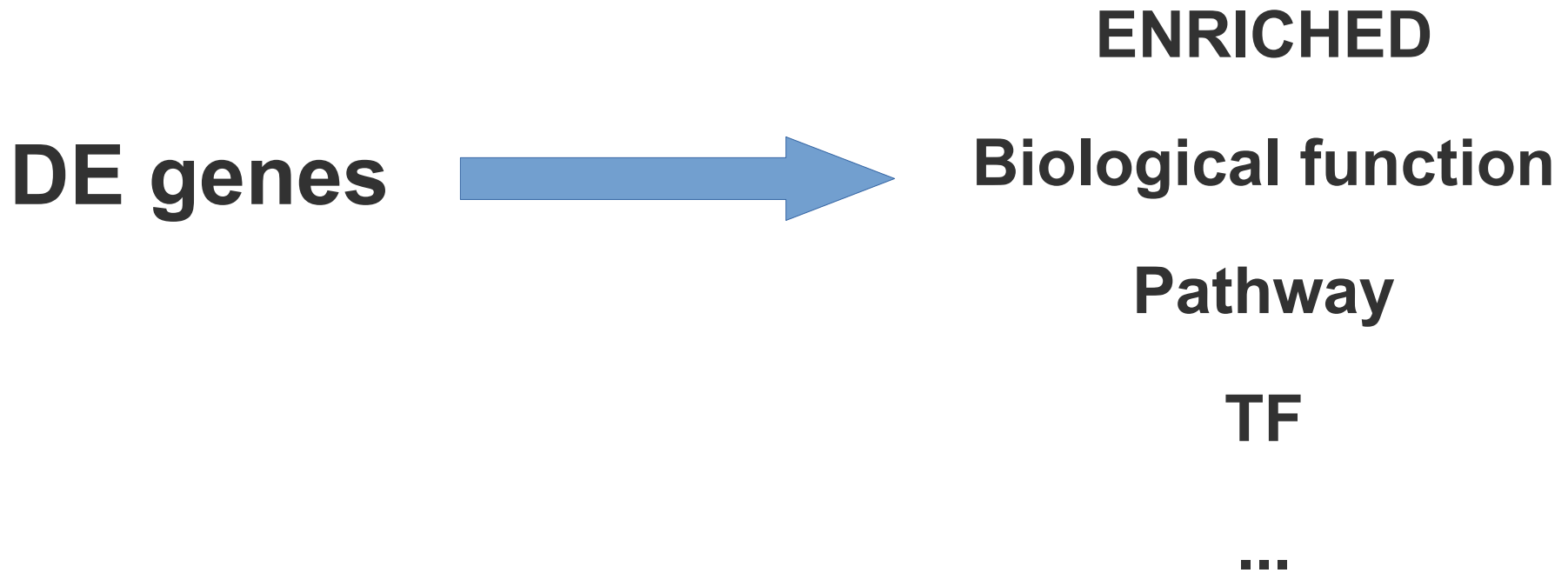# Introduction to RNA-Seq – Enrichment analysis

Wandrille Duchemin

Enrichment analysis

- **What to do with your list of differentially expressed gene ?**

- **Interpretation can be difficult, especially when many genes are DE**

Enrichment analysis

- **What to do with your list of differentially expressed gene ?**

- **Interpretation can be difficult, especially when many genes are DE**

DE genes → **ENRICHED**

Biological function

Pathway

TF

...

