DiagBIOsis : A bioinformatic tool for high-throughput data analysis and visualization

Mathieu Blanchet[1], Elina Alaterre[1], Angélique Bruyer[1]
1. Diag2tec, Montpellier, France

JOBIM 2024 25>28 JUIN Toulouse

Background

The analyses of high-throughput sequencing data, including transcriptomic and whole exome sequencing data, are complex processes that require bioinformatic knowledge. Indeed, each analysis (differential expression analysis, survival analysis, mutation identification, etc.) is based on a complex pipeline using different programming languages.

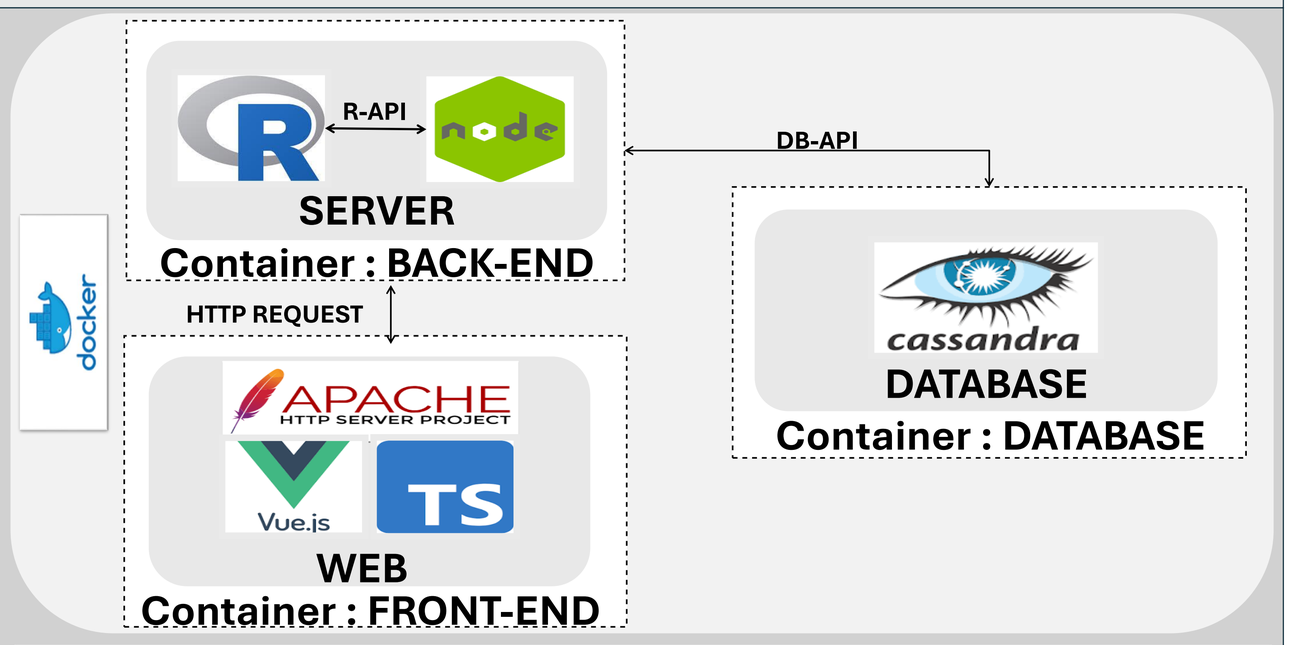
Large amounts of publicly available sequencing data are accumulating but are poorly exploited by scientists due to a lack of easy-to-use bioinformatics resources.

To make bioinformatics analyses accessible and simplify the visualization of sequencing data, Diag2Tec developed the “DiagBIOsis” web tool.

DiagBIOsis provides five types of analysis: **gene expression analysis, survival analysis, principal component analysis, differential expression analysis, and mutation identification and comparison** using normalized or non-normalized data from high-throughput sequencing. The tool automatically selects the most appropriate normalization for each analysis.
In addition, the user can normalize its raw counts according to the method (DESeq2, TPM or FPKM) or the type of analysis selected.

Finally, this tool allows users to easily perform bioinformatic analyses and generate **personalized plots**.

