UNIVERSITÀ
DI PAVIA

**Dipartimento di Biologia e Biotecnologie**

**“Lazzaro Spallanzani”**

Laurea Magistralis in Molecular Biology and Genetics

# Bulk RNA sequencing and Spatial Transcriptomic Analysis in Early-Relapsing and Long-Responders Patients with Epithelioid Peritoneal Mesothelioma after CRS-HIPEC

Supervisor:

*Prof.ssa Mariangela Santorsola*

Co-supervisor:

*Dott. Luca Agnelli*

Experimental thesis by

*Giorgia Brognara*

Academic Year 2024/2025


UNIVERSITÀ
DI PAVIA

**Dipartimento di Biologia e Biotecnologie**

**"Lazzaro Spallanzani"**

Laurea Magistralis in Molecular Biology and Genetics

# Bulk RNA sequencing and Spatial Transcriptomic Analysis in Early-Relapsing and Long-Responders Patients with Epithelioid Peritoneal Mesothelioma after CRS-HIPEC

**Analisi Trascrittomica Bulk e Spaziale del Mesotelioma Peritoneale Epitelioide in Pazienti con Recidiva Precoce o Tardiva dopo CRS-HIPEC**

Supervisor:

*Prof.ssa Mariangela Santorsola*

Co-supervisor:

*Dott. Luca Agnelli*

Experimental thesis by

*Giorgia Brognara*

Academic Year 2024/2025

# List of abbreviations

| | |
|---|---|
| **1A4** | Clone for alpha-SMA (Marker for activated fibroblasts/stroma) |
| **18F-FDG** | 18F-fluorodeoxyglucose (radiopharmaceutical used in PET) |
| **ACSL3** | Acyl-CoA Synthetase Long Chain Family Member 3 |
| **ADAM9/15** | ADAM Metallopeptidase Domain 9 / 15 |
| **ALDOA** | Aldolase, Fructose-Bisphosphate A (Glycolytic enzyme) |
| **AOI** | Area of Illumination |
| **ATG3** | Autophagy Related 3 |
| **ATP5MC2** | ATP Synthase Membrane Subunit C Locus 2 |
| **ATR** | Ataxia Telangiectasia and Rad3-related protein |
| **BAM** | Binary Alignment Map |
| **BAP1** | BRCA1-Associated Protein 1 (Deubiquitinating enzyme) |
| **BCL** | Base Call (raw sequence files from Illumina) |
| **BDNF-AS** | Brain-Derived Neurotrophic Factor Antisense RNA |
| **bp** | Base pairs |
| **BST2** | Bone Marrow Stromal Cell Antigen 2 |
| **CAF** | Cancer-Associated Fibroblast (iCAF: inflammatory; myCAF: myofibroblastic) |
| **CC-score** | Completeness of Cytoreduction score |
| **CCNB1** | Cyclin B1 |
| **CD68** | Cluster of Differentiation 68 (Macrophage marker) |
| **CDKN2A/B** | Cyclin-Dependent Kinase Inhibitor 2A/B |
| **cDNA** | Complementary DNA |
| **CEA** | Carcinoembryonic Antigen |
| **CET** | Comitato Etico Territoriale (Local Ethics Committee) |
| **CI** | Confidence Interval |
| **CK5/6** | Cytokeratin 5/6 |
| **$CK^+$ / PanCK** | Cytokeratin positive (Epithelial tumor cell marker) |
| **CLI** | Command-Line Interface |
| **c-Met** | Tyrosine-protein kinase Met (Hepatocyte Growth Factor Receptor) |
| **COL18A1** | Collagen Type XVIII Alpha 1 Chain (Precursor to endostatin) |
| **COX5B** | Cytochrome C Oxidase Subunit 5B |
| **CPU** | Central Processing Unit |
| **CRS-HIPEC** | Cytoreductive Surgery with Hyperthermic Intraperitoneal Chemotherapy |
| **CT** | Computed Tomography |
| **CTNNB1** | Catenin Beta 1 (Key component of Wnt signaling) |
| **D2-40** | Podoplanin (mesothelial marker) |
| **DCC** | Digital Count Conversion (GeoMx count data format) |
| **DDX3X** | DEAD-Box Helicase 3 X-Linked |
| **DE / DEG** | Differential Expression / Differentially Expressed Gene |
| **DESeq2** | Differential Expression Sequencing 2 |
| **DFS** | Disease-Free Survival |
| **DMD** | Digital Micromirror Device |
| **DMPeM / DMPM** | Diffuse Malignant Peritoneal Mesothelioma |
| **dNTP** | Deoxynucleoside Triphosphate |
| **DOD** | Dead of Disease |
| **DSP** | Digital Spatial Profiler |
| **DV200** | Percentage of RNA fragments >200 nucleotides (Quality metric) |
| **E2F** | E2 Factor (Transcription factor family involved in cell cycle) |
| **ECM** | Extra-cellular Matrix |
| **e-DMPeM** | Epithelioid Diffuse Malignant Peritoneal Mesothelioma |
| **EEF1A1** | Eukaryotic Translation Elongation Factor 1 Alpha 1 |
| **EGFR** | Epidermal Growth Factor Receptor |
| **EIF2** | Eukaryotic Initiation Factor 2 (Master regulator of protein synthesis) |
| **EMT** | Epithelial-Mesenchymal Transition (including E→S: Epithelial-to-Sarcomatoid) |
| **ER** | Early Recurrence |
| **EZH2** | Enhancer of Zeste Homolog 2 |
| **FASTA / FASTQ** | Text-based formats for storing biological sequences and quality scores |
| **FDR** | False Discovery Rate |
| **FFPE** | Formalin-Fixed Paraffin-Embedded |
| **FTH1 / FTL** | Ferritin Heavy Chain 1 / Ferritin Light Chain |
| **GAS1** | Growth Arrest Specific 1 |
| **GLM** | Generalized Linear Model |
| **GLOBOCAN** | Global Cancer Observatory |
| **GO** | Gene Ontology (BP: Biological Process; CC: Cellular Component; MF: Molecular Function) |

| | |
|---|---|
| **GTF** | Gene Transfer Format |
| **HIF1A** | Hypoxia Inducible Factor 1 Alpha |
| **HLA-A** | Human Leukocyte Antigen – A (MHC Class I) |
| **HPC** | High-Performance Computing |
| **HR** | Hazard Ratio |
| **H-score** | Histology score (IHC quantification) |
| **HSP90** | Heat Shock Protein 90 |
| **HTSeq** | High-Throughput Sequencing |
| **IARC** | International Agency for Research on Cancer |
| **IHC** | Immunohistochemistry |
| **IL6/JAK/STAT3** | Interleukin 6 / Janus Kinase / Signal Transducer and Activator of Transcription 3 |
| **INT** | Istituto Nazionale dei Tumori |
| **IPA** | Ingenuity Pathway Analysis |
| **IQR** | Interquartile Range |
| **IRCCS** | Istituto di Ricovero e Cura a Carattere Scientifico |
| **ITGB1** | Integrin Subunit Beta 1 |
| **KEGG** | Kyoto Encyclopedia of Genes and Genomes |
| **Ki-67** | Proliferation marker protein |
| **KM** | Kaplan-Meier (Survival curves) |
| **log2FC** | Log2 Fold Change |
| **LR** | Long Responders |
| **LS** | Lesion Size (score) |
| **MAPK8IP1** | Mitogen-Activated Protein Kinase 8 Interacting Protein 1 |
| **MRI** | Magnetic Resonance Imaging |
| **MSigDB** | Molecular Signatures Database |
| **mTORC1** | Mammalian Target of Rapamycin Complex 1 |
| **MZT2B** | Mitotic Spindle Organizing Protein 2B |
| **NF2** | Neurofibromatosis Type 2 (Merlin) |
| **NGS** | Next-Generation Sequencing |
| **noSV** | No Surrogate Variables (uncorrected statistical model) |
| **nt** | Nucleotides |
| **NYNRIN** | NYN Domain and Retroviral Integrase Containing |
| **OS** | Overall Survival |
| **padj** | Adjusted P-value (Multiple-testing corrected) |
| **PARP** | Poly (ADP-ribose) Polymerase |
| **PAX8** | Paired Box gene 8 |
| **PBRM1** | Polybromo 1 |
| **PCA** | Principal Component Analysis (PC1 / PC2: Principal Components) |
| **PCI** | Peritoneal Cancer Index |
| **PET/CT** | Positron Emission Tomography / Computed Tomography |
| **PFS** | Progression-Free Survival |
| **PKC** | Probe Kit Configuration (GeoMx mapping file) |
| **QC** | Quality Control (e.g., FastQC, MultiQC) |
| **RLE** | Relative Log Expression |
| **RMRP** | RNA Component of Mitochondrial RNA Processing Endoribonuclease |
| **RN7SL1** | RNA Component of Signal Recognition Particle 7SL1 |
| **ROI** | Region of Interest |
| **ROS** | Reactive Oxygen Species |
| **RPL / RPS** | Ribosomal Protein Large / Small subunit |
| **RUV4** | Remove Unwanted Variation 4 (Batch correction algorithm) |
| **SAM** | Sequence Alignment Map |
| **SBS** | Sequencing-by-Synthesis |
| **scRNA-seq** | Single-cell RNA Sequencing |
| **SCYL2** | SCY1 Like Pseudokinase 2 |
| **SETD2** | SET Domain Containing 2 (Histone methyltransferase) |
| **SLURM** | Simple Linux Utility for Resource Management |
| **SMAD3** | SMAD Family Member 3 |
| **SPNS2** | Sphingolipid Transporter 2 |
| **SSR1** | Signal Sequence Receptor Subunit 1 |
| **STAR** | Spliced Transcripts Alignment to a Reference |
| **STUB1 (CHIP)** | STIP1 Homology and U-Box Containing Protein 1 |
| **SVA / SV** | Surrogate Variable Analysis / Surrogate Variable |
| **SYTO13** | Fluorescent nucleic acid stain (Nuclear marker) |
| **TGF-β** | Transforming Growth Factor Beta |
| **TILs** | Tumor-Infiltrating Lymphocytes |
| **TME** | Tumor Microenvironment |
| **TMM** | Trimmed Mean of M-values (Normalization method) |
| **TPM3** | Tropomyosin 3 |
| **TTP** | Time-To-Progression |
| **TTC21A** | Tetratricopeptide Repeat Domain 21A |
| **TTF-1** | Thyroid Transcription Factor-1 |
| **UMI** | Unique Molecular Identifier |
| **UV** | Ultraviolet |
| **WEE1** | WEE1 G2 Checkpoint Kinase |
| **WHO** | World Health Organization |
| **WT1** | Wilms' Tumor 1 |
| **WTA** | Whole Transcriptome Atlas |

# Abstract

**Introduction and Aims:** Epithelioid diffuse malignant peritoneal mesothelioma (e-DMPeM) is a rare, aggressive cancer of the abdominal mesothelium with highly heterogeneous characteristics. Cytoreductive surgery combined with hyperthermic chemotherapy (CRS-HIPEC) is the standard of care, but the biological determinants of recurrence after this treatment are unknown. This study aimed to define bulk and spatial transcriptomic signatures discriminating early recurrence (ER) versus long responders (LR) patients (pts) and to assess how tumor and microenvironment organization influence relapse.

**Materials and Methods**: 36 e-DMPeM pts treated with CRS-HIPEC were classified as ER (progression ≤12 months) or LR (progression-free ≥36 months). Bulk RNA-seq profiles, generated on the Illumina platform, were analyzed by DEA, GSEA, IPA, immune deconvolution, and CAF-signature enrichment. *GAS1*, the top differentially expressed gene, was validated by IHC, and patients were further stratified into *GAS1*-high/low within ER and LR. In 22 patients, spatial transcriptomics (DSP GeoMx) was used to assess tissue heterogeneity across four compartments: tumor core, macrophage niches, fibroblasts, and lymphoid structures, enabling tissue-wide and cell-specific signature analysis.

**Results and Conclusion**: Bulk RNA-seq revealed a sharp molecular split between ER and LR. ER tumors showed proliferative, metabolic, and ECM-remodeling programs in an immunosuppressed, fibroblast-activated microenvironment, whereas LR tumors displayed a quiescent profile with diminished oncogenic signaling and enriched immune pathways. High *GAS1* expression marked a less aggressive state, with reduced proliferation, metabolic inhibition, and attenuated angiogenic signaling. Spatial transcriptomics uncovered compartment-specific organization: ER tumors exhibited coordinated pro-tumorigenic signatures in $CK^+$ cells and $CD68^+$ macrophages, while LR retained functionally active lymphoid niches enriched for antigen presentation and metabolically competent lymphocytes. This first spatial transcriptomic study in e-DMPeM delineates how tumor, immune, and stromal niches jointly shape prognosis and paves the way for future investigations into niche-specific crosstalk.

**Introduzione e Obiettivi:** Il mesotelioma peritoneale diffuso maligno di tipo epitelioide è una neoplasia rara e aggressiva del rivestimento mesoteliale addominale, con caratteristiche per diversi aspetti, eterogenee. La combinazione CRS-HIPEC è lo standard terapeutico, ma i determinanti biologici che distinguono le recidive precoci dalle risposte prolungate non sono definiti. Questo studio mira a identificare firme trascrittomiche associate alla recidiva precoce (ER) o tardiva (LR) dopo CRS-HIPEC, e a comprendere come l'organizzazione del tumore e del microambiente contribuisca a questa diversità.

**Materiali e Metodi:** 36 pazienti con e-DMPeM trattati con CRS-HIPEC sono stati suddivisi in ER (progressione ≤12 mesi) e LR (sopravvivenza libera da progressione ≥36 mesi). Il profilo trascrizionale bulk ottenuto tramite RNA-seq con Illumina, è stato valutato con le seguenti analisi: DEA, GSEA, IPA, deconvoluzione immunitaria e analisi di firme CAF. *GAS1* è risultato come principale gene differenzialmente espresso. È stato validato tramite immunoistochimica e ha permesso un'ulteriore stratificazione in *GAS1*-high/low all'interno dei gruppi ER e LR. In un sottogruppo di 22 pazienti, la trascrittomica spaziale (GeoMx DSP) è stata utilizzata per valutare l'eterogeneità tissutale in quattro compartimenti (tumore, nicchie macrofagiche, fibroblasti, e strutture linfoidi), consentendo un'analisi specifica per tipo cellulare.

**Risultati e Conclusioni:** L'RNA-seq bulk ha evidenziato una netta differenza molecolare tra ER e LR: i tumori ER mostrano vie proliferative, metaboliche e di rimodellamento della matrice extracellulare in un microambiente immunosoppresso e ricco di fibroblasti attivati, mentre i tumori LR presentano un profilo quiescente, con soppressione dei segnali oncogenici. *GAS1* emerge come determinante di uno stato tumorale meno aggressivo, associato a minore proliferazione, inibizione metabolica e ridotta segnalazione angiogenica. Dall'analisi di trascrittomica spaziale i pazienti ER mostrano firme pro-tumorigeniche coordinate tra cellule $CK^+$ e macrofagi $CD68^+$, mentre i LR conservano nicchie linfoidi funzionalmente attive, arricchite per presentazione dell'antigene e linfociti metabolicamente competenti. Questo lavoro ha permesso una prima applicazione della trascrittomica spaziale al mesotelioma peritoneale diffuso maligno. I nostri dati forniscono una base per future indagini sul crosstalk tra cellule tumorali e TME, facilitando la comprensione della complessa eterogeneità biologica della malattia.

# Table of contents

# 1. Introduction

## A Heterogeneous Disease: Diffuse Epithelioid Peritoneal Mesothelioma

Mesothelioma is an uncommon tumor malignancy arising from the mesothelial lining of serosal cavities, including the pleura, peritoneum, pericardium, and tunica vaginalis of the testis. Globally, it remains a rare cancer, with 30,633 new cases reported in 2022 according to WHO/IARC GLOBOCAN estimates.

The most studied and common type is pleural malignant mesothelioma, which typically presents with chest pain, dyspnoea, and breathlessness (Moore et al., 2008). Among the various types of mesotheliomas, diffuse malignant peritoneal mesothelioma (DMPM) accounts for only a minority of cases (15-20%). It is a biologically distinct and uncommon neoplasm, with an estimated prevalence of 1-9 cases per 100,000 individuals. The mean age at diagnosis of malignant peritoneal mesothelioma (MPeM) is about 69 years. One-year overall survival is approximately 46% in patients with MPeM, and 5-year overall survival is about 20%; cure is very rare (Ullah et al., 2022). Malignant peritoneal mesothelioma is equally distributed between men and women, whereas pleural mesothelioma predominates in men (Greenbaum & Alexander, 2020). Asbestos remains the primary established cause of mesothelioma, though its association with MPeM is weaker, with only a minority of patients reporting clear exposure (Boffetta, 2007; Moolgavkar et al., 2009). A large population-based case-control study conducted in Italy between 2000 and 2021 demonstrated a strong association between asbestos exposure and peritoneal mesothelioma (Consonni et al., 2025). In peritoneal disease, asbestos fibers likely reach the cavity indirectly, via lymphatic spread or ingestion, initiating inflammation, oxidative stress, and genetic damage that drive malignant transformation (Neumann et al., 2013). The prolonged latency between exposure and disease onset suggests that the disease's incidence will continue to rise over time. Additional environmental and iatrogenic factors have been implicated, including exposure to toxins, genetic susceptibility, and conditions such as Hodgkin’s disease, chronic peritonitis, or prior therapeutic radiation (Carbone et al., 2012).

The peritoneum is a serous membrane composed of a single layer of mesothelial cells, of mesodermal origin, with complex apical and basal interactions. It forms a peritoneal sac that envelops the abdominal organs, providing structural support, vascular and lymphatic supply, and a protective barrier against injury and infection. As shown in Figure 1, the visceral

peritoneum covers the intraperitoneal organs, whereas the parietal peritoneum lines the inner surface of the abdominal wall (Bootsma et al., 2023).

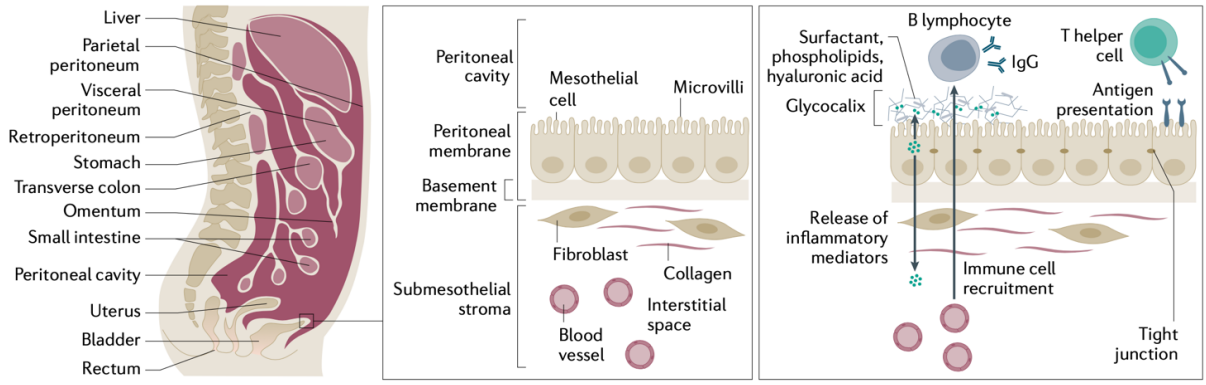


***Figure 1. Peritoneal anatomy and physiology.***

Mesothelial cells are enriched with apical microvilli that facilitate absorption, secretion, and smooth organ mobility, and they rest on a basement membrane supported by a submesothelial stroma composed of fibroblasts, extracellular matrix, blood vessels, and immune cells. Despite the presence of tight junctions, immune cells can access the mesothelial interface, rendering the peritoneum an immunologically active membrane that allows bidirectional exchange of fluids and signaling molecules and shapes local inflammatory responses (Kipps et al., 2013).

Within the tumor microenvironment, mesothelial cells actively interact with tumor and immune cells, contributing to tumor initiation and progression. Accumulating evidence indicates that they can promote an immunosuppressive milieu by secreting cytokines, chemokines, and exosome-derived factors that recruit or modulate immunosuppressive cell populations, such as tumor-associated macrophages, regulatory T cells, and myeloid-derived suppressor cells. Through these interactions, mesothelial cells facilitate immune evasion and support tumor growth and dissemination (Sheng et al., 2025). DMPM usually begins with vague, non-specific, insidious symptoms, and many individuals are asymptomatic early on, leading to a median diagnostic delay of about four months. The most common clinical features were ascites, abdominal pain, asthenia, weight loss, night sweats, fever, anorexia, and, less frequently, an abdominal mass (Manzini et al., 2010). Cancer cell spread from the primary tumor to the peritoneum can occur via different routes: direct seeding from the primary tumor into the peritoneal cavity, via the lymphatic system, via the hematogenous system, or reseeding from existing PM lesions (Bootsma et al., 2023).

As DMPM progresses, small peritoneal nodules may coalesce into larger masses, and tumor infiltration of the greater omentum can lead to diffuse thickening and replacement of normal omental fat by tumor tissue (Stamou et al., 2015). Progressive mesenteric involvement may distort bowel loops and, in some cases, result in obstruction. Tumor dissemination can further extend to the upper abdominal regions, pelvic sidewalls, and retroperitoneal lymph nodes. On computed tomography (CT) imaging, peritoneal lesions usually appear hypodense and may show calcifications; in more localized forms, the disease may present as a heterogeneous solid mass with direct invasion of adjacent organs (Ambrosetti et al., 2025).

Histopathologically, according to the World Health Organization classification, DMPM comprises three major subtypes (Brcic & Kern, 2020), as shown in Figure 2, which are:

(i) Epithelioid mesothelioma (≈70-75%), the most common form, and the one in this study, which is characterized by cohesive epithelioid cells arranged in tubulopapillary, trabecular, or solid patterns, and associated with the most favorable prognosis.

(ii) Sarcomatoid mesothelioma, a rare variant composed of spindle-shaped cells resembling sarcoma, is associated with poor prognosis and limited therapeutic responsiveness.

(iii) Biphasic mesothelioma, containing both epithelioid and sarcomatoid components, with prognosis determined by their relative proportions (Dacic, 2022).

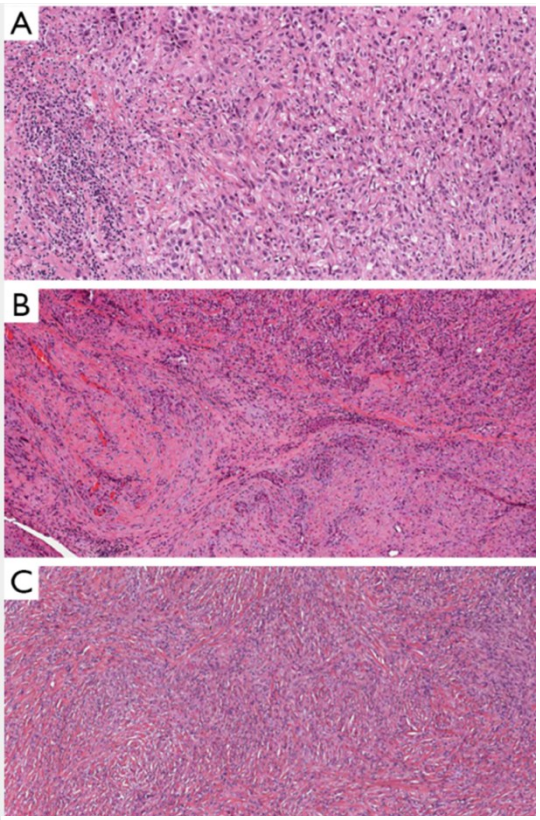


***Figure 2. Histologic presentation of 3 major malignant mesothelioma types; epithelioid (A), biphasic (B) and sarcomatoid (C). (H&E staining; A: objective ×10, B and C: objective ×5).***

Radiologically, DMPM shows three patterns: a wet type with ascites and diffuse peritoneal thickening, typically linked to epithelioid disease; a dry type with large solid masses and minimal ascites, more often associated with sarcomatoid mesothelioma; and a mixed type combining both features (Bonde et al., 2023; Carlson et al., 2022). DMPM is historically

difficult to diagnose, and later recognition showed frequent misclassification with metastatic adenocarcinoma, prompting the development of a diagnostic multimodal approach combining clinical assessment, imaging, surgical exploration, and histopathological confirmation (Winslow & Taylor, 1960). These criteria are currently used in clinical practice:

| | |
|---|---|
| CT (first line) | Detects nodules, thickening, omental caking, ascites, masses; may underestimate small-volume disease; does not limit surgery (Low et al., 2015). |
| MRI | Diffusion-weighted MRI provides better soft-tissue contrast and more accurate PCI estimation (Low et al., 2021). |
| PET/CT | Differentiates benign vs malignant lesions via 18F-FDG uptake; improves lymph node staging and recurrence detection; sensitivity 86%, specificity 89%, accuracy 87% (Dubreuil et al., 2017). |
| Cytology | Not recommended alone (30-75% sensitivity); cannot assess architectural invasion or Ki-67 (Kusamura et al., 2023). |
| Histopathology IHC | Requires ≥ 2 positive mesothelial markers (calretinin, CK5/6, WT1, D2-40) and ≥ 2 negative carcinoma markers (claudin 4, CEA, TTF-1, PAX8); Ki-67 and mitotic rate assessed, where Ki-67 represents the percentage of tumor cells actively proliferating (Husain et al., 2024). |

This comprehensive assessment is essential not only for establishing the diagnosis but also for selecting appropriate candidates for cytoreductive surgery, which represents the current standard of care for this disease.