Enrichment analysis

- **What to do with your list of differentially expressed gene ?**

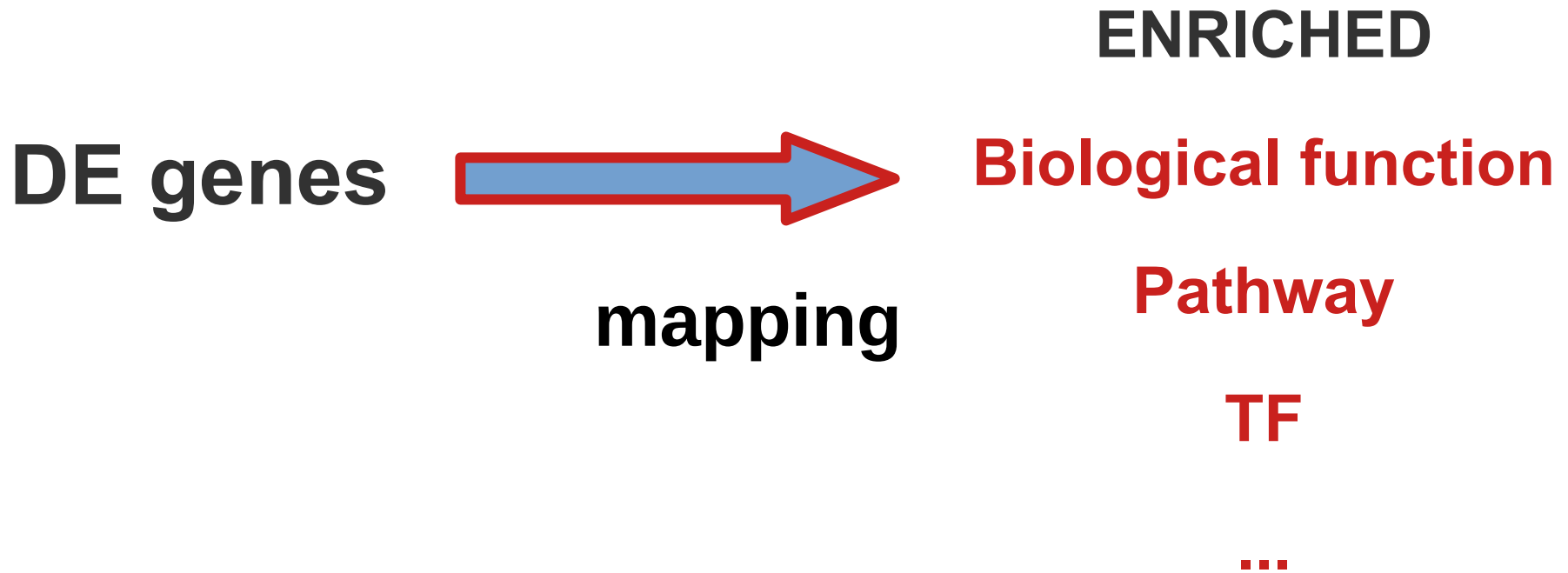- **Interpretation can be difficult, especially when many genes are DE**

