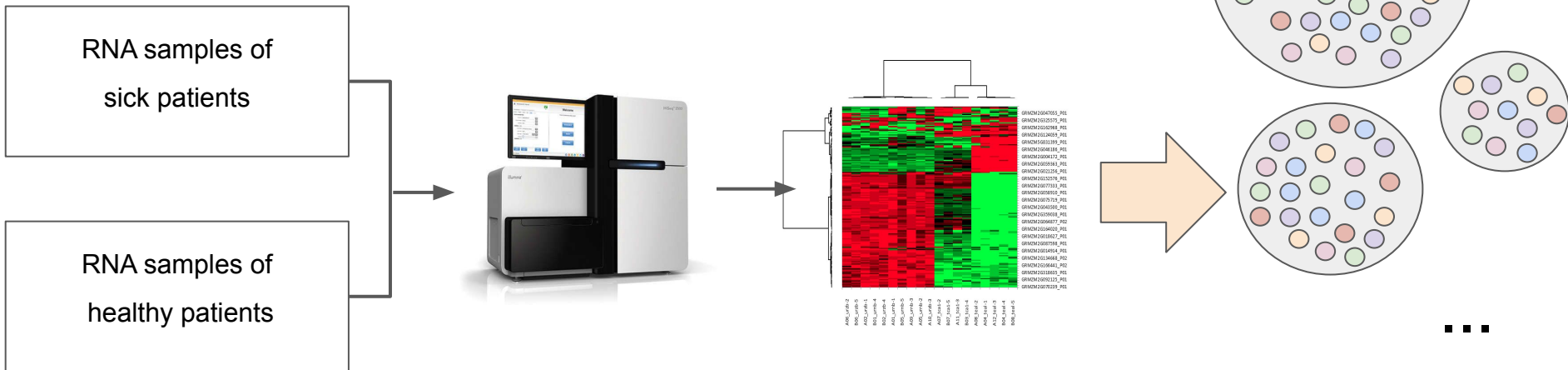


rGSAn: An R package dedicated to the gene set analysis using semantic similarity measures

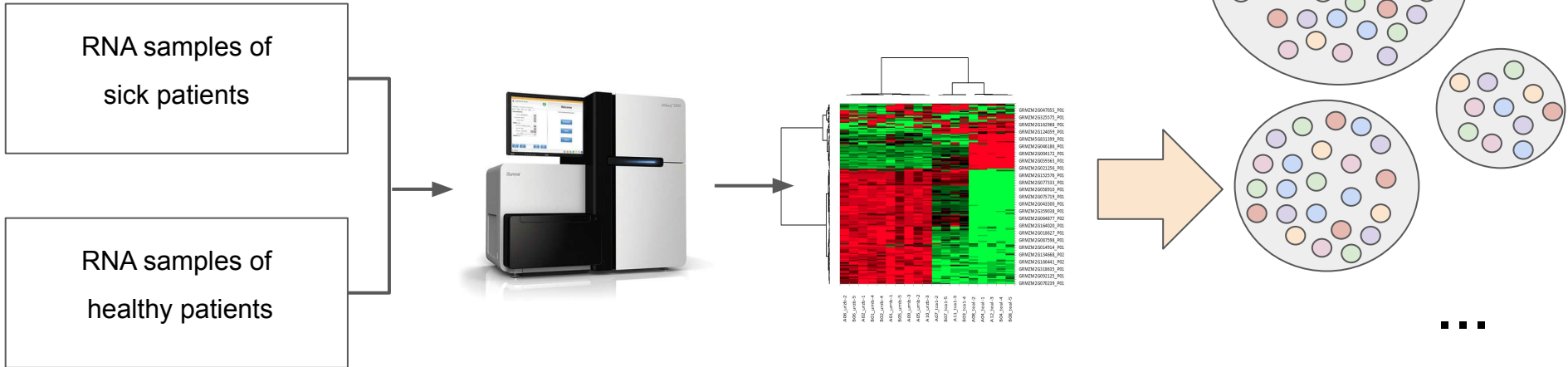
Aarón Ayllón Benítez, Fleur Mougin, Patricia Thébault



...