## Cytoreductive Surgery combined with Perioperative Intraperitoneal Chemotherapy as Standard of Care

Systemic therapies for malignant peritoneal mesothelioma have historically demonstrated limited efficacy, with no significant improvements in either overall survival or disease-free survival. In most retrospective series, median survival under conventional treatment rarely exceeds one year. The introduction of cytoreductive surgery (CRS) combined with perioperative intraperitoneal chemotherapy, most effectively delivered as heated intraoperative intraperitoneal chemotherapy (HIPEC), has markedly altered this prognosis (Stamou et al., 2015). CRS-HIPEC has emerged as the preferred initial treatment in selected DMPM patients. Appropriate patient selection is the most critical determinant for treatment response. The Peritoneal Cancer Index (PCI) is the most accurate tool to quantify disease extent: low PCI values support curative-intent surgery, whereas high PCI values indicate that only palliative debulking is appropriate (Figure 3). The Completeness of Cytoreduction (CC-score) is the strongest prognostic factor: CC-0 or CC-1 (no visible disease or nodules <2.5 mm) favorably denotes complete cytoreduction, whereas CC-2/3 indicates unresectable disease or partial cytoreduction with poorer outcome.

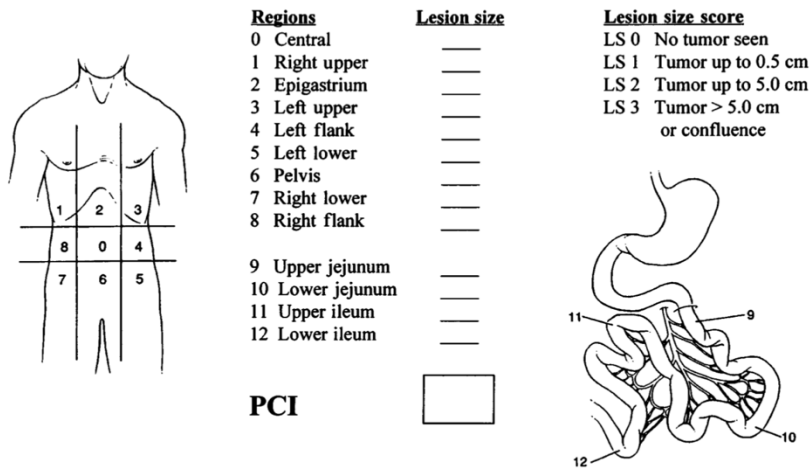


***Figure 3. The PCI is calculated by assigning a lesion size score (LS 0-3) to each of the 13 abdominopelvic regions, based on the largest tumor nodule: LS-0 (no disease), LS-1 (<0.5 cm), LS-2 (0.5-5 cm), LS-3 (>5 cm or confluent). The PCI is the sum of all LS scores, with a maximum of 39.***

Guidelines recommend CRS-HIPEC only for patients in good clinical condition, without extra-peritoneal metastases, and with disease amenable to CC-0/1. Absolute contraindications include sarcomatoid histology, massive, small-bowel serosal involvement, pleural disease, and retroperitoneal or cardiophrenic lymph-node metastases. Relative contraindications include biphasic histology, a Ki-67 index >9%, PCI >17, extensive mesenteric or diaphragmatic involvement, and any situation in which complete cytoreduction is unlikely; in these cases, the

physician will take the surgical choice (Kusamura et al., 2021; Sun et al., 2023). CRS-HIPEC carries rates of severe complications that range from 30% to 41% and rates of post-operative mortality ranging from 0% to 6% (Bootsma et al., 2023).

Once the cytoreductive surgery has been completed and all visible disease has been removed, the most effective HIPEC regimen is the platinum-based combination of HIPEC with high-dose chemotherapeutic agents, circulating in all the regions of abdominal and pelvic cavities, under sustained hyperthermia that directly damages cancer cells, enhances drug penetration into tissues, and increases the cytotoxic effect of the drug, a synergy on peritoneal surface. As shown in Figure 4, catheters and temperature probes are positioned, and the abdomen is temporarily closed using the "coliseum technique," which creates a heated, sterile chamber for manual manipulation during perfusion. Chemotherapy is circulated for about 90 minutes, with inflow temperatures of 44-46°C to maintain an intraperitoneal temperature of 42-43°C. Continuous manual movement of the viscera ensures uniform exposure of all peritoneal surfaces (Sugarbaker, 1999). Although CRS-HIPEC has significantly improved outcomes, its benefits are not consistently reproducible across all patients. This variability underscores the need to identify the factors that influence treatment efficacy. In this context, transcriptomic technologies have become essential tools for understanding the biological determinants that may underlie therapeutic variability.

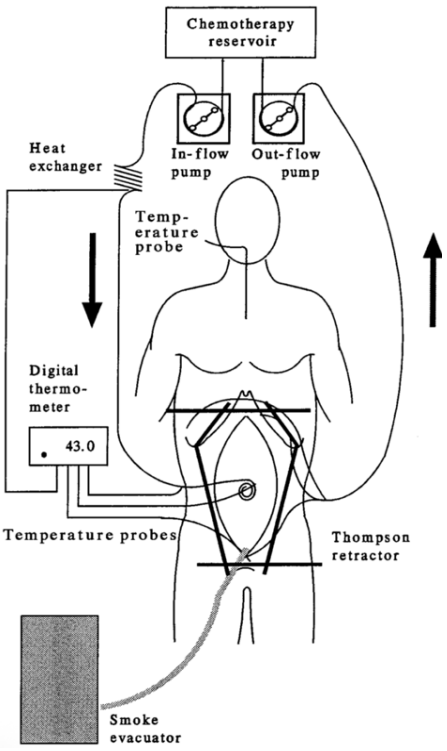


***Figure 4. Apparatus required for heated intraoperative intraperitoneal chemotherapy.***

## From RNAseq bulk to Spatial Transcriptomic: A New Framework for Understanding Mesothelioma

Beyond genetic mutations affecting oncogenes and tumor suppressor genes, cancer cells frequently exhibit aberrant transcriptional activation or epigenetic silencing of key regulatory pathways, promoting uncontrolled proliferation, metastasis, and therapeutic resistance (Han et al., 2018). Advances in molecular biology technologies, including highly sensitive detection, amplification methods, and the emergence of omics approaches, have substantially expanded our understanding of RNA function across diverse cellular contexts. Transcriptomic technologies have evolved through successive phases, each overcoming key limitations of the previous approaches (Hong et al., 2020). A major technological shift occurred in the mid-2000s with the advent of next-generation sequencing (NGS), which enabled massively parallel sequencing of millions of nucleic acid fragments without prior sequence knowledge.

Illumina short-read sequencing currently represents one of the most widely used technologies. After extraction and purification of RNA, enrichment of targeted RNA and library preparation, cDNA fragments are bound to a flow cell and undergo bridge amplification to form clonal clusters (Goodwin et al., 2016). In this case, sequencing is performed through sequencing-by-synthesis (SBS), in which single-stranded templates are read as the complementary strand is generated. In each cycle, a single fluorescently labelled and 3′-blocked deoxy nucleoside triphosphate (dNTP) is incorporated into the growing DNA strand (Figure 5).

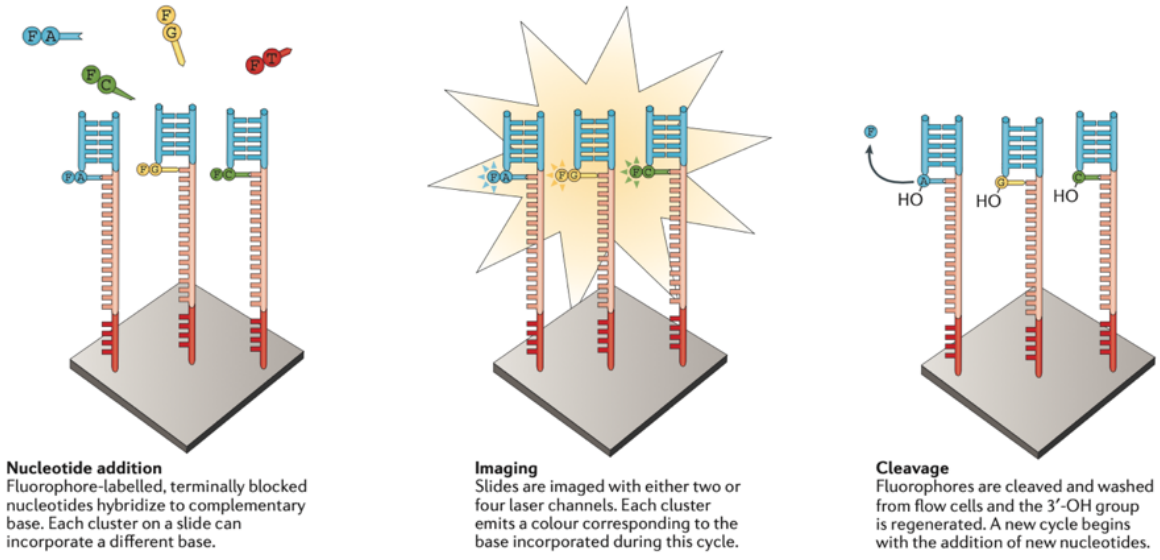


***Figure 5. Sequencing by synthesis: cyclic reversible termination approach with Illumina.***

The fluorescent label acts as a reversible terminator, preventing the addition of more than one nucleotide at a time. After imaging the incorporated base, the fluorescent label is enzymatically

cleaved, allowing the next dNTP to bind. Base calls are inferred directly from the measured fluorescent signal intensity.

The sequencing output consists of millions of short reads that require dedicated bioinformatic processing, the focus of this project. This typically includes quality assessment, trimming of low-quality bases and adapter sequences, and alignment or quasi-mapping to a reference genome or transcriptome. Subsequent analyses depend on the goals of the study and may involve quantification of gene or transcript abundance, differential expression analysis, variant detection, or functional interpretation of the resulting gene expression profiles.

This workflow produces reads of 50-500 base pairs, providing highly accurate quantification of gene expression and enabling robust transcriptome profiling even in degraded RNA samples. Short-read sequencing has limited ability to resolve complex transcript isoforms, since over 90% of human genes undergo alternative splicing, and shared exons can cause ambiguous read mapping. Junction-spanning reads improve discrimination, but regions common to multiple transcripts remain difficult to resolve. Despite this limitation, Illumina short-read sequencing remains the foundation of most RNA-seq applications, including bulk, single-cell, spatial, and small-RNA profiling.

The progressive refinement of transcriptomic technologies has expanded the ability to characterize gene expression heterogeneity across tissues and disease. In this context, genomic and transcriptomic studies have reshaped the understanding of peritoneal mesothelioma biology. Large genomic studies have delineated the molecular landscape of pleural and peritoneal mesothelioma, revealing recurrent alterations in several key tumor-suppressor genes (Hiltbrunner et al., 2022). For example, *BAP1* is mutated in approximately 47.9% of peritoneal mesotheliomas; it encodes a deubiquitinase involved in gene transcription, DNA-damage repair, and cell-cycle control, and its loss represents one of the defining molecular features of this disease. Additional recurrent alterations include *NF2* (26.5%), a regulator of Hippo signaling; *CDKN2A* (25.9%) and *CDKN2B* (19.5%), both central to G1/S checkpoint control; and *SETD2*, a histone methyltransferase mutated in roughly 10% of both pleural and peritoneal tumors. Moreover, alterations in *PBRM1* (present in 15.8% of peritoneal mesotheliomas) are particularly noteworthy, as they have been associated with a less immunogenic tumor microenvironment. Some of these alterations also carry potential therapeutic relevance: *PBRM1*-deficient cells show increased sensitivity to PARP and ATR inhibition, while *SETD2* loss has been associated with susceptibility to WEE1 inhibitors currently under clinical

evaluation (Möhrmann et al., 2025). In addition, recurrent alterations affecting genes involved in RNA metabolism and cellular stress responses, including *DDX3X* (Joseph et al., 2017), have been identified in a subset of cases. Moreover, preclinical evidence supports the therapeutic potential of inhibitors of epigenetic modifiers, including histone deacetylases and *EZH2* (Hong et al., 2020).

RNA sequencing-based analyses have further clarified the functional consequences of genomic alterations in mesothelioma. In a large integrated study combining whole-exome sequencing and transcriptomic profiling of pleural and peritoneal mesothelioma, researchers identified a 48-gene prognostic signature, enriched in cell-cycle regulation and DNA repair pathways, which robustly predicted overall survival across independent cohorts. Notably, *CCNB1* emerged as the most informative single-gene marker (the protein encoded by this gene is a regulatory protein involved in mitosis), underscoring a proliferative transcriptional state as a key determinant of aggressive disease. In addition, transcriptomic data were used to infer the tumor immune microenvironment and to support RNA-based prediction of response to immunotherapy and chemotherapy, highlighting the clinical utility of RNA-seq beyond genomic profiling alone (Nair et al., 2023).

A recent *in-silico* systems-level interactome analysis has expanded the protein-protein interaction landscape of malignant peritoneal mesothelioma by integrating known interactions with 417 newly predicted protein-protein interactions. This network-based approach suggested the complex molecular connectivity underlying the disease and putative functional modules involved in transcriptional regulation, cell cycle control, cellular stress responses, and cytokine signaling (Karunakaran & Ganapathiraju, 2024).

While transformative, bulk RNA-seq is limited by its inability to resolve cellular heterogeneity. By providing an averaged expression profile across diverse cell populations, it obscures the distinct cellular architecture of solid tumors and precludes the analysis of specific cell-cell interactions. To overcome this limitation, single-cell RNA sequencing (scRNA-seq) was introduced in 2009, enabling transcriptomic profiling at single-cell resolution and revealing rare cell populations, dynamic cellular states, and lineage relationships. However, scRNA-seq requires enzymatic or mechanical dissociation of tissue, a process that disrupts native architecture and eliminates the spatial information fundamental for a deep understanding of the tumor *milieu*. The need to preserve tissue organization drove the development of spatial transcriptomic technologies, first reported in 2016.

Since then, spatial transcriptomics technologies have evolved into two main groups. In the first, imaging-based platforms use in situ hybridization and sequential imaging to detect RNA with single-cell or even subcellular resolution. Among the most widely used, Xenium, Merscope, and CosMx SMI offer exceptional spatial precision but rely on predefined gene panels.

In contrast, sequencing-based platforms capture mRNA with spatially barcoded oligos and reconstruct transcript location through next-generation sequencing (NGS). While 10x Visium offers whole-transcriptome profiling at a 55 µm spot-level resolution, the next generation of platforms significantly refines this scale: Visium HD enhances resolution to 2 µm, and Stereo-seq achieves a sub-cellular 0.22 µm resolution using DNA nanoballs, currently the highest spatial precision among NGS-based methods (Lim et al., 2025).

Amidst these advancements, the GeoMx Digital Spatial Profiler (DSP) remains a unique high-plex, ROI-based (region of interest) strategy. Introduced in 2019, it combines morphological imaging with UV-cleavable probes to quantify transcripts within anatomically defined regions. It represents a hybrid approach: while it does not offer single-cell resolution like imaging platforms, it provides whole-transcriptome data anchored to the tissue's native architecture by allowing for targeted ROI selection and segmentation (Merritt et al., 2020). Briefly, DSP probes hybridize to RNA targets in fresh or FFPE tissue sections mounted on slides, which are simultaneously stained with fluorescent markers to define tissue morphology and guide the selection of ROIs. ROIs can be segmented into compartments

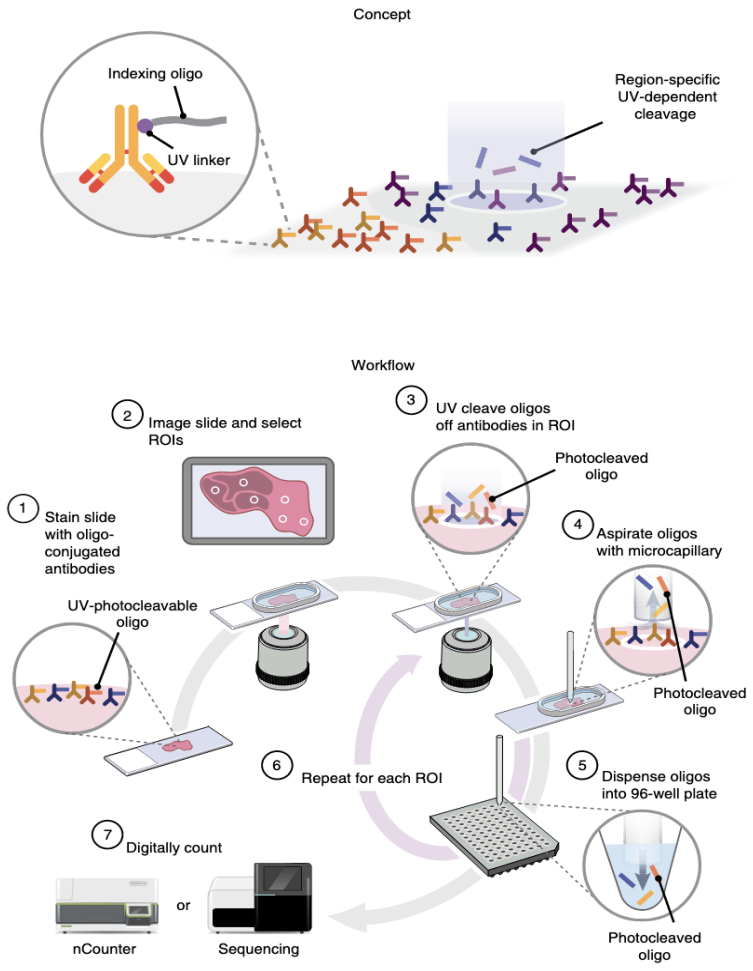


***Figure 6: Schematic representation of the GeoMx DSP methodology and operational workflow.***

(e.g., tumor vs stroma or specific immune cell populations) for targeted analysis. Selected ROIs are illuminated with UV light using a digital micromirror device (DMD), releasing the oligonucleotide barcodes. These are collected via microcapillary aspiration, hybridized to fluorescent barcodes, and quantified using either the nCounter system or standard NGS workflows. Data are processed through quality control, normalization, and visualization (Hernandez et al., 2022). The whole experimental workflow is shown in Figure 6.

# 2. Rationale and aims of the study

Evolution from bulk to spatial informative analyses is important for MPeM, where complex tumor heterogeneity demands spatial approaches. Consequently, recent omics studies reshaped our understanding of mesothelioma biology.

In recent years, spatial transcriptomics has begun to transform the study of malignant pleural mesothelioma, revealing layers of biological complexity previously inaccessible to bulk RNA sequencing. The GeoMx DSP was applied to 139 morphologically defined regions of biphasic malignant pleural mesothelioma, enabling precise segmentation of epithelioid, sarcomatoid, and transitional areas directly on FFPE tissue. This approach revealed that the epithelial-to-sarcomatoid (E→S) transition occurs as a spatially continuous process rather than a discrete histological shift. Spatial transcriptomics identified early activation of epithelial to mesenchymal (EMT) transition-related transcription factors, including *SNAI2* and *ZEB1*, accompanied by extensive extracellular matrix remodeling and fibroblast activation. EMT is characterized by the loss of epithelial traits, including polarity and cell-cell adhesion, and the acquisition of mesenchymal properties that enhance motility and invasiveness. In parallel, spatial mapping uncovered a highly inflamed yet immunosuppressed microenvironment in sarcomatoid regions, enriched for M2-polarized macrophages, neutrophils, plasmacytoid dendritic cells, and exhausted T-cell signatures with upregulated immune checkpoints (Torricelli et al., 2023). Despite its impact on pleural mesothelioma, spatial transcriptomics has never been applied to DMPM. Addressing this unmet need requires comprehensive whole-transcriptome analyses, including both bulk and spatial approaches, capable of capturing the tumor microenvironment complexity.

Moreover, although CRS-HIPEC has significantly revolutionized the e-DMPeM treatment, some crucial aspects remain to be solved: i) a substantial number of patients encounter an early relapse after therapy (Gilani et al., 2018) ii) biological characteristics of DMPeMs, helpful for CRS-HIPEC treatment, are limited to PCI and Ki67 indicators, iii) no effective treatments can be offered to pts who are not eligible for CRS-HIPEC or with metastatic disease (Shao et al., 2020). Therefore, it remains essential to understand which aspects of tumor biology are associated with early recurrence and which, on the other hand, support a more lasting clinical response after CRS-HIPEC. This task is further complicated by the peculiar biology of epithelioid-DMPeM cells, which display a highly plastic, intermediate epithelial-mesenchymal

phenotype, as previously demonstrated by researchers at Fondazione IRCCS Istituto Nazionale Tumori of Milan.

Progression toward a more mesenchymal state (progressed e-DMPeM) is marked by i) increased *EZH2* expression (Bozzi et al., 2016); ii) development of an immunosuppressive tissue microenvironment (Tazzari et al., 2018); iii) acquisition of mutations in the SEMA domain of the EMT (*c-Met*) proto-oncogene receptor (Belfiore et al., 2019). Together, these alterations reflect a progressive mesenchymal drift that may profoundly influence treatment response, suggesting that epithelial-mesenchymal dynamics are central to understanding e-DMPeM.

**The main purposes of the present study are:**

- To understand whether **distinct transcriptional signatures differentiate patients with early disease recurrence from those with markedly longer outcomes.** This was addressed through a comprehensive bioinformatic analysis of bulk RNA-seq data generated on an Illumina platform from e-DMPeM samples, including differential gene expression, enrichment of gene sets, and functional analyses.
- To evaluate whether a **spatial perspective** on the tumor microenvironment may influence tumor behavior and contribute to explaining early relapse or prolonged response. This was explored using GeoMx Digital Spatial Profiling

This experimental thesis was conducted at the **Fondazione IRCCS Istituto Nazionale Tumori of Milan**, within the Department of Pathology 2 and the Bioinformatics Unit, in collaboration with the Colon and Rectal Surgery Unit and the Department of Diagnostic Innovation. The initial concept of the experimental design for this work is illustrated in Figure 7.

# Experimental Design

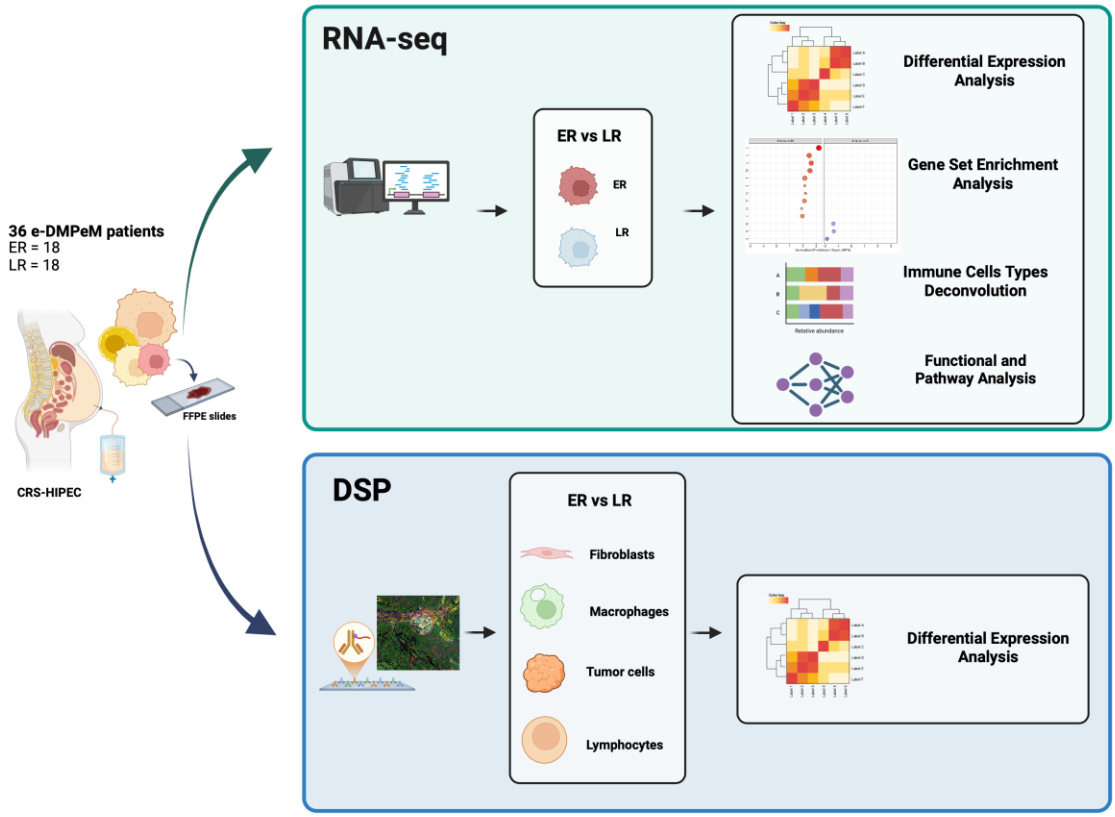


***Figure 7. Schematic representation of the initial study design, illustrating bulk RNA-seq and spatial transcriptomic profiling used to dissect molecular and microenvironmental differences between early-relapsing (ER) and long-responder patients (LR).***

# 3. Methods and Materials

## 3.1 Patient's population, Ethical Approval, and Clinical-Pathological Metadata

For this study, **36 primary epithelioid malignant peritoneal mesothelioma** patients were considered and collected during cytoreductive surgery and before HIPEC **between 2012 and 2022**. They were stratified according to **recurrence patterns.**

In detail, the stratification included two groups:

**Early recurrence** (ER): progression ≤12 months after CRS-HIPEC.

**Long responders** (LR): progression-free survival ≥36 months with complete follow-up.

Clinical and pathological variables were collected from electronic medical records and from the prospectively maintained **database of the Peritoneal Surface Malignancy Unit.**

The study was approved by the **local ethics committee (CET Lombardia 4, INT 93/24)** and conducted in accordance with the **Declaration of Helsinki** and current regulations regarding data privacy and security.

## 3.2 RNA-seq Experimental Procedures

### RNA Extraction

Total RNA was extracted from 3-4 sections 8 μM thick cut from Formalin Fixed Paraffin Embedded (FFPE) blocks. Briefly, sections were de-waxed and digested using proteinase k provided with the RSC RNA FFPE Extraction kit (Promega Corporation, Madison, WI, USA) following the manufacturer's instructions. RNA obtained was purified and collected through a semi-automated extractor (Maxwell®, Promega Corporation, Madison, WI, USA). To proceed with the RNA-seq library, total RNA was quantified at the Qubit fluorometer (Invitrogen, Waltham, MA, USA) and integrity (DV200) was measured through the Tape Station 4200c (Agilent Technologies, Santa Clara, CA, USA). All samples met the quality criteria for RNA-seq library generation.

