



# Lheuristic-A Shiny app to select genes potentially regulated by methylation

Sánchez-Pla, Alex<sup>1,3</sup>, Castellano, Pol<sup>1</sup>, Miró, Berta<sup>2</sup>, Carmona, Francesc<sup>1</sup>

<sup>1</sup> Departament de Genètica Microbiologia i Estadística, Universitat de Barcelona

<sup>2</sup> International Rice Research Institute, (IRRI), Philippines

<sup>3</sup> Statistic and Bioinformatics Unit. Vall d'Hebron Research Institute. (VHIR). Barcelona



## 1 Introduction

- Methylation of genes involved in the oncogenic process is a key process contributing to tumor initiation and/or progression[4].
- Finding *Genes Regulated by Methylation* or GRM can lead to a better understanding and be a guide to finding new drug targets.
- This study originates in a work searching for colon cancer biomarkers [1]. Cell lines with increasing sensitivity to a chemotherapy drug, were analyzed with Expression and Methylation arrays. Finding GRM was used to search of candidate genes for new therapies.
- In cancer-related genes it is common to observe a decrease in gene expression associated with hypermethylation. Methylation is often described as a binary on-off signal ([2]) that is, when methylation is “off” the gene can express normally and its expression will be low or high, whereas when methylation is “on”, the expression of the gene will be *repressed* and its values will tend to be low.
- As a consequence of this *high-methylation/low-expression* and *low-methylation/high-expression* relation plots depicting methylation and expression will show L-shape patterns so the strategy adopted will be to mine such plots and select those that have such a shape.

## 2 What does the app do?

Visually explore the relation between expression and methylation -omic results to potentially identify genes regulated by methylation.

The app Lheuristic will:

- Two different methods of analysis
- Step-by-step parameter selection
- Instant data visualization
- Table and graph export options and customization

## 3 App Usage

- Data upload
- Setting of parameters
- Analysis
  - Method choice (Correlation, L-heuristic)
- Intersection

The screenshot shows the 'Selection of L-shaped genes using a heuristic algorithm' app interface. It includes a sidebar with navigation options (Home, Upload Data, Analysis, Correlations, L-heuristic, Intersection, Help). The main panel is divided into 'Choose files' (with demo data and upload options), 'Choose input files' (with a file browser), and 'Set format parameters of your methylation data file' (with options for separator, decimal, quote, and quote style). Below this is the 'Expression Data' table, which shows a list of genes (ABCG2, ADH6, AKR1C4, BCL11B, CA9) and their methylation values across various cell lines (GAC02, COLO205, COLO208, COLO209, DLD1, HCC2998, HCT116, HCT15).

Data uploading

### 3.1 Correlation method

a simple absolute correlation can be applied to the list of genes. The adjustable parameters are:

- Correlation type (Pearson, Spearman, Kendall)
- Correlation coefficient
- Correlation p-value

Optionally, a linear model fit can be overlayed on the scatterplot graph.