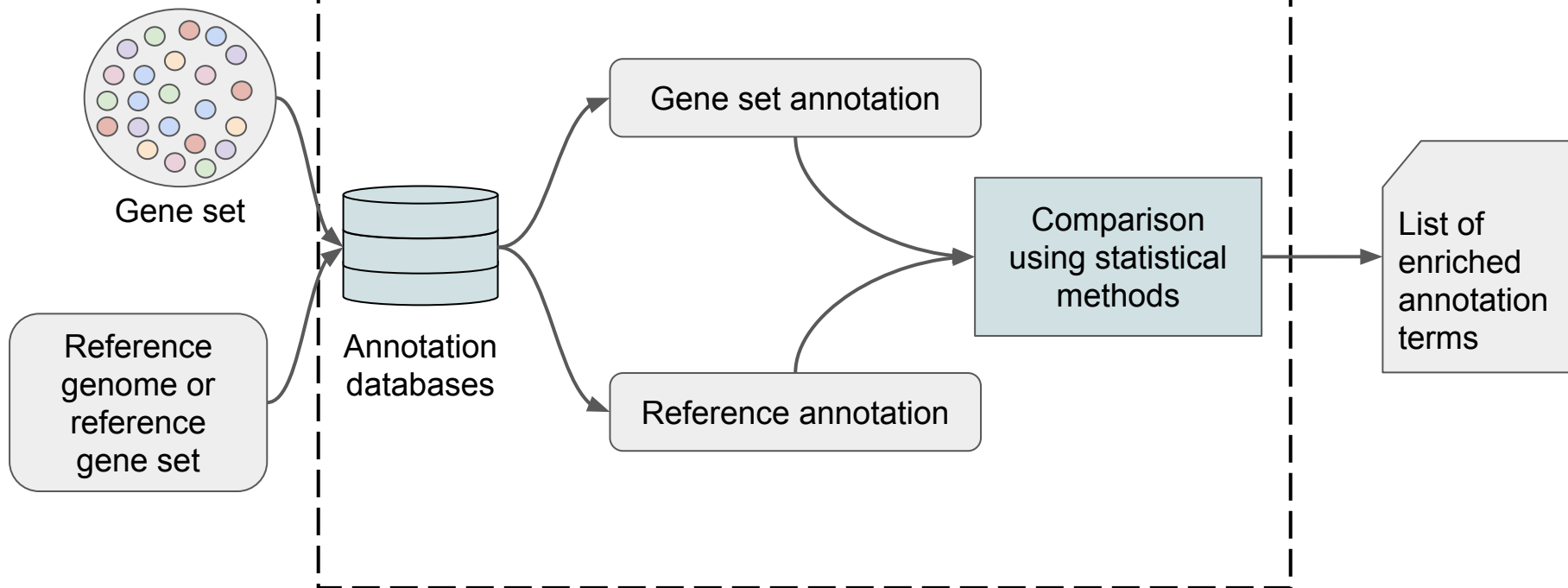
rGSAn: An R package dedicated to the gene set analysis using semantic similarity measures

Aarón Ayllón Benítez, Fleur Mougín, Patricia Thébault



...

Statistical enrichment analysis tool



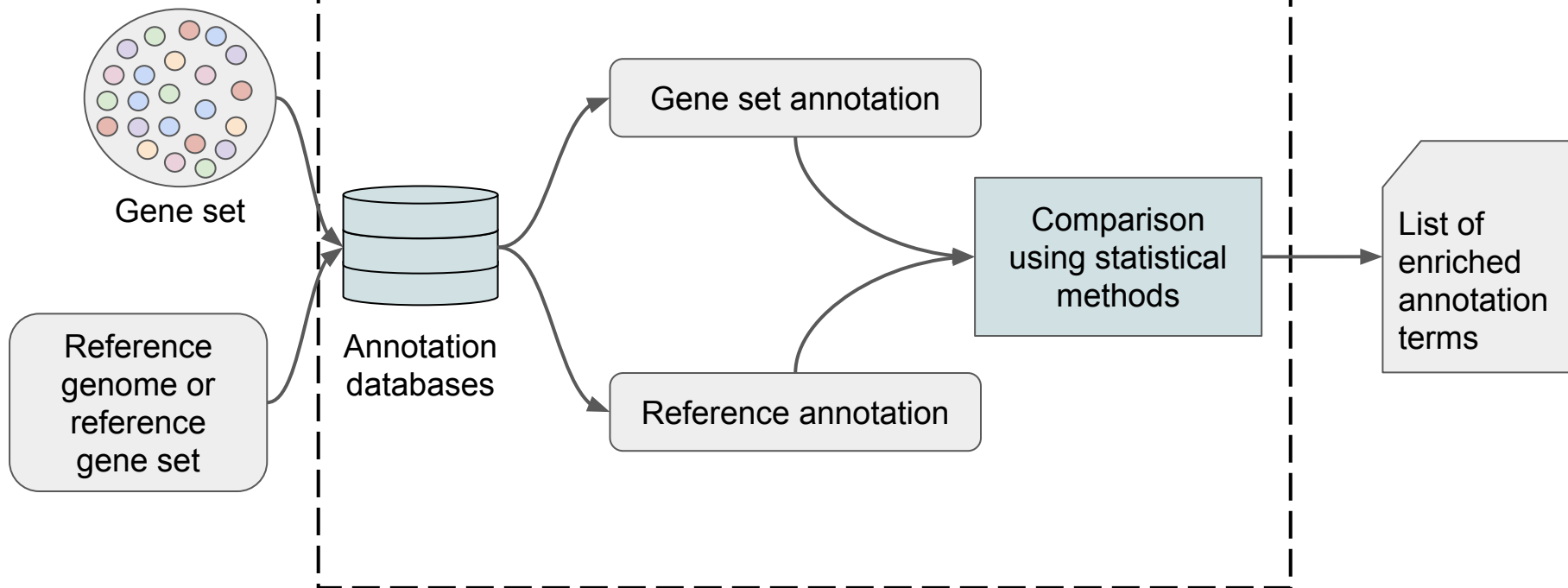
R tools

- **gProfileR**
(Reimand *et al.* 2007)
- **RDAVIDWebService**
(Fresno *et al.* 2013)
- **Cluster profiler**
(Yu *et al.* 2012)
- **topGO**
(Alexa *et al.* 2010)