DE genes

mapping

ENRICHED

Biological function

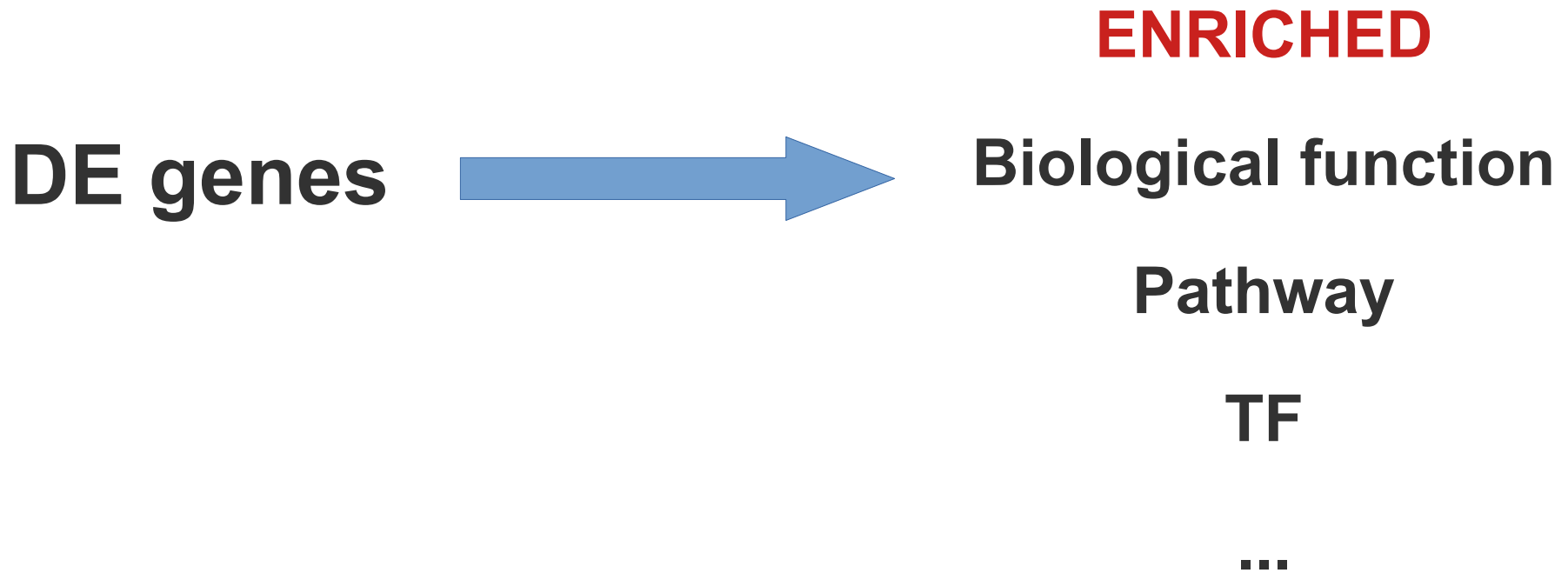**Pathway**

TF

...

Enrichment analysis

- **What to do with your list of differentially expressed gene ?**

- **Interpretation can be difficult, especially when many genes are DE**

ENRICHED

DE genes ➡️ **Biological function**

Pathway

TF

...