### RNA-seq Libraries Preparation and Sequencing

RNA-seq libraries were generated from 100 ng of total RNA using the Illumina RNA Prep with (L) Tagmentation kit of Illumina (Illumina, San Diego, CA, USA) following the manufacturer's instructions. Briefly, after RNA denaturation and double-strand cDNA synthesis, the resulting cDNA fragments were tagmented and purified. Libraries normalization was performed using 200 ng of total mass in 7,5 µl of resuspension buffer. Probes hybridization and capturing were performed on normalized libraries with subsequent amplification. Finally, libraries were checked for quantity at the Qubit fluorometer (Invitrogen, Waltham, MA, USA), and fragment length was measured through the Tape Station 4200 (Agilent Technologies, Santa Clara, CA, USA). Sequencing was performed to collect at least 25,0mln reads for each sample (2x75nt read length) through the **Illumina NovaSeq6000Dx platform** (Illumina, San Diego, CA, USA) following manufacturer's recommendations.

## 3.3 Statistical Testing of Clinical Variables

Clinical metadata were imported and processed in **R (version 4.5.2)**. The analysis began with a comparison of clinical and pathological variables between ER and LR patients, to characterize the two groups and **identify potential differences**.

Continuous variables (tumor cell score, tumor infiltrating lymphocytes, Ki67 index, age) were cleaned, converted to numeric format, and checked for missing values, while categorical variables (sex, pre-operative chemotherapy, completeness of cytoreduction) were recoded as factors. To obtain an initial descriptive overview, **continuous** variables were visualized through **boxplots** and **categorical** variables through **barplots**, allowing a simple comparison of distributions and group proportions. Data wrangling and visualization were performed using the `tidyverse` system, including `ggplot2`.

To evaluate statistical significance, appropriate tests were applied according to the type and distribution of each variable. Normality of continuous variables was evaluated within each DFS (Disease Free Survival) class with **Shapiro-Wilk test**, and variables were compared using either **t-tests** (for approximately normal distributions) or **Wilcoxon rank-sum tests** (for non-normal distributions). Categorical variables were analyzed using **Fisher's Exact test**.

## 3.4 Survival Analysis and Prognostic Factors

### Cox Regression Modelling

Survival analysis was conducted to determine whether **differences in PCI and Ki67** were linked to a **higher risk of disease progression**. The goal was to identify clinically relevant variables with **potential independent prognostic significance**. `Survival` and `survminer` packages were used. Metadata were organized by standardizing column names and converting relevant variables into proper numerical formats; Ki67 values reported as "<5%" were recoded to 2.5% to allow for quantitative analysis, and PCI was converted to a numeric format. Time-To-Progression (TTP) was defined as the period from CRS-HIPEC to clinical progression. The event indicator coded progression as 1 and censoring as 0. The **Cox proportional hazards model** was chosen because it is specifically designed for time-to-event data, handles variable follow-up times, and properly manages censored observations. Cox models were estimated using the `coxph()` function, with the survival object defined through `Surv(TTP, event)`. The `Surv()` function sets up the time-to-event data, while `coxph()` fits the proportional hazards model using partial likelihood, providing regression coefficients, hazard ratios, 95% confidence intervals, and p-values. Univariate Cox models were first used to evaluate the individual association of Ki67 and PCI with TTP. Then, a multivariable Cox model including both variables was built to assess their independent prognostic effects. The proportional hazards assumption was checked with Schoenfeld residuals using `cox.zph()`.

### Log-rank Testing

**Kaplan-Meier curves** were generated using the `survfit()` function to **explore differences in progression-free survival** (PFS) **and overall survival** (OS) between ER and LR patients. PFS was defined as the interval between treatment initiation and the first documented progression; because ER and LR groups were derived from progression timing, PFS comparisons were considered descriptive and exploratory. OS was defined as the time from treatment initiation to death attributable to disease, with events identified through the final status variable and censoring applied according to follow-up in months. Group comparisons were performed using the **Log-rank test** (`survdiff()`), which compares event patterns over the entire follow-up period.

## 3.5 RNA-seq Pipeline

The computational analyses were conducted on a **High-Performance Computing platform** (HPC) at the **Fondazione IRCCS Istituto Nazionale dei Tumori**. Due to the size and complexity of the RNA-seq dataset, resources exceeding those of a typical workstation were necessary; the HPC environment provided the essential CPU, memory, and storage capacity to run the entire workflow smoothly (Reed et al., 2022). Access to the cluster was established via a **command-line interface (CLI)** using **the Bash shell** and **SSH connection**. All workflow steps were executed through **Unix shell scripts** and submitted to the **SLURM** (Simple Linux Utility for Resource Management) **workload manager**. Briefly, any SLURM script begins with a series of `#SBATCH` directives, which define the computational resources requested from the SLURM scheduler. The first line (`#!/bin/bash`) specifies that the script should be executed using the Bash shell. The `--job-name` option assigns a recognizable name to the job, while `--account` specifies the user account. The directives `--mem`, `--nodes`, and `--cpus-per-task` define the requested computational resources, and `--partition` selects the queue in which the job will run. The `--error` and `--output` flags indicate where SLURM should write log files. Finally, `--mail-user` and `--mail-type` enable e-mail notifications for job status updates.

```
#!/bin/bash

#SBATCH --job-name=bcl2fastq
#SBATCH --account=brognaragiorgia
#SBATCH --mem=64GB
#SBATCH --nodes=1
#SBATCH --cpus-per-task=8
#SBATCH --partition=ricerca
#SBATCH --error="/share/home/brognaragiorgia/AutomatedPipeline.Gio.err"
#SBATCH --output="/share/home/brognaragiorgia/AutomatedPipeline.Gio.out"
#SBATCH --mail-user=giorgia.brognara@istitutotumori.mi.it
#SBATCH --mail-type=ALL
```

Then as in all scripts, the appropriate conda environment is activated at the start and deactivated at the end. Custom scripts for data processing, quality control, trimming, alignment, and read counting were created directly on the cluster, using nano text editor, made executable with `chmod +x`, and submitted with `sbatch`. Job dependencies were handled through SLURM's `--dependency=afterok:<jobID>` option to ensure proper execution order.

## Step 1 – Conversion from BCL to FASTQ using Bcl2fastq

Illumina sequencing produces raw **base call files** (BCL), which must be converted into **FASTQ format** and **demultiplexed** before downstream RNA-seq processing. This step assigns each read to the correct sample based on the index sequences specified in the **SampleSheet.csv created with the Illumina BaseSpace** cloud service. Conversion and demultiplexing were performed using bcl2fastq (Illumina Inc., 2017). As it is shown in the SLURM code block below, the environment containing bcl2fastq was loaded before running the conversion and closed upon completion. The script creates the output directory for the FASTQ files and then runs bcl2fastq, specifying the run folder, the SampleSheet used for demultiplexing, and the output path. The `--no-lane-splitting` option was used to merge reads from different lanes of the same sample into a single FASTQ file per read direction. The script ends by deactivating the environment and printing a completion message.

```
echo "activating conda env"
source /share/data/apps/anaconda3/bin/activate tools   # Activate environment with bcl2fastq

mkdir -p /share/home/brognaragiorgia/scriptgio/fastq     # Create output directory

echo "starting bcl to fastq conversion"
bcl2fastq --runfolder-dir /share/data/ricerca/research_runs/250320_ADX00148_0045_AHLHG3DRX5 \
         # Input BCL run folder

        --sample-sheet /share/data/ricerca/research_runs/250320_ADX00148_0045_AHLHG3DRX5/Samplesheet.csv \
        # Sample sheet

        --output-dir /share/home/brognaragiorgia/scriptgio/fastq \ # Output FASTQ directory

        --no-lane-splitting # Prevent Lane splitting

conda deactivate                                          # Deactivate environment
echo "bcl2fastq completed successfully”
```

## Step 2 – FASTQ Quality Check using FastQC

Following FASTQ generation, a **comprehensive quality control** (QC) assessment was performed to evaluate the overall quality of the sequencing reads before downstream processing. QC was carried out using FastQC, a quality control tool for high-throughput sequence data (Leggett et al., 2013), and the resulting reports were aggregated using MultiQC, an open-source tool to aggregate bioinformatic analysis results, to do a comparative evaluation across samples (Ewels et al., 2016). A dedicated output directory was created (`mkdir -p qc_results`), and FastQC was executed on all FASTQ files in multithreaded mode (`fastqc -t 4 -o qc_results fastq/*.fastq.gz`). The individual reports were then consolidated and preliminary checked using MultiQC, as detailed below. This step ensured that all libraries met the expected quality standards before proceeding to trimming.

## Step 3 – Trimming with Trimmomatic

Raw paired-end FASTQ reads were trimmed using Trimmomatic (a flexible trimmer for Illumina sequence data (Bolger et al., 2014) to **remove** residual **adapter** sequences and low-quality bases, particularly at the read ends. A Bash loop was used to automatically pair each R1 file with its corresponding R2 mate, verify the files presence, and generate the trimmed outputs. For each sample, Trimmomatic was executed in **paired-end mode** (`trimmomatic PE -threads 16`) using the Nextera adapter file (`ILLUMINACLIP:NexteraPE-PE.fa:2:30:10`) and discarding reads shorter than 35 bp (MINLEN:35). This approach is optimized for paired-end data and improves adapter detection. Only paired reads that survived filtering were retained, while unpaired reads were discarded. This ensured that downstream alignment and quantification were performed exclusively on high-quality, adapter-free read pairs.

## Step 4 – Alignment with STAR

Sequencing reads were aligned to the human reference genome using STAR (Spliced Transcripts Alignment to a Reference), a splice-aware aligner optimized for RNA-seq data (Dobin et al., 2013). STAR performs alignment using a seed-based strategy, identifying maximal exact matches along each read and then clustering them by genomic position to reconstruct full spliced alignments. Splice-junction detection was guided by the Ensembl **GRCh38.115 GTF annotation file**, which provides genomic coordinates and hierarchical relationships for genes, transcripts, exons, coding regions, and untranslated regions. Before

alignment, a STAR genome index was generated using the reference FASTA and GTF files. The **GRCh38 genome** and the annotation file were downloaded from the Ensembl database. The index was generated using the following command:

```
STAR --runThreadN 16 \                              # Use 16 threads
     --runMode genomeGenerate \                     # Indexing mode
     --genomeDir "$GENOME_INDEX_DIR" \              # Output directory
     --genomeFastaFiles "$GENOME_FA" \              # Input FASTA
     --sjdbGTFfile "$GTF_FILE" \                    # Input GTF
     --sjdbOverhang 99                              # Read length - 1
```

The `--sjdbOverhang` parameter (read length minus one) was set to 99 to optimize splice-junction detection for 100 bp paired-end reads. Trimmed paired-end reads (100 bp) were then aligned to the GRCh38 genome using STAR. Alignments were produced in coordinate-sorted BAM format, and STAR was instructed to generate gene-level read counts during alignment.

```
    STAR --runThreadN 16 \                          # Use 16 CPU threads
     --genomeDir "$GENOME_INDEX_DIR" \              # Path to the STAR genome index directory
     --readFilesIn "$R1" "$R2" \                    # Paired-end input FASTQ files
     --readFilesCommand zcat \                      # Decompress .gz FASTQ files on the fly
     --sjdbGTFfile "$GTF_FILE" \                    # Gene annotation file (GTF)
     --outSAMtype BAM SortedByCoordinate \          # Output alignment in BAM format
     --quantMode GeneCounts \                       # Generate gene-level read counts
     --outReadsUnmapped None \                      # Do not save unmapped reads
     --outFileNamePrefix "${OUTPUT_DIR}/${base}_"   # Prefix for all output files
```

## Step 5 – Counting with HTSeq

STAR configuration produced compressed, indexed BAM files suitable for downstream analyses. Gene-level quantification was performed using HTSeq-count, which assigns **aligned reads to annotated genomic features** to generate a gene-by-sample count matrix, by evaluating whether each read overlaps an exon, whether it is uniquely mapped, whether its strand orientation matches the library protocol, and whether it can be unambiguously attributed to a single gene (Anders et al., 2015). Because HTSeq requires BAM files sorted by read name, the coordinate-sorted BAM files produced by STAR were first converted using `samtools sort -n` (H. Li et al., 2009)

For each sample, HTSeq-count was executed with this set-up:

```
 "$HTSEQ_COUNT" \                                          # Run HTSeq-count
       -f bam \                                              # Input format is BAM
       -r name \                                             # Reads are sorted by name
       -s reverse \                                          # Library is reverse-stranded
       -t exon \                                             # Count only exon features
       -i gene_id \                                          # Use gene_id attribute from GTF
       "$BAM_SORTED" "$GTF_FILE" > "$COUNT_FILE"        # Generate count table for this sample
rm "$BAM_SORTED"                                        # Remove temporary name-sorted BAM
```

Read counting was performed with HTSeq-count using the Ensembl GRCh38.115 GTF in **reverse-stranded mode**, reflecting the strand-specific library preparation. Reads were counted at the exon level, grouped by gene ID, using the **default union overlap mode** to maximize read retention while limiting ambiguous assignments. HTSeq-count produced one count table per sample plus summary categories (no feature, ambiguous, alignment not unique, low quality, and not aligned), providing an overview of assignment efficiency and supporting the assessment of quantification reliability.

## Step 6 – Final MultiQC Report

A global quality summary was finally generated by re-running the MultiQC v1.33 version (`multiqc qc_results -o qc_results`), which aggregates outputs from multiple bioinformatics tools into a single interactive report. A dedicated SLURM script was used to run MultiQC. The tool was executed on the directories containing the results of the previous steps, including raw and trimmed FastQC reports, STAR alignment logs, and HTSeq-count output files.

## 3.6 RNAseq Downstream and Integrative Analysis

### RNA-seq Counts exploration Analysis

After generating 34 gene-level count files on the HPC cluster, they were downloaded and imported into RStudio. Sample identifiers were extracted from filenames to build a unified annotation table, which was then merged with the previously imported metadata through a standardized sample-ID field. Quality control metrics from the MultiQC report were also imported, cleaned, and aggregated by biological sample before being integrated into the main annotation table. The **clinical variable of interest** (**DFS class**) was recorded into a two-level factor ("Early Recurrence" vs "Long Responders") and used as the condition for downstream differential expression analysis. A **DESeqDataSet objec**t was created using the `DESeqDataSetFromHTSeqCount()` function, specifying the metadata table, the directory containing the raw HTSeq-count files, and the design formula (~ condition), indicating that gene expression was modelled as a function of the DFS class. Exploratory analyses of raw counts were performed to inspect gene-level expression distributions (histograms, boxplots, density plots), guiding the filtering strategy. Low-count genes were removed by retaining only those with at least 10 counts in at least 20% of samples, a threshold chosen to eliminate genes with insufficient or unreliable expression while preserving those with meaningful biological signals. The filtered dataset was saved and re-evaluated to confirm the effectiveness of the filtering step before normalization and differential expression analysis.

### Principal Component and Surrogate Variable Analysis

Principal Component Analysis (PCA) was conducted on the transformed expression matrix using the `plotPCA(ntd, returnData = TRUE)` function. The main aims were **to identify major sources of transcriptomic variability**, assess potential clustering by recurrence pattern, and detect possible confounding factors. PCA identifies orthogonal directions capturing the largest sources of variance in the dataset, allowing high-dimensional gene expression profiles to be visualized in a reduced space (Jolliffe & Cadima, 2016). Samples were plotted in the PC1-PC2 space and annotated according to the condition of interest. Additional PCA plots were generated by coloring samples according to continuous variables (tumor cell score, TILs, Ki67, age) or categorical variables (sex, pre-operative chemotherapy, cytoreduction completeness, clinical status), using gradient or discrete color palettes. All plots were generated using `ggplot2` and `RColorBrewer`.

**Heterogeneity and latent variables are well-known sources of unwanted variability in high-throughput experiments**. To account for these unmeasured factors, Surrogate Variable Analysis (SVA) was applied using `sva` (Leek et al., 2012). Normalized gene-level counts were extracted from the DESeq2 object and transformed with a $\log_2$ scale to stabilize variance and reduce the influence of highly expressed genes. Two linear models were constructed using `model.matrix()`. SVA works by comparing these two models: the **full model**, which contains the biological signal that must be preserved `(~ condition + sex)`, and the **null model**, which contains all known sources of variation that are not the primary focus of the study `(~ sex)`, except the one of interest. This structure ensures that SVA isolates only unmeasured sources of variation, without absorbing the biological effect of interest or confounding variables such as sex. The theoretical number of surrogate variables was estimated using `num.sv()`, and significant SVs were computed with `sva()`. Surrogate variables were appended to the sample metadata and retained only if weakly correlated with the biological condition ($|r| < 0.3$), ensuring that only unwanted technical or latent variation was modelled in the DESeq2 design formula. For exploratory visualization, a corrected expression matrix was generated using `removeBatchEffect()`, regressing out the effects of sex and the selected SVs while preserving the biological condition. This corrected matrix was used exclusively for PCA to evaluate whether removing known and latent confounders improved sample clustering and clarified the separation between recurrence groups.

## Differential Expression Analysis

Differential expression analysis was performed on the original count data using `DESeq2` (Love et al., 2014), a statistical framework specifically designed for **high-throughput sequencing count data**. It models RNA-seq data with a negative binomial GLM and accounts for differences in sequencing depth and gene-specific variability. The filtered count matrix and the complete metadata table (including sex and the surrogate variables) were used to construct the final DESeqDataSet:

```
dds_final <- DESeqDataSetFromMatrix(
  countData = round(counts_matrix_filtered),
  colData = metadata,
  design = design_formula)
```

The full workflow was then initiated with: `dds_final <- DESeq(dds_final)`. This command performs size-factor estimation, dispersion estimation, GLM fitting according to the design formula ( `~ sex + SV1 + SV2 + SV3 + SV4 + SV5 + SV6 + SV7 + condition`), $\log_2$ fold-change estimation with shrinkage, Wald testing, multiple-testing correction (FDR < 0.05), and independent filtering.

Differential expression results were extracted with:

```
res <- results(dds_final,
contrast = c("condition", "Early_Recurrence", "Long_Responders"),
alpha = 0.05)
```

Significant genes (padj < 0.05) were annotated using `org.Hs.eg.db` and exported in tsv/csv formats. To evaluate the impact of surrogate variables, three DESeq2 models (noSV, allSV, strictSV) were compared in terms of the number of significant genes and the stability of $\log_2$ fold-change estimates. Exploratory visualizations included a heatmap of significant genes and a volcano plot of DEGs, highlighting up- and down-regulated genes (considering a padj < 0.05 and a log2 fold-change threshold of $|\log2FC| \geq 1$). These plots were generated with `pheatmap`, `matrixStats`, `ggplot2`, `ggrepel`, and `dplyr`.

## Gene Set Enrichment Analysis

Gene Set Enrichment Analysis was performed to **identify pathways associated with transcriptional differences between ER and LR**. `GSEA` is a computational method that determines whether an a priori defined set of genes shows statistically significant, concordant differences between two biological states. It evaluates the entire ranked gene list (Subramanian et al., 2005). Genes were ranked using the Wald statistic (i.e., the ratio between the $\log_2$FC and the standard deviation for each gene) returned by `DESeq2`, which incorporates both effect size and its uncertainty, providing a stable and informative ordering for enrichment testing. The ranked vector was constructed by pairing each Wald statistic with its ENSEMBL gene ID, and sorting in decreasing order. GSEA was run across multiple pathway collections such as GO (BP, CC, MF), KEGG, Reactome, and MSigDB Hallmark. Sets were restricted to sizes between 10 and 500 genes, and statistical significance was assessed using 10,000 permutations. Normalized Enrichment Scores (NES), p-values, and FDR-adjusted q-values were computed, applying an FDR threshold of 0.05. Visualizations included dotplots, barplots, and enrichment curves summarizing NES values and significance patterns. Analyses were performed using `clusterProfiler` and `fgsea`, with visualization through `enrichplot`, `ggplot2`, and `msigdbr` for Hallmark gene sets.

## Cancer-associated Fibroblasts Signature Analysis

CAF signatures were curated from multiple reference datasets to capture the major fibroblast programs described in the literature. The CAF-S1 panel was derived from the study "Single-Cell Analysis Reveals Fibroblast Clusters Linked to Immunotherapy Resistance in Cancer" (Kieffer et al., 2020) while two additional CAF collections (CAF-1 and CAF-2) were provided by collaborators. Signatures correspond to iCAF, myCAF, CAF-S1, and Normal Fibroblasts. CAF signatures enrichment was assessed using `FGSEA` multilevel, applied to the `DESeq2`-ranked gene list. Significance was set at padj $< 0.05$. This approach tests whether genes belonging to each signature accumulate preferentially at the top or bottom of the ranked distribution, providing a directional measure of enrichment. Enrichment patterns were visualized using barplots and dot plots.

## Immune Cell Types Deconvolution

Immune cell composition was estimated using `CIBERSORTx` (Newman et al., 2015), which infers the **abundance of immune populations from bulk RNA-seq data.** Deconvolution was performed using the LM22 reference (Chen et al., 2018)(disabling quantile normalization and computing Monte Carlo p-values). The input matrix was generated from normalized counts, and gene identifiers were converted from ENSEMBL to HUGO symbols to ensure compatibility with the LM22 signature matrix. `CIBERSORTx` returned absolute immune scores, deconvolution p-values, Pearson correlations, and RMSE values for each sample. Absolute immune scores were imported into R for downstream analysis. Boxplots were used to visualize differences between ER and LR groups. Statistical comparisons were performed using the Wilcoxon rank-sum test for each of the 22 LM22 cell types, with p-values adjusted using the Benjamini-Hochberg FDR method.

## Functional and Network Analysis

Ingenuity Pathway Analysis (IPA, QIAGEN) was used **to functionally interpret the differentially expressed genes** and to reconstruct regulatory and interaction networks. For each gene set resulting from `Deseq2` analysis, ENSEMBL IDs, $\log_2$ fold changes, and p-values were uploaded and mapped to the QIAGEN Knowledge Base. Analyses were performed using default IPA settings, considering Homo sapiens, experimentally validated direct and indirect interactions, and the "Ingenuity Knowledge Base (Genes Only)" as reference. IPA modules applied included: Canonical Pathway Analysis, Upstream Regulator Analysis, Causal Network Analysis, Diseases and Bio Functions, Network Generation, and Regulator Effects (when sufficient molecules were available). Results were considered significant at $p < 0.05$ (Fisher's Exact Test) and absolute activation Z-score $\geq 2$.

## 3.7 Protein-Level Validation by Immunohistochemistry

Sectioning and staining of FFPE e-MPeM specimens collected during CRS, before HIPEC administration, were carried out following standard procedures by colleagues. Briefly, five consecutive 3-µm sections were cut using a Leica RM2265 microtome (Leica Biosystems, Wetzlar, Germany) and dried for 1 hour at 60 °C. Antigen retrieval was performed using an automated BenchMark Ultra immunostainer (Ventana Medical Systems, Roche, Basel, Switzerland) according to the manufacturer's instructions. Following the differential expression analysis, immunohistochemistry (IHC) for GAS1 was performed on the BenchMark Ultra platform using the primary antibody α-GAS1 (clone DF4098, 1:500; Affinity Biosciences, Cincinnati, OH, USA). Images were digitized using the Aperio 450DX automated slide scanner (Leica Biosystems, Wetzlar, Germany). GAS1 expression was quantified using the QuPath positive cell detection tool, which classified positive cells into four categories: negative (0), weak (1+), moderate (2+), and strong (3+). The reliability of the QuPath-derived H-score was confirmed by histopathological evaluation performed by expert MPeMs pathologists.

### Stratified analyses for *GAS1* within the ER and LR

Following IHC quantification, patients were stratified into *GAS1*-high and *GAS1*-low groups within each DFS class, using the median H-score in ER (130) and LR (190) patients as thresholds. All downstream analyses described in Sections 3.4 and 3.6 were repeated within each subgroup (ER *GAS1* high vs low; LR *GAS1* high vs low) to **identify genes and pathways influenced by *GAS1* in tumors that recur early or late.** In this context, survival analyses (KM curves and log-rank tests) were performed with an inferential purpose, to assess whether *GAS1* expression levels were associated with statistically significant differences in progression-free survival overall and within the classification of its level in ER and LR patients.

## 3.8 Spatial Transcriptomics

### Experimental Procedure GeoMx DSP

Spatial transcriptomic profiling was performed with the **GeoMx Digital Spatial Profiler** (DSP; Bruker) and the **Human Whole Transcriptome Atlas** (HuWTA v1.0), enabling high-plex analysis of ~18,000 targets from a single 4-µm FFPE section. We profiled **22 patients (11 ER, 11 LR)** selected from the bulk RNA-seq cohort based on FFPE block quality and the presence of clearly identifiable tumor and microenvironmental compartments.

The workflow began with the deparaffinization and rehydration of the sections, followed by an overnight hybridization at 37°C with the HuWTA probe set. These RNA probes, which carry photocleavable DNA oligonucleotides, were applied alongside fluorescent morphology markers to enable immunofluorescence imaging and the precise selection of **regions of interest**. To resolve distinct tissue compartments within each ROI, **areas of illumination (AOIs)** were segmented based on specific marker expression. This allowed for the isolation of the **tumor** (nuclei/SYTO13$^{+}$ PanCK$^{+}$), **lymphoid structures** (nuclei/SYTO13$^{+}$ PanCK$^{-}$ CD68$^{-}$ 1A4$^{-}$), **fibroblasts** (1A4$^{+}$), and **macrophages** (nuclei/SYTO13$^{+}$ CD68$^{+}$).

Following segmentation, the barcoded oligonucleotides were released via **UV-mediated photocleavage**, aspirated, and transferred into 96-well plates for library preparation. To ensure experimental balance and technical consistency, AOIs from individual patients were distributed across multiple slides, with each slide containing a mix of samples from both the ER and LR groups. Finally, sequencing libraries were generated by PCR using unique i7/i5 dual indices and processed on an **Illumina NovaSeq 6000**. By utilizing Unique Molecular Identifier (UMIs) and probe-specific barcodes for digital molecule counting, this method provided an accurate and spatially resolved quantification of gene expression across all targeted tissue compartments.