LIMITATIONS

- **Loss of information: Loss of annotated genes**
- **Redundant information that have to be minimized**

Statistical enrichment analysis tool



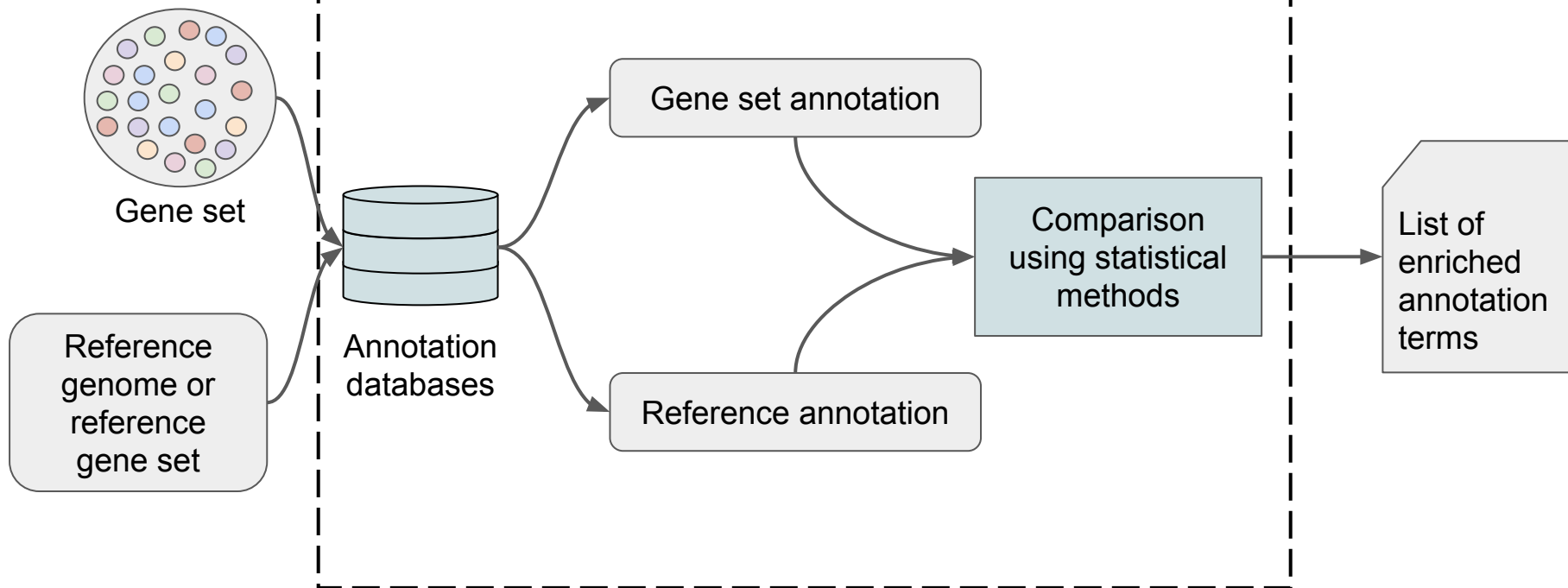
R tools

- **gProfileR**
(Reimand *et al.* 2007)
- **RDAVIDWebService**
(Fresno *et al.* 2013)
- **Cluster profiler**
(Yu *et al.* 2012)
- **topGO**
(Alexa *et al.* 2010)

LIMITATIONS

- **Loss of information: Loss of annotated genes**
- **Redundant information that have to be minimized**

Statistical enrichment analysis tool



R tools

- **gProfileR**
(Reimand *et al.* 2007)
- **RDAVIDWebService**
(Fresno *et al.* 2013)
- **Cluster profiler**
(Yu *et al.* 2012)
- **topGO**
(Alexa *et al.* 2010)

LIMITATIONS

- **Loss of information: Loss of annotated genes**
- **Redundant information that have to be minimized**