Enrichment analysis - mapping

- **The goal is to regroup certain genes together in meaningful sets**

- **Genes involved in the same pathway (eg. DNA repair)**

- **Genes located in the same biological compartment**

- **Genes with a similar molecular function**

- **Genes regulated by the same transcription factor**

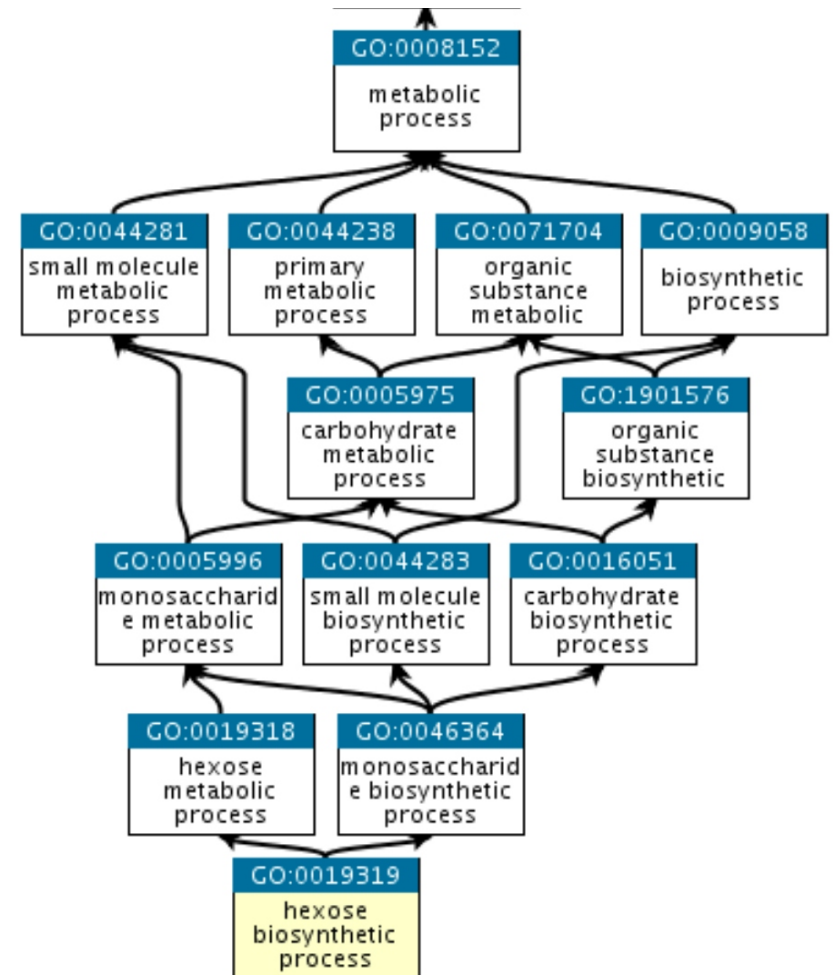- **…**

Enrichment analysis - mapping

- **A few of the possibilities :**

- **Gene Ontology**

- **Reactome**

- **KEGG**

- **MSigDB**

- **Custom set**

- **…**
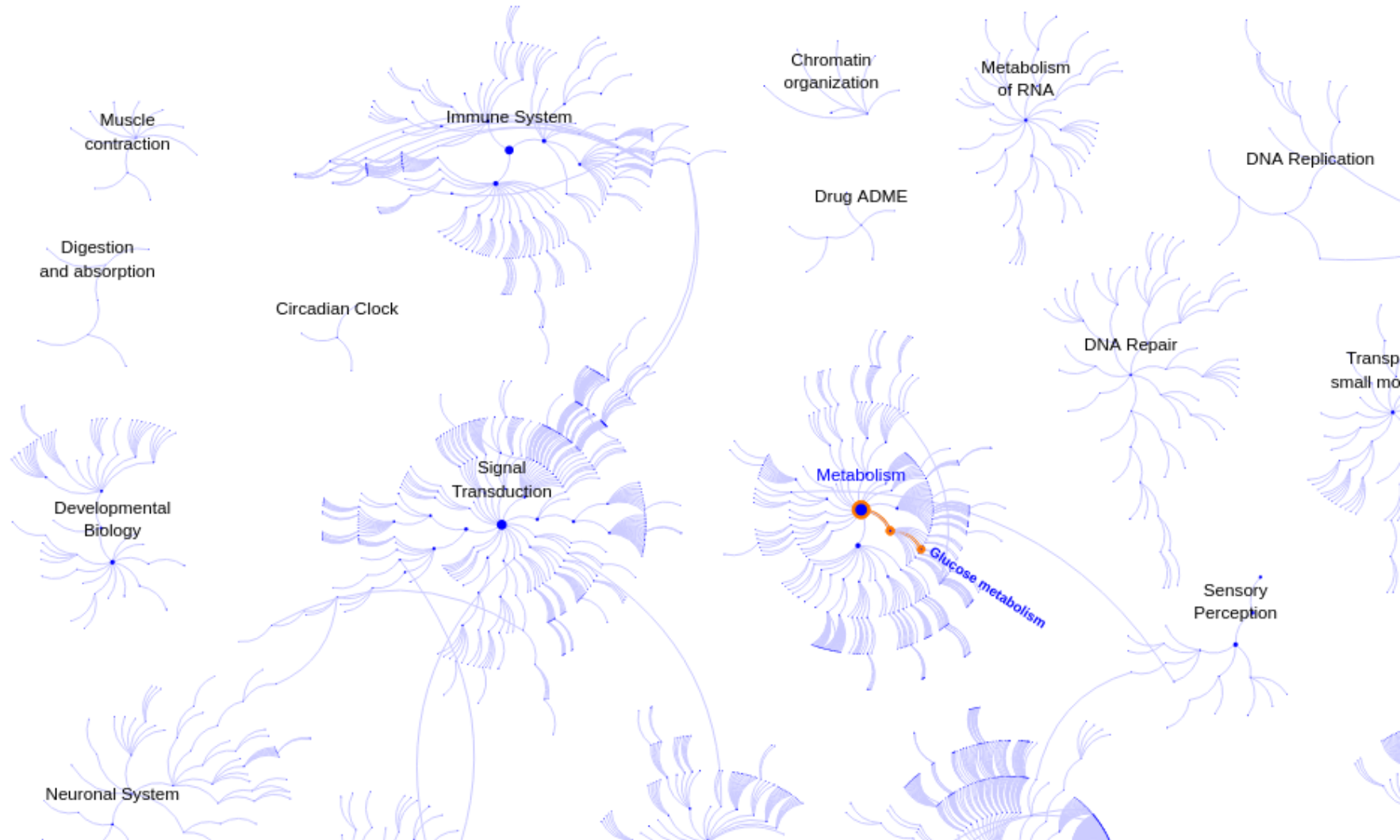
Enrichment analysis - mapping

Gene ontology : http://geneontology.org

3 domains of nested terms:

- **Molecular Function**

- **Cellular Component**

- **Biological Process**

Enrichment analysis - mapping

Reactome : https://reactome.org

Enrichment analysis - mapping

MSigDB : http://www.gsea-msigdb.org/gsea/msigdb/index.jsp

Human mouse and rat only

H	**hallmark gene sets** are coherently expressed signatures derived by aggregating many MSigDB gene sets to represent well-defined biological states or processes.

C1	**positional gene sets** for each human chromosome and cytogenetic band.

C2	**curated gene sets** from online pathway databases, publications in PubMed, and knowledge of domain experts.

C3	**regulatory target gene sets** based on gene target predictions for microRNA seed sequences and predicted transcription factor binding sites.

C4	**computational gene sets** defined by mining large collections of cancer-oriented microarray data.

