

Biological data mining

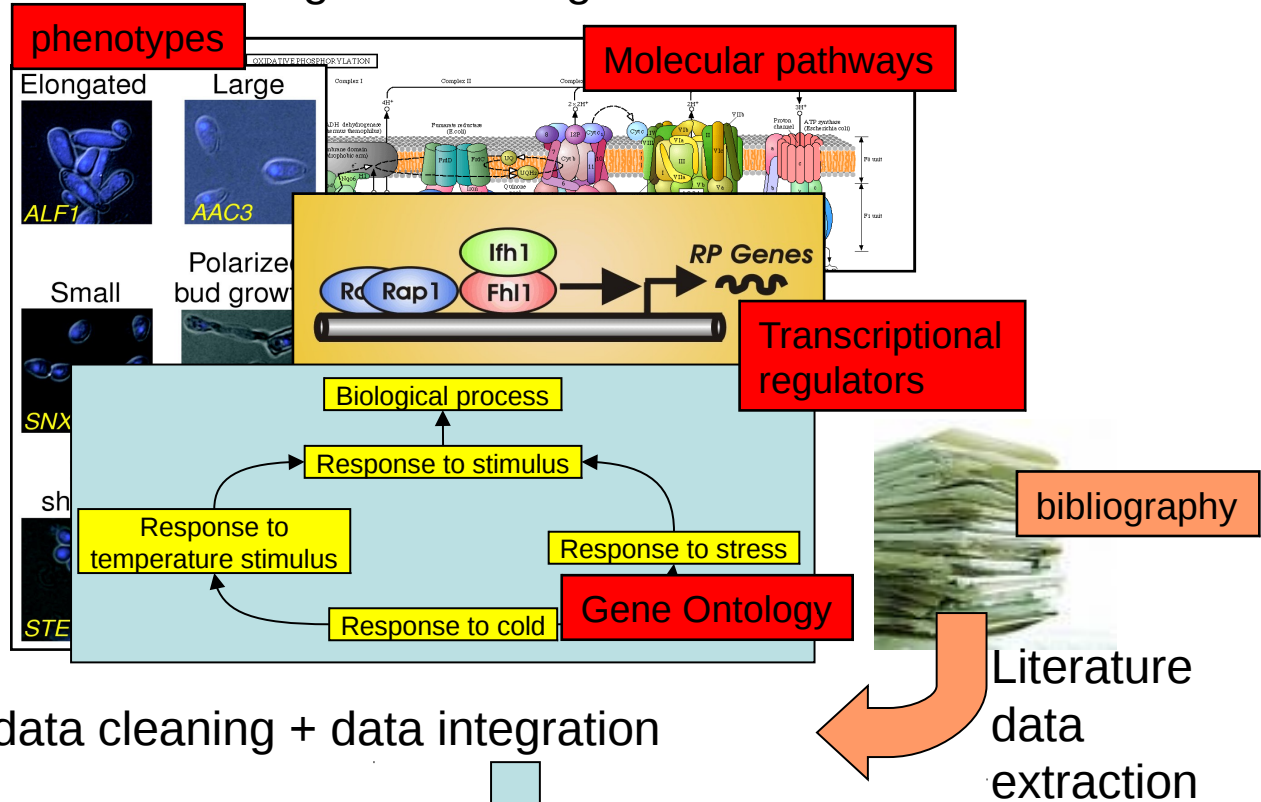
Claude Pasquier
cpasquie@unice.fr
SPARKS / I3S

Information generation from data

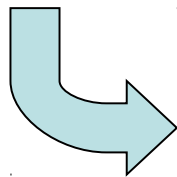
Measures



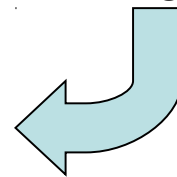
Background biological data



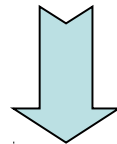
Data retrieval + data cleaning + data integration



