G.L.A.S.S

Gene

List

   Analysis using

   Shiny

    Server

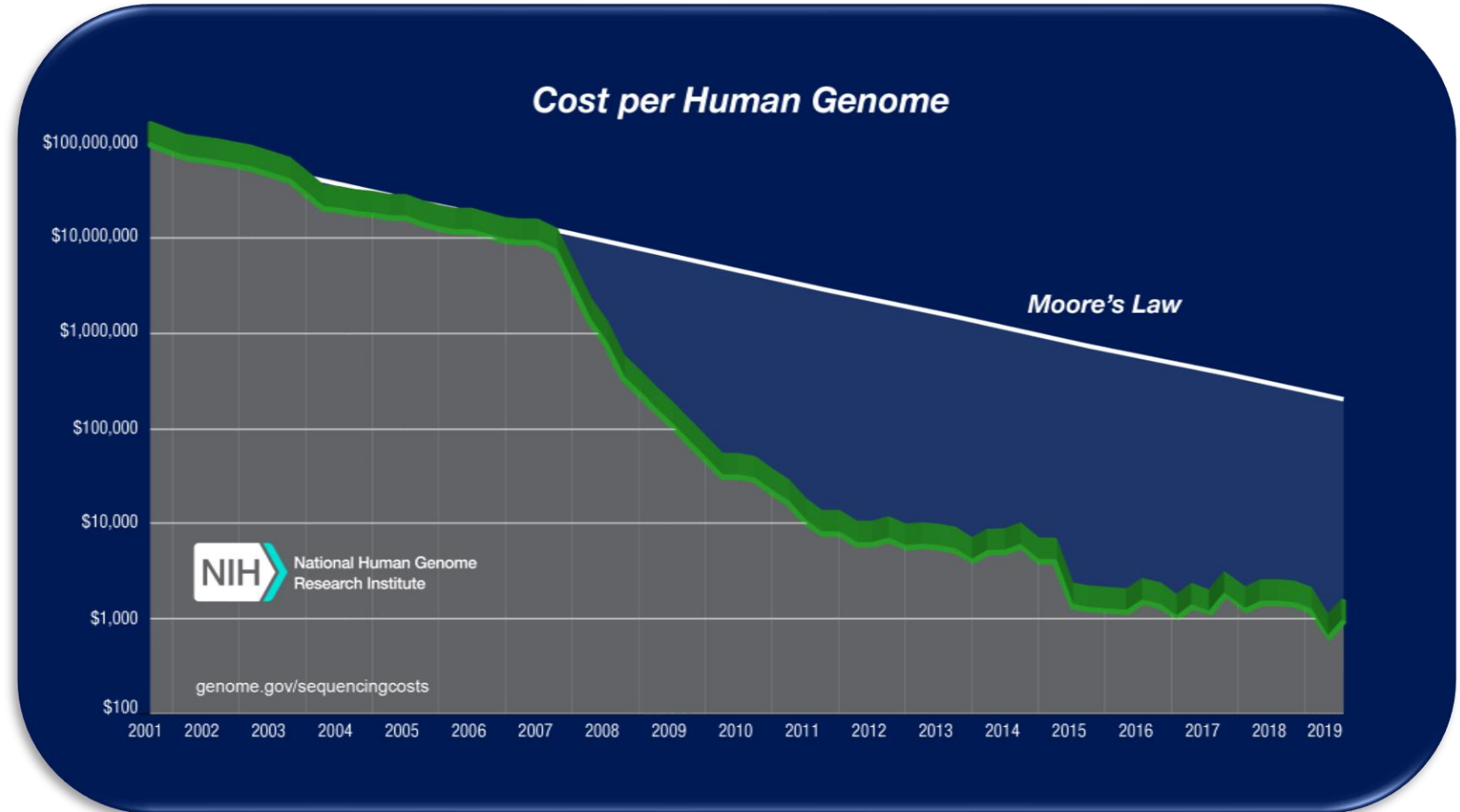LONG PROJECT

Coralie CAPRON   2019/2020   Tutor : Costas BOUYIOUKOS

Introduction

- Next Generation Sequencing (NGS)

- Result :
    Gene List of interest

Introduction

- How to analyze a gene list to get functional information?
  - First option : Analyze each gene manually ..