C5	**ontology gene sets** consist of genes annotated by the same ontology term.

C6	**oncogenic signature gene sets** defined directly from microarray gene expression data from cancer gene perturbations.

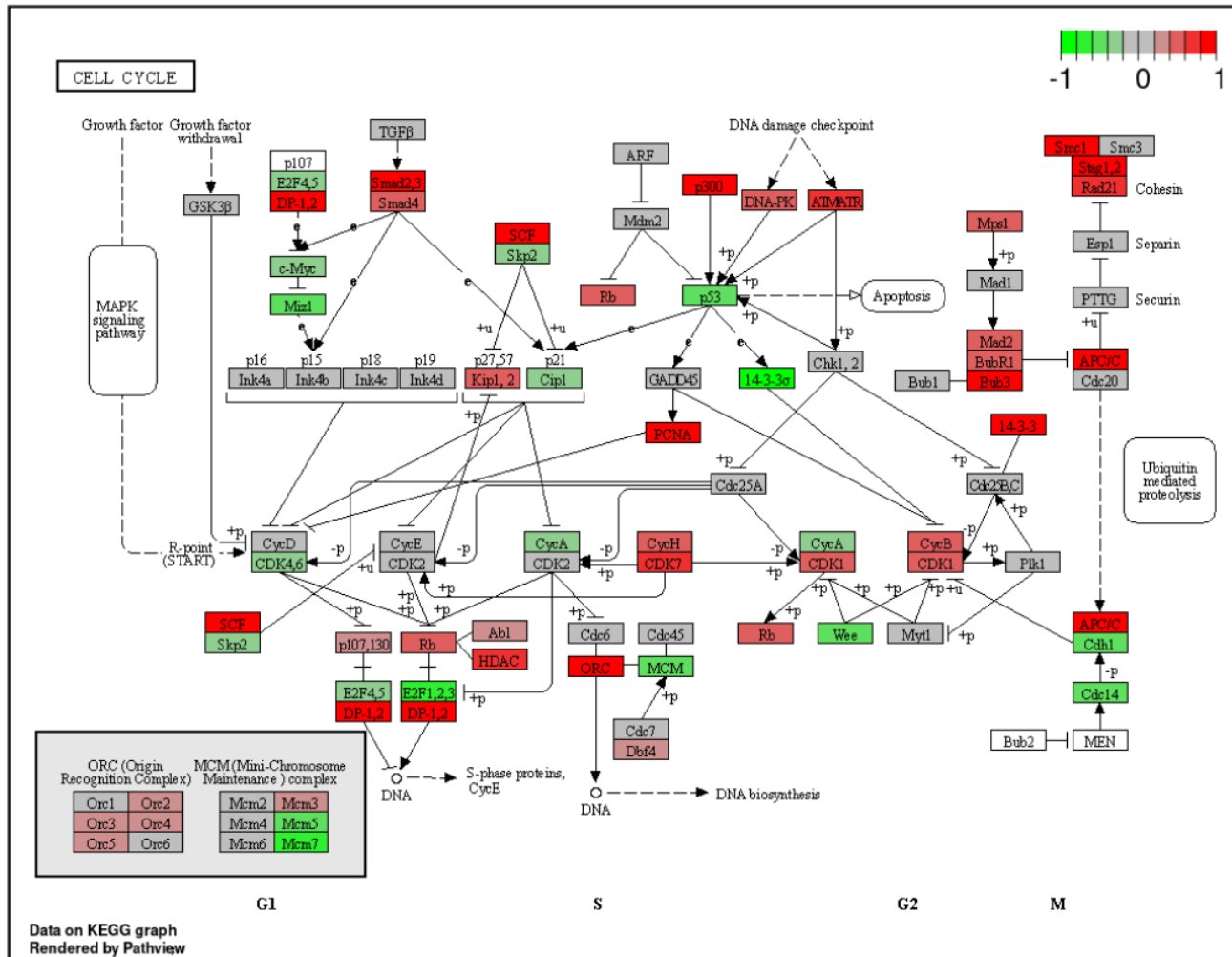|	
C7	**immunologic signature gene sets** represent cell states and perturbations within the immune system.

C8	**cell type signature gene sets** curated from cluster markers identified in single-cell sequencing studies of human tissue.