Data interpretation



Information

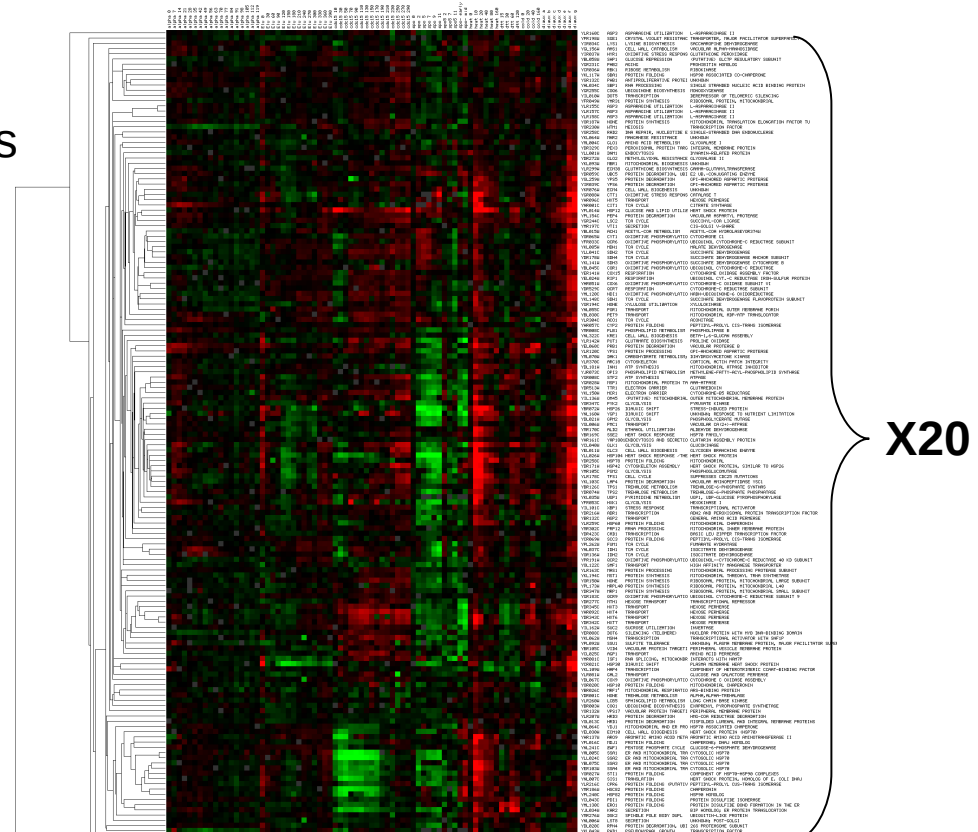


Gene expression mining

Several technologies allow the measurement of the expression of thousands of genes at once. This creates a global picture of genome activity at a given time.

Several data mining approaches:

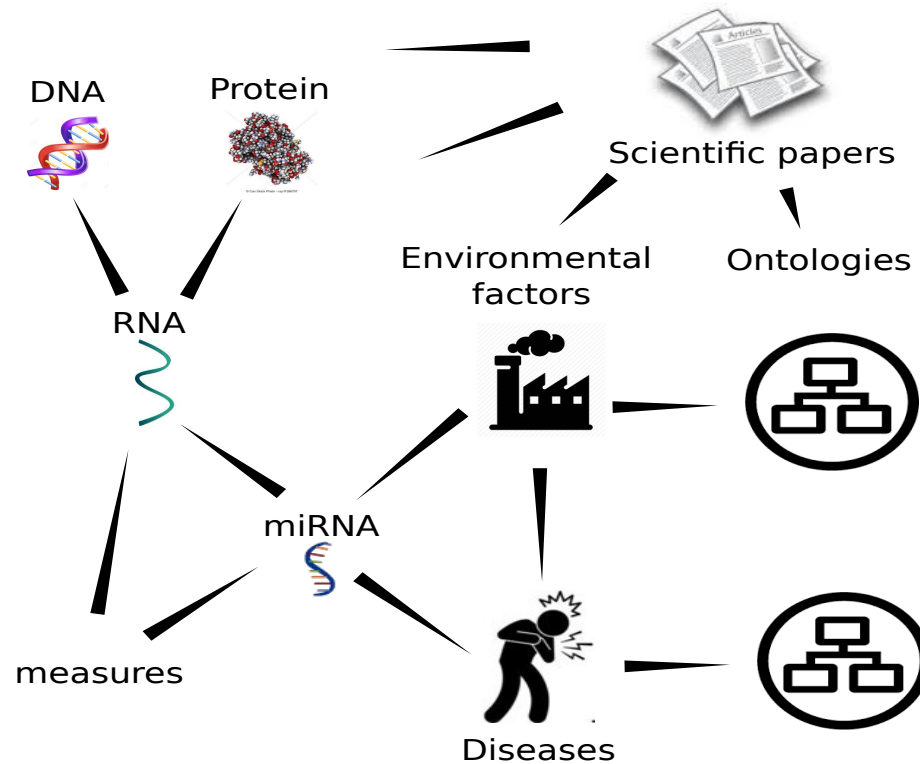
- Clustering
- Biclustering,
- Association rules