  - Second option : Use a tool designed to automate this task

2003

Overlaps between user-supplied
gene lists and the curated databases

2005

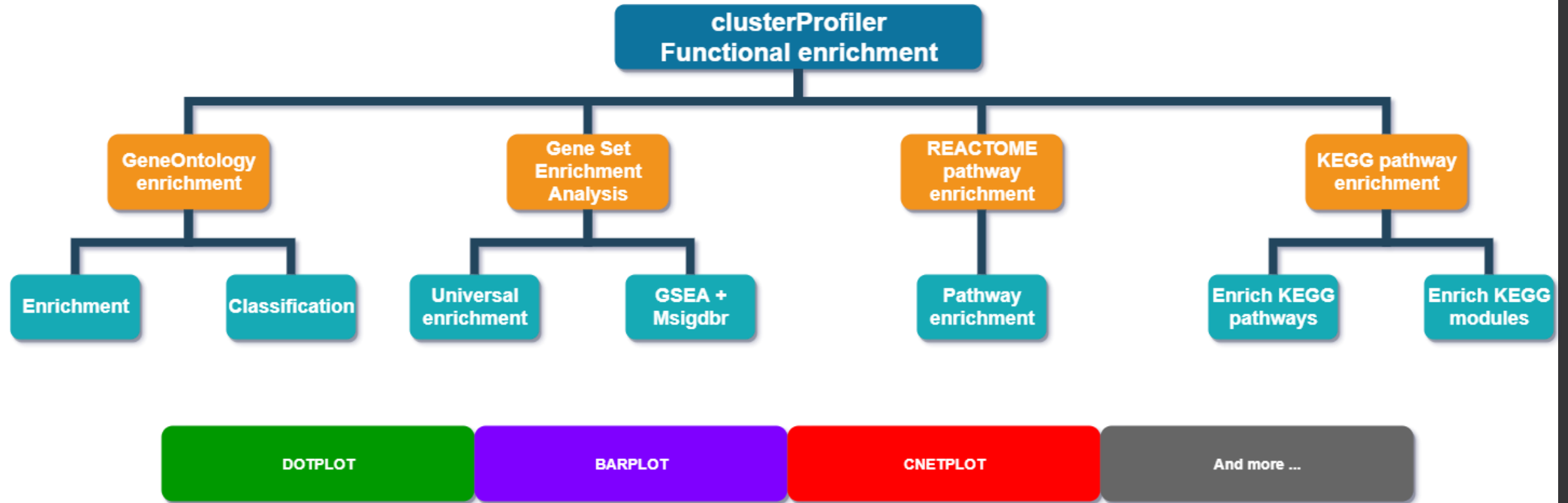Uses every datapoint in
its statistical algorithm