Enrichment analysis - mapping

KEGG : https://www.genome.jp/kegg/ - KEGG PATHWAY

Projection of DE data
onto a KEGG pathway map
With R package pathview

Enrichment analysis - mapping

- **Custom gene sets :**

- **Derived from specialized litterature**

- **Tentative annotation**

- **Understudied organisms**

- **...**

Enrichment analysis – computing enrichment

- **2 possible approaches (among many!)**

- **Over-representation analysis**

- **Gene Set Enrichment Analysis**

Enrichment analysis – computing enrichment

- **2 possible approaches (among many!)**

- **Over-representation analysis**

  - **Fisher's exact test (with p-val correction)**

	DE	Not DE
in gene set	A	B
not in gene set	C	D

N = A+B+C+D   **# total genes**
M = A+B   **# genes in set**

n = A+C   **# DE genes**
k = A   **# DE genes in set**

$$p = 1 - \sum_{i=0}^{k-1} \frac{\binom{M}{i}\binom{N-M}{n-i}}{\binom{N}{n}}$$

Enrichment analysis – computing enrichment

- **2 possible approaches (among many!)**

- **Gene Set Enrichment Analysis**

  - **Do not rely on 0/1 DE status, but on a continuous measurement (eg. log2FC)**

  - **Computes Enrichment Score from list of ranked genes**

  - **Estimates significance using permutations**

Enrichment analysis – computing enrichment
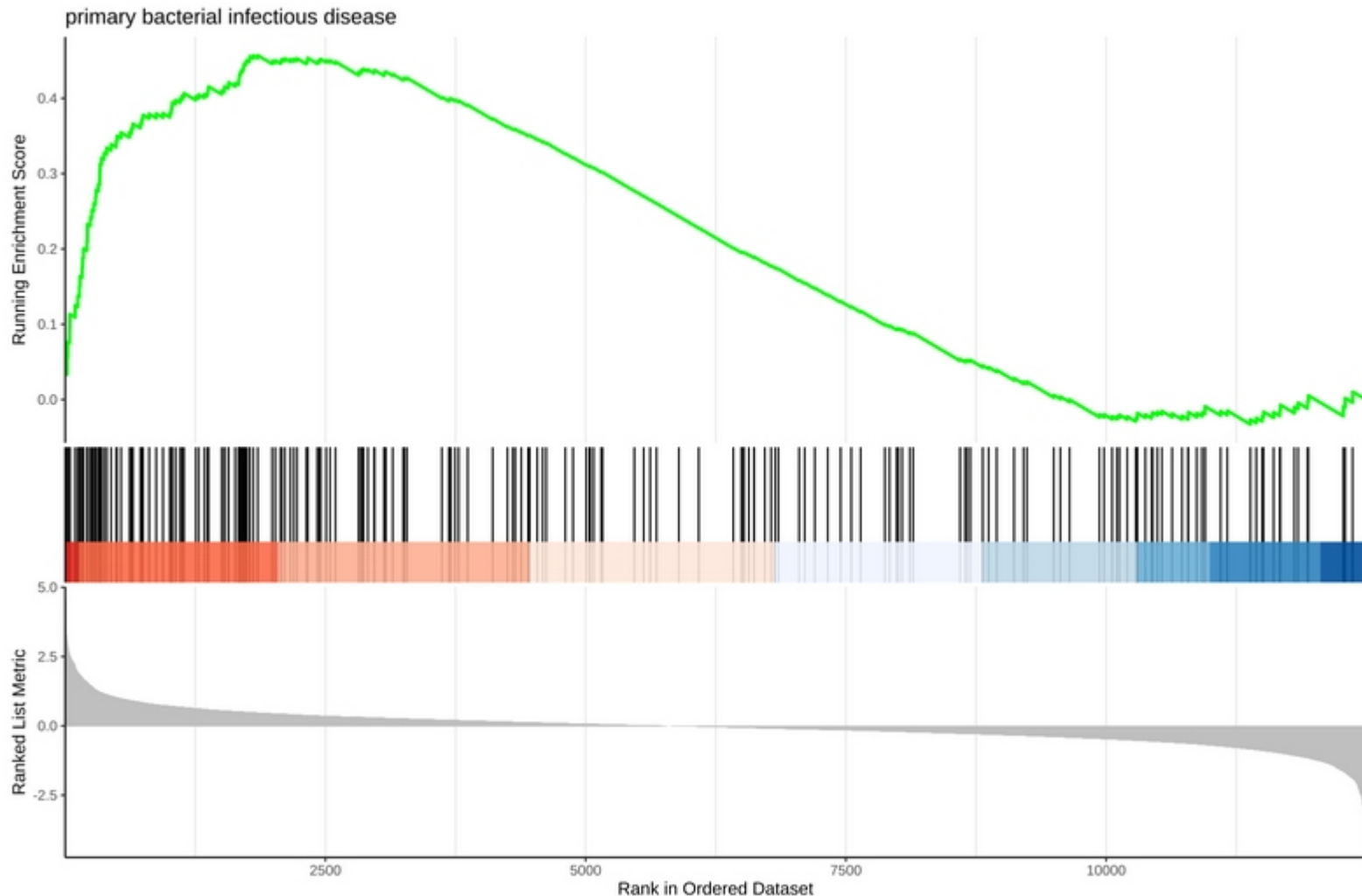
Enrichment Score visualized using functions from the R package enrichplot
http://www.bioconductor.org/packages/release/bioc/html/enrichplot.html

