

Introduction

- Gene Set Enrichment Analysis (GSEA) is a classical computational method in transcriptomic analysis.
- GSEA determines whether an *a priori* defined geneset (**GS**) shows statistically significant and concordant differences between two conditions [1].
- We developed a new metabolomic signature (collection of annotated GS) for Multi-omics analysis.
- We illustrated our signature using RNA-seq data from the TCGA cholangiocarcinoma dataset [2].

Methods

- Extract metabolite-genes interactions from HMDB (file: « *hmdb_metabolites.xml.gz* »).
- Create GS for each metabolite associated with its HGNC (HUGO Gene Nomenclature Committee) gene symbol.
- Gene Matrix Format (GMT) export.
- GS filtration:
 - Exclude metabolites with no gene name association.
 - Redundant GS (same list of genes) are merged into one class.
 - A dictionary file is provided with the description of class.
- GMT Metabolite Signature is imported in the GSEA software for analysis.

Ressources

- Human Metabolome Database (HMDB, v.5.0) [3]: <https://hmdb.ca>.
- Molecular signature Database (MSig DB) and its GSEA software: <https://www.gsea-msigdb.org/gsea>.
- The Cancer Genome Atlas (TCGA) for cholangiocarcinoma (TCGA-CHOL): <https://portal.gdc.cancer.gov>.

Results

- The filtration step is crucial and mandatory (figure A1 & A2):
 - From **271 921 Metabolites** to **571 GS** (figure A2).
- Application to cholangiocarcinoma (TCGA-CHOL):
 - **198 GS** upregulated in Solid_Tissue_Normal (FDR < 25%).
 - **13 GS** enriched in Primary_Tumor (FDR < 25%).
- Exemple of **Estradiol**:
 - Positively correlated to normal tissue (Enrichment Score ES = 0.8, figure C).
 - Inhibits the proliferation of hepatocellular cancer cell lines and increased apoptosis for liver cancer [4].
 - Heatmap of gene expression (figure D).