2012

Analysis module and visualization
module were combined

Problematic

Develop an R-Shiny application for the downstream treatment and analysis of a list of "genes of interest"
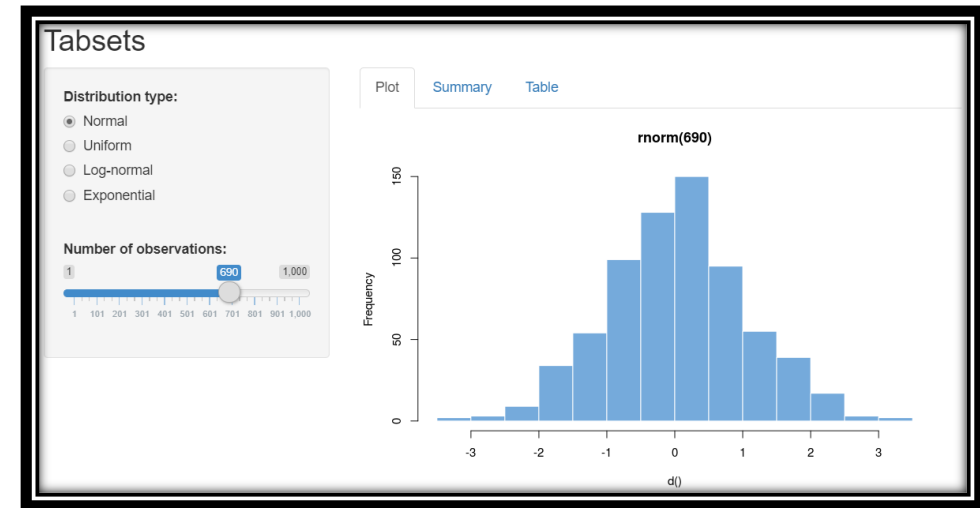
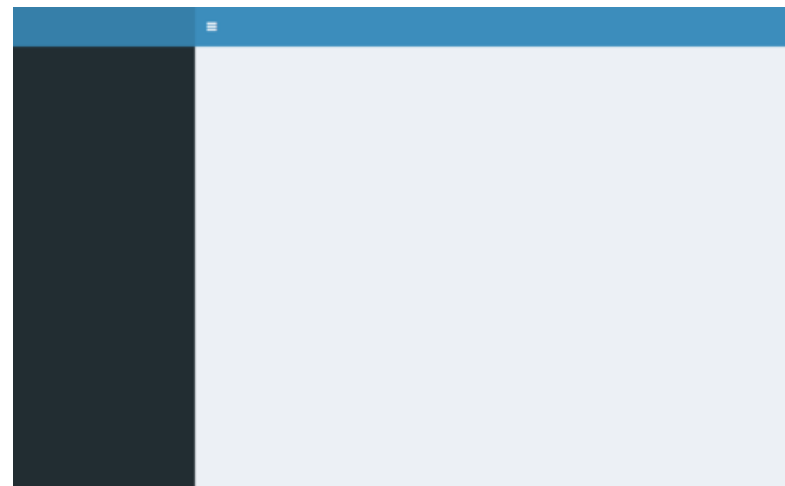# Materials and methods