Enrichment analysis – computing enrichment

- **Many more methods or implementations:**
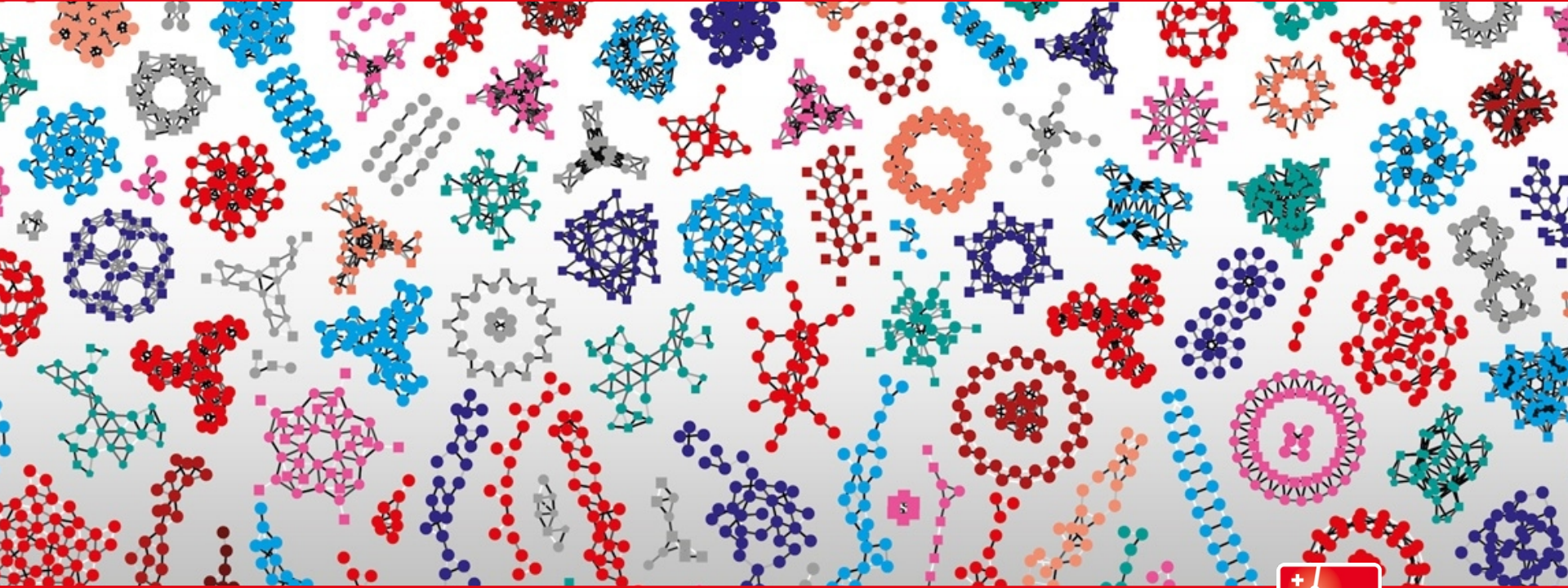
- **Signaling Pathway Impact Analysis**

  - **https://bioconductor.org/packages/release/bioc/html/SPIA.html**

- **ISMARA**

  - **https://ismara.unibas.ch/mara/**

- **...**

Practical

- **Go to the website and follow the Enrichment practical**

Contributors:

Geoffrey Fucile
Walid Gharib
Irene Keller
Pablo Escobar Lopez
Charlotte Soneson