The aim of this work

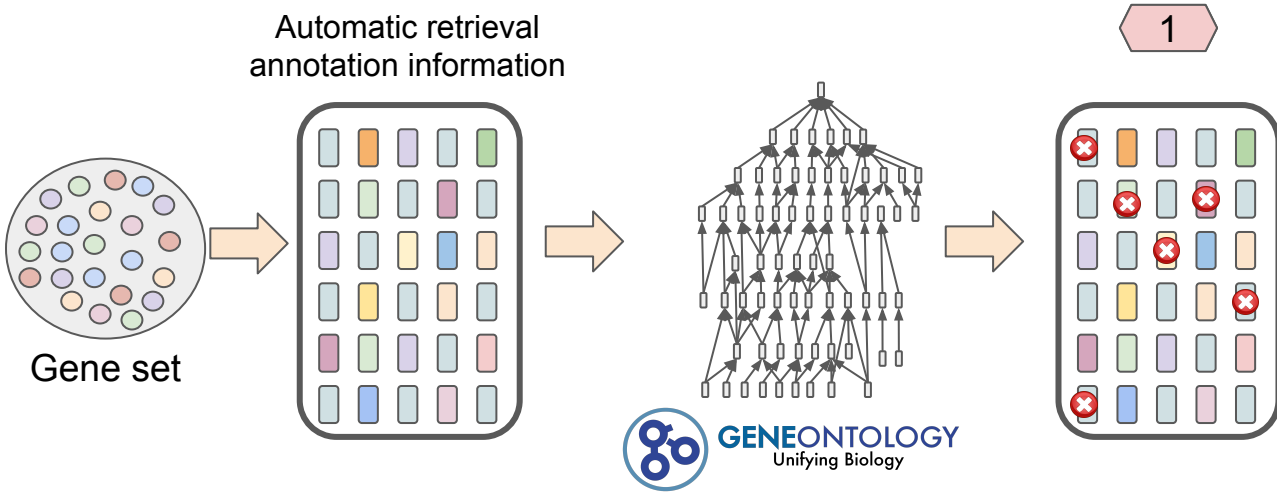
To **interpret** and analyze **gene sets** grouped according to a phenotype of interest.

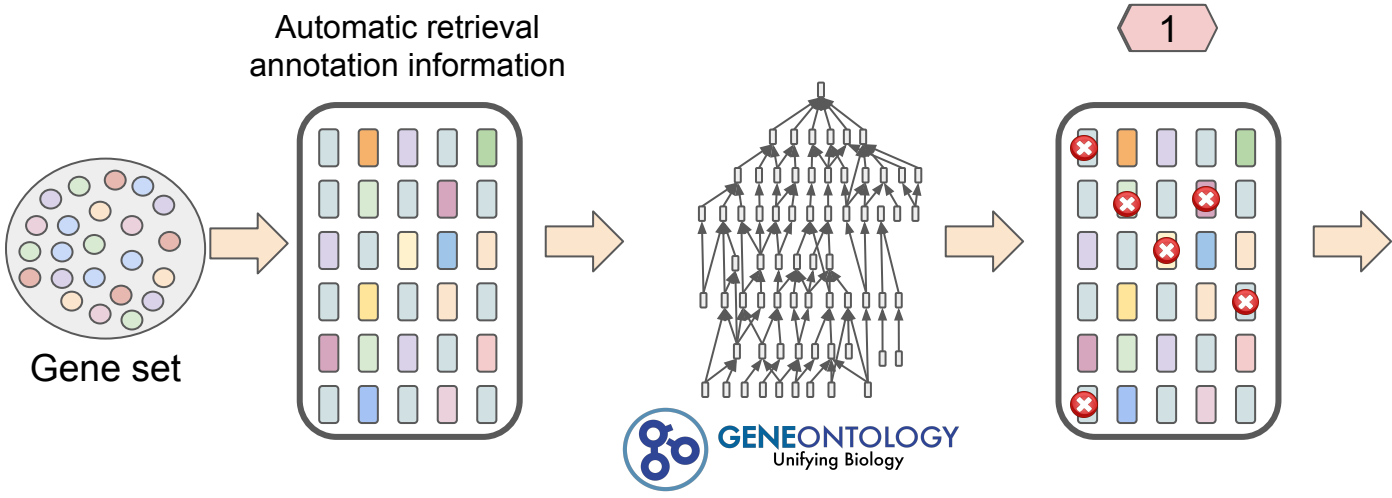
As an alternative to enrichment statical methods, **we investigated**:

- **semantic similarity measures**
- **graph theory methods**

The **main objectives** are to identify:

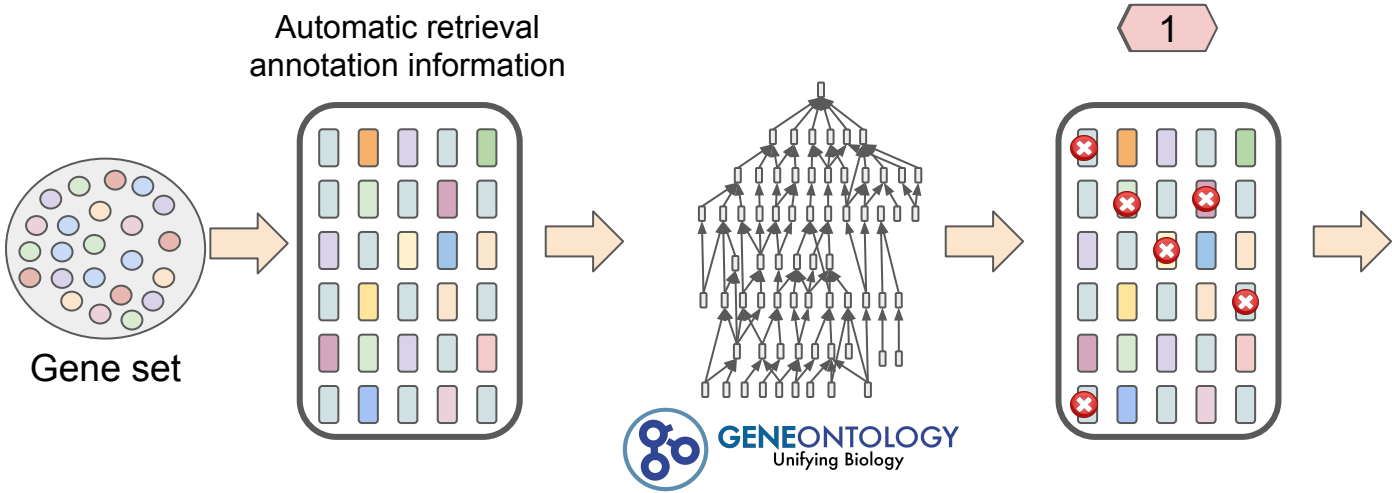
- the best compromise between the number of **retained annotation terms** that has to be **drastically reduced**
- the **number of related genes** that has to be as **large as possible**
- to propose **original interactive visualization facilities** dedicated to the multi-scale analysis of gene set annotations