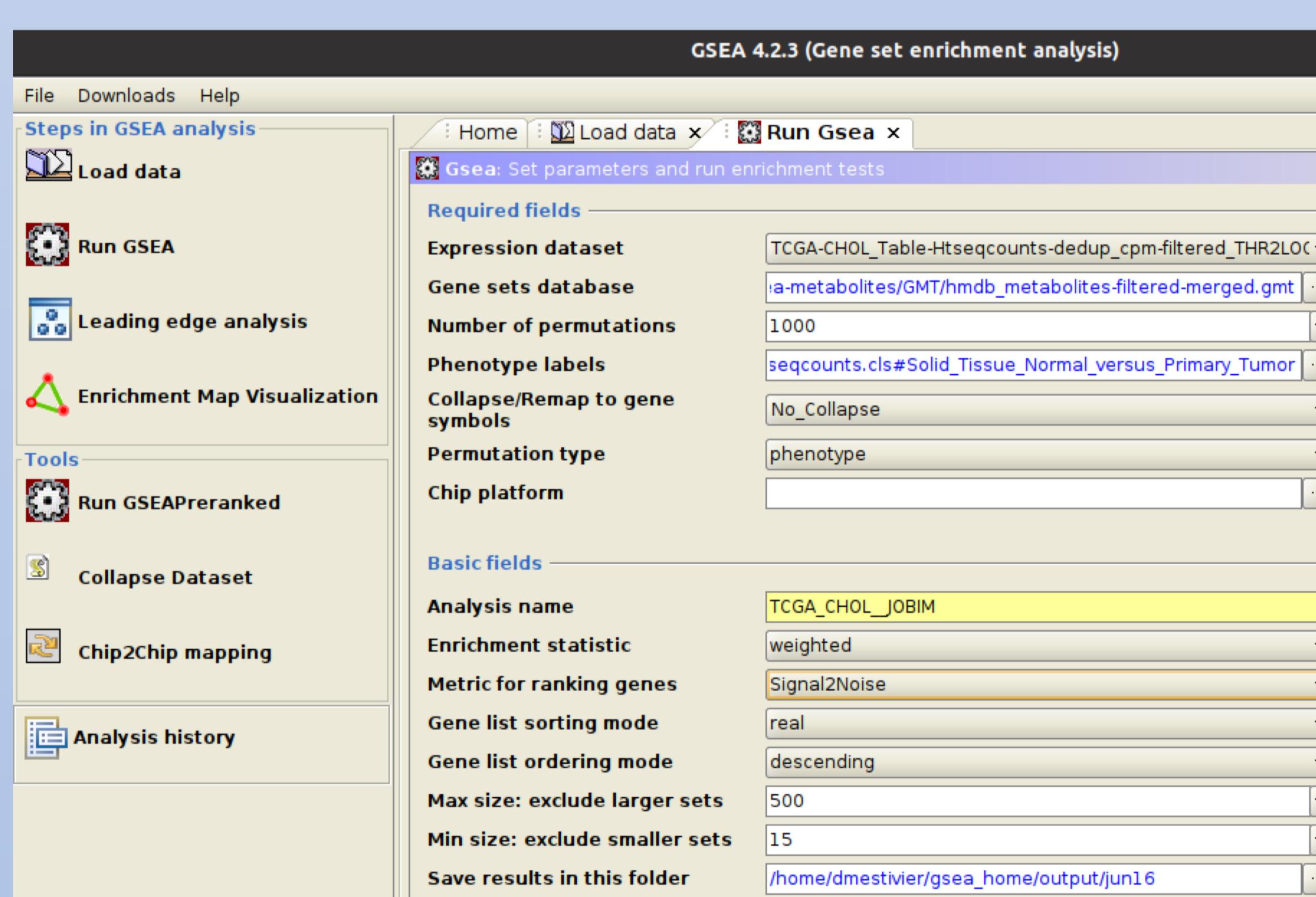
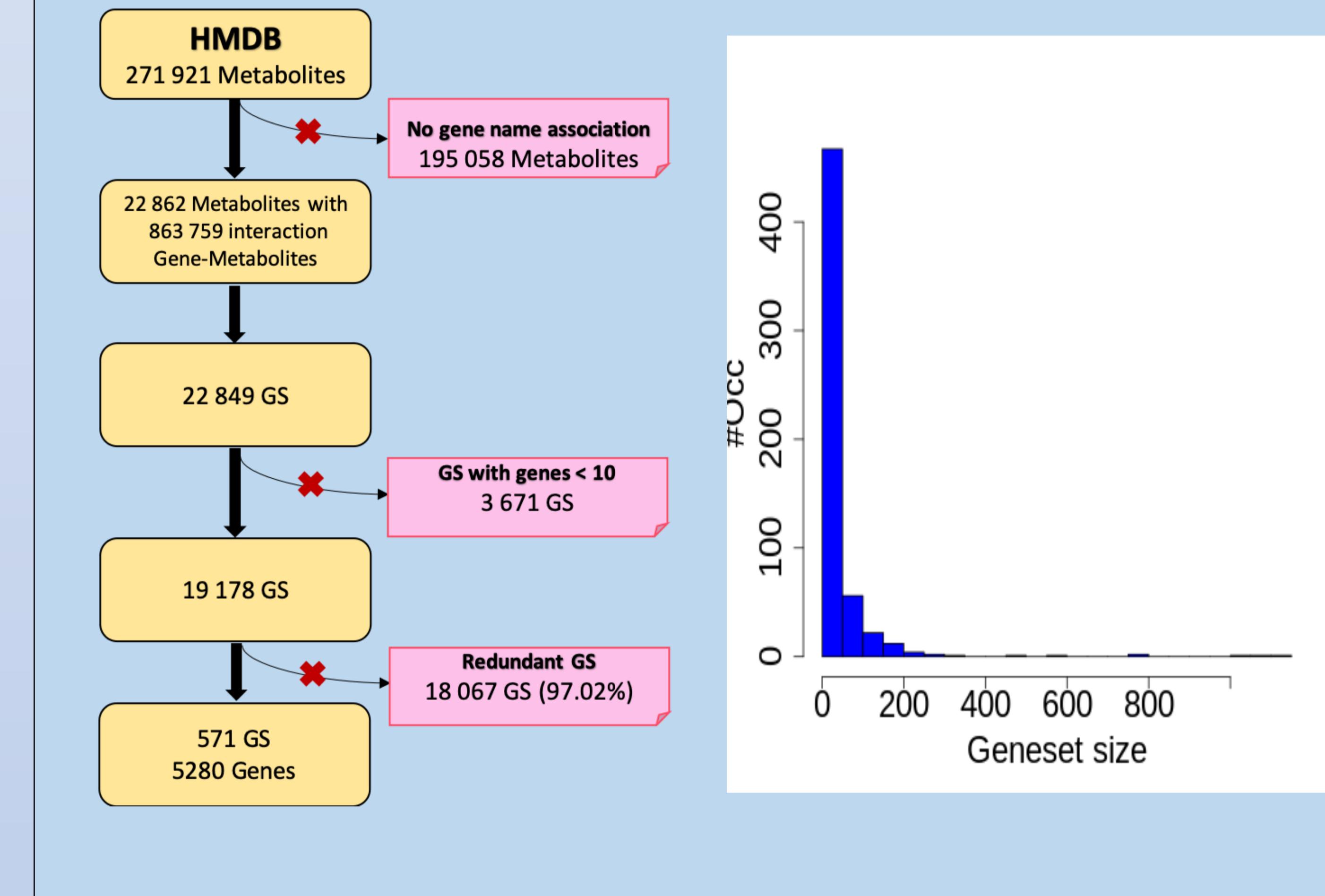


Figure B: Graphic interface of the software GSEA.

