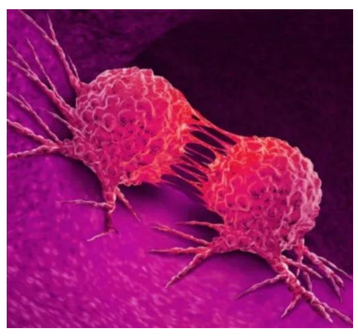
# A FEATURE SELECTION AND ASSOCIATION RULE APPROACH TO IDENTIFY GENES ASSOCIATED WITH METASTASIS AND LOW SURVIVAL IN SARCOMA

**M. Martínez-Ballesteros** (mariamartinez@us.es),
**M. Lourdes Linares-Barrera, J. Riquelme**
Dept. Computer Languages and Systems, University of Seville
**J.M. García-Heredia**
Dept. Vegetal Biochemistry and Molecular Biology, University of Seville

JABI JORNADAS ANDALUZAS DE BIOINFORMÁTICA
SEVILLA, 3-5 DE JULIO 2023

## 1. MOTIVATION

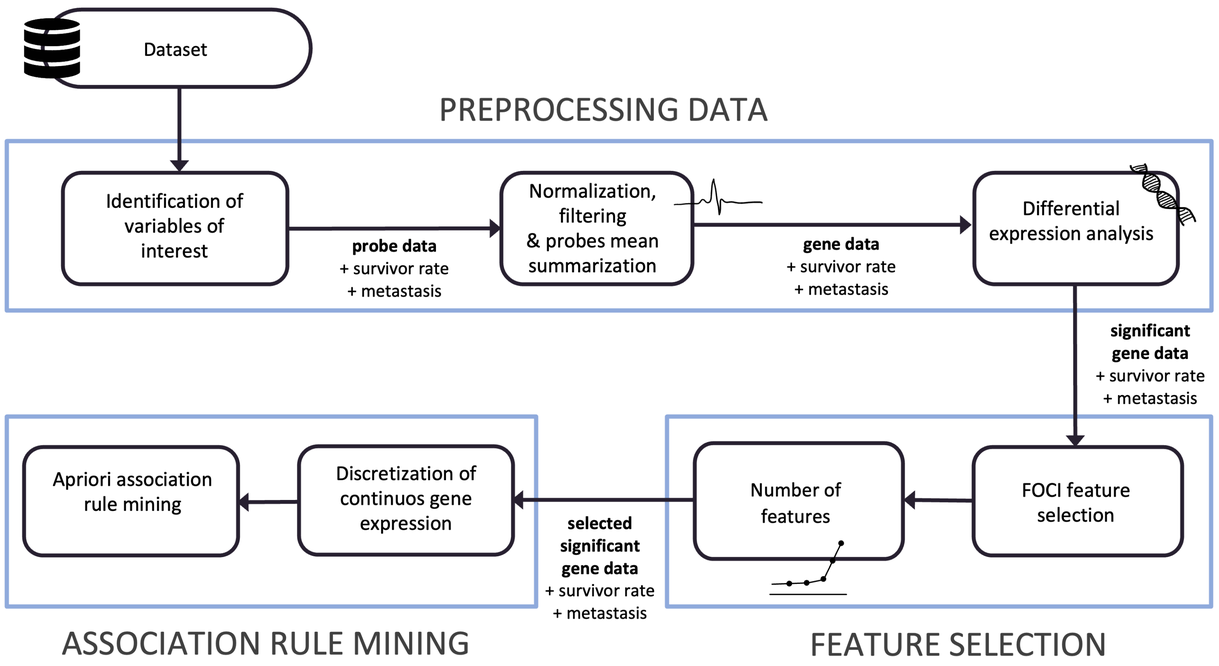


### Problem:

- Identification of relevant genes to metastatic sarcoma.
- Discovery of significant relationships between metastasis-associated genes and < 5-year patient survival.