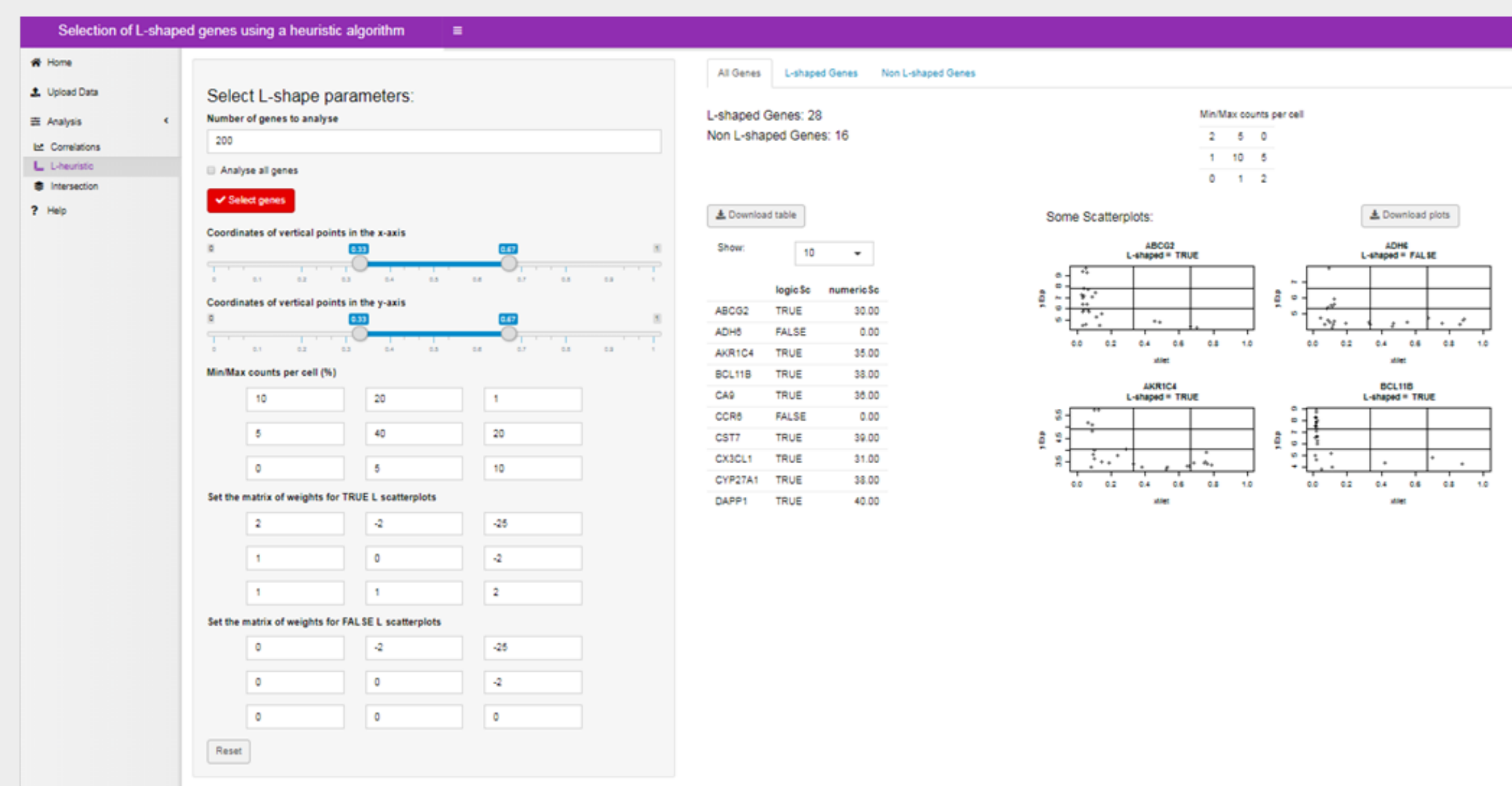


Correlation method

### 3.2 L-heuristic method

A heuristic method has been developed basing on imitating the visual selection of L-shapes by overimposing a  $3 \times 3$  grid on the scatterplot (ref F1000).

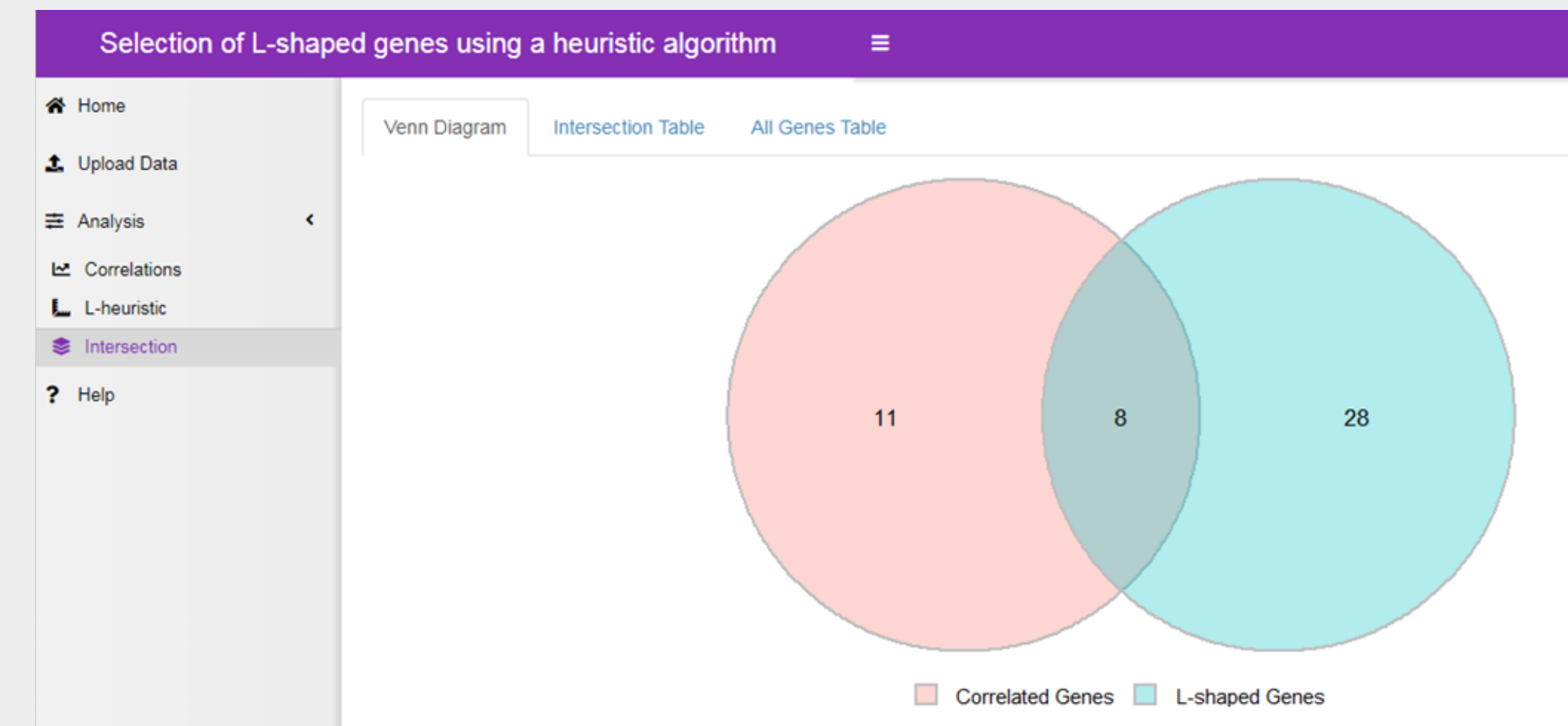
- Parameter selection
  - Grid size (x and y axis)
  - Percentage of points per grid cell
  - Weights matrix defining true and false gene shapes on the grid