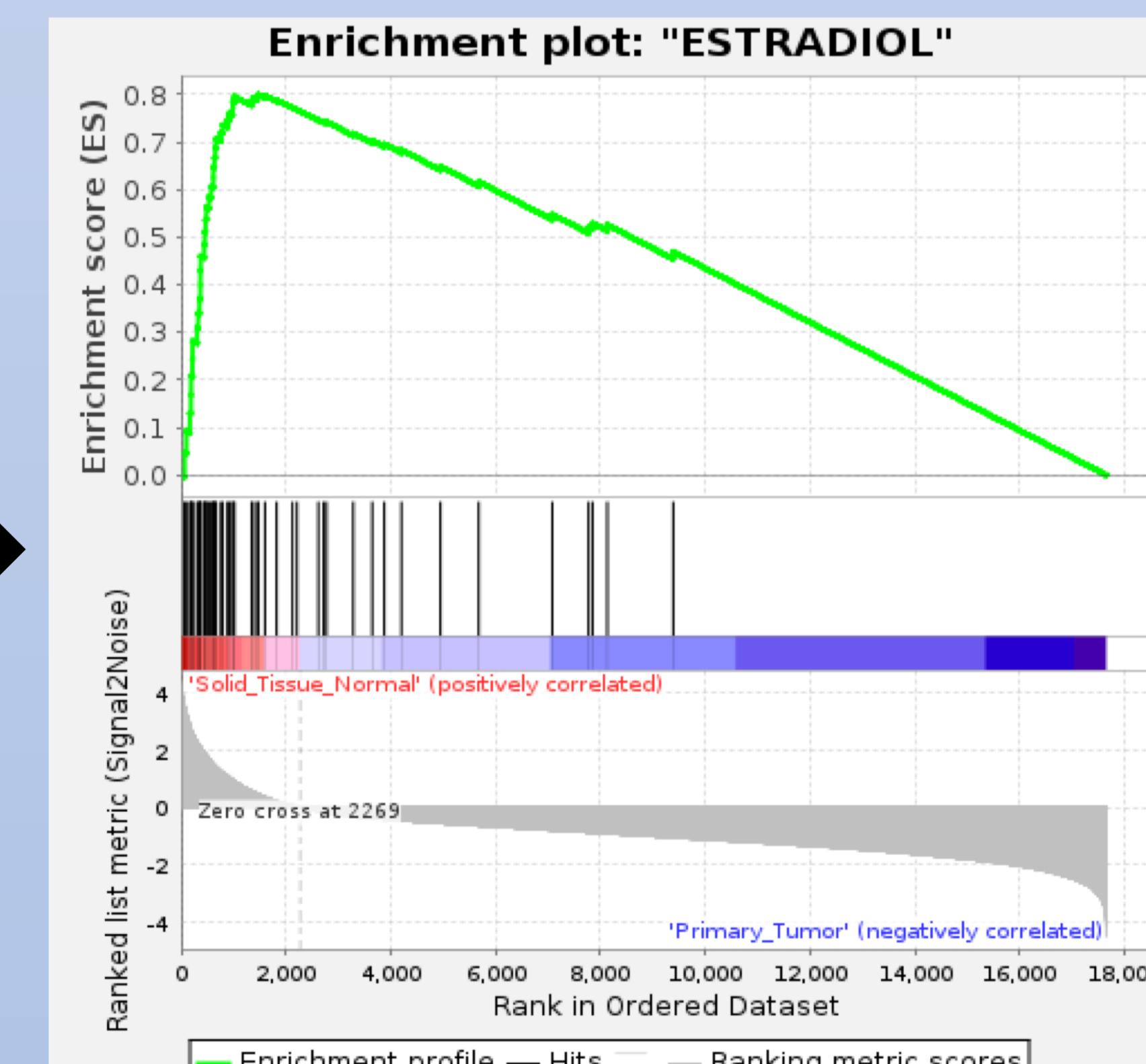


Figure C: Enrichement plot of Estradiol between healthy and cancerous tissues.