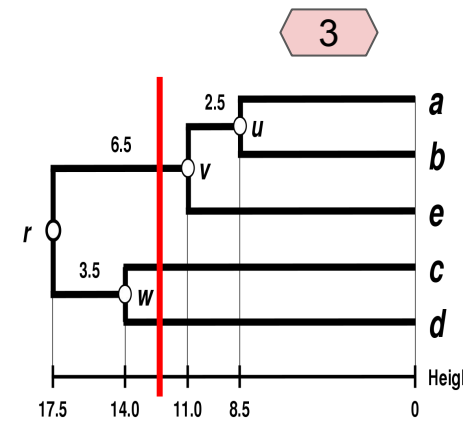
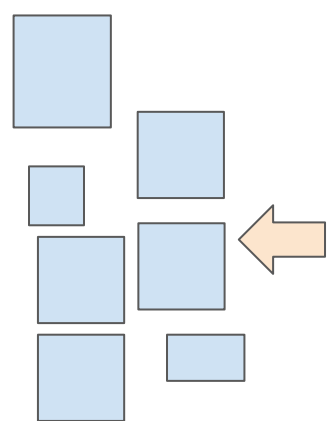
2

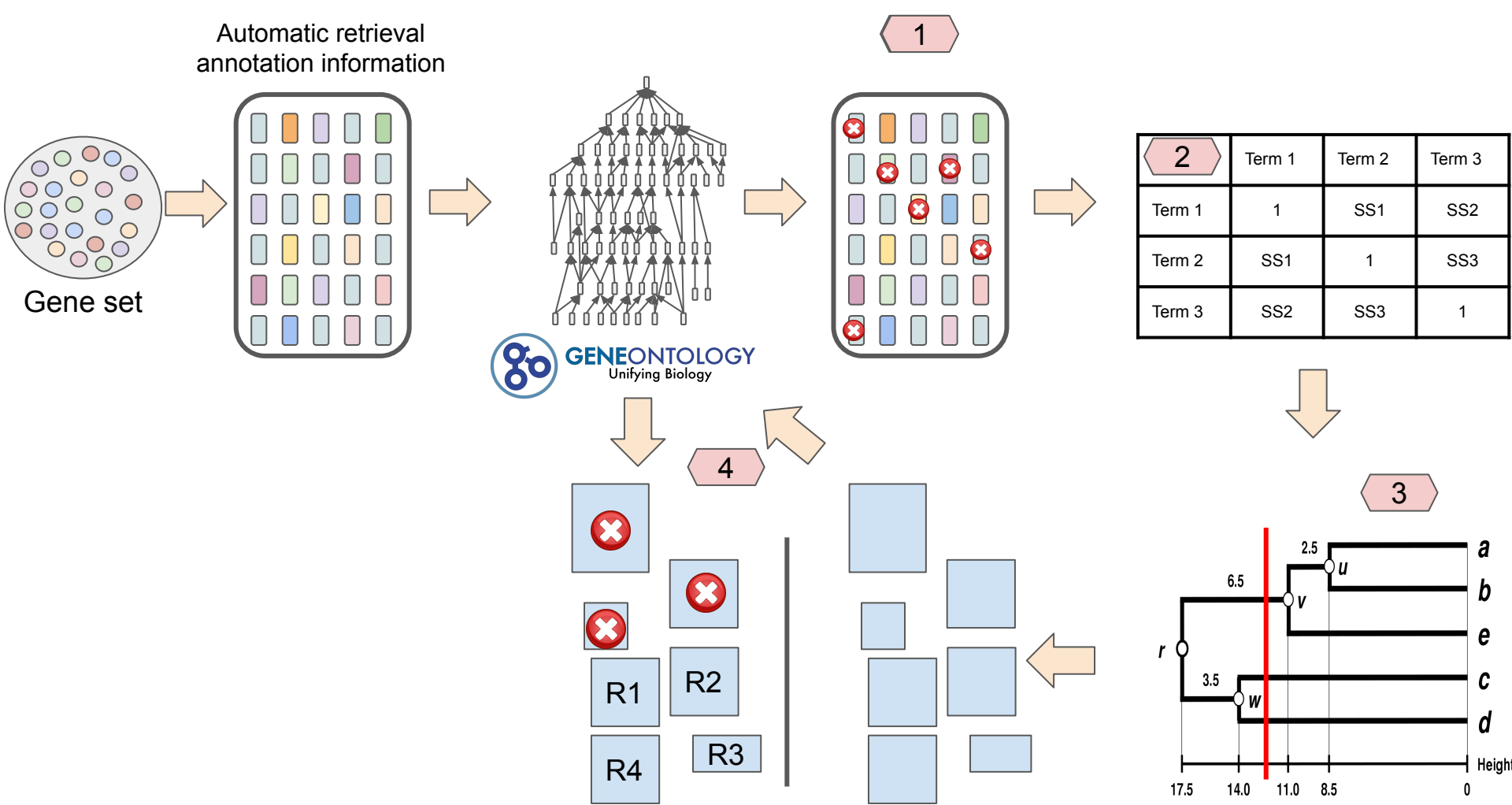
	Term 1	Term 2	Term 3
Term 1	1	SS1	SS2
Term 2	SS1	1	SS3
Term 3	SS2	SS3	1