Msigdbr : annotated gene sets for use with GSEA

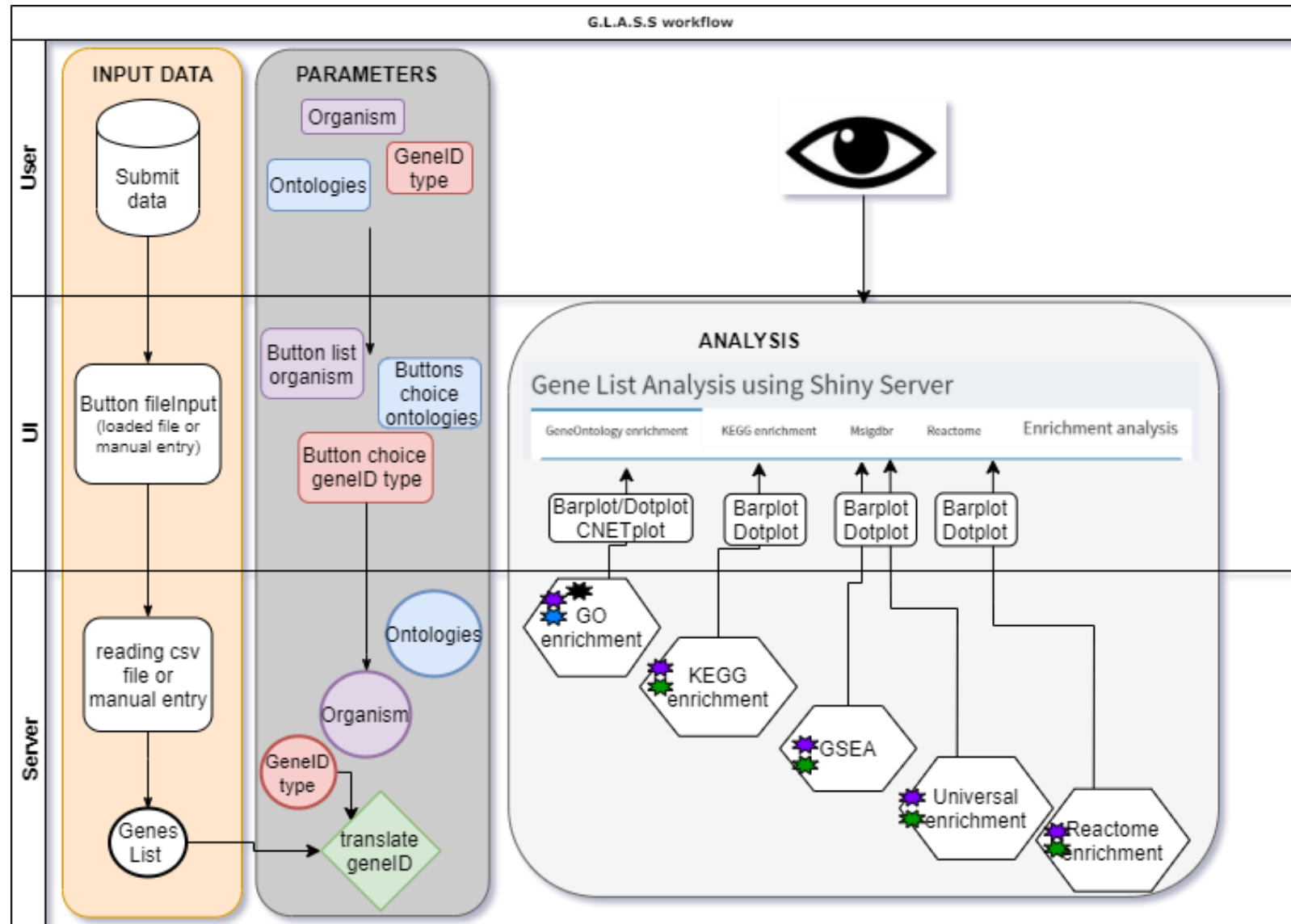# Materials and methods

- Shiny

- Shiny Dashboard

Materials and methods

Results

- Example file : Genes impacted by the chemical compound Perfluorooctane sulfonate (PFOS)
  - PFOS is persistent, bioaccumulative and toxic to mammalian species and was used to waterproof materials

- First step : Import your data

Results

- Second step : Choose your enrichments parameters and global parameters
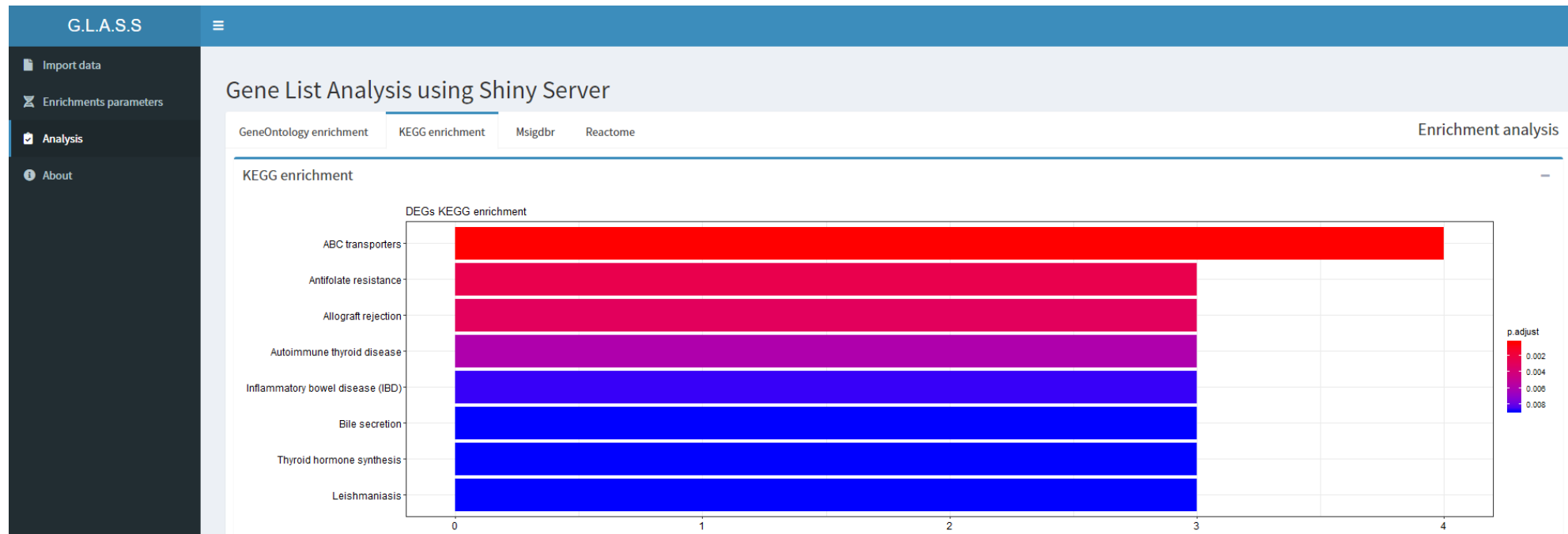
Results