A schematic overview of the workflow used for this study (from slide preparation and ROI selection to library construction, sequencing, and generation of DCC files) is shown in Figure 8.

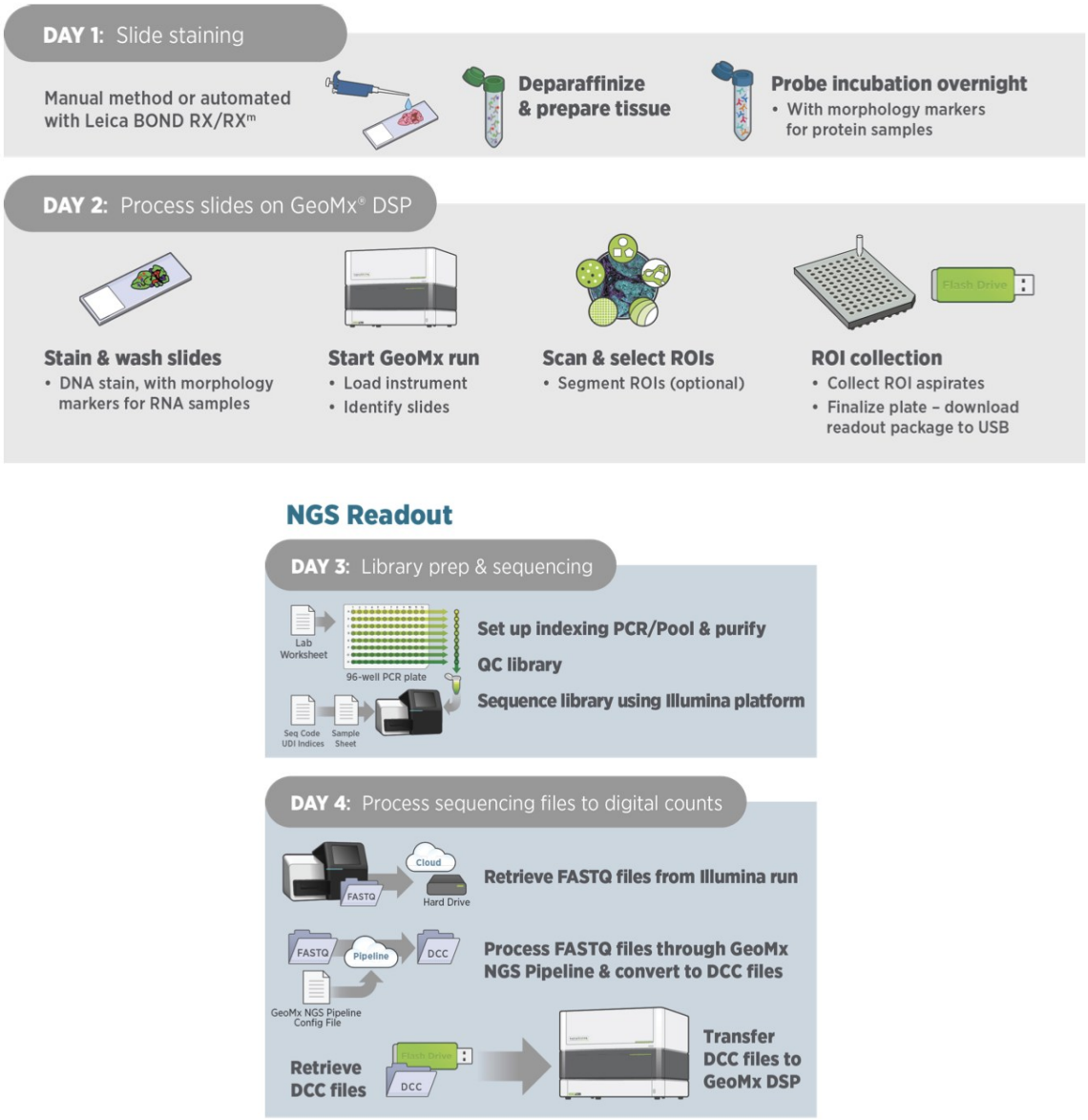


***Figure 8. Overview of the four-day GeoMx DSP workflow for spatial transcriptomics (©2022 NanoString Technologies, Inc).***

## GeoMx DSP Data Processing and Analysis Workflow

All data processing steps adhered to the official GeoMx DSP guidelines provided by NanoString. The analysis pipeline comprised two main stages: raw data generation and downstream processing in R (v4.5.2). Sequencing data were generated in **BCL forma**t and converted to **FASTQ files** using Illumina BCL Convert v4.4.6 on the BaseSpace cloud platform.

To build the **GeoMxSet object** in R, three inputs were combined:

1. **DCC files** containing processed counts (Figure 9). FASTQ files were processed with the NanoString pipeline to generate DCC files, which contain digital counts for each target in each AOI. This step required both the FASTQ files and a merged configuration file preserving run information.
2. The **PKC file** provides probe-target relationships for the GeoMx Human Whole Transcriptome Atlas, which was downloaded from the NanoString website.
3. A curated **annotation file** linking each AOI to sample metadata, which was assembled by combining information from the LabWorksheet with the clinical database.

```
<Header>
FileVersion,0.02
SoftwareVersion,"GeoMx_NGS_Pipeline_2.3.3.10"
Date,2025-5-28
</Header>

<Scan_Attributes>
ID,DSP-1001660039031-B-C06
Plate_ID,1001660039031
Well,C06
</Scan_Attributes>

<NGS_Processing_Attributes>
SeqSetId,ADX00148:52:HNJHYDRX2
Raw,16669026
Trimmed,16479994
Stitched,16267802
Aligned,15810294
umiQ30,0.9872
rtsQ30,0.9861
</NGS_Processing_Attributes>

<Code_Summary>
RTS0020877,267
RTS0020878,70
RTS0020879,147
RTS0020880,139
RTS0020881,215
RTS0020882,97
RTS0020883,171
RTS0020884,104
RTS0020885,168
RTS0020886,92
RTS0020887,80
RTS0020888,79
RTS0020890,93
RTS0020892,127
RTS0020893,263
RTS0020894,105
RTS0020895,106
RTS0020896,133
RTS0020897,221
RTS0020898,210
RTS0020899,104
RTS0020902,86
RTS0020904,186
RTS0020905,105
RTS0020906,214
RTS0020907,262
RTS0020908,93
RTS0020909,99
RTS0020910,111
RTS0020911,97
RTS0020912,90
RTS0020913,81
RTS0020914,103
RTS0020915,1684
RTS0020916,428
RTS0020917,1213
```

***Figure 9. Example of a Digital Count Conversion file generated by the GeoMx NGS Pipeline. The file includes metadata (software version, plate ID, and well ID), sequencing metrics (raw, trimmed, stitched, and aligned reads), internal quality scores (UMI and base-calling quality), and a code summary reporting counts per probe.***

## 3.9 GeoMx DSP Downstream Analysis

All downstream analyses were performed in R using the *standR* workflow (Figure 10). This pipeline was selected for its **robust quality-control** procedures and its ability to **address batch effects** (Y. Liu et al., 2025).

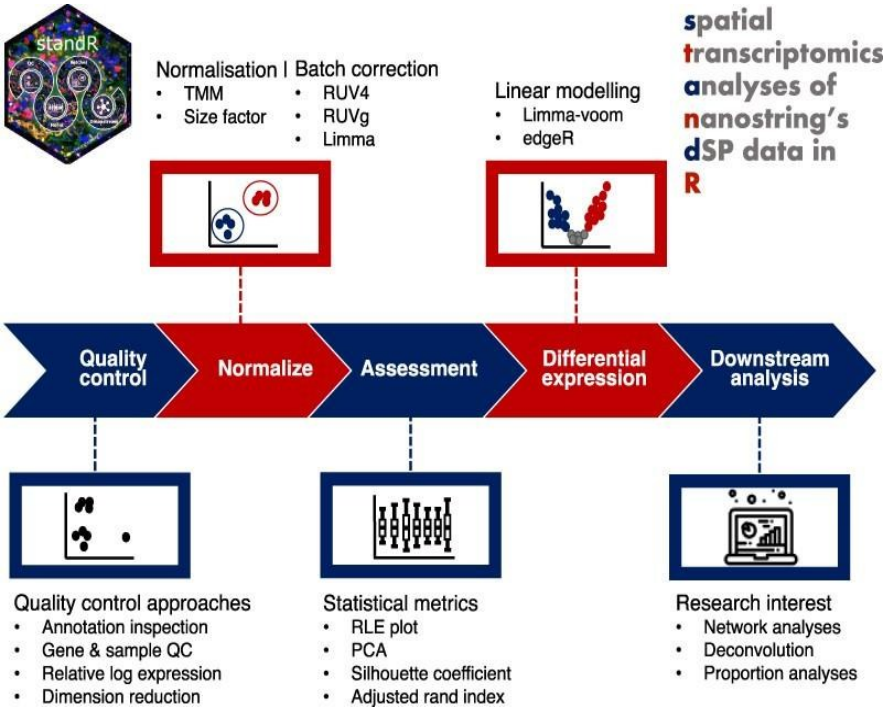


***Figure 10. standR workflow principal steps.***

### Data Import and Quality Assessment

DCC files, the WTA v1.0 PKC file, and a curated annotation table were imported and integrated into a unified data structure that contains raw counts, gene annotations, and ROI-level metadata. Metadata included patient and sample identifiers, slide and scan information, clinical group, segment type ($CK^+$, $CD68^+$, $1A4^+$, Lymph), segmented area, multiplex immunofluorescence markers, estimated nuclei count, spatial coordinates, and technical variables such as scan dimensions and sequencing run ID. Because the WTA panel contains multiple probes per gene, probe-level counts were aggregated to gene-level expression using `GeomxTools`. Spatial coordinates were standardized by renaming column names and converting them to a numeric format. Each AOI was assigned a unique identifier corresponding to its DCC filename to ensure correct alignment between count matrices and metadata. Gene-level counts, ROI metadata, and probe annotations were then integrated into a SpatialExperiment object using `readGeoMx()`, explicitly specifying the columns corresponding to gene and AOI identifiers to ensure consistent mapping across all components. AOIs with

zero total counts, typically arising from segmentation failures or annotation mismatches, were removed before QC. Per-AOI QC metrics (library size, number of detected genes, proportion of low-expression genes) were computed with `addPerROIQC()`. AOIs with insufficient nuclei or with a library size <50,000 counts were excluded. Diagnostic visualizations, including Relative Log Expression (RLE) plots stratified by slide and patient, and AOI area-library size relationships, were used to assess dataset integrity, detect outliers, and identify slide-specific batch effects. After QC filtering and metadata standardization (including anonymization of patient and slide identifiers), the processed dataset was stored as a SingleCellExperiment object for downstream analyses.

## Unsupervised Clustering Analysis for DSP data

**To investigate the intrinsic transcriptional structure of the dataset**, unsupervised clustering was performed on the log-transformed expression matrix derived from the top 500 highly variable genes (HVGs), identified using `modelGeneVar()` and `getTopHVGs()` (`scran`). These HVGs capture the dominant sources of biological variation and provide a robust basis for exploratory analysis. The optimal number of clusters was estimated using the `NbClust` package, testing k = 2-10 and selecting the silhouette index as the internal validation metric. This procedure was applied independently to AOIs (columns) and genes (rows), enabling the identification of coherent transcriptional modules and AOI subgroups. Hierarchical clustering was then performed using Euclidean distance and Ward.D2 linkage. Before clustering, gene-wise z-score scaling was applied to normalize expression ranges across genes. A quantile-based color palette (1st, 50th, 99th percentiles) was used to enhance visualization of expression gradients and reduce the influence of extreme values. Heatmaps were generated using the `ComplexHeatmap` package, incorporating multiple layers of metadata (patient ID, clinical group, sex, segment type, and slide identifier) to contextualize clustering patterns before and subsequently batch correction.

## Batch Correction with Ruv4

**DSP datasets often display pronounced slide-specific batch effects** that can mask true biological differences. Therefore, batch correction was a vital step to ensure that subsequent comparisons between AOIs reflected biological rather than technical variation. Batch correction was performed using the `RUV4` method, which estimates and removes latent unwanted factors while preserving the biological signal of interest (Gagnon-Bartsch & Speed, 2012). `RUV4` requires a set of Negative Control Genes (NCGs) to model technical variation. In this dataset, the main batch effect was linked to slide identity; thus, NCGs were identified using the `findNCGs()` function (`standR`), which selects the 300 least variable genes across slides based on their coefficient of variation. These genes were presumed to be stable across biological conditions and used to estimate slide-related unwanted factors. `RUV4` correction needs three inputs: (i) the biological variable of interest (clinical group), (ii) the list of NCGs, and (iii) the number of unwanted factors (*k*) to remove. To determine an appropriate k value, multiple candidate values ($k$ = 1-5) were tested following `RUV4` guidelines. For each *k*, a corrected dataset was generated using `geomxBatchCorrection()`, and PCA plots were examined to assess the reduction of slide-driven variation. Additionally, the optimal k was estimated with the `findBestK()` function, which first quantifies the variance attributable to batch versus biological factors and then evaluates cluster coherence using silhouette score*s*. This combined qualitative-quantitative approach ensured that the chosen *k* effectively reduced slide-specific effects without overcorrecting or distorting the underlying biological structure. RUV4 outputs both a corrected expression matrix and a set of latent weights (W) that quantify the influence of unwanted technical factors on each AOI. Finally, in the `limma-voom` differential expression pipeline, these W weights were incorporated as covariates in the linear model, allowing technical variability to be removed without altering the count data's structure.

## Differential Expression Analysis

DEA was performed independently for each of the four identified tissue compartments: **tumor (CK$^{+}$)**, **macrophages (CD68$^{+}$)**, **fibroblasts (1A4$^{+}$)**, and **lymphocytes**. This compartmentalized approach allowed for the identification of cell-type-specific transcriptional shifts associated with relapse outcomes (ER vs LR) within the tumor microenvironment. For each compartment, AOIs were identified using curated lists of DCC filenames to subset the SpatialExperiment object. Raw counts were extracted and converted into a `DGEList` object using the `edgeR` package.

To ensure robust statistical power, lowly expressed genes were removed using the `filterByExpr()` function, and library sizes were normalized via the trimmed mean of M-values (TMM) method (`calcNormFactors()`). The differential expression framework was based on the `limma-voom` pipeline, which accounts for the mean-variance relationship inherent in RNA-seq count data (Ritchie et al., 2015).

A design matrix was constructed to enable direct contrasts between the ER and LR groups within each specific cell type. To account for the hierarchical structure of the dataset, where multiple AOIs were collected from the same patient, the `duplicateCorrelation()` function was employed. This step estimated the intra-patient correlation, which was then incorporated into the linear modelling to prevent the inflation of statistical significance and ensure that the patient, rather than the individual AOI, served as the effective unit of biological replication.

The voom-transformed data were fitted using `lmFit()`, followed by empirical Bayes moderation (`eBayes()`) to stabilize variance estimates. DEGs were identified based on a Benjamini-Hochberg False Discovery Rate (FDR < 0.05). Positive log-fold change (logFC > 1) values indicated up-regulation in the ER group compared to the LR group (volcano plots were created to identify differentially expressed genes). Significant genes for each compartment were visualized using heatmaps created with `ComplexHeatmap`.

This systematic comparison across all segmented compartments provided a comprehensive map of the transcriptional differences between early recurrence and long responders pts, spanning from epithelial tumor signaling to stromal and immune-mediated responses.

# 4. Results and Discussion

## 4.1 Principal Characteristics of the Cohort

The cohort included **36 patients with epithelioid peritoneal mesothelioma** treated with CRS-HIPEC, divided into **two clinical groups based on disease-free survival**: early recurrence (**ER, n = 18**) and long responders (**LR, n = 18**).

All tissue samples were **epithelioid.** The **median age** of the entire population was **53 years** (range 34-75). ER patients were slightly older (median 56 years) compared to LR patients (median 52 years), with similar interquartile ranges. There was a **female predominance overall** (20 F vs 16 M), more evident in the LR group (12 F vs 6 M), while the ER group had a more balanced distribution (8 F vs 10 M). **ER patients** had a **median time to recurrence of 10.7 months** (IQR 7.4-11.75), whereas **LR patients** had a median of **45.1 months** (IQR 42.5-71.2). The cohort had a **median follow-up of 50.5 months**, as estimated with the reverse Kaplan-Meier method. Observation time differed markedly between groups: all ER patients experienced disease progression, whereas only 6 out of 18 LR patients recurred.

**Disease-related death** (DOD) further confirmed the prognostic divergence, occurring in 11/36 patients, all in the ER group ($p < 0.001$). **Preoperative chemotherapy** was administered to 12 patients: 8 in the ER group and 4 in the LR group. Among the 24 patients who did not receive preoperative chemotherapy, 10 were ER and 14 were LR. Regarding tumor burden, ER patients were more often in higher PCI categories, whereas LR patients were mainly in the lowest range (0-16). **Cytoreduction** was complete (CC0-1) in nearly all cases; the only two CC2 occurred in the ER group. Regarding the **Ki67 proliferation index,** 17 patients had Ki67 ≤10% and 19 had Ki67 >10%, with similar distributions in ER (≤10%: 8; >10%: 10) and LR (≤10%: 9; >10%: 9).

## 4.2 Statistical Testing of Clinical Variables

Before exploring molecular and transcriptomic differences between early recurrence and long responder patients, the analysis first assessed whether the two groups differed in their clinical characteristics. This step is crucial in peritoneal mesothelioma, where prognosis depends on multiple factors such as tumor burden, response to systemic therapy, and completeness of cytoreduction. For this reason, the evaluation included treatment exposure, immune contexture, surgical outcome, demographic features, as each could plausibly influence DFS. For example, long responders were expected to show higher tumor-infiltrating lymphocytes, reflecting a more active immune microenvironment, and patients receiving pre-operative chemotherapy were supposed to follow a more favorable trajectory. Similarly, a more complete cytoreduction was assumed to reduce the likelihood of immediate recurrence.

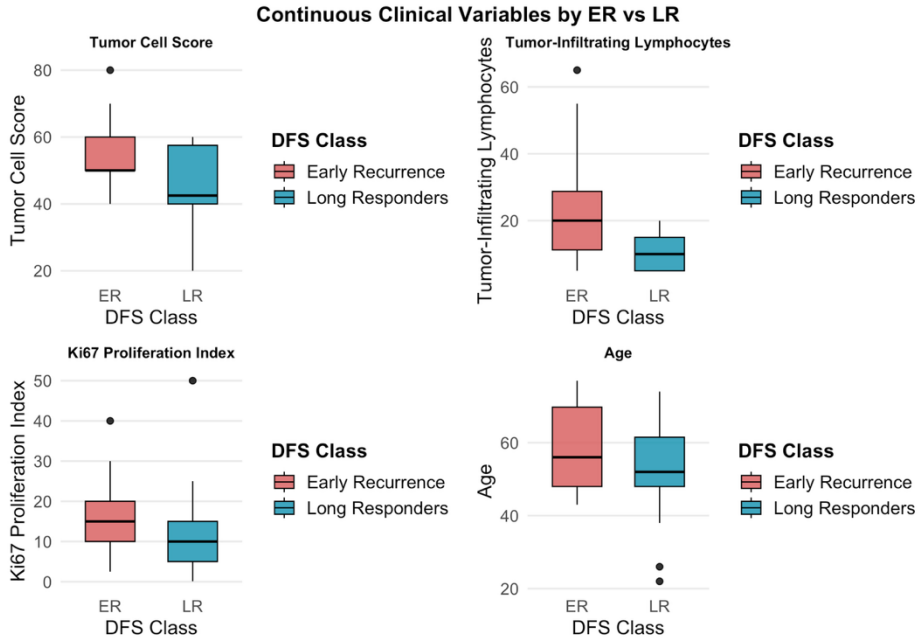


***Figure 11. Boxplots of continuous clinical variables in early recurrence ER and LR patients. Each panel shows the distribution of a clinical variable across the two DFS classes (x-axis), with the measured value on the y-axis. Colores indicate group membership (red = ER, blue = LR). Tumor cell score and age were approximately normally distributed and compared using a t-test, whereas TILs and Ki67 did not meet normality assumptions and were assessed with the Wilcoxon test.***

The exploratory comparison of continuous clinical variables revealed a small set of features that clearly distinguished the two groups. As it is represented in Figure 11, **tumor cell score was significantly higher in ER patients** ($p = 0.0272$); **TILs levels**, against initial expectations, **were elevated in the ER group** ($p = 0.0024$). In contrast, Ki67 and age showed similar median values and overlapping distributions, with no significant differences between groups (Ki67: $p = 0.1287$; age: $p = 0.2045$).

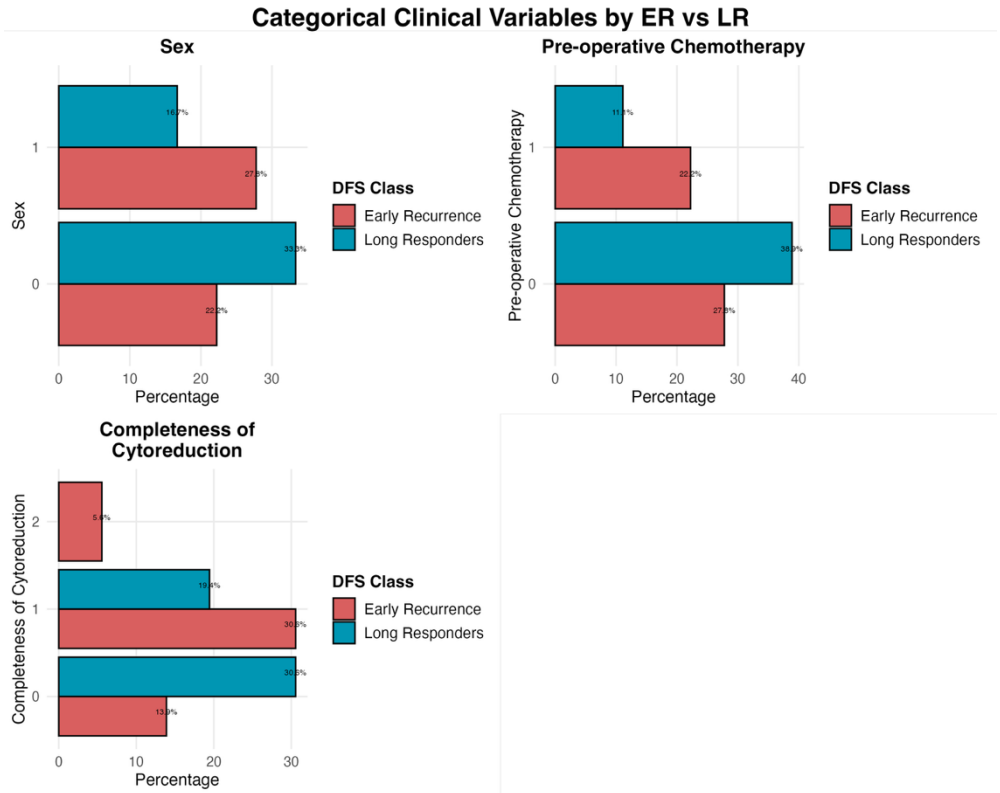


***Figure 12. Barplots of categorical clinical variables in ER and LR patients. Each panel shows the percentage distribution of ER (red) and LR (blue) cases across sex, pre-operative chemotherapy, and completeness of cytoreduction. Two CC2 cases occurred only in the ER group; however, their progression times fell within the median ER range, so they were retained in the cohort. Due to low expected counts, group comparisons were evaluated using Fisher's exact test.***

The barplots of categorical variables shown in Figure 12 display the distribution of sex, pre-operative chemotherapy, and completeness of cytoreduction across the two groups. Sex was not significantly associated with DFS class in this dataset (male: 16.7% LR vs. 27.8% ER; female: 33.3% LR vs. 22.2% ER; p = 0.1797). Pre-operative chemotherapy also showed no significant association (11.1% LR vs. 22.2% ER; p = 0.289). Both groups showed high rates of optimal cytoreduction, yet CC2 procedures were observed only in ER patients (5.6%), although the association did not reach statistical significance (p = 0.0565).

In conclusion, based on the available data, only a limited subset of features was significantly associated with the probability of belonging to the early recurrence or long responder group. Although tumor cell scores and TILs appeared higher in patients with early recurrence, these observations do not allow for definitive biological conclusions. They may suggest that a greater tumor burden and a more inflamed, yet functionally ineffective, immune microenvironment are statistically associated with an increased probability of early relapse. However, this study is limited by the relatively small cohort size, and the TIL metric used here captures only a global

lymphocytic infiltrate without distinguishing specific immune cell subsets (e.g., $CD8^+$ cytotoxic T cells, $CD4^+$ T cells, regulatory T cells, B cells, or NK cells).

Studies based on much larger datasets, of 3,636 peritoneal mesothelioma patients carried from the National Cancer Database, have demonstrated that age, sex, comorbidity burden, histological subtype, surgical margins, treatment modality, and care at high-volume centers are all significant predictors of long-term survival. For example, patients under 50 years of age showed nearly a seven-fold higher likelihood of surviving beyond five years, women had better outcomes than men, and negative surgical margins increased the probability of long-term survival more than four-fold (Bhatt et al., 2026).

In the present study, the absence of broader clinical associations might be due to the small sample size and the high degree of homogeneity within the CRS-HIPEC cohort selected for transcriptomic analysis, rather than a true lack of clinical relevance.

## 4.3 Outcome Analysis

The association of PCI and Ki67 with an increased risk of disease progression after treatment was first evaluated. In the univariate models, Ki67 did not show a significant association with progression risk (HR = 1.014, 95% CI 0.982-1.048, p = 0.384), indicating that, within this cohort, proliferative index alone did not meaningfully influence time to progression.

In contrast, PCI demonstrated a strong and statistically significant association with time to progression. Each one-point increase in **PCI was associated with an approximately 6% increase in the hazard of progression** (HR = 1.064, 95% CI 1.025-1.105, p = 0.001). This finding suggests that greater peritoneal tumor burden at the time of surgery is directly associated with an increased likelihood of earlier recurrence.