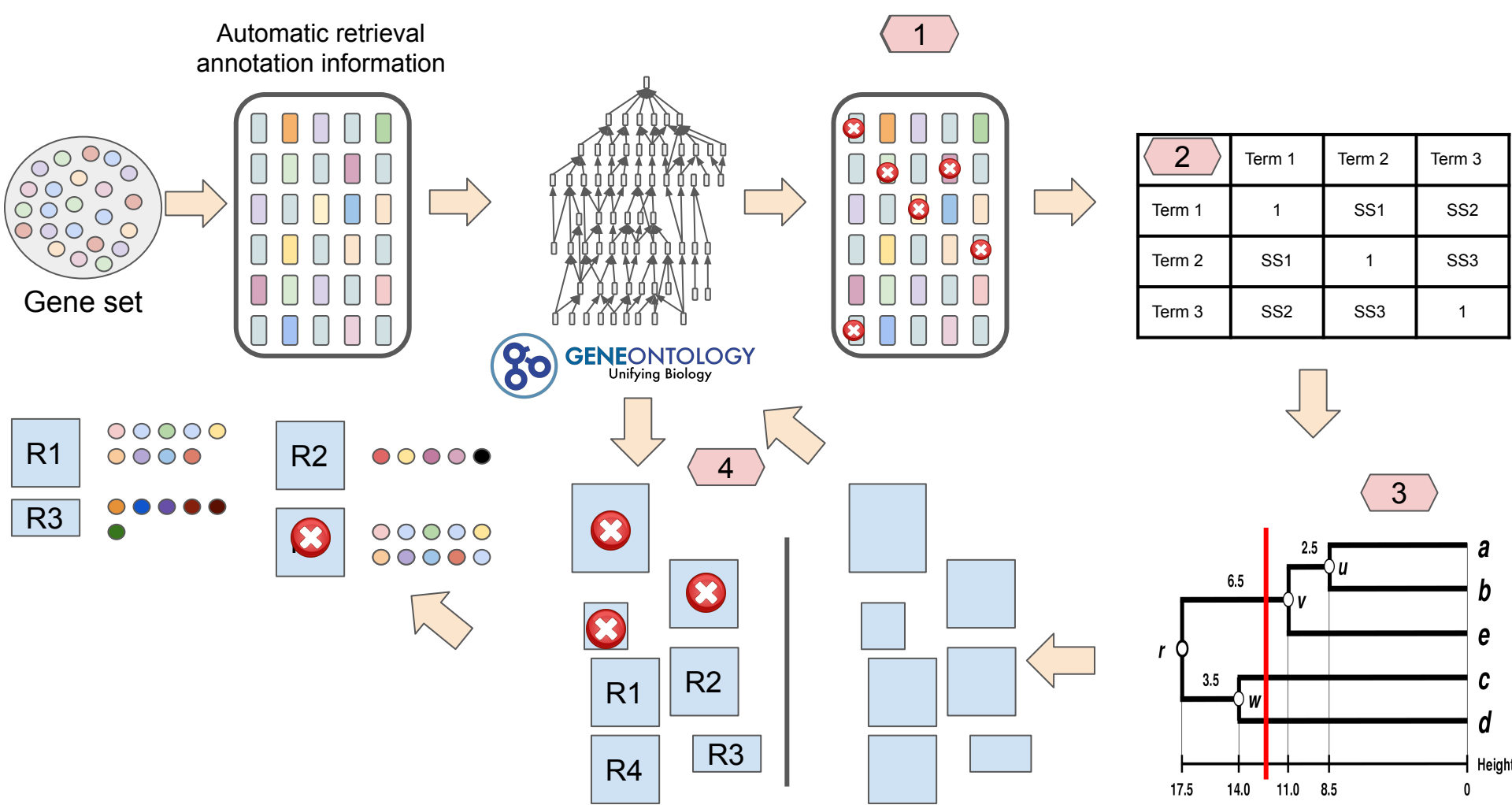


2

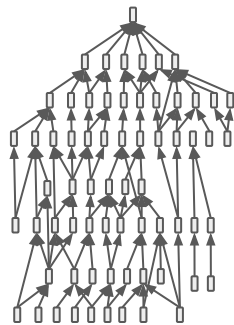
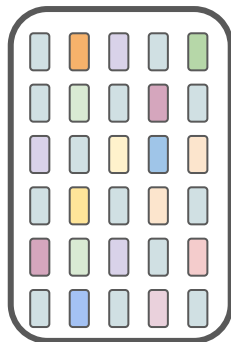
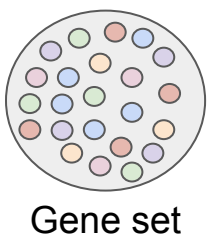
	Term 1	Term 2	Term 3
Term 1	1	SS1	SS2
Term 2	SS1	1	SS3
Term 3	SS2	SS3	1



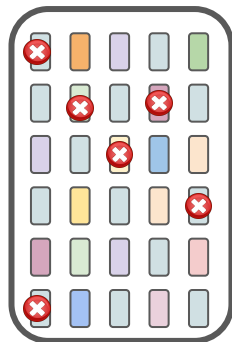




Automatic retrieval annotation information

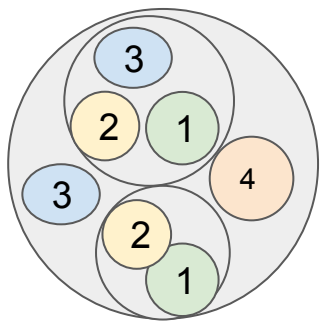
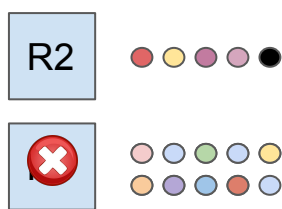
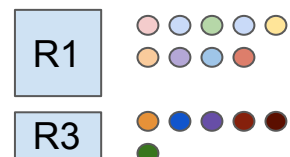


1