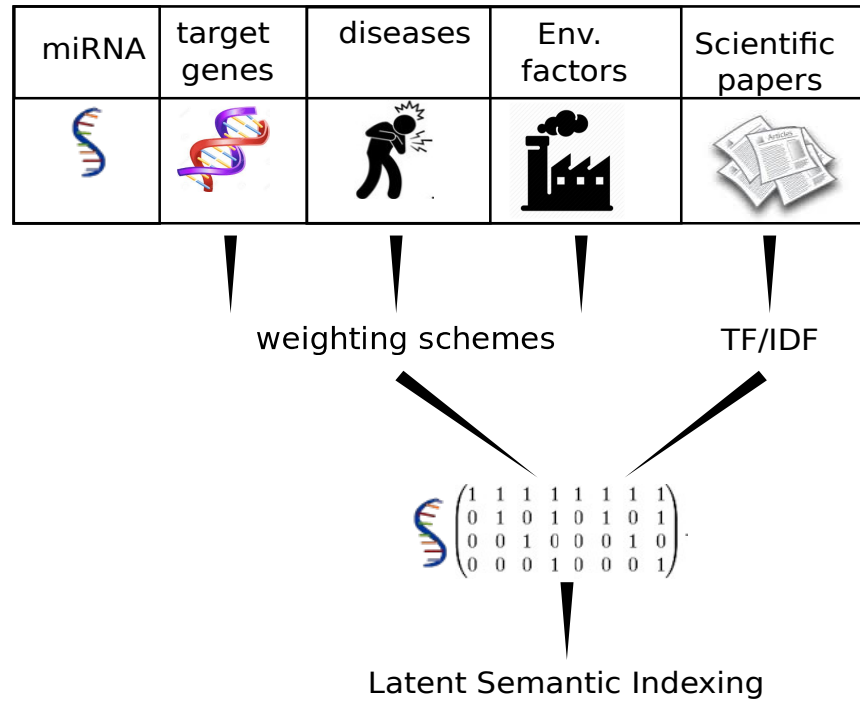
6228 genes X 80 experiments

Data from: Eisen, M.B. et al. (1998) Cluster analysis and display of genome-wide expression patterns. Proc Natl Acad Sci U S A. 1998 Dec 8;95(25):14863-8.

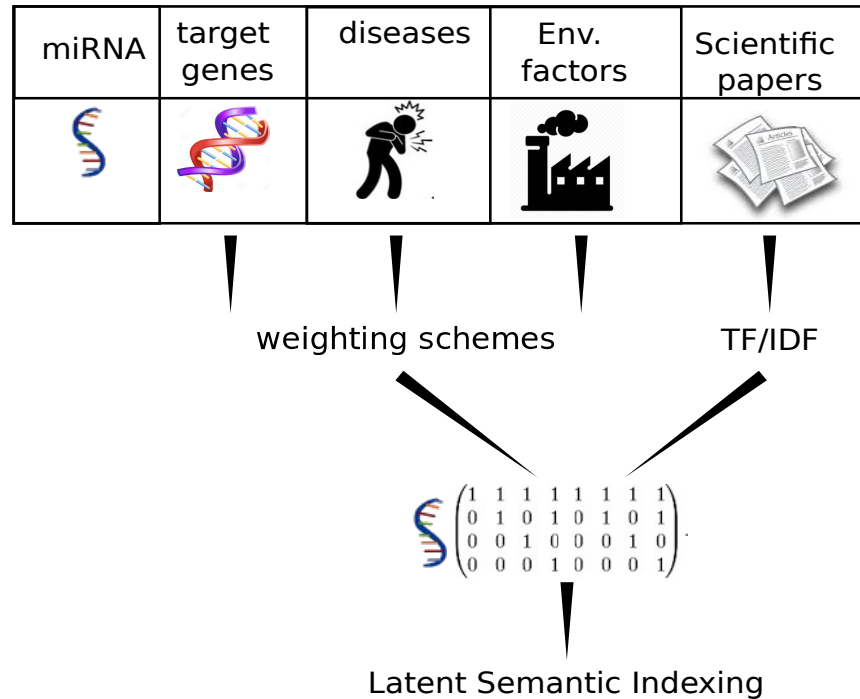
Work on miRNA – disease associations



Method



Method



Query = "lung cancer" → {list of miRNA}
Query = "cigarette smoke" → {list of miRNA}
Query = "cigarette smoke" → {list of diseases}

Query on multiple terms

- With LSI, we can perform queries with multiple terms
- What could be the meaning of multiple terms queries ?
- For example:
 - query on “Glucose” returns $D1 = \{\text{list of diseases}\}$
 - query on “Cigarette Smoke” returns $D2 = \{\text{list of diseases}\}$
 - query on “Glucose” + “Cigarette smoke” returns $D3 = ?$