L-heuristic method

### 3.3 Intersection analysis

A Venn diagram plots the relation of L-shaped genes selected between the correlation and the L-heuristic methods.



Intersection of L-shaped genes selected through the correlation and the L-heuristic methods

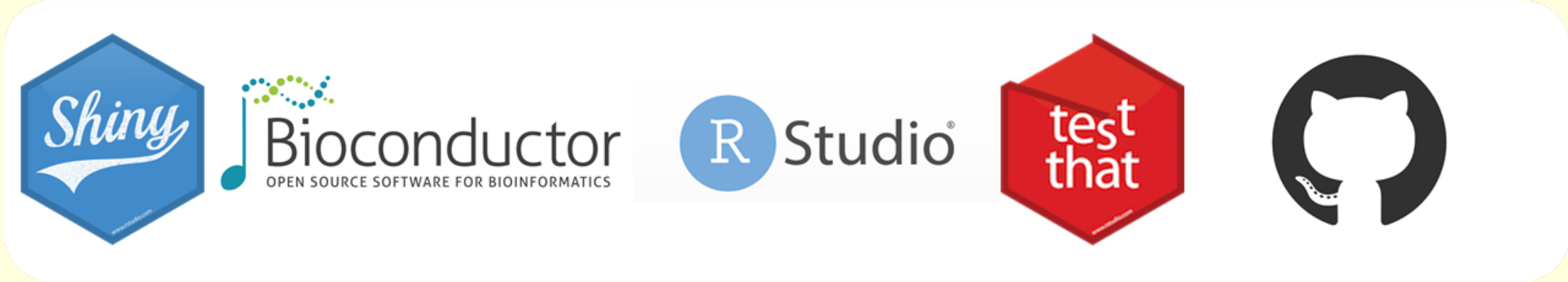


## 4 App structure and technology

### Packages used

- Shiny, DT, shinydashboard, ggplot2, dashboardthemes, corrplot, Hmisc, tidyverse, ggforce

### Technologies used



[1] Sarah Bazzocco, Hafid Alazzouzi, M. Carne Ruiz de Villa, Alex Sanchez-Pla, John M. Mariadason, Diego Arango (2013) *Genome-Wide Analysis of DNA Methylation in Colorectal Cancer*. Submitted.

[2] Yihua Liu and Peng Qiu. (2012) *Integrative analysis of methylation and gene expression data in TCGA* IEEE International Workshop on Genomic Signal Processing and Statistics (GENSIPS)

[3] Jeffrey Racine. (2012) A primer on regression splines.  
[http://cran.r-project.org/web/packages/crs/vignettes/spline\\_primer.pdf](http://cran.r-project.org/web/packages/crs/vignettes/spline_primer.pdf)

[4] B Sadikovic, K Al-Romaih, J.A Squire, and M Zielenska. Cause and Consequences of Genetic and Epigenetic Alterations in Human Cancer. *Current Genomics*, 9(6):394–408, September 2008.