- Third step : Time to analyze !
  - GeneOntology enrichment : terms enriched related to metabolic process, cellular process
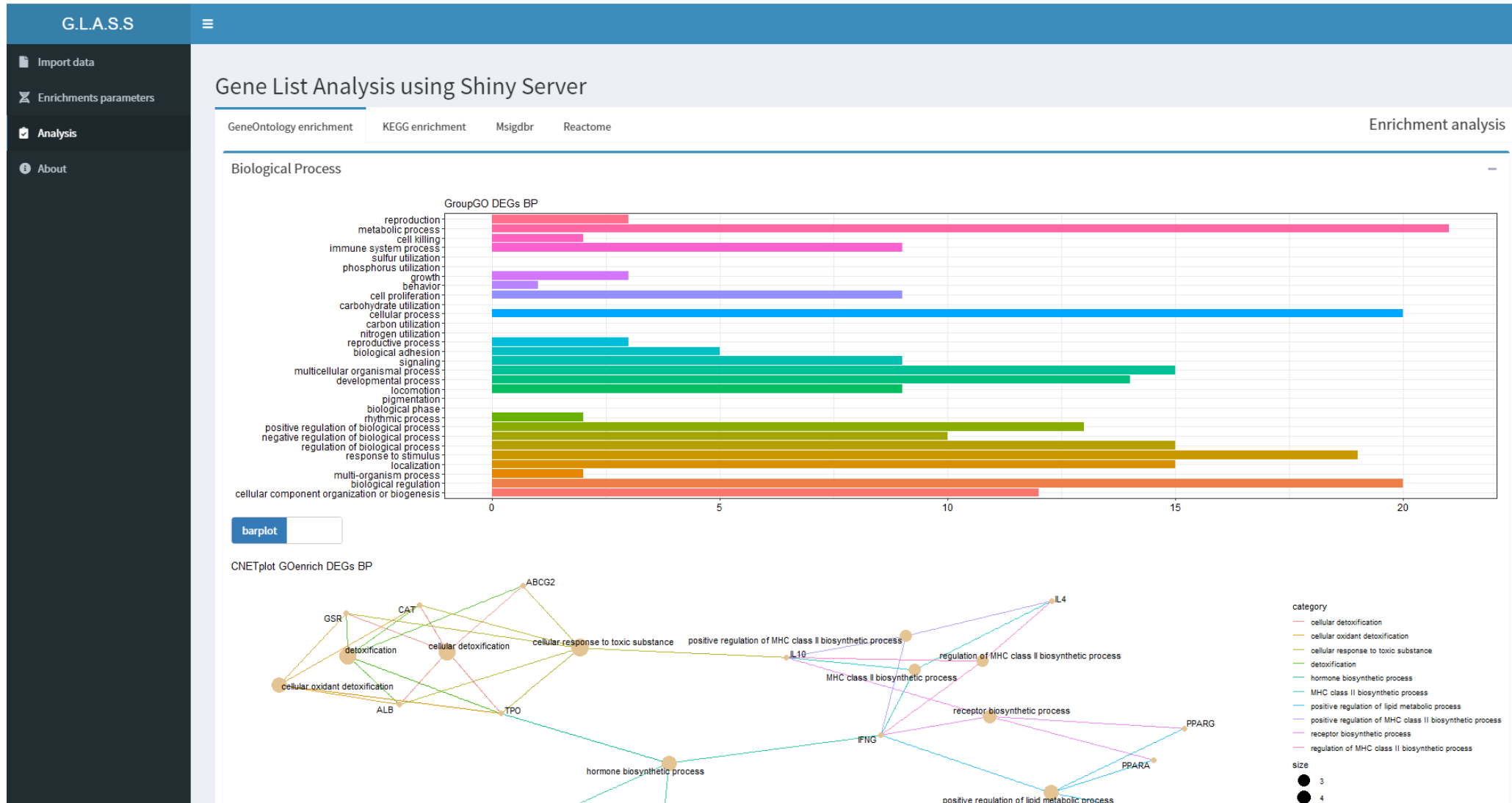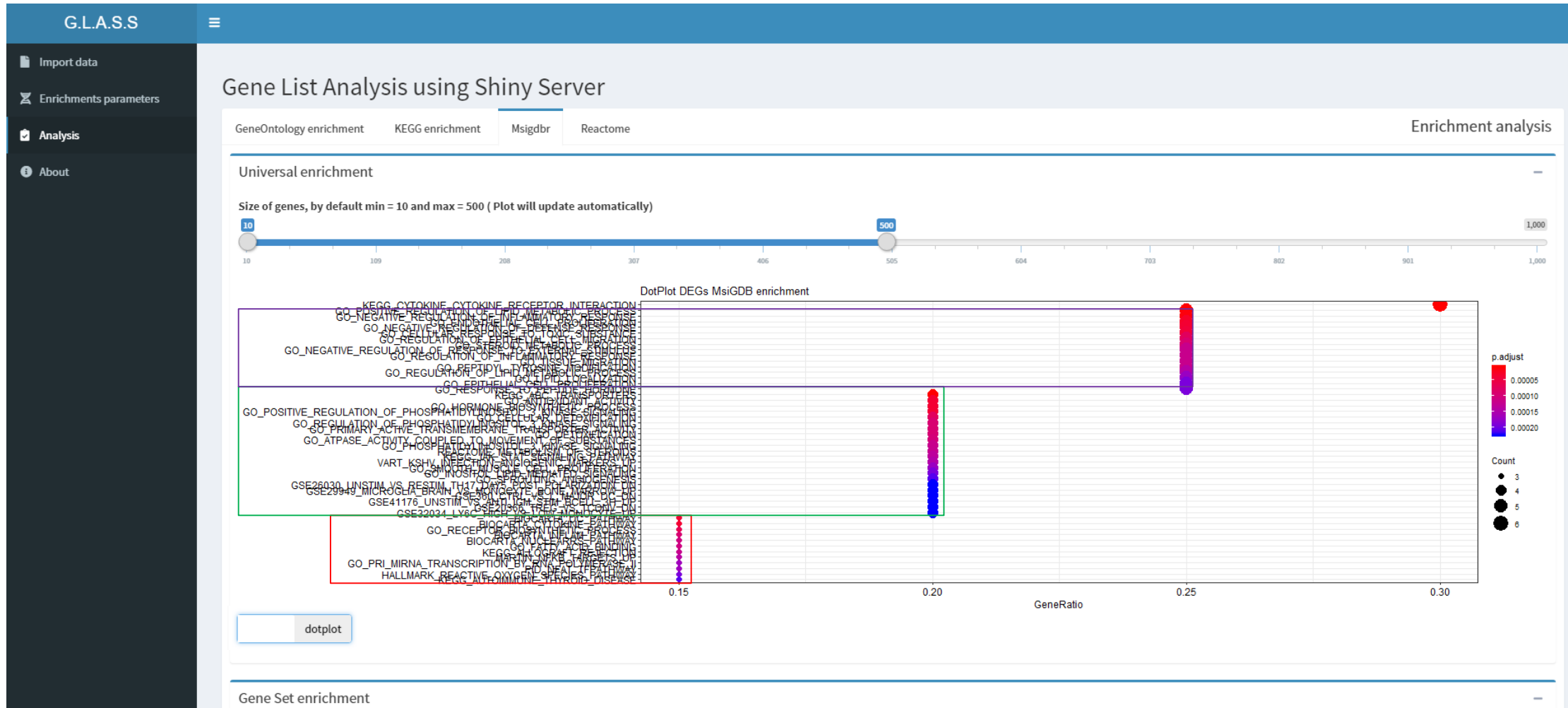
Results