When both variables were included in the multivariable model, **PCI remained significantly associated with progression risk** (HR = 1.069, p = 0.0009), confirming its role as an **independent prognostic factor** in this cohort. Ki67 again showed no significant association (p = 0.226). Biologically, this highlights that the overall extent of peritoneal dissemination could be a major factor driving recurrence after cytoreductive surgery.

As expected, survival curves (Figures 13-14) show a marked divergence between the two groups, with **ER patients experiencing early events** and **LR patients maintaining prolonged disease control.** The log-rank test statistically supports this divergence. For PFS, ER patients had far more events than expected (18 vs 5.84), while LR patients had far fewer (6 vs 18.16), yielding a very large chi-square value ($\chi^2 = 40.8$, $p = 2\times10^{-10}$). For OS, all deaths occurred in the ER group (11 observed vs 4.16 expected), whereas no deaths were observed in the LR group despite 6.84 expected ($\chi^2 = 18.8$, $p \approx 1\times10^{-5}$).

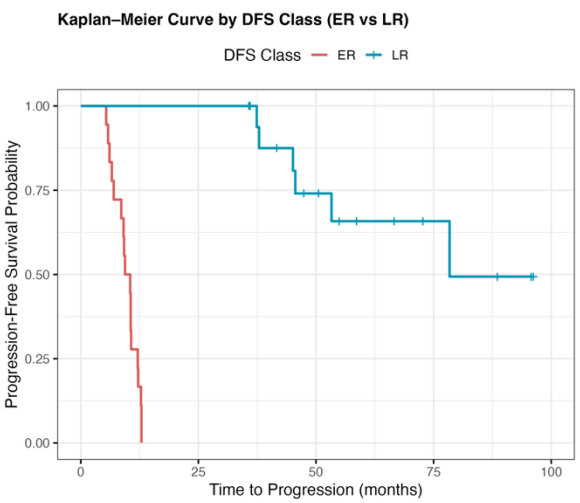


*Figure 13. KM curves for PFS in the ER and LR groups. The x-axis reports time to progression (months), and the y-axis shows progression-free survival probability. The ER curve (red) drops steeply in the first months, while the LR curve (blue) remains high and declines only gradually.*

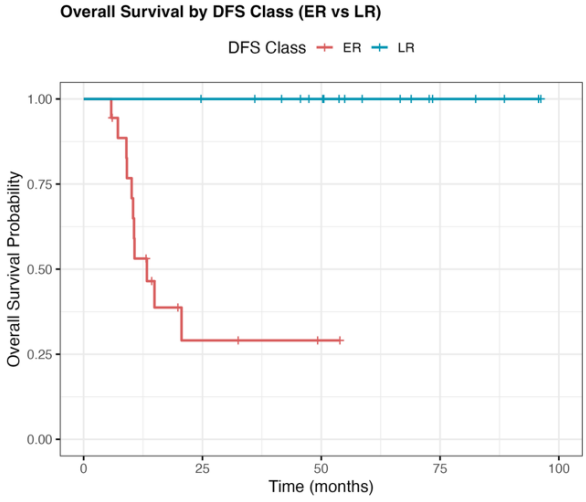


*Figure 14. KM curve for OS in the ER group. The x-axis shows time from surgery (months), and the y-axis indicates overall survival probability. The ER curve (red) declines sharply within the first 20-25 months before reaching a plateau.*

## 4.4 RNA-seq Pipeline

### Conversion from BCL to FASTQ

A total of **34 samples (ER, n = 17; LR, n = 17)** out of 36 successfully **passed RNA sequencing**, and their corresponding BCL files were demultiplexed to generate compressed FASTQ files (. fastq.gz). For paired-end sequencing, two files were produced per sample: one for Read 1 (R1) and one for Read 2 (R2), following the standard Illumina naming structure: samplename_S1_L001_R1_001.fastq.gz. The conversion and demultiplexing process completed successfully, producing the expected **FASTQ files** for all samples and providing a clean and structured dataset suitable for downstream quality control and alignment.

### FASTQ Quality Check

FastQC evaluated several key metrics, including per-base sequence quality, GC content, read length distribution, duplication levels, and the presence of adapter sequences. Quality control metrics extracted from the General Stats section of the FastQC reports and aggregated by MultiQC showed the following ranges: **duplication levels** varied **between 43.7% and 81.2**%, **GC content** ranged from **49% to 54%**, and **read length** was consistently **100 nt** across all samples. The **mean number of reads per sampl**e ranged from **34.6 to 85.3 million**, indicating a robust sequencing depth suitable for downstream analyses. Per-base sequence quality profiles demonstrated high Phred scores throughout the read length, with only slight decreases at the beginning and end of the reads, as expected in Illumina sequencing. **Overall, quality scores remained well above 30**, confirming high base-calling accuracy (Figure 15).

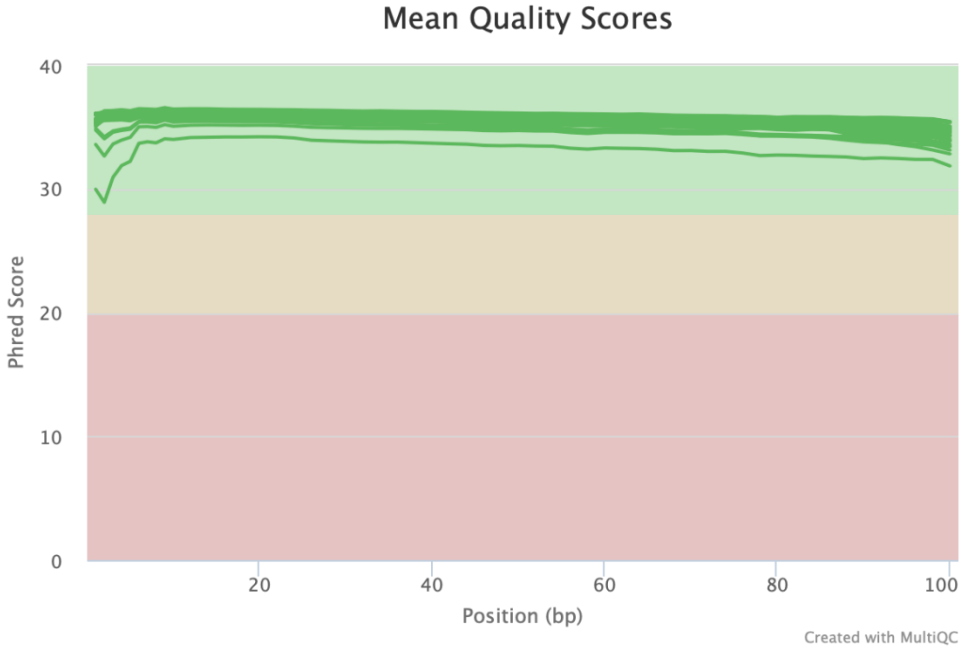


***Figure 15. Exemplar graphs to check quality: the plot shows the average Phred quality score (y-axis) at each base position along the read length (x-axis, 1-100 bp). Each green line represents a different sample.***

## Trimming and QC

Trimming successfully **removed residual Nextera adapter sequences and low-quality bases** from the raw paired-end reads. After filtering, only paired reads longer than 35 bp were retained. Across samples, the number of **surviving reads ranged from 17 to 55 million**, reflecting the expected reduction due to adapter clipping and quality trimming. Post-trimming reports confirmed the effective removal of adapter contamination, as shown in Figure 16. Quality control metrics obtained from FastQC and summarized by MultiQC indicated that trimmed reads maintained high overall quality. Read lengths were consistent with the trimming threshold (≥35 bp), **GC content** remained **stable** across samples (≈48-55%), and **duplication levels** ranged from **48.5% to 83.2%**. The **total number of reads per sample after trimming** ranged from **17.3 to 61.8 million,** providing sufficient depth for downstream alignment and quantification. Moreover, per-base Phred quality scores remained high across the entire read length, confirming the suitability of the trimmed data for subsequent analyses.

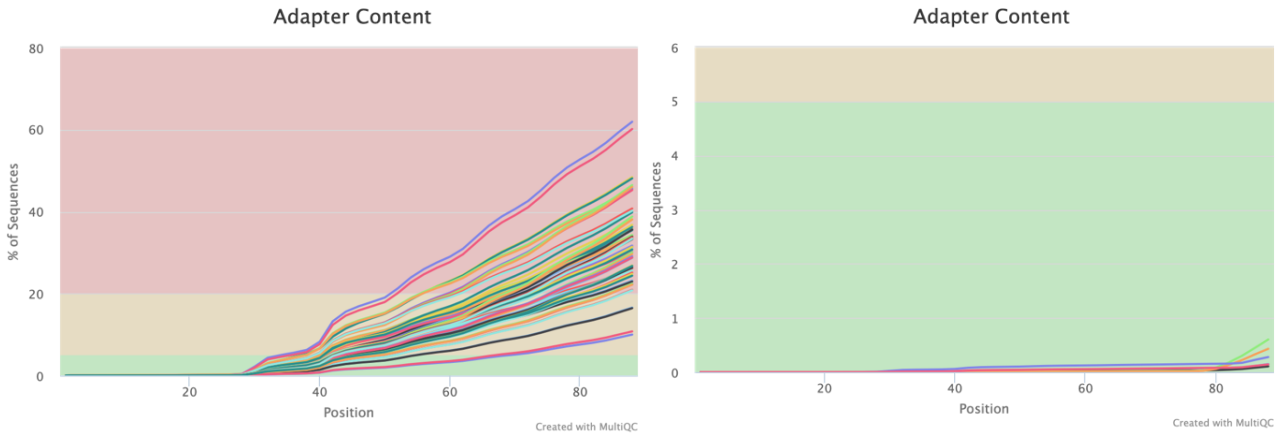


***Figure 16. Adapter content across read positions: the plots show the percentage of reads containing adapter sequences (y-axis) at each nucleotide position along the read length (x-axis). Adapter contamination increases sharply after ~35 bp in the raw data, indicating the presence of residual adapter sequence (graph on the right). After trimming, adapter content becomes negligible across all positions, confirming the effectiveness of the trimming step (graph on the left).***

## STAR alignment

STAR successfully **aligned trimmed paired-end reads to the GRCh38 reference genome**, generating coordinate-sorted BAM files, splice-junction tables, and Log.final.out summaries, all aggregated in the MultiQC report (Figure 17).

Alignment performance was consistently high across samples. Overall **alignment rates** ranged **from 82.4% to 95.3**%, with most samples above 90%. The highest rate was observed in S20-7465_S29 (95.3%), while the lowest occurred in S18-144_S31 (82.4%). **Uniquely mapped reads** ranged **from 76.5% to 89.3%**, again with S20-7465_S29 showing the best performance (89.3%) and S18-144_S31 the lowest (76.5%). A few samples (e.g., S17-7099_S27, S16-2090_S26) displayed slightly reduced alignment efficiency, but all values remained within acceptable ranges for RNA-seq.

**Average mapped read length** was highly consistent across samples **(~194-199 bp)**, confirming correct paired-end reconstruction after trimming. The number of detected splice junctions varied substantially (0.1-13.5 million), reflecting differences in sequencing depth and transcriptome complexity. Base-level accuracy was excellent: **mismatch rates** remained **low** (0.3-0.6%), while **insertion** and **deletion** rates were **negligible**, with **stable indel lengths** (~1.5-2.1 bp).

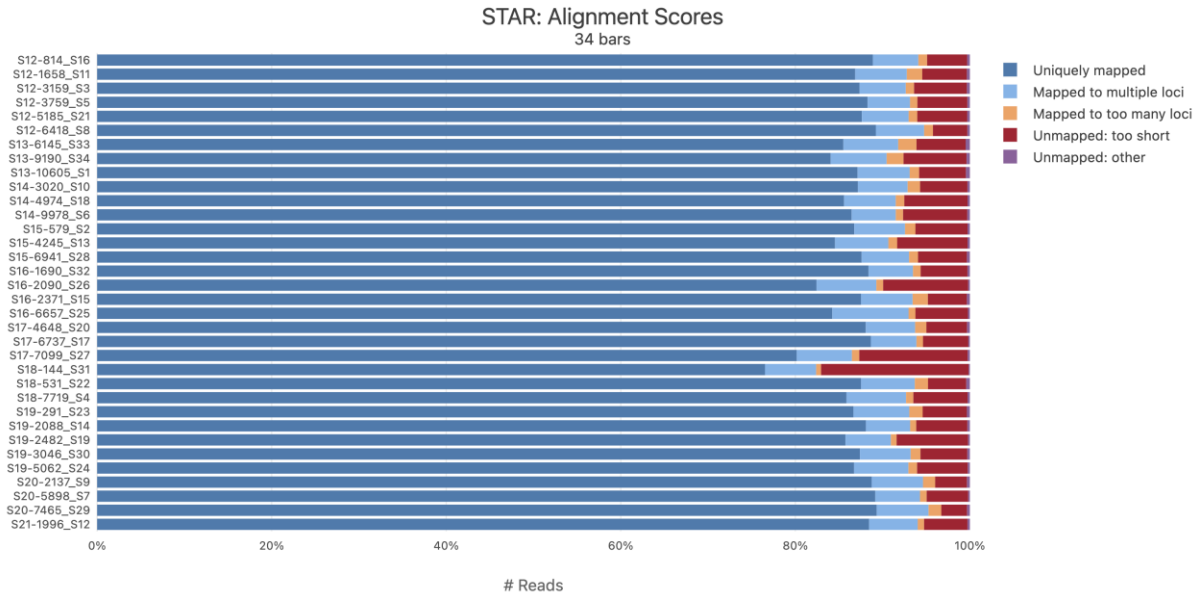


***Figure 17. STAR alignment summary across samples: horizontal stacked bar plots showing, for each sample (y-axis), the proportion of reads falling into different alignment categories (x-axis), including uniquely mapped reads, multi-mapped reads, reads mapped to too many loci, and unmapped reads.***

## HTSeq Counting

HTSeq-count successfully **generated a count file for each sample**, containing **raw counts for all annotated genes in the GRCh38.115 reference**, along with summary categories. These metrics were parsed and visualized by the MultiQC report, allowing an integrated evaluation of read-assignment quality and annotation coverage across all samples (Figure 18). Across the 34 samples, the proportion of **reads** successfully **assigned** to annotated genes ranged from **32.5% to 47.4%,** consistent with human RNA-seq processed in exon-based union mode. Indeed, this is expected in human datasets, where many reads originate from intronic or intergenic regions, incompletely annotated loci, or pre-mRNA molecules that do not overlap annotated exons. One sample (S17-7099_S27) showed a markedly higher assignment rate (61.6%), reflecting a larger fraction of reads mapping to well-annotated exonic regions and fewer ambiguous or multimapping events. The **ambiguous category remained low** and stable (2-4%), indicating that only a small fraction of reads overlapped multiple genes. The **alignment not unique category** accounted for **5-7%** of reads, reflecting multimapping in repetitive or duplicated genomic regions. Both **too low aQual** and **not aligned categories** were **0%**, confirming that **no reads were discarded due to low alignment quality** and that BAM files contained only mapped reads. Overall, HTSeq read-assignment metrics confirm that the dataset is well-suited for downstream and differential expression analysis.

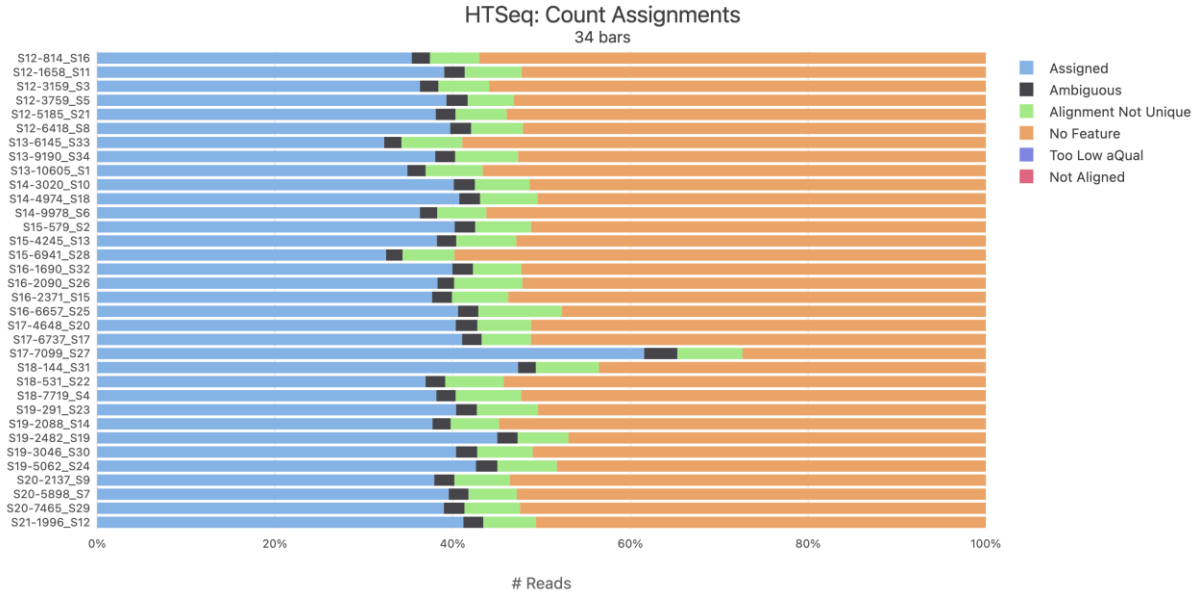


***Figure 18. HTSeq read assignment summary across samples: stacked horizontal bar plots showing, for each sample (y-axis), the proportion of reads assigned to annotated genes or classified into diagnostic categories (x-axis), including assigned, ambiguous, alignment not unique, no feature, too low aqual, and not aligned.***

## 4.5 Comprehensive Transcriptomic and Functional Analysis

To reduce noise and avoid unstable variance and dispersion estimates in DESeq2, low-count genes were filtered out. This improves the sensitivity of differential expression testing and focuses the analysis on genes with meaningful expression across samples. **After filtering, 42,143 genes were removed, and 36,756 genes were retained for downstream analyses.** Although low-count genes were removed, the filtered dataset remained right-skewed, as expected in RNA-seq, where a small subset of genes is expressed at very high levels. These highly expressed genes elevate the mean while leaving the median lower, producing a large mean-median difference even after filtering. Histograms and density plots (log10 scale) showed that the dominant low-count peak in the raw distribution was substantially reduced, and the filtered distribution shifted toward higher expression values, with the highest concentration of genes around log10 ≈ 4 (~10,000 total counts). A direct comparison of raw (red) and filtered (purple) distributions in Figure 19 clearly illustrates the impact of filtering: lowly expressed genes were effectively removed, yielding a cleaner and more informative dataset suitable for normalization and differential expression analysis.

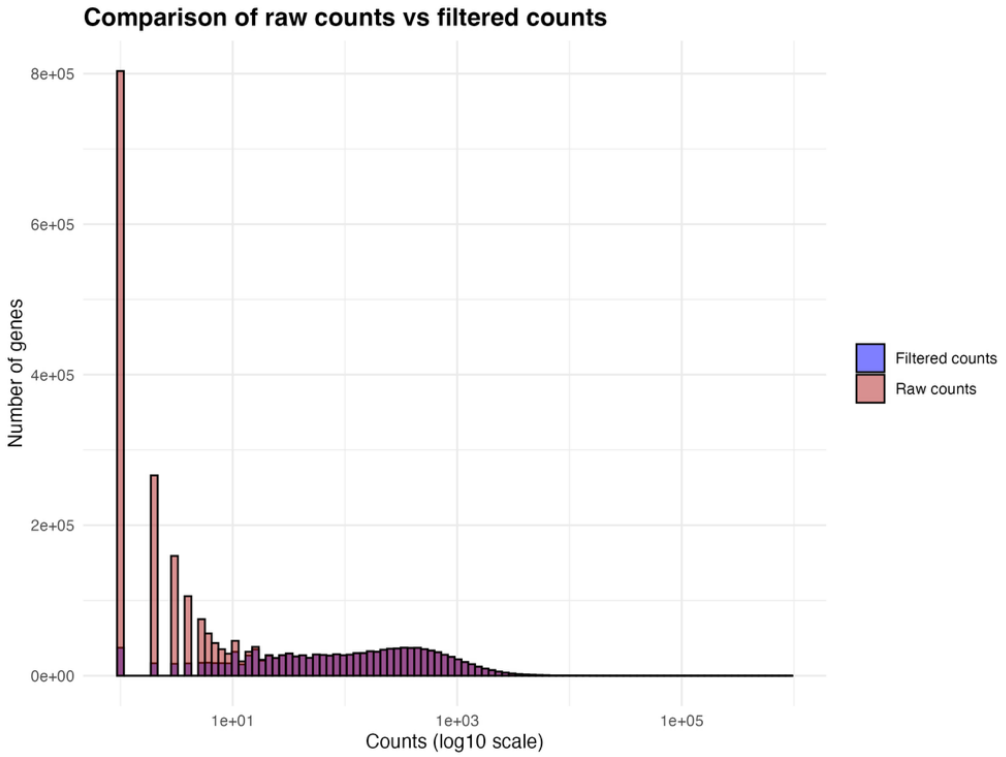


***Figure 19. Exemplar histogram comparing the distribution of gene expression counts before and after filtering. The x-axis shows gene counts on a log10 scale, and the y-axis reports the number of genes. Raw counts are shown in red, while filtered counts are shown in purple.***

## Principal Component Analysis

Each sample, originally represented by more than 36,000 gene expression values, is visualized as a point in a two-dimensional space defined by PC1 and PC2, which **explained 18% and 16% of the total variance, respectively. The samples did not show a clear separation between ER and LR.** The two groups were largely intermixed along PC1, indicating that the DFS class is not the primary driver of global transcriptomic variation in this dataset. Along PC2, a separation is visible, but it is **not associated** with relapse outcome (Figure 20). To investigate alternative sources of variation, the PCA was colored according to metadata variables. When samples were colored by **sex** (Figure 21), a clear separation emerged: males and females formed two distinct clusters along the principal axes. This indicates that **sex** is a **major contributor** to the variance captured by the PCA

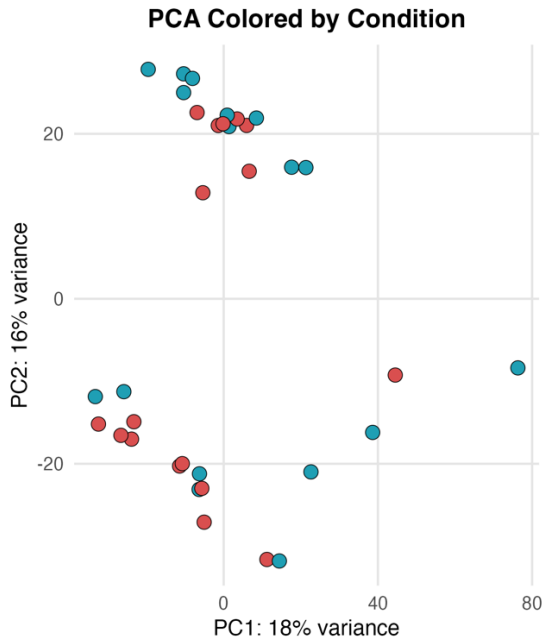


*Figure 20. PCA colored by clinical outcome showing each sample projected onto PC1 (x-axis) and PC2 (y-axis). Samples are colored by DFS class: Early Recurrence (red) and Long Responders (blue).*

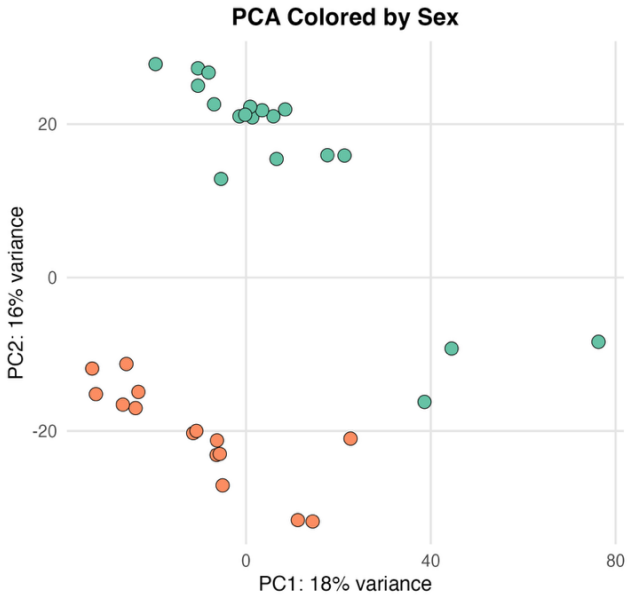


*Figure 21. PCA colored by sex showing each sample projected onto PC1 (x-axis) and PC2 (y-axis). Samples are colored by sex: female = green-aqua and male = orange.*

The observed differences likely reflect broad sex-related transcriptional programs involving hormonal regulation, immune activity, and systemic biological processes, extending beyond sex-chromosome genes.

In addition, recent clinical evidence shows that **sex-based biological differences significantly influence survival in MpeM.** Large-scale analyses have demonstrated that **female** patients,

particularly those in the **premenopausal age** range, experience markedly **improved outcomes** compared with males and postmenopausal females, suggesting a hormonal contribution to tumor behavior (Cummins et al., 2026).

To further assess whether the effect of early recurrence differed between males and females, an interaction model (sex × DFS class) was tested in DESeq2. No genes reached statistical significance, indicating that the ER-LR transcriptional contrast does not differ between sexes in our dataset.

Because sex explains a substantial portion of the variability (much larger than the expected differences between ER and LR), **failing to adjust for it in the design formula would risk confounding the differential expression analysis.** Sex-driven signals could otherwise be misinterpreted as condition-specific effects, potentially generating false positives.

## Surrogate Variable Analysis

SVA identified **seven surrogate variables** (SVs) as hidden sources of variation that could interfere with global transcriptomic patterns and downstream differential expression analysis. All seven SVs showed weak correlation with the biological condition ($|r| < 0.3$), indicating that none of them captured true ER-LR biological signals and can be included in the design formula for DE analysis.

