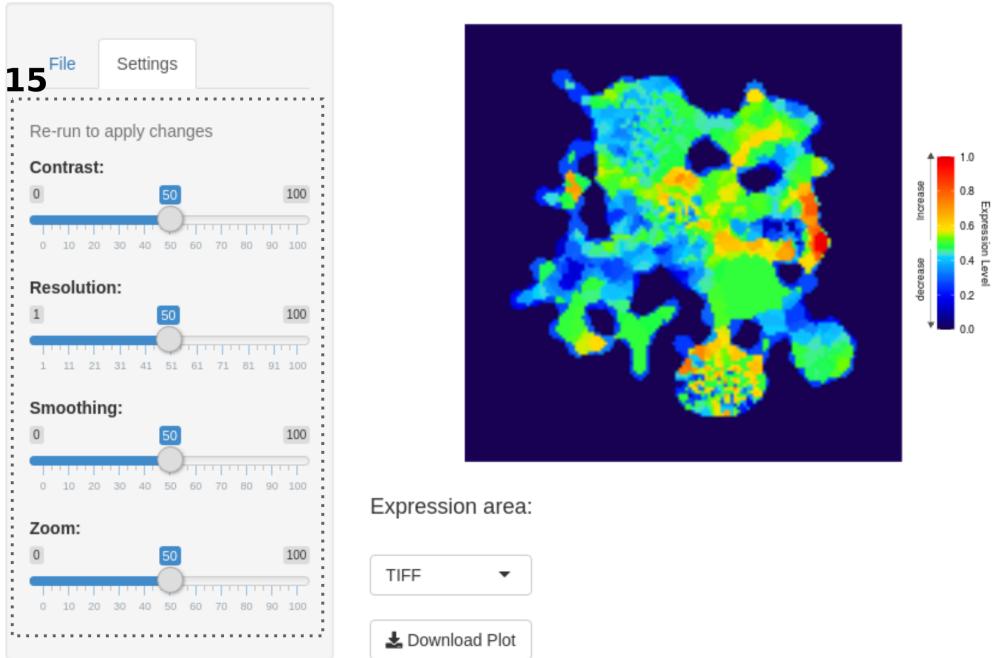


Figure 6: The second tab from levi application. It contains settings for contrast, resolution, smoothing and zoom (15).

References

- Castro, M. A. A. et al. (2012). Reder: R/bioconductor package for representing modular structures, nested networks and multiple levels of hierarchical associations. *Genome Biology*, **13**(4), R29.

- Landi, M. T. *et al.* (2008). Gene expression signature of cigarette smoking and its role in lung adenocarcinoma development and survival. *PloS one*, **3**, e1651.
- Mrvar, A. and Batagelj, V. (2016). Analysis and visualization of large networks with program package pajek. *Complex Adaptive Systems Modeling*, **4**(1), 6.
- Pavlopoulos, G. A. *et al.* (2011). Medusa: A tool for exploring and clustering biological networks. *BMC Research Notes*, **4**(1), 384.
- Szklarczyk, D. *et al.* (2016). Stitch 5: augmenting protein-chemical interaction networks with tissue and affinity data. *Nucleic acids research*, **44**, D380–D384.
- Szklarczyk, D. *et al.* (2019). String v11: protein-protein association networks with increased coverage, supporting functional discovery in genome-wide experimental datasets. *Nucleic acids research*, **47**, D607–D613.