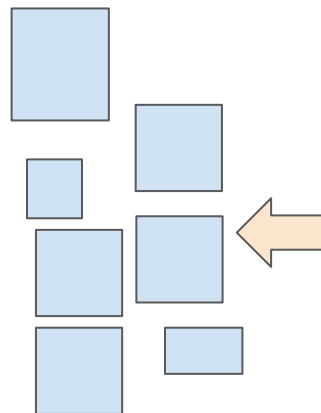
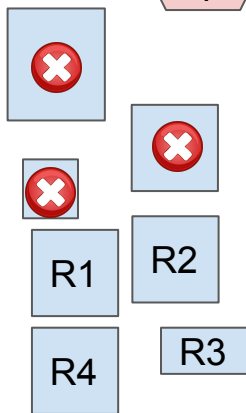


2

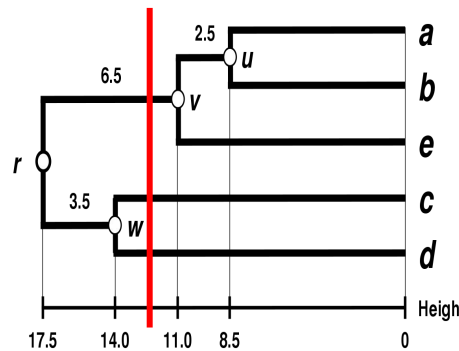
	Term 1	Term 2	Term 3
Term 1	1	SS1	SS2
Term 2	SS1	1	SS3
Term 3	SS2	SS3	1



4



3



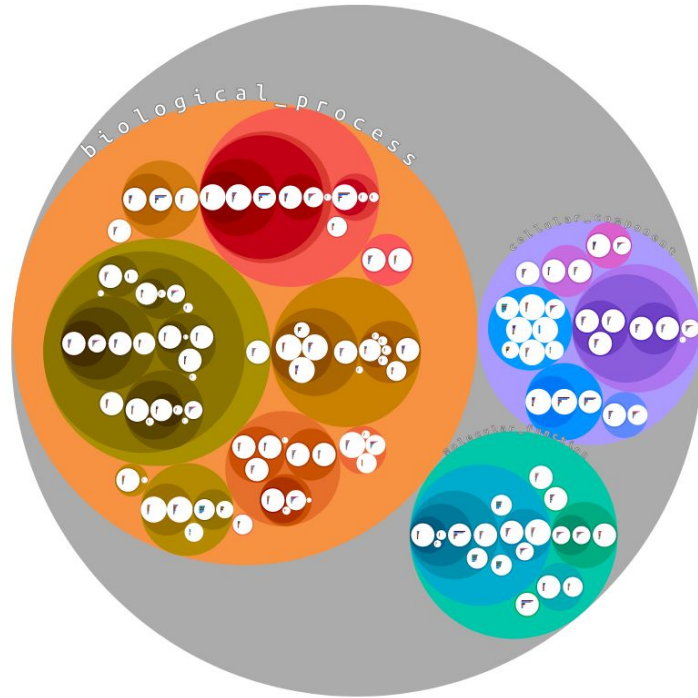
Gene_Ontology

molecular_function

cellular_component

biological_process

- DAG → TREE
- Combination of two tree visualization metaphors
 - Circular tree map
 - Indented tree
- Interactive facilities to explore the results