- KEGG enrichment
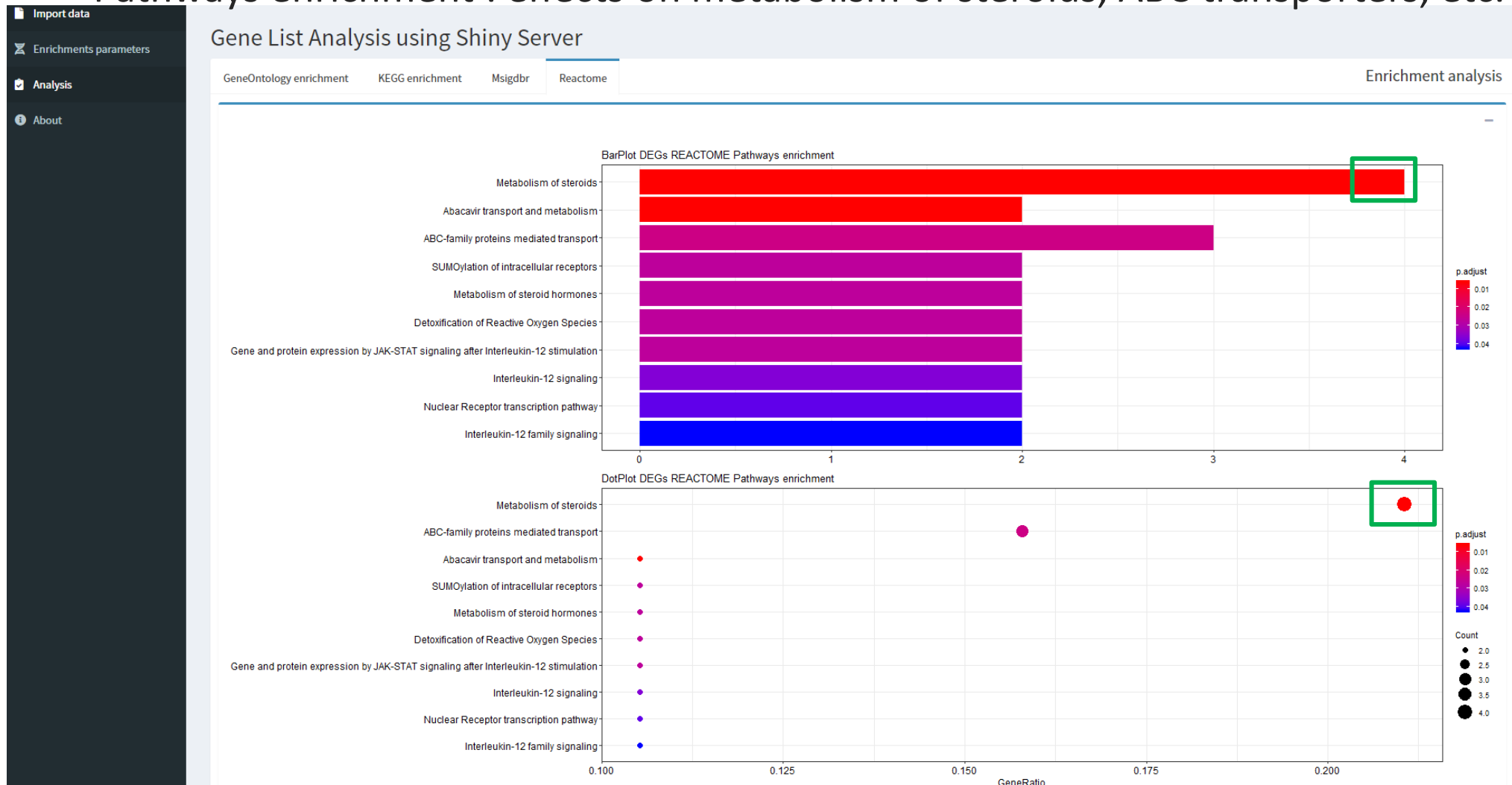 - Biological pathways enriched : transporters, diseases, etc.