- Sarcomas are rare and heterogeneous mesodermal tumors that primarily affect children and adolescents (over 20% of all pediatric tumors).
- The relative 5-year survival rate for patients with metastatic sarcoma is usually low.
- Investigating the genetic basis of these tumors through genome-wide analysis is crucial due to their rarity and late diagnosis.

## 2. METHODOLOGY




### Methodology phases:

1. **Preprocessing of data:** filtering, summarization and analysis of differential gene expression.
2. **Feature selection:** subset of relevant genes for metastatic sarcoma using FOCI method.
3. **Association rule mining:** discovering interesting relationships between genes, leading to the identification of potential biomarkers related to metastastic sarcoma and low survival time.

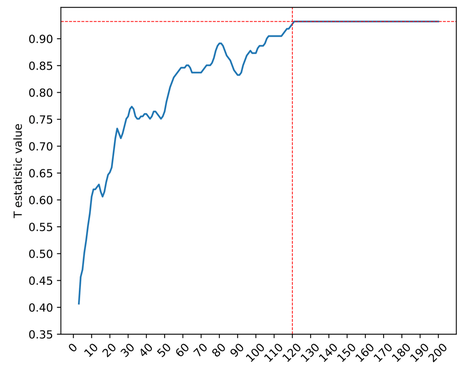
### Input data

- Gene expression dataset GSE21050 - French Sarcoma Group (FSG) database available in Gene Expression Omnibus (GEO).
- 309 patients:
  - Time Survivor
  - Metastasis status: Yes (121) + No (188).
  - 54,613 probes (*Affymetrix Human Genome U133 Plus 2.0*).

## 3. RESULTS

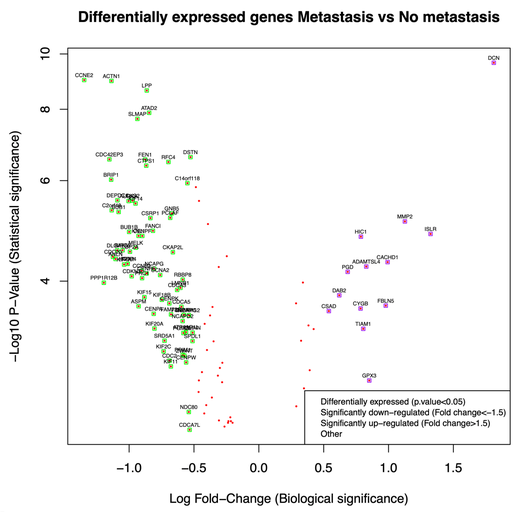
### Preprocessing results

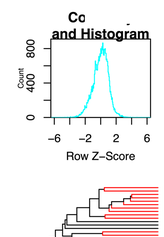
- Filtering and summarization process: 11,901 genes.
- Differential expression analysis: 1,516 genes.