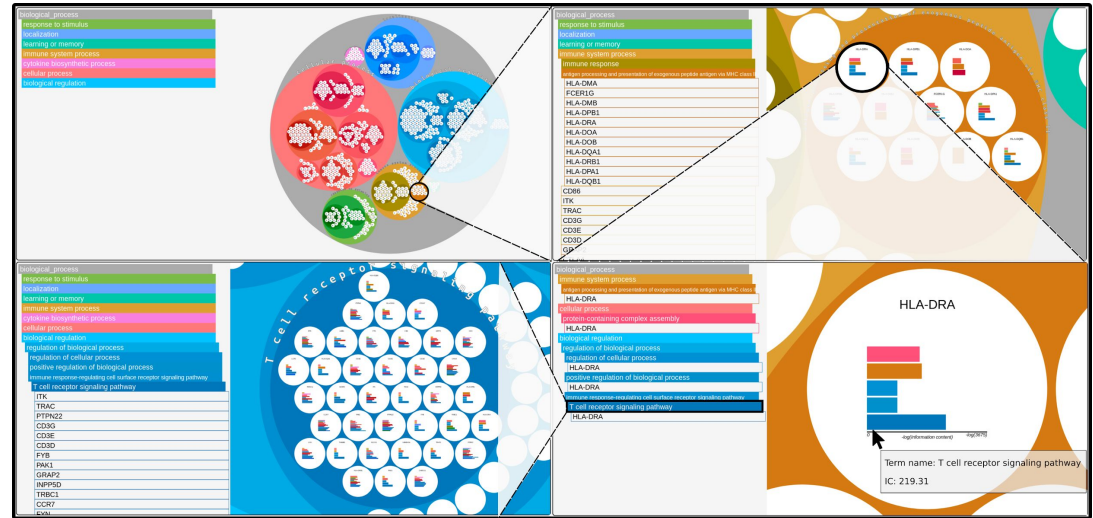


rGSAn:

- R package
- With native functions in C++
- Import ontology (in OBO format)
- Insert annotation
 - From *Go3AnnDbBimap* R Class
 - From GAF file using **flavin** package
- Run the GSAn method

Output:

- Annotation results in a data frame for further purpose
 - Export files in CSV format
 - Interactive visualization using **htmltools**



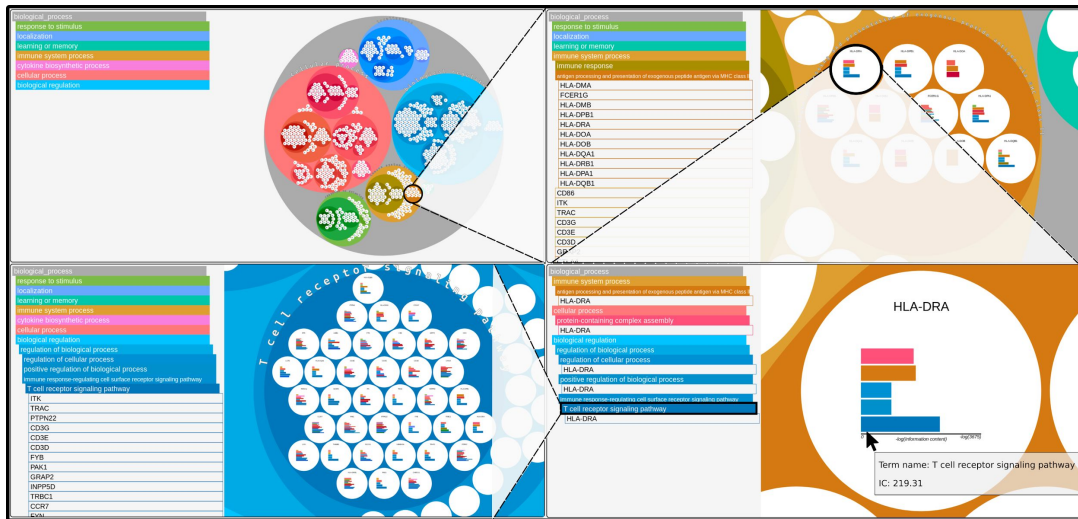


rGSAn:

- R package
- With native functions in C++
- Import ontology (in OBO format)
- Insert annotation
 - From *Go3AnnDbBimap* R Class
 - From GAF file using **flavin** package
- Run the GSAn method

Output:

- Annotation results in a data frame for further purpose
 - Export files in CSV format
 - Interactive visualization using **htmltools**



Thanks

Git repository:

<https://bitbucket.org/Ayllonbe/rgsan>

Contact:

aaron.ayllon-benitez@u-bordeaux.fr