Results

- Universal enrichment and GSEA :
  - Observe 4 different clusters

Results

- Reactome enrichment:
  - Pathways enrichment : effects on metabolism of steroids, ABC transporters, etc.

Conclusion and prospects

G.L.A.S.S application :

- Local Shiny server used to analyze a gene list of interest :
    - Gene Ontologies enrichment
    - Gene Set Enrichments Analysis (in case an ordered list is provided)
    - REACTOME pathway enrichment
    - KEGG pathway enrichment
- Some improvements can be done :
    - Decrease the processing time by modifying the library clusterProfiler itself
    - Add some features like interactive view in dotplots, add fold change values, etc.
    - Put the server online

References

- Ashburner, Michael, Catherine A. Ball, Judith A. Blake, David Botstein, Heather Butler, J. Michael Cherry, Allan P. Davis, et al. 2000. « Gene Ontology: tool for the unification of biology ». *Nature genetics* 25 (1): 25-29. https://doi.org/10.1038/75556.

- Huang, Da Wei, Brad T. Sherman, et Richard A. Lempicki. 2009. « Systematic and Integrative Analysis of Large Gene Lists Using DAVID Bioinformatics Resources ». *Nature Protocols* 4 (1): 44-57. https://doi.org/10.1038/nprot.2008.211.

- Huber,W, V.J. Carey, R. Gentleman. «Orchestrating high-throughput genomic analysis with Bioconductor». Morgan Nature Methods, 2015:12, 115.

- Kanehisa, Minoru, et Susumu Goto. 2000. « KEGG: Kyoto Encyclopedia of Genes and Genomes ». *Nucleic Acids Research* 28 (1): 27-30.

- Mardis, Elaine R. 2011. « A Decade's Perspective on DNA Sequencing Technology ». *Nature* 470 (7333): 198-203. https://doi.org/10.1038/nature09796.

- Metzker, Michael L. 2010. « Sequencing Technologies - the next Generation ». *Nature Reviews. Genetics* 11 (1): 31-46. https://doi.org/10.1038/nrg2626.

- Subramanian, Aravind, Pablo Tamayo, Vamsi K. Mootha, Sayan Mukherjee, Benjamin L. Ebert, Michael A. Gillette, Amanda Paulovich, et al. 2005. « Gene Set Enrichment Analysis: A Knowledge-Based Approach for Interpreting Genome-Wide Expression Profiles ». *Proceedings of the National Academy of Sciences of the United States of America* 102 (43): 15545-50. https://doi.org/10.1073/pnas.0506580102.

- « Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. - PubMed - NCBI ». s. d. Consulté le 16 janvier 2020. https://www.ncbi.nlm.nih.gov/pubmed/19131956?dopt=Abstract.

- Wetterstrand, KA. s. d. « DNA Sequencing Costs: Data ». Genome.Gov. Consulté le 11 janvier 2020. https://www.genome.gov/about-genomics/fact-sheets/DNA-Sequencing-Costs-Data.

- Winston Chang, Joe Cheng, JJ Allaire, Yihui Xie and Jonathan McPherson (2019). shiny: Web Application Framework for R

- Yu, Guangchuang, Li-Gen Wang, Yanyan Han, et Qing-Yu He. 2012. « clusterProfiler: an R Package for Comparing Biological Themes Among Gene Clusters ». *OMICS : a Journal of Integrative Biology* 16 (5): 284-87. https://doi.org/10.1089/omi.2011.0118.

Thank you for your attention

Coralie CAPRON

2019/2020

Annex

- GO enrichment window, « bloc organization »

Annex

- GO enrichment Biological Process dotplot

Annex

- GSEA message error

Annex

- Universal enrichment (barplot)

Annex

- About section

Annex

- Hiden sidebar

Annex

- First implementation of G.L.A.S.S without using shiny dashboard