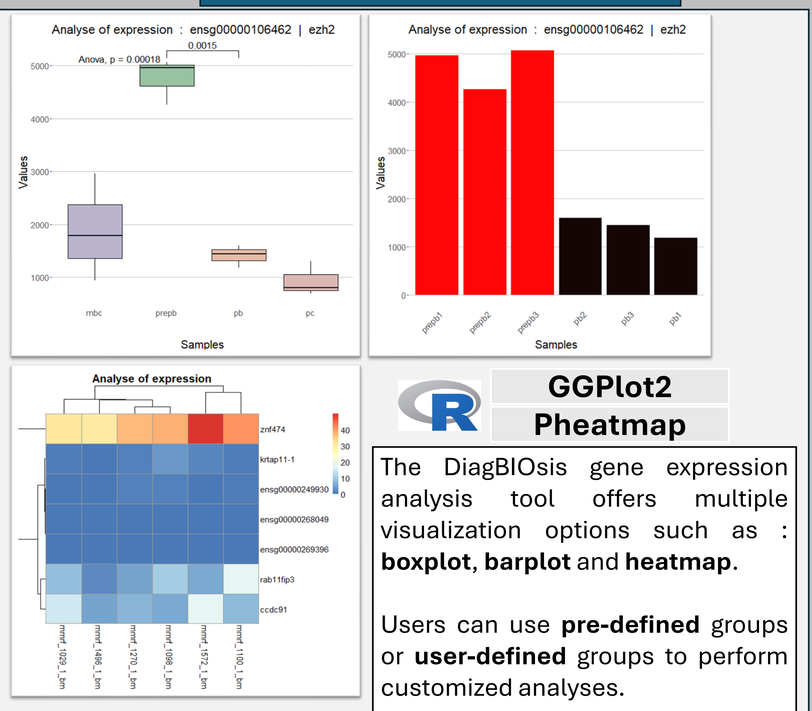
Technical architecture



The DiagBIOsis web application uses **server-side NodeJS technology** to **call R scripts** that analyse the data and provide visualization.
Using an **http request**, the front-end can receive the output data from R scripts and display it for users.
To store all data related to high-throughput sequencing data, we use a **database** managed with the DBMS Cassandra. All these technologies are deployed on a VPS using **Docker** to contenerize all parts of the application.

Bioinformatic tools

Gene expression analysis



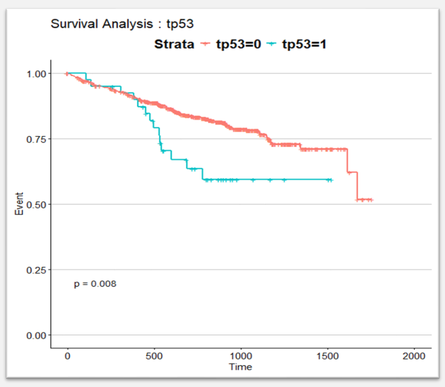
GGPlot2
Pheatmap

The DiagBIOsis gene expression analysis tool offers multiple visualization options such as : **boxplot**, **barplot** and **heatmap**.

Users can use **pre-defined** groups or **user-defined** groups to perform customized analyses.

Users can select several statistical tests and custumize their display.

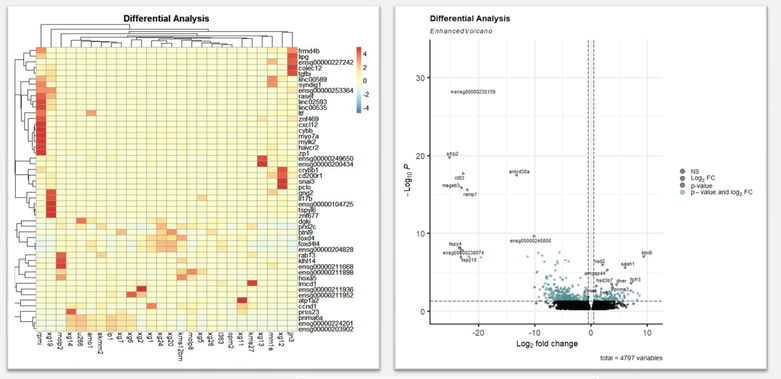
Survival analysis



Survminer
Survival
GGPlot2
Maxstat

Using clinical data**,** DiagBIOsis survival analysis supports **binary** entries (e.g. mutation or cytogenetic abnormality data), but also **decimal** entries (e.g. gene expression data) using the **Maxstat** algorithm**.**