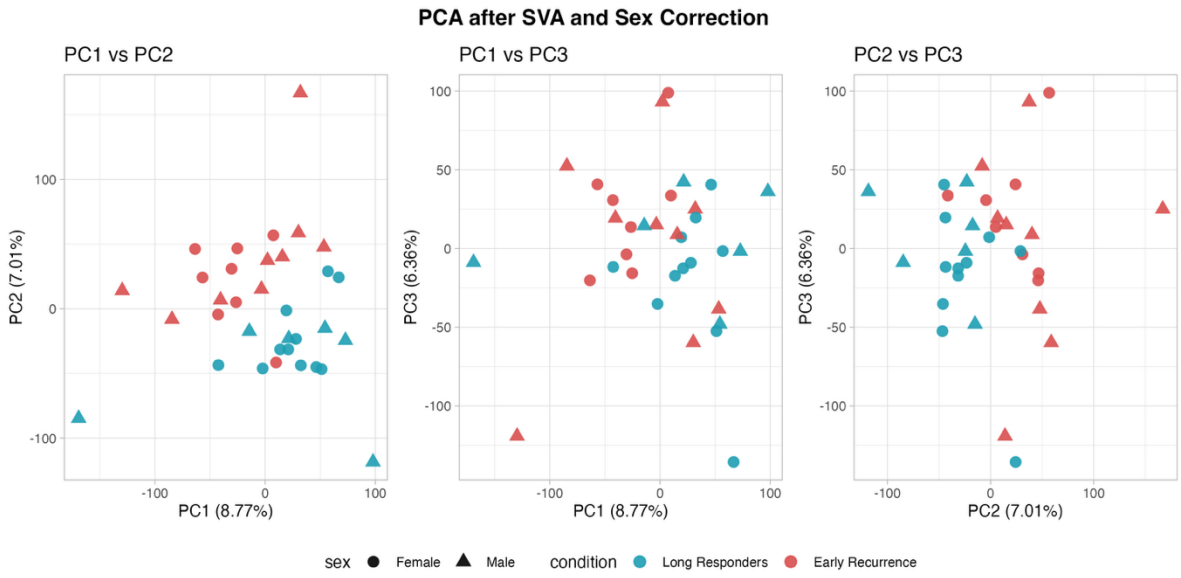


***Figure 22. PCA after SVA- and sex-correction showing sample projections onto PC1 vs PC2 (left), PC1 vs PC3 (middle), and PC2 vs PC3 (right). Axes report the percentage of variance explained by each PC. Samples are colored by relapse outcome (blue = long responders, red = early recurrence) and shaped by sex (circles = females, triangles = males).***

Just for visualization, after correcting the expression matrix for both sex and the seven SVs, PCA plots (Figure 22) showed that samples no longer clustered by sex, confirming successful

removal of sex-related variation. ER and LR samples remained largely intermixed across all principal components, indicating that recurrence status is not a major driver of global transcriptomic structure. A slight tendency toward separation was visible in some projections (PC1-PC2 and PC2-PC3), but the pattern was weak and did not align with the main axes of variance (PC1 = 8.77%, PC2 = 7.01%, PC3 = 6.36%).

**Overall, these results indicate that ER-LR differences do not manifest at the level of global transcriptomic variation and are instead detectable only through gene-specific differential expression analysis.**

## Differential Expression Analysis and IHC Validation

Across the dataset, 36,756 genes were tested. The impact of SVA was substantial: the uncorrected model (noSV) failed to identify any significant genes (padj <0.05), confirming that technical noise and variability in RNA quality initially masked the underlying biological signal. By incorporating these latent factors into the model, 15 significant genes (padj < 0.05) were identified. The global distribution of effect sizes and statistical significance is visualized in the volcano plot (Figure 23), where a dual threshold of padj < 0.05 and $|\log_2 FC| \geq 1$ was applied to prioritize biologically relevant differentially expressed candidates. Genes with significant p-values are positioned on the upper side of the plot.

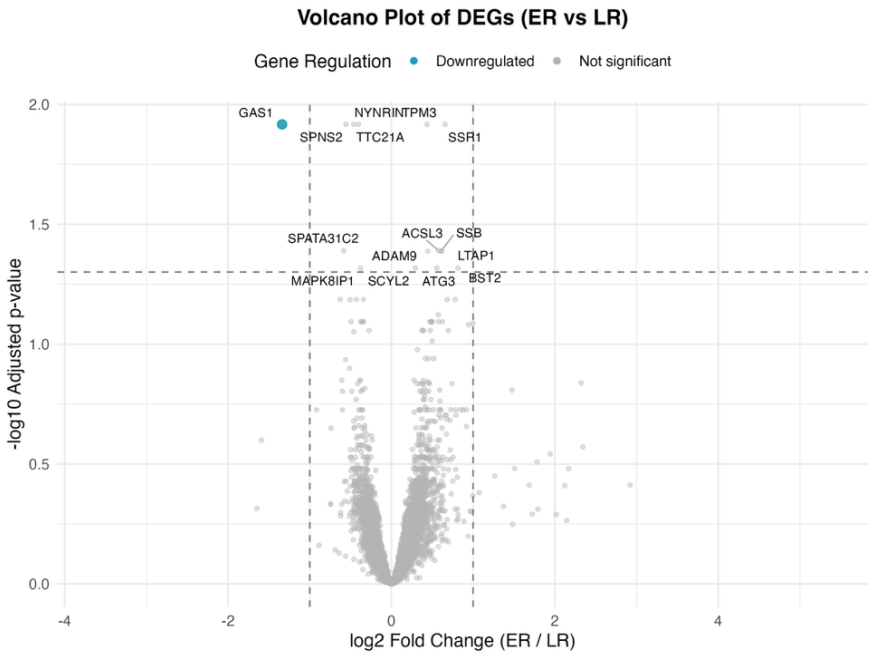


***Figure 23. Volcano plot of DEGs (ER vs LR): it displays the differential expression results comparing ER and LR. The vertical dashed line marks the $|\log_2 FC| = 1$ threshold, and the horizontal dashed line indicates FDR = 0.05. Axes represent $\log_2$ fold change (ER vs LR) on the x-axis and $-\log_{10}$(FDR) on the y-axis. Most genes cluster around $\log_2 FC \approx 0$ and are not significant as DEGs (gray points). Gene which appears significantly downregulated in ER is identified with a blue point.***

Genes such as *SSR1*, *ACSL3*, *ADAM9*, *ATG3*, *BST2*, *SCYL2*, *SSB*, and *LTAP1* exhibited a statistically significant, albeit moderate, increase in expression in ER patients ($log_2$FC between +0.29 and +0.82). From a biological perspective**, ER samples tend to exhibit moderately higher expression of genes whose primary functions involve membrane dynamics, protein trafficking, and cytoskeletal organization**. Conversely, several genes, such as *TTC21A, SPNS2, NYNRIN, MAPK8IP1,* and *SPATA31C2* (with log2FC between –0.34 and –0.58) showed moderately negative log2FC values, indicating reduced expression in the ER group. Therefore, **LR samples show relatively higher expression of genes functionally linked to growth restraint, epithelial structure, and cellular homeostasis.**

**However, the most remarkable finding in this group was the expression profile of *GAS1* (Growth Arrest Specific 1).** It is the only gene in the entire dataset to meet both the rigorous statistical significance and the substantial fold-change threshold **(log2FC = –1.34)**, effectively isolating it as a candidate driver of the observed clinical divergence (blue dot in Figure 23).

The expression patterns of the 15 genes with the lowest padj value are shown in the heatmap in Figure 24, which highlights two main transcriptional trends: a group of genes more highly expressed in ER samples and a second preferentially expressed in LR samples.

*GAS1* is a putative tumor suppressor gene that plays a role in growth suppression, blocking entry to S phase, and preventing cycling of normal and transformed cells. From a functional perspective, the suppression of *GAS1* in ER patients could help explain their rapid disease progression.

**Undoubtedly, it represented the most suitable gene for immunohistochemical validation (IHC),** given its strong downregulation in ER samples and its potential as a prognostic marker. IHC (Figure 25) confirmed that this pattern was also evident at the protein level, with LR tumors showing stronger GAS1 staining than ER tumors. The concordance between transcriptomic and protein-level data indicates that *GAS1* downregulation in **ER** tumors is biologically meaningful, **reflecting a true reduction in the tumor-suppressive protein**.

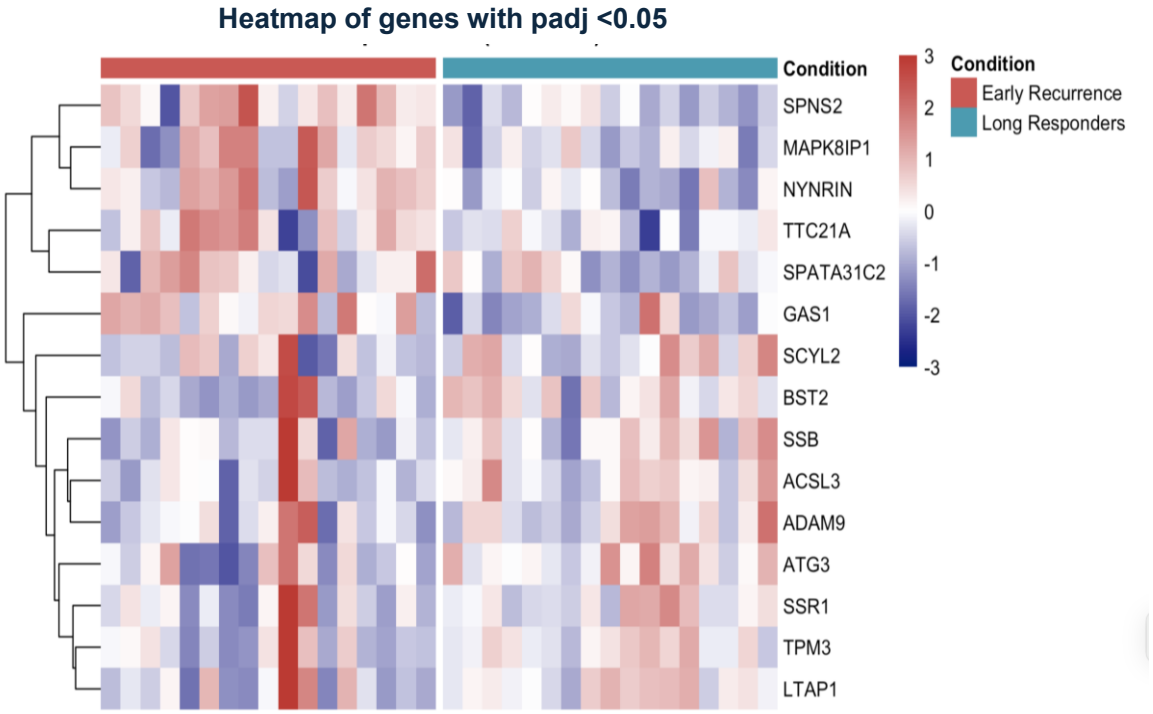


*Figure 24. Heatmap of significant genes: it shows normalized expression levels of the 15 with padj < 0.05, across all samples. Rows represent genes and columns correspond to individual samples (ER in red, LR in blue). Expression values are scaled by gene (z-score) where red corresponds to upregulation and blue to downregulation.*

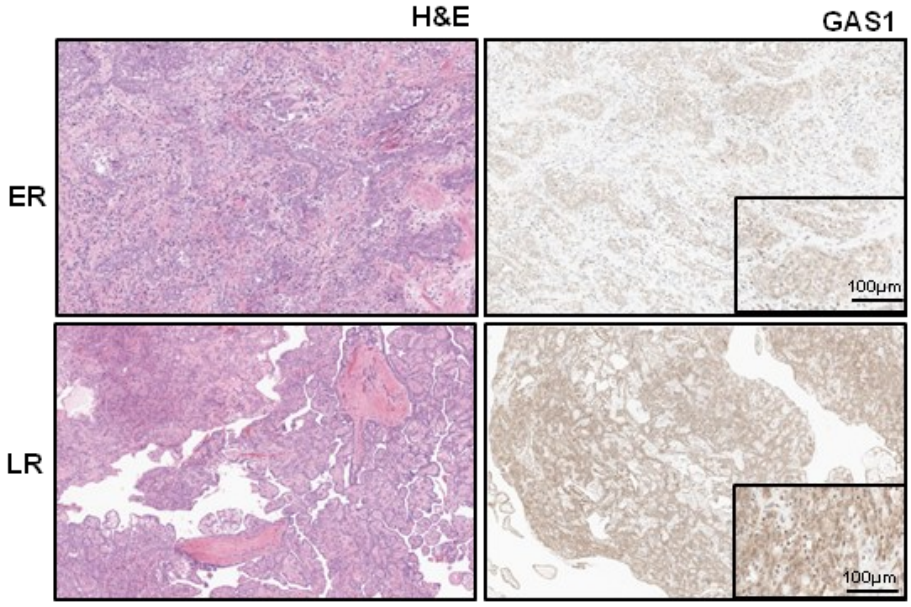


*Figure 25. GAS1 expression in LR and ER tumors. H&E (left) shows morphology; IHC (right) shows more intensive GAS1 staining in LR tumors. Insets (100 µm) highlight the difference.*

## Gene Set Enrichment Analysis

Among GSEA outputs, Hallmark results are the most interesting, because they showed the most biologically informative differences between ER and LR (Figure 26).

The **top ER-enriched pathways** included **MYC targets** (NES = 3.20), **mTORC1 signaling** (NES = 2.63), **oxidative phosphorylation** (NES = 2.53), **E2F targets** (NES = 2.48), and **interferon α/γ responses** (NES = 2.19 and 2.11), indicating a strongly proliferative and biosynthetically active phenotype.

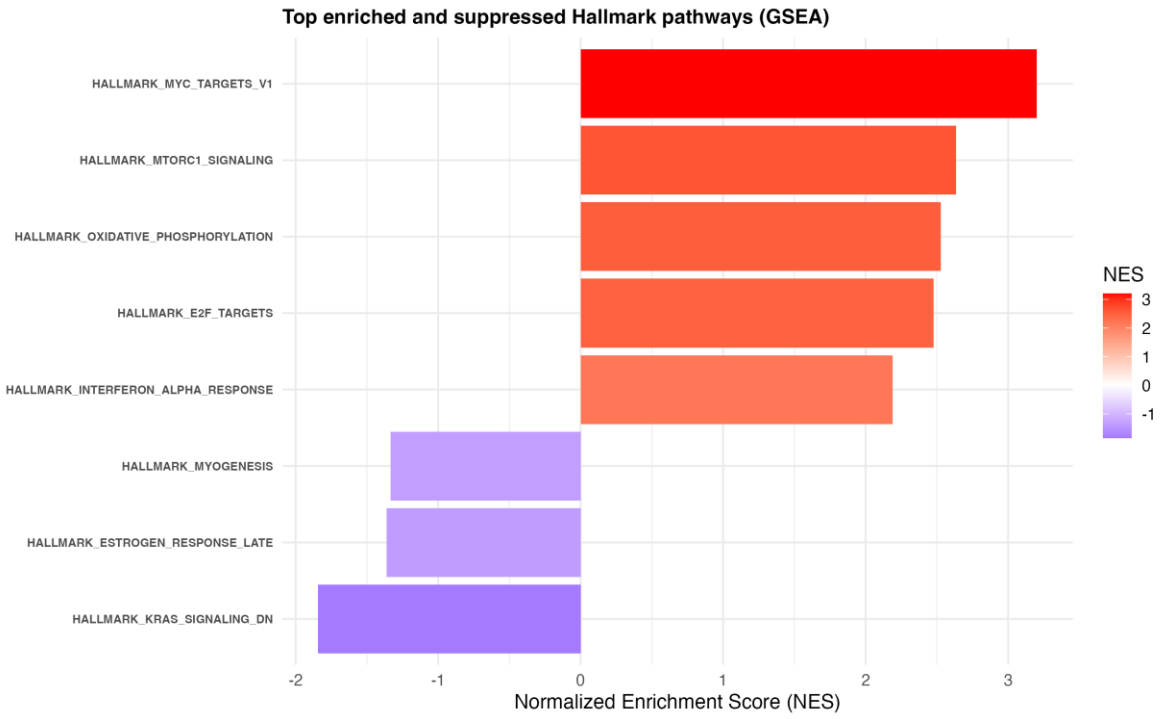


***Figure 26. Exemplar horizontal barplot showing the Normalized Enrichment Score (NES) for the most significantly enriched Hallmark pathways in the ER vs LR. The x-axis represents the NES, with positive values indicating enrichment in ER and negative values indicating enrichment in LR. The y-axis lists the corresponding Hallmark gene sets. Bar colors reflect the magnitude and direction of enrichment (red = enriched in ER; purple = enriched in LR).***

MYC targets enrichment likely reflects activation of a transcriptional program that drives growth, biosynthesis, and DNA replication, whereas E2F targets enrichment was associated with highly active G1/S transition; mTORC1 signaling implies enhanced anabolic metabolism and protein synthesis supporting cellular expansion, while oxidative phosphorylation enrichment indicates increased mitochondrial ATP production to fuel these energy-intensive processes; the interferon-α response suggests intrinsic inflammatory and stress-related signaling, rather than effective anti-tumor immunity.

Additional enrichment in **ER** samples was observed for the **G2M checkpoint** (NES = 2.14), **unfolded protein response** (NES = 1.90), **cholesterol homeostasis**, **PI3K-AKT–mTOR signaling, apoptosis, ROS pathway**, **complement**, **angiogenesis**, and **EMT** (exemplar enrichment plot in Figure 27).

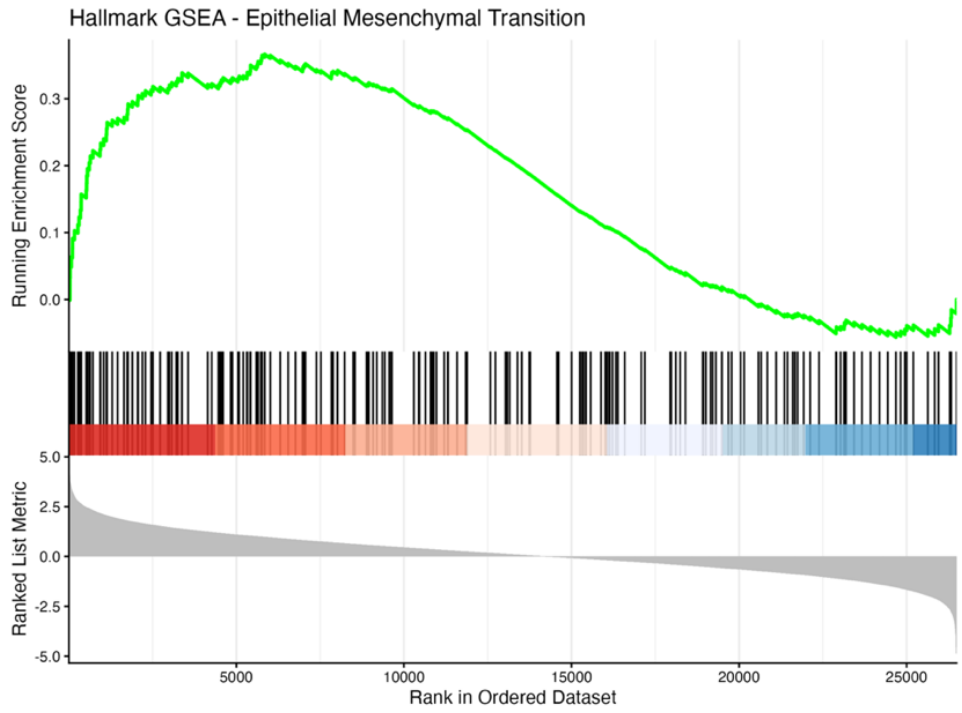


***Figure 27. Enrichment plot for Hallmark EMT pathway, showing the distribution of the EMT gene set across the ranked list. The enrichment score (top, green curve) rises sharply at the start, indicating EMT genes are enriched among those most positively linked to the ER phenotype. Black bars (middle) mark EMT genes and cluster at the list's start, showing coordinated upregulation in ER tumors. The metric (bottom) shows strong positive correlations in this region, confirming EMT enrichment in ER.***

Figure 27 shows the GSEA enrichment plot of genes involved in epithelial-to-mesenchymal transition (EMT). The enrichment of the **EMT** gene set in **ER tumors** indicates possible activation of a mesenchymal program associated with **invasiveness, extracellular matrix remodeling, and poorer prognosis**, in line with previous research.

In contrast, only three pathways were significantly enriched in **LR tumors: KRAS signaling down** (NES = -1.84), which denotes lower proliferative activity**, estrogen response late** (NES = -1.36)**,** and **myogenesis** (NES = -1.33). In this context, myogenesis reflects genes involved in cytoskeletal organization and cell adhesion, consistent with a less invasive phenotype. A late estrogen response is typically associated with epithelial differentiation; its enrichment in LR tumors is also consistent with the previously discussed evidence, which suggests that estrogen-related pathways may contribute to a less aggressive tumor phenotype in MPeM in females.

## Cancer-Associated Fibroblasts Signature Analysis

The tumor microenvironment constitutes a dynamic ecosystem shaped by the reciprocal interactions between non-malignant stromal cells (such as immune cells, endothelial cells, and fibroblasts) and cancer cells. These bidirectional exchanges are pivotal in driving tumor initiation, progression, and therapeutic resistance, underscoring that cancer behavior is determined not only by tumor-intrinsic alterations but also by the multifaceted and context-dependent stromal milieu. Among the stromal components, cancer-associated fibroblasts (CAFs) represent one of the most prevalent and adaptable populations. Because **EMT, angiogenesis, and inflammatory/TGFβ signaling are typically associated with stromal activation, the Hallmark GSEA suggested a possible involvement of CAFs in ER tumors.** This prompted the evaluation of CAF signatures. The CAF gene sets included 121 iCAF genes, 37 myCAF genes, and 128 Normal Fibroblast genes, of which 81, 16, and 126, respectively, overlapped with the ranked list used for FGSEA.

Therefore, FGSEA analysis revealed distinct enrichment patterns of CAF types between ER and LR (Figure 28).

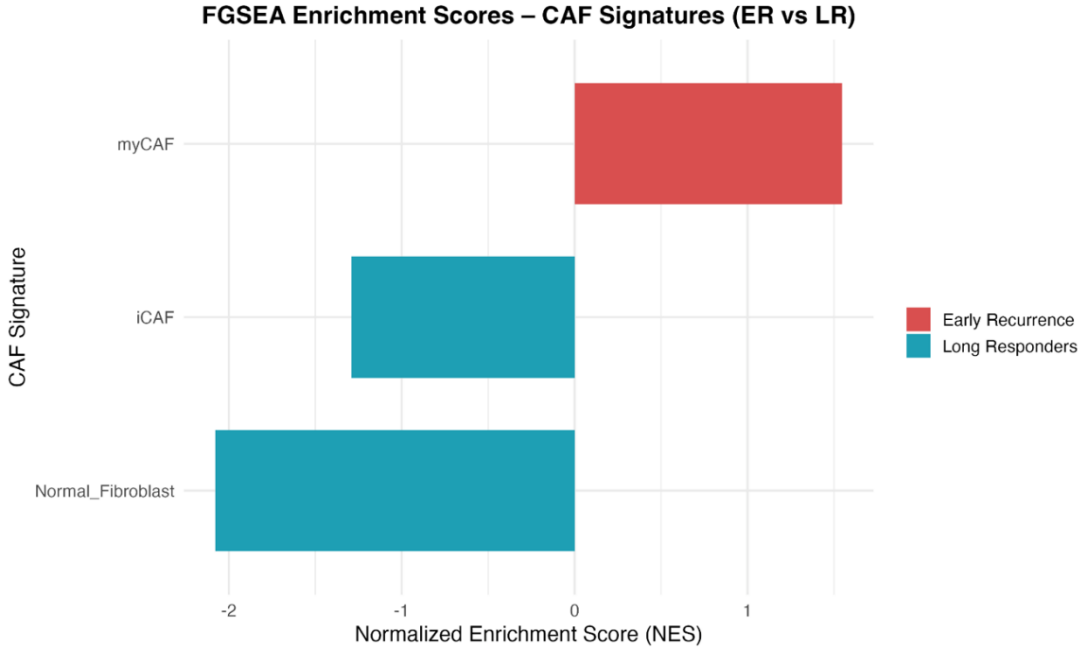


***Figure 28. Exemplar horizontal barplot showing the NES for CAF-related signatures in the ER vs LR comparison. The x-axis reports the NES values, where positive scores indicate enrichment in ER samples and negative scores indicate enrichment in LR samples. The y-axis lists the CAF signatures included in the analysis (myCAF, iCAF, Normal Fibroblast). Bar colors represent the direction of enrichment: red for signatures enriched in ER and blue for signatures enriched in LR.***

**Myofibroblastic CAFs signatures** were enriched in **ER** samples. This finding aligns with the biological role of myCAF, which is driven by TGF-β/SMAD3 signaling, expresses α-SMA,

and remodels the extracellular matrix through ECM deposition, supporting invasive behavior. Conversely, **inflammatory CAF** and **Normal Fibroblast** signatures displayed **negative NES values**, suggesting higher activity in **LR** pts, consistent with a more inflamed but less activated stromal context (Adusumalli & Banala, 2025).

## Immune Cell Types Deconvolution

Then, I tested whether any divergence in bulk transcriptional profiles between ER and LR might reflect shifts in immune-cell abundance, rather than differences in transcriptional patterns or signaling programs.

Based on the deconvolution analysis, the divergence between ER and LR tumors could not be attributed to major shifts in immune cell composition. As shown in Figure 29, the estimated fractions of 22 immune cell types did not show any statistically significant differences between groups when compared using Wilcoxon tests followed by FDR correction (all padj > 0.05).