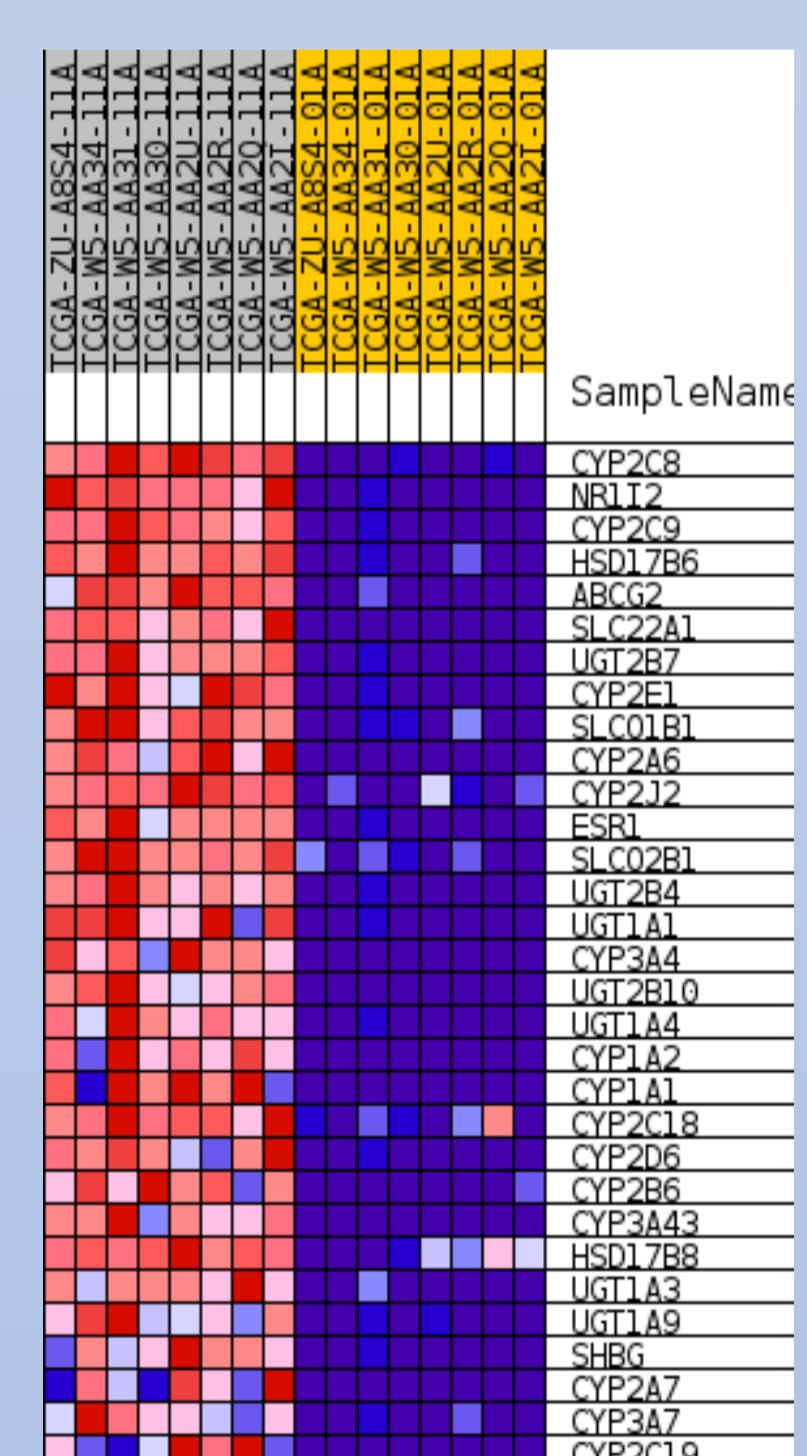


Figure D: Gene expression heatmap between the two tissues.

Conclusion

- We created a metabolomic signature for GSEA/RNA-seq analysis.
- We provide a GMT format file easily used with standard (R and GSEA software for example).
- Our approach can be extended to compartments of HMDB (saliva, urine, feces ...).
- Filtering and « high quality » GS is mandatory (eg. for the multiple correction statistics).
- Scripts and files are available at the **GITHUB** : <https://github.com/dmestivier/gsea-hmdb>.

REFERENCES

- Subramanian, A. et al. Gene set enrichment analysis: A knowledge-based approach for interpreting genome-wide expression profiles. Proc. Natl. Acad. Sci. U. S. A. 102, 15545–15550 (2005).
- Farshidfar, F. et al. Integrative Genomic Analysis of Cholangiocarcinoma Identifies Distinct IDH-Mutant Molecular Profiles. Cell Rep. 18, 2780–2794 (2017).
- Wishart, D. S. et al. HMDB 5.0: The Human Metabolome Database for 2022. Nucleic Acids Res. 50, D622–D631 (2022).
- Wang, R. et al. Estradiol is significantly associated with prognosis in non-surgical liver cancer patients: from bench to bedside. Aging (Albany, NY). 13, 3483–3500 (2021).