Differential expression analysis

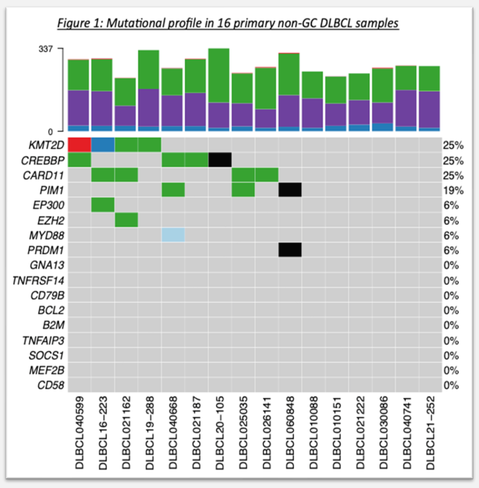


DESeq2
Pheatmap
Enhanced Volcano

Users can perform a **differential expression analysis**, using a dataset from the **database** or by **importing** a **rawcount** dataset and then proceed to the **DESeq2** differential analysis.

This analysis generates a **DESeq_Result** CSV file, used to **visualize results** using a volcano or a heatmap plot, thus separating the analysis from the visualization.

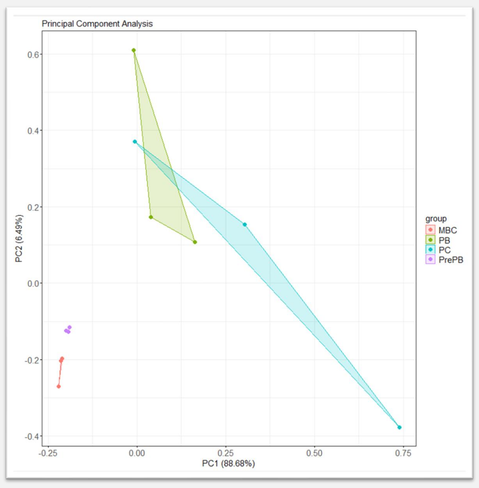
Mutation research



Maftools
Pheatmap

The mutation research tool allows users to **visualize all mutations** for a given mutation dataset by generating an oncoplot. The tool can also generate **mutation comparisons between groups.**

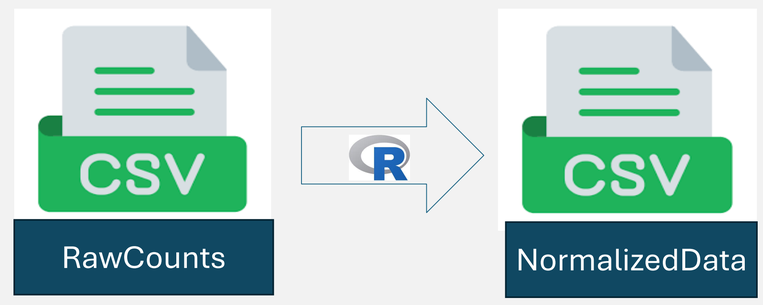
Principal component analysis



GGFortify
GGPlot2
DESeq2

Users can use individual data, **pre-defined** or **user-defined** groups to perform principal component analysis and generate a scatter plot.

RawCount normalization tool



DESeq2

This tool offers the possibility to normalized data with a chosen algorithm (DESeq2, FPKM, TPM, VST ...) by sending **a RawCount dataset**. Once done, normalized data can be exported as CSV files.

Importation tool

Authorized users can **import their own dataset** by using the importation tool. To import their dataset, the user need to precise multiple parameters such as : a **dataset name**, the **type of data** (gene expression data, survival data, mutation data...) and the **type of normalization** used (TPM,VST...).
The importation can also be used to defined « pre-defined groups » on available dataset.

Conclusion

DiagBIOsis is a **bioinformatics web tool** developed by **Diag2Tec** to **simplify access to bioinformatics analysis** for all biologists. DiagBIOsis provides **5 types of analysis and visualization tools**, and gives access to a **rawcount normalization tool** and a **data import tool** to allow users to use their own dataset and generate **exportable, customized plots**.

Contacts : m.blanchet@diag2tec.com | e.alaterre@diag2tec.com | a.bruyer@diag2tec.com

Submission : 87