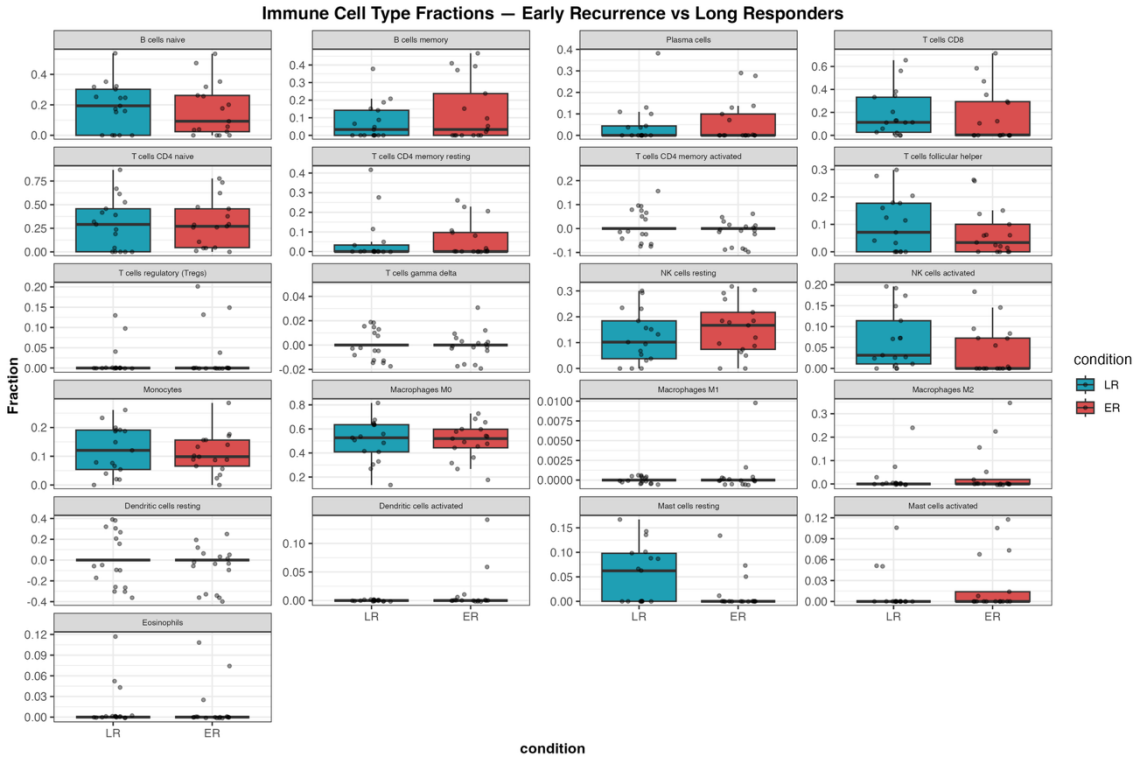


***Figure 29. Immune cell fractions inferred by CIBERSORTx, on the x-axis, there are clinical conditions of patients (LR = long responders; ER= early recurrence), on the y-axis, there is the estimated fraction of each immune cell type.***

## Functional and Network Analysis

Finally, to better characterize the regulatory mechanisms driving the ER-LR transcriptional divergence, a causal network analysis in Ingenuity Pathway Analysis software was performed. By integrating direct, experimentally demonstrated, and inferred interactions, this method reconstructs regulatory networks that have the highest probability to explain the observed expression patterns. Figure 30 depicts one of the most relevant outputs of the analysis.

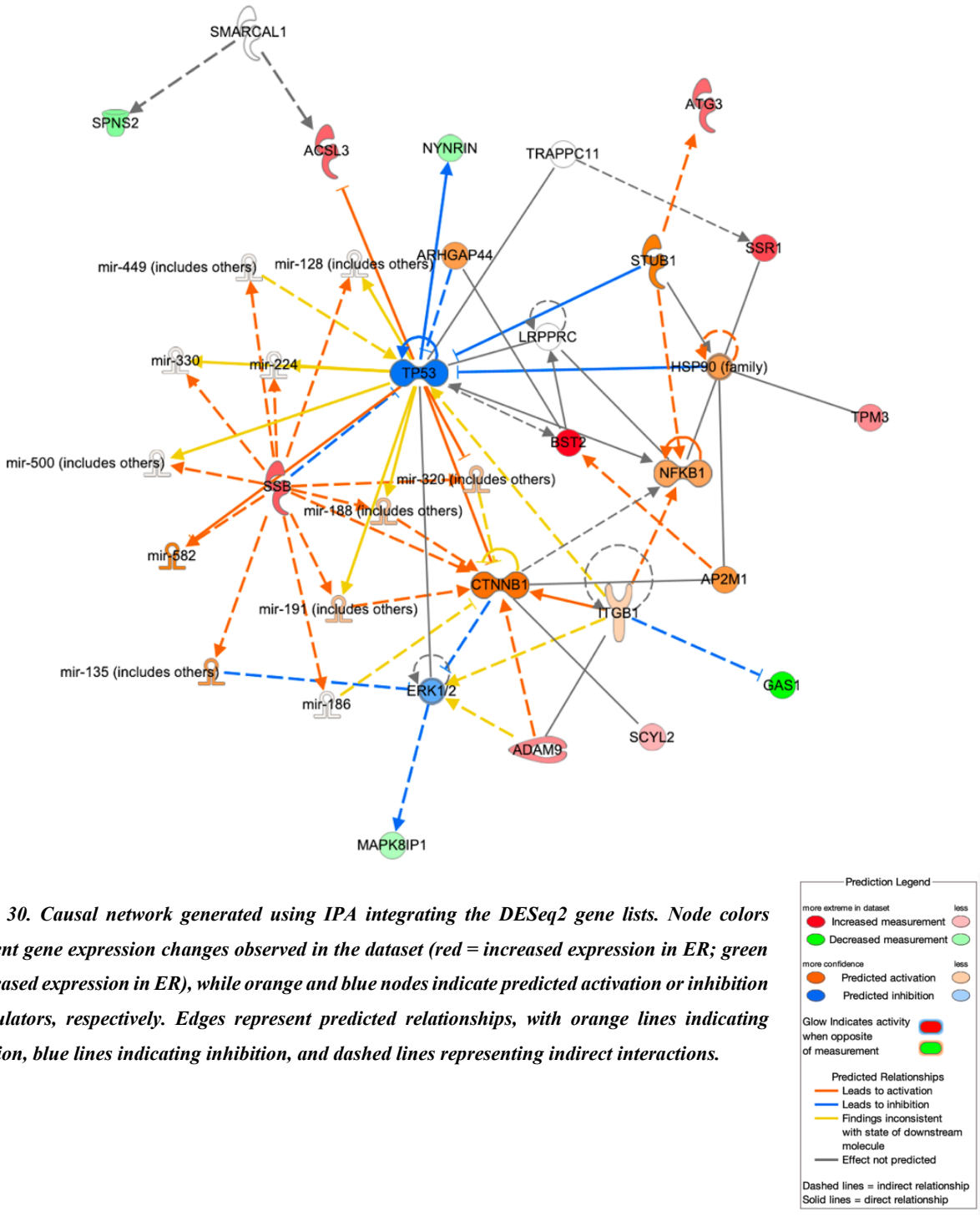


***Figure 30. Causal network generated using IPA integrating the DESeq2 gene lists. Node colors represent gene expression changes observed in the dataset (red = increased expression in ER; green = decreased expression in ER), while orange and blue nodes indicate predicted activation or inhibition of regulators, respectively. Edges represent predicted relationships, with orange lines indicating activation, blue lines indicating inhibition, and dashed lines representing indirect interactions.***

In detail: ***TP53*** occupies the central position in the network, connecting multiple downstream targets identified in the dataset. Given its role in **regulating cell-cycle arrest, apoptosis, and genomic stability**, it appears to be functionally modulated in ER tumors, suggesting that its tumor-suppressor activity may be inhibited compared to LR tumors, contributing to enhanced proliferation and cellular plasticity (Kastenhuber & Lowe, 2017). The network highlights that this inhibition is reinforced by several components involved in protein homeostasis, including **HSP90** and the E3 ligase **STUB1**. Their predicted activation suggests that ER tumors rely on protein-quality control and degradation, **which can reduce TP53 activity and support tumor survival** (Muller et al., 2008). Another key component is ***CTNNB1*** **(**Catenin Beta-1**)**, the main effector of the **Wnt/β-catenin** pathway. **Activation of β-catenin** is associated with **increased proliferation, transcriptional plasticity, and epithelial-mesenchymal transition** (Zhan et al., 2017). Its presence aligns with the enrichment of EMT-related and proliferative pathways seen in the previous gene set enrichment analysis. Notably, ***ITGB1*** **(Integrin β1)** appears to **inhibit *GAS1***. This might causally explain *GAS1* suppression and the potential consequent **reduced growth-inhibitory signals** (Cooper & Giancotti, 2019) **and further prompt the study of *GAS1*-related patterns to understand the favorable biological behaviors of LR patients.**

## 4.6 Analysis stratified by *GAS1* Expression

As described in the 4.5 section above, after observing that *GAS1* expression differed markedly between the two clinical subgroups (ER and LR), this distinction was also confirmed at the protein level through IHC. Within each clinical group, however, *GAS1* expression was not completely uniform: some patients displayed relatively higher levels, whereas others showed lower levels. This internal variability suggested that *GAS1* might influence tumor biology differently even within the same subgroup, making an additional stratification useful. This observation prompted the stratification of both ER and LR patients according to *GAS1* expression (***GAS1*-high** and ***GAS1*-low** based on IHC-defined threshold; 130 for ER; 190 for LR), performing a differential expression analysis to identify whether specific gene patterns were present in relation to *GAS1* levels and, if so, which biological pathways could be evidenced.

In the ER subgroup, *GAS1* expression is globally low and shows limited variability. Accordingly, GAS1-based stratification yields only 9 significant genes (padj < 0.05) from DESeq2, with no coherent functional pattern. GSEA identifies enrichments in proliferation-related pathways (e.g., MYC targets, E2F targets, G2M checkpoint), but these likely reflect the intrinsic aggressiveness of ER tumors rather than a *GAS1*-specific effect. Given the minimal transcriptional impact of *GAS1* in the ER, subsequent analyses focus on the LR subgroup, where *GAS1* variability is biologically meaningful, revealing a much stronger and interpretable signal.

Indeed, transcriptomic analysis in the **LR subgroup discovered a strong impact of *GAS1* stratification.** As shown in the volcano plot (Figure 31), the *GAS1*-high phenotype is characterized by extensive transcriptional repression (222 downregulated genes vs. only 7 upregulated). A major component of this repressed signature is the **drastic downregulation of ribosomal proteins from both the large (RPL) and small (RPS) subunits.** The coordinated decline of these structural constituents, alongside essential non-coding RNAs for ribosome biogenesis **(*RN7SL1*, *RMRP*)** and key translation elongation factors **(*EEF1A1*,** Eukaryotic Translation Elongation Factor 1 Alpha 1), points to a **systemic collapse of the cell's protein synthesis machinery.** This could indicate that ***GAS1*-high** tumors **adopt a state of metabolic quiescence**, characterized by a "metabolically cold" profile.

This state is further reinforced by the significant **downregulation of ferritin genes**, specifically ***FTH1*** **(Ferritin Heavy Chain 1)** and ***FTL*** **(Ferritin Light Chain)**, suggesting a **sharp reduction in iron-dependent biosynthesis and mitochondrial energy consumption.** Conversely, among the very few **upregulated transcripts** emerged ***BDNF-AS*** (BDNF Antisense RNA). This antisense RNA is a known **epigenetic suppressor** that represses BDNF by recruiting repressive chromatin-modifying complexes, leading to transcriptional silencing (Modarresi et al., 2012). The upregulation of BDNF-AS in GAS1-high tumors may indicate the presence of an additional regulatory layer contributing to growth restraint. In this context, BDNF-AS could be associated with the **suppression of growth-promoting pathways**, potentially reinforcing a less aggressive transcriptional state.

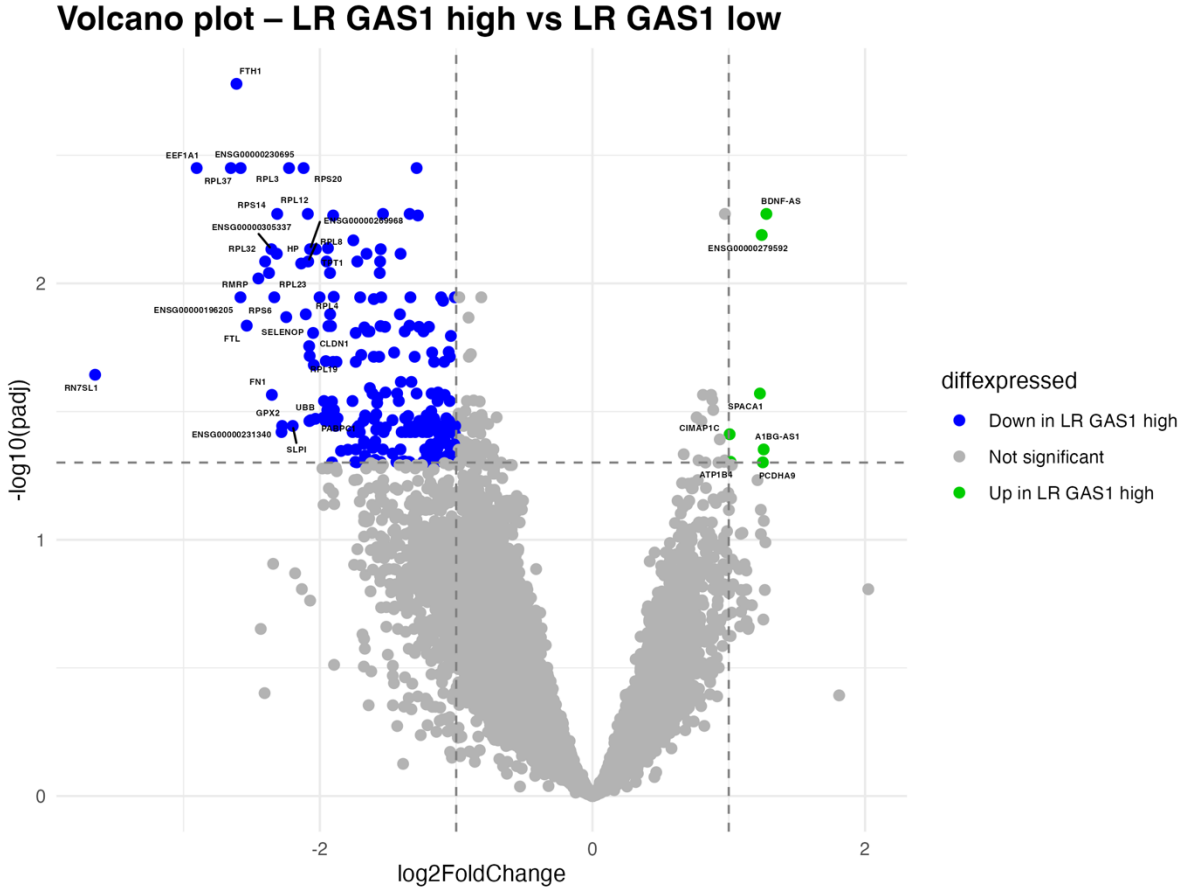


***Figure 31. Volcano plot of DEGs (LR GAS1-high and GAS1-low tumors). Axes represent $log_2$ fold change (ER vs LR) on the x-axis and $–log_{10}$(FDR) on the y-axis. The vertical dashed line marks the $|log_2FC| = 1$ threshold, and the horizontal dashed line indicates FDR = 0.05. Blue dots indicate the 222 genes significantly downregulated in LR GAS1-high and the green dots the 7 genes upregulated in GAS1-high tumors; grey dots represent non-significant genes. Top selected genes with strong statistical significance and large effect sizes are labelled.***

The heatmap (Figure 32) displays the top 309 genes ranked by their significance (padj < 0.05) from DESeq2, focusing on those with the most robust statistical association despite modest fold-change variation. The samples clearly separate into distinct transcriptional clusters

according to *GAS1* levels. Two major groups emerge, corresponding to *GAS1*-low and *GAS1*-high tumors, with a small intermediate cluster showing partial overlap with both states, which suggests that *GAS1*-associated transcriptional programs may operate along a continuum rather than as a strict binary switch, **reinforcing the idea that *GAS1* modulates tumor behavior in a graded manner within the LR subgroup**.

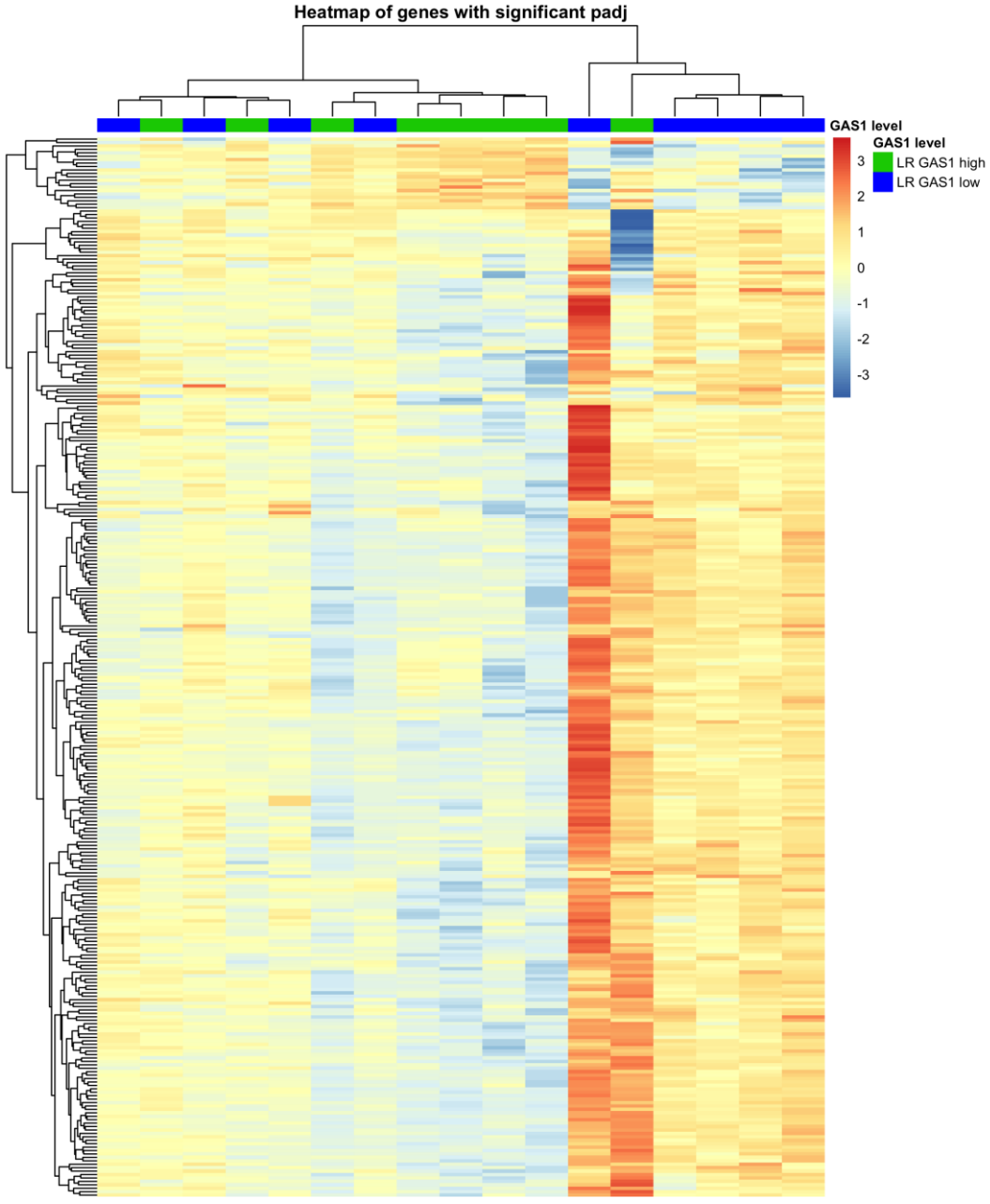


***Figure 32. Heatmap of the 309 genes showing modest but statistically significant expression changes (in LR GAS1-high vs GAS1-low tumors). The x-axis represents individual patient samples, and the y-axis corresponds to the differentially expressed genes. The top annotation bar indicates GAS1 level (blue: GAS1-low; green: GAS1-high), while the color scale reflects gene expression levels (from low, blue, to high, red).***

**Overall, *GAS1*-high LR tumors display a transcriptionally "colder" and more quiescent profile, supporting the probable role of *GAS1* as a favorable biological factor associated with delayed recurrence and reduced tumor aggressiveness.**

GSEA was performed to identify biological pathways associated with different levels of *GAS1* expression. As it is summarize in Figure 33, **LR *GAS1*-high tumors showed reduced enrichment of proliferative, metabolic and inflammatory pathways**, such as **MYC targets** (NES = –2.65), **mTORC1 signaling** (NES = –2.39), **oxidative phosphorylation** (NES = –2.17), G2M checkpoint (NES = –2.07), E2F targets (NES = –2.21), IL6/JAK/STAT3 signaling (NES = –1.75) and EMT (NES = –1.58).

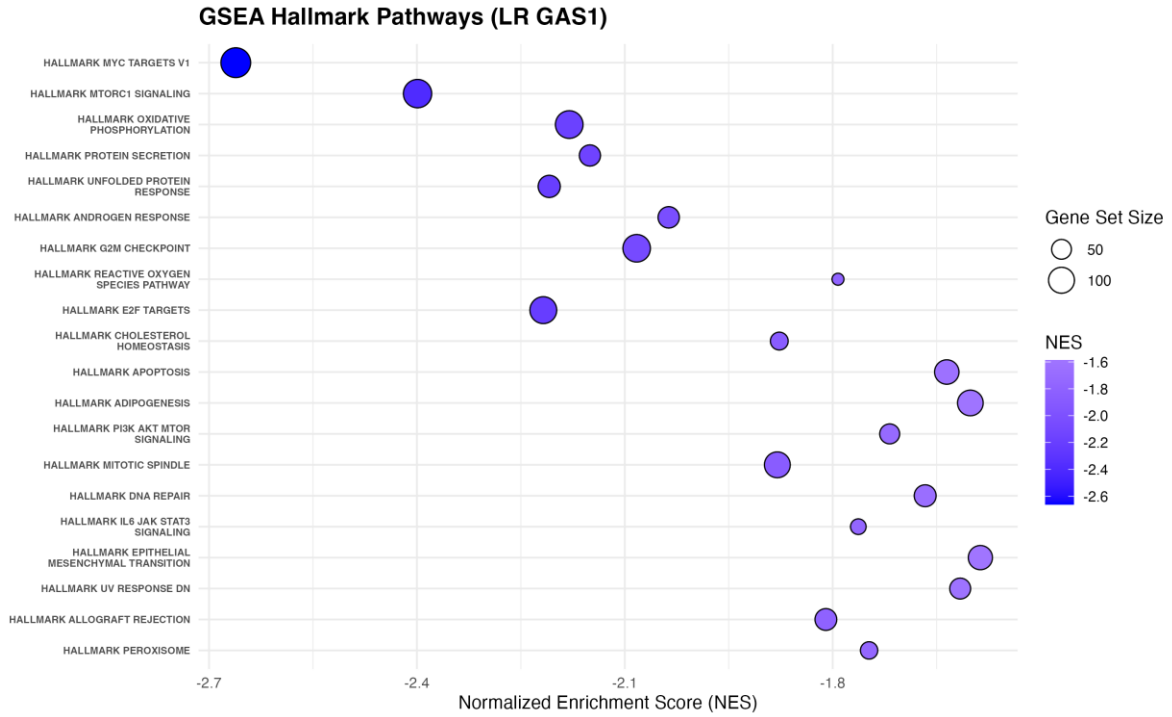


***Figure 33. Dotplot of Hallmark pathways enriched in LR GAS1-high vs GAS1-low tumors. The x-axis shows the NES, while the y-axis lists the Hallmark pathways. Bubble size indicates gene set size, and bubble color reflects the NES value.***

Additionally, minimal pathway enrichment is found, limited to KRAS signaling DN (NES = +1.43), as illustrated in Figure 34.

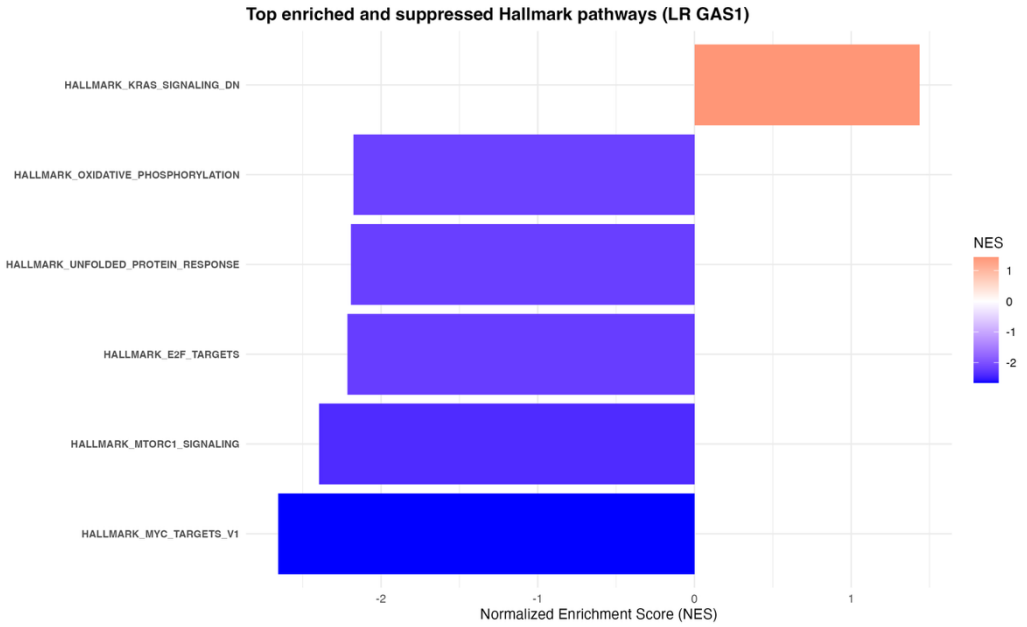


***Figure 34. Barplot of the top enriched and suppressed Hallmark pathways in LR GAS1-high vs GAS1-low tumors. The x-axis represents NES, and the y-axis lists the selected pathways. Bar color indicates the NES value (blue: negative NES; red: positive NES).***

IPA analysis on DEGs in LR tumors stratified by *GAS1* reveals that, while LR tumors with low *GAS1* expression mirror the aggressive profile of early recurrences, ***GAS1*-high tumors exhibit a highly coordinated suppression of pro-tumorigenic pathways.**

In more detail, a key discovery of this analysis is the significant suppression of **EIF2 signaling in LR *GAS1*-high patients.** This indicates that elevated *GAS1* levels effectively impose a strong constraint on the tumor's biosynthetic capacity. This translational brake is demonstrated by the substantial downregulation of genes crucial for protein production, including elongation factors (e.g., *EEF1A1*, Log Ratio -2.9) and several ribosomal proteins (e.g., RPL37, RPL32, RPL23; all Log Ratios < -2.3). By decreasing ribosome formation and translation initiation, *GAS1*-high tumors enter a metabolically dormant state, in contrast to the biosynthetically active and aggressive profile seen in *GAS1*-low tumors.

Upstream regulator analysis further confirms this shutdown of the core machinery driving cancer hallmarks (Table 1). The IPA summary network in Figure 35 synthesizes how *GAS1* high levels shape the functional state of LR tumors.

| Upstream Regulator | Molecule Type | Predicted State (in GAS1-high) | z-score | p-value | Biological Role |
|---|---|---|---|---|---|
| **MYC** | Transcription regulator | **Inhibited** | -4.768 | 1.50×10^−13 | Central driver of cell cycle and metabolism. |
| **TGFB1** | Growth factor | **Inhibited** | -2.697 | 2.00×10^−10 | Primary mediator of EMT and metastasis. |
| **EGFR** | Kinase | **Inhibited** | -2.595 | 6.93×10^−4 | Key receptor for growth and survival. |
| **HIF1A** | Transcription regulator | **Inhibited** | -2.022 | 6.45×10^−7 | Orchestrates hypoxia and angiogenesis. |
| **COL18A1** | Enzyme | **Activated** | 2.000 | 2.44×10^−2 | Endostatin precursor; anti-angiogenic factor. |

***Table 1. Key Upstream Regulators Predicted in LR GAS1-high vs. GAS1-low tumors.***

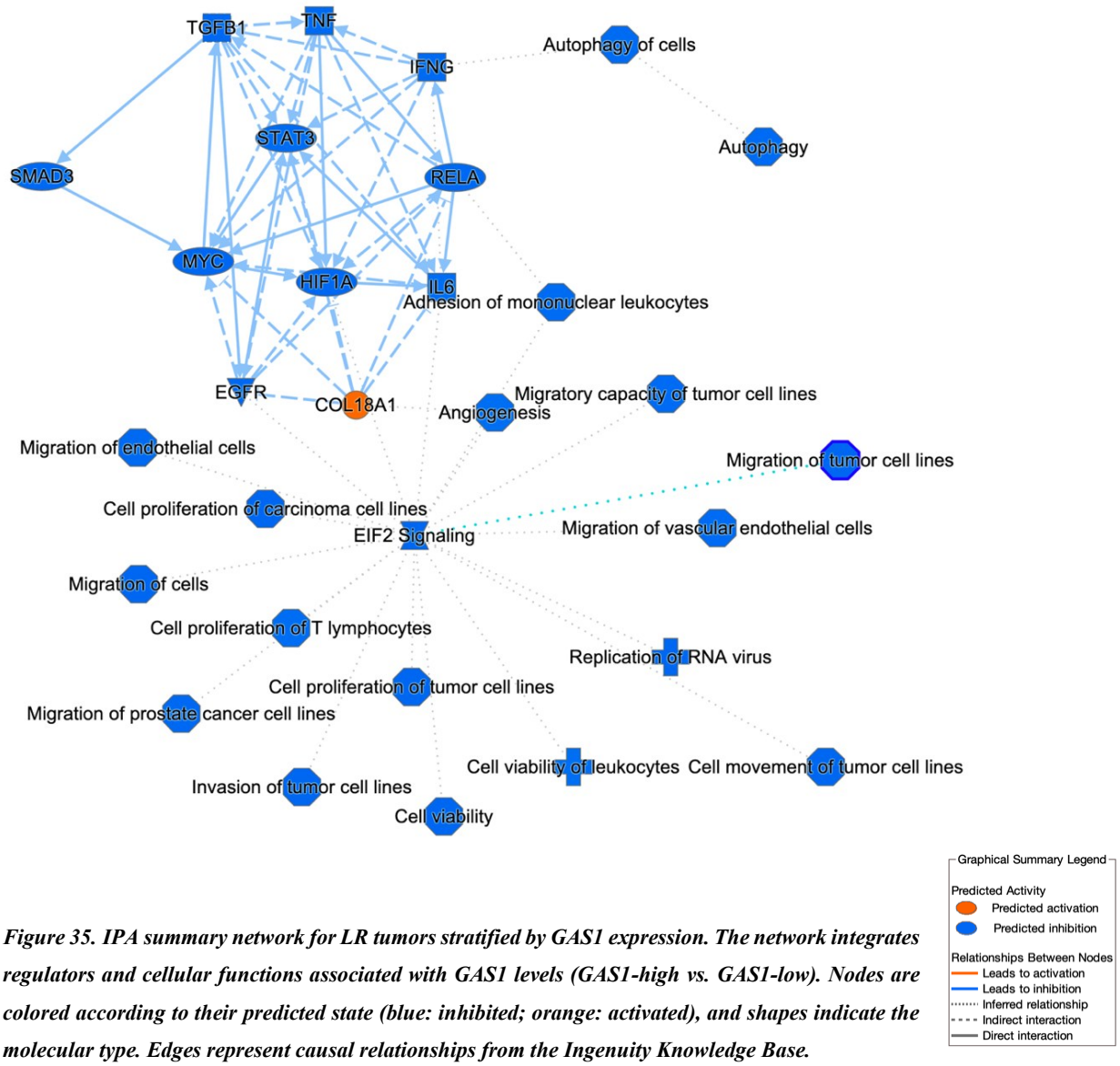


***Figure 35. IPA summary network for LR tumors stratified by GAS1 expression. The network integrates regulators and cellular functions associated with GAS1 levels (GAS1-high vs. GAS1-low). Nodes are colored according to their predicted state (blue: inhibited; orange: activated), and shapes indicate the molecular type. Edges represent causal relationships from the Ingenuity Knowledge Base.***

The network is dominated by inhibited nodes (blue), reflecting a widespread **dampening of aggressive hallmarks** in ***GAS1*-high tumors**, including **migration, invasion, EMT, angiogenesis, proliferation,** and **cell survival**. At the center of the network lies the **inhibition** of **EIF2 signaling**. Its extensive connections to **suppressed downstream functions suggest that reduced biosynthetic activity is a primary driver of dampened tumor motility and proliferation**. Surrounding this central hub, several major master regulators, including ***TGFB1*, *MYC*, *EGFR*, *STAT3***, and ***HIF1A***, are also predicted to be inhibited, consistent with a **reduction in inflammation and tissue remodeling**. *MYC* and *EGFR* are well-established drivers of cell cycle progression and metabolic reprogramming; their downregulation is consistent with a shift toward a quiescent, **less proliferative state** (Gabay et al., 2014; Sigismund et al., 2018). Furthermore, the concomitant inhibition of ***TGFB1*** and ***STAT3*** indicates a significant **reduction in the tumor's ability to undergo epithelial-to-mesenchymal transition and to orchestrate a pro-tumorigenic inflammatory microenvironmen**t (Yu et al., 2009;). Finally, the suppression of ***HIF1A***, a master regulator of the hypoxic response, suggests a **diminished capacity for neo-angiogenesis and metabolic adaptation to low-oxygen conditions** (Semenza, 2013). Conversely, the **activation of *COL18A1*** stands out as the sole orange node; as a well-known **precursor of the anti-angiogenic factor** endostatin, its activation reinforces a stable, **less invasive tissue microenvironment** (Heljasvaara et al., 2017). Overall, this network delineates a clear functional division: while *GAS1*-low tumors engage pro-migratory and anabolic programs, ***GAS1*-high tumors exhibit a coordinated lockdown of tumor-promoting pathways coupled with an active anti-angiogenic signal**, explaining their superior clinical stability. A similar anti-proliferative and anti-migratory role of *GAS1* has also been suggested in ovarian cancer, where *GAS1* was identified as a hub prognostic gene within angiogenesis-related networks. In that study, functional assays showed that its upregulation tended to reduce proliferation and migration of ovarian cancer cells (Zhai et al., 2025). These observations point toward a possible involvement of *GAS1* in restraining angiogenesis-linked tumor progression, a scenario that aligns with the attenuated angiogenic features observed in *GAS1*-high LR tumors.

Finally, I tested whether *GAS1*-high and *GAS1*-low patients presented differences in progression-free survival in either the LR or ER group. Neither analysis revealed any significant differences in the outcome of the two groups.

## 4.7 GeoMx DSP Downstream Analysis

### Data Preprocessing and Technical Variability Assessment in GeoMx DSP Spatial Transcriptomics

The spatial transcriptomic dataset was successfully generated and processed, providing a comprehensive overview of gene expression across multiple tissue regions and cellular compartments. During the initial data import, three AOIs were excluded because they had no detectable counts, suggesting a technical failure. Following probe-level aggregation to gene-level expression, the dataset comprised **18,676 genes across 528 areas of interest**, representing multiple patients, tissue regions, and distinct cellular compartments, including epithelial tumors ($CK^+$), macrophages ($CD68^+$), fibroblasts ($1A4^+$), and lymphoid segments (Figure 36). The distribution of AOIs across LR and ER groups within each morphological segment is visualized in Figure 37, through a Sankey plot, highlighting the relative contribution of each group to the $CK^+$, $CD68^+$, $1A4^+$, and lymphoid compartments.

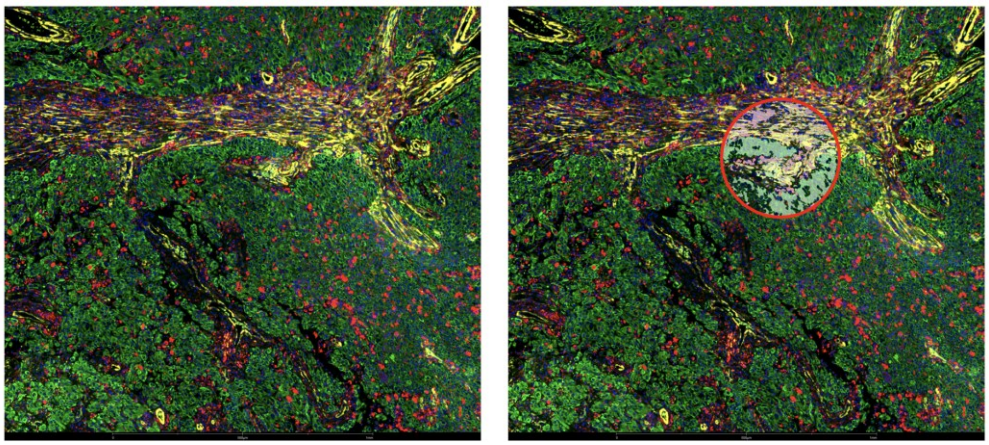

***Figure 36. Exemplar morphological segmentation of AOIs based on immunofluorescence markers. In this example, nuclear staining with SYTO13 (FITC/525 nm, blue) identifies all cells. PanCK (Cy3/568 nm, Alexa 532, green) marks epithelial tumor cells. CD68 (Texas Red/615 nm, Alexa 594, red) highlights macrophages, while alpha SMA_1A4 (Cy5/666 nm, Alexa 647, yellow) identifies fibroblasts. The right panel shows a selected AOI (red circle) used for transcriptomic profiling.***

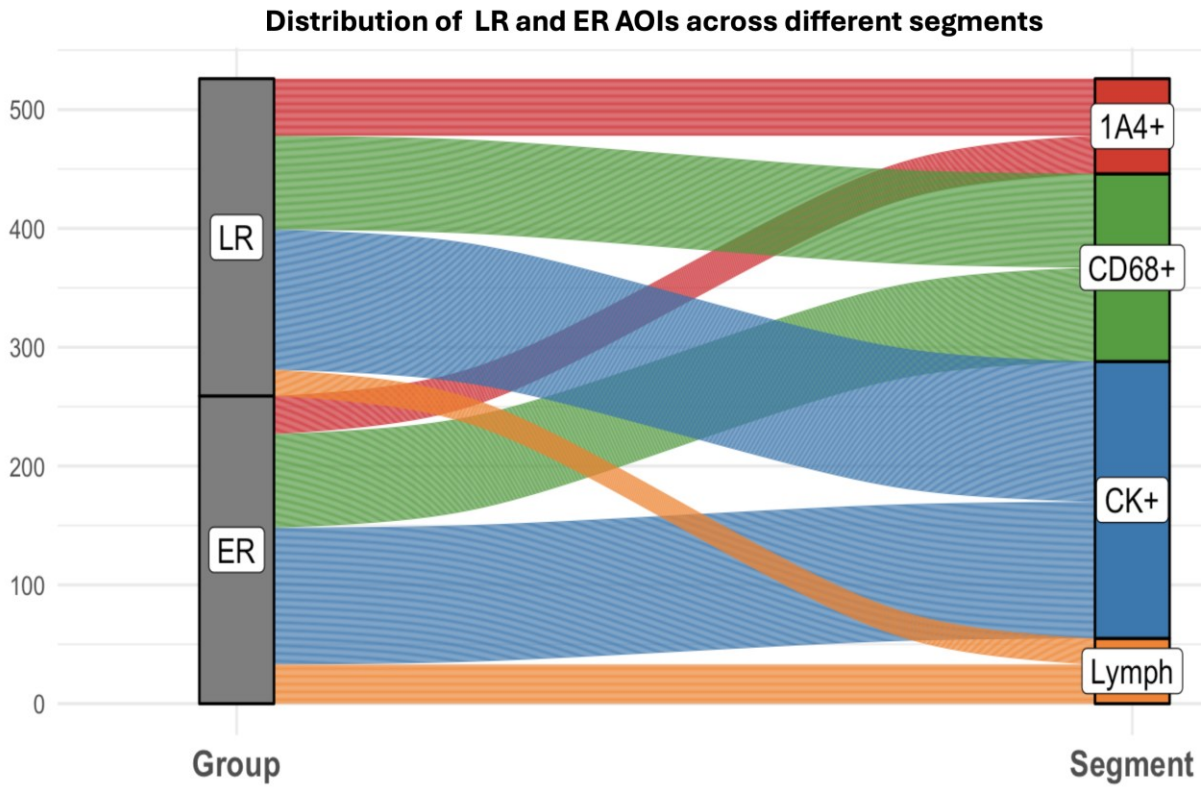


***Figure 37. Sankey plot illustrating the distribution of AOIs from the two transcriptional groups (LR and ER) across the four morphologically defined segments. LR ROIs (left, grey) flow into epithelial CK⁺ segments (blue, n = 118), macrophage CD68⁺ segments (green, n = 79), fibroblasts 1A4⁺ segments (red, n = 48), and lymphoid segments (orange, n = 22). ER ROIs (left, dark grey) show a comparable distribution, contributing 115 CK⁺, 79 CD68⁺, 32 1A4⁺, and 33 lymphoid ROIs. The color scheme reflects the segment identity: CK⁺ (blue), CD68⁺ (green), 1A4⁺ (red), and lymphoid (orange).***

A comprehensive quality control analysis was subsequently performed at the ROI level. Several metrics were evaluated, including library size, number of detected genes, and the proportion of lowly expressed genes. Based on these criteria, regions with insufficient sequencing depth (library size <50,000 counts) or low nuclear content were excluded. After filtering, **526 high-quality AOIs were retained,** indicating that most of the dataset met the required quality standards. In addition, the relationship between AOI surface area and library size was evaluated. As expected, a positive correlation was observed, reflecting the fact that larger regions contain more cells and therefore yield higher RNA counts.

To assess potential technical variability and batch effects, RLE distributions were examined (Figure 38). It revealed that, although overall variability was comparable across samples, RLE values differed slightly between slides, indicating the presence of marginal **slide-associated technical effects**. These systematic shifts are consistent with the independent processing of each slide in the GeoMx workflow and support the use of slide identity as a batch factor to be

corrected. Importantly, in the dataset, slides do not correspond to individual patients or clinical groups. Each slide contains AOIs originating from multiple patients, and each patient is represented across more than one slide. For example, slide S162371_S1310605 includes AOIs from both PT_09 (early recurrence) and PT_10 (long responder), while slide S154245_S123159 contains AOIs from PT_11 (ER) and PT_12 (LR). Clinical groups are therefore interspersed across slides, and sequencing runs also include AOIs from multiple slides and patients.

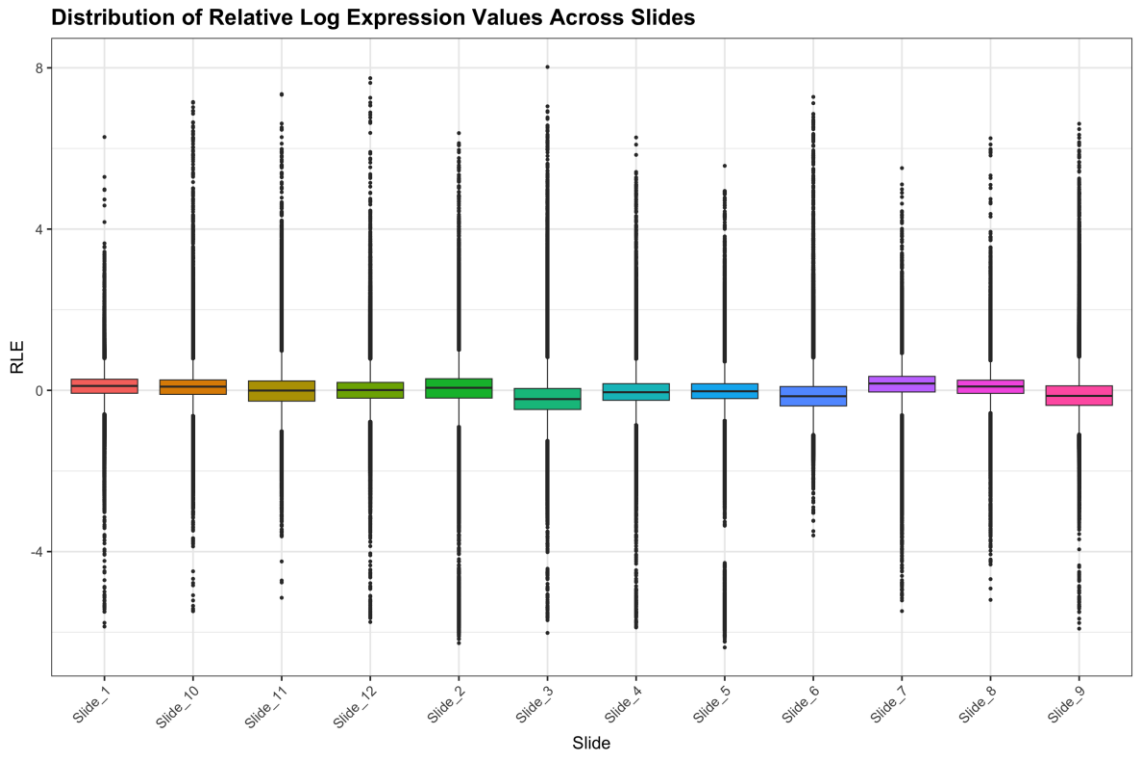


***Figure 38. Boxplots showing the distribution of RLE values for each slide (x-axis: slide; y-axis: RLE).***

To further investigate whether the potential batch effect of slides could hamper analysis results, an unsupervised analysis was performed to check the natural grouping of samples based on their transcriptional patterns.

## Unsupervised Clustering Visualization

Hierarchical agglomerative clustering of the top highly variable genes, shown in Figure 39, revealed a complex continuous structure among AOIs. While the column dendrogram exhibited multiple branching points suggesting three or more potential subclusters, a two-cluster solution (k = 2) provided the optimal balance between within-cluster cohesion and between-cluster separation. Consequently, the dendrogram was partitioned at k = 2, resulting in two broad clusters of unequal density, one notably more homogeneous than the other. When we look at how the samples are grouped, they do not follow the relapse status; ER and LR groups are mixed, meaning there is no clear transcriptomic separation between them at this stage. Other factors, such as sex, also show no relevant influence on the overall grouping. While we see some small clusters based on cell types (such as CK+ or CD68+), these represent only local patterns. In other words, although similar cells sometimes group in small branches, they are not the main factor defining the two large clusters in the dendrogram. In contrast, slide identity determined a clear and consistent influence; in fact, AOIs originating from the same slide tended to cluster together, forming visual blocks along the heatmap. This slide-driven grouping represented the strongest and most systematic pattern in the unsupervised structure and mirrored the mild but detectable slide-specific variation observed in the RLE analysis. Overall, **the heatmap highlighted the presence of slide-related batch effects and showed that, although some local biological patterns were present, technical variation remained the primary driver of clustering**. This motivated the application of batch correction, here performed using RUV4 with an empirically determined value of k, before re-evaluating the clustering structure and drawing biologically meaningful conclusions.

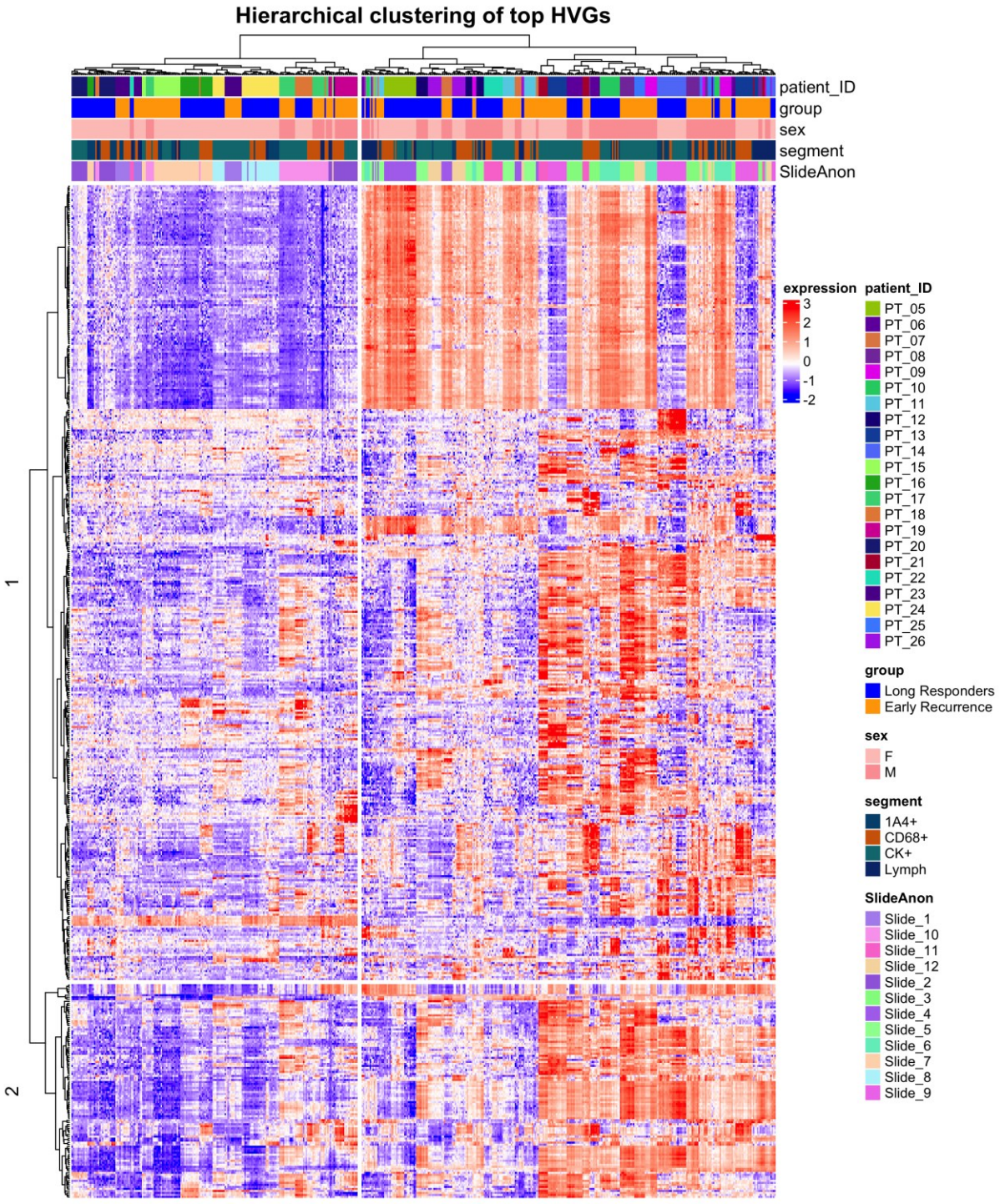


***Figure 39. Unsupervised hierarchical clustering of top HVG: It shows a heatmap of variable gene profiles across 526 AOIs. The x-axis depicts AOIs; the y-axis lists clustered genes. The top dendrogram divides samples into two clusters (k=2) based on expression. Metadata bars show patient ID, relapse status (LR vs. ER), sex, tissue segment (CK+, CD68+, 1A4+, Lymph), and slide ID (SlideAnon). The color scale indicates Z-score expression from low (blue) to high (red).***

## PCA-based Assessment of Slide-Associated Batch Effects and RUV4 Correction

As previously described, in the GeoMx DSP platform, each patient can be represented across multiple slides, and each slide contains only a portion of that patient's tissue. However, even though these slides originate from the same biological sample, they are not technically equivalent. Each slide undergoes an independent experimental cycle (including separate WTA hybridization, morphology staining, imaging, UV photocleavage, and processing under slightly different temperature, humidity, and laser conditions). Slides also differ in background autofluorescence and staining quality. Even small variations across these steps accumulate and generate systematic patterns that reflect technical processing rather than biological differences. These slide-specific effects are also clearly visible in the pre-correction PCA (Figure 40).