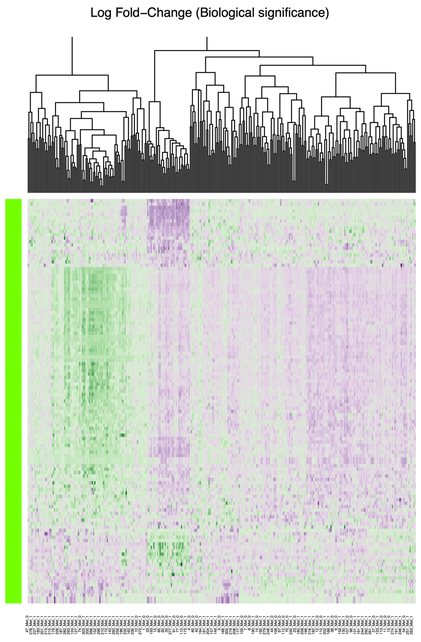


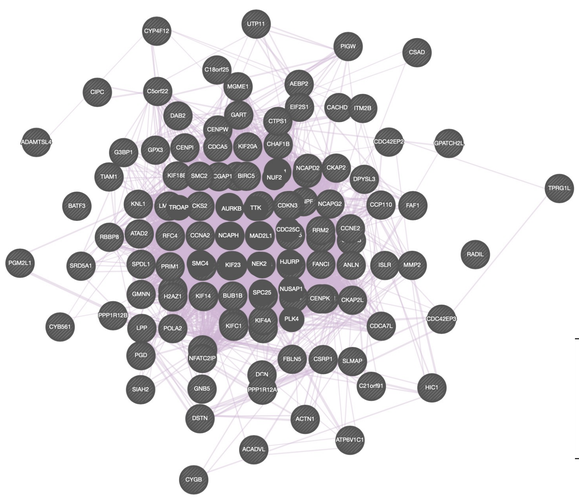

### Feature selection results

- Number of genes selected using FOCI that best explains the response variable (metastasis): 120
- Hierarchical clustering and Volcano Plot to analyze the selected genes expression levels.
- Coexpression network.





### Association rules mining results

- Example of associations between genes with low (0) and high (1) expression levels in patients with <5-year survival (t_survival = 0) and metastasis.

| Antecedent | Consequent | Sup | Lift | Conf |
|---|---|---|---|---|
| CKAP2=1 ∧ GART=0 ∧ H2AZ1=0 ∧ t_survival=0 | metastasis | 0.055 | 1.64 | 1.00 |
| ADAMTSL4=1 ∧ CDC42EP2=0 ∧ DSTN=0 ∧ t_survival=0 | metastasis | 0.058 | 1.64 | 1.00 |
| ADAMTSL4=1 ∧ DSTN=0 ∧ GMNN=1 ∧ t_survival=0 | metastasis | 0.055 | 1.64 | 1.00 |
| ADAMTSL4=1 ∧ ISLR=1 ∧ TOP2A=1 ∧ t_survival=0 | metastasis | 0.052 | 1.55 | 0.94 |
| ACADVL=1 ∧ C2orf48=1 ∧ FBLN5=1 ∧ t_survival=0 | metastasis | 0.058 | 1.48 | 0.90 |
| ACADVL=1 ∧ SPDL1=1 ∧ TIAM1=1 ∧ t_survival=0 | metastasis | 0.061 | 1.49 | 0.90 |
| ATP6AP1L=0 ∧ DPYSL3=0 ∧ NCAPG2=0 ∧ t_survival=0 | metastasis | 0.061 | 1.56 | 0.95 |
| DPYSL3=0 ∧ NCAPG2=0 ∧ POLA2=0 ∧ t_survival=0 | metastasis | 0.074 | 1.51 | 0.92 |
| CDC42EP2=0 ∧ DPYSL3=0 ∧ NCAPG2=0 ∧ t_survival=0 | metastasis | 0.074 | 1.51 | 0.92 |

## 4. CONCLUSIONS

- Relevant group of genes associated with metastasis and low survival rates in sarcoma*.
- Potential new treatments for metastasic sarcoma patients integrating drug repurposing strategies,, association rules, and gene-drug associations**.
- **Future works:** heterogeneous omic data and other diseases, advanced algorithms for identifying relevant association rules.

***Acknowledgements:*** *The authors would like to thank the Spanish Ministry of Science and Innovation for the support under PID2020-117954RB-C22 project.*

*M.L. Linares-Barrera, M. Martínez-Ballesteros, J.M. García-Heredia, J.C. Riquelme. A Feature Selection and Association Rule Approach to Identify Genes Associated with Metastasis and Low Survival in Sarcoma. 18th International Conference on Hybrid Artificial Intelligence Systems, 2023 (In press).

**J.M. García-Heredia, M. Pérez, E.M. Verdugo-Sivianes, M. Martínez-Ballesteros, S.M. Ortega-Campos A. Carnero. A new treatment for sarcoma extracted from combination of miRNA deregulation and gene association rules. Signal Transduction and Targeted Therapy, 8:1, 231, 2023, https://doi.org/10.1038/s41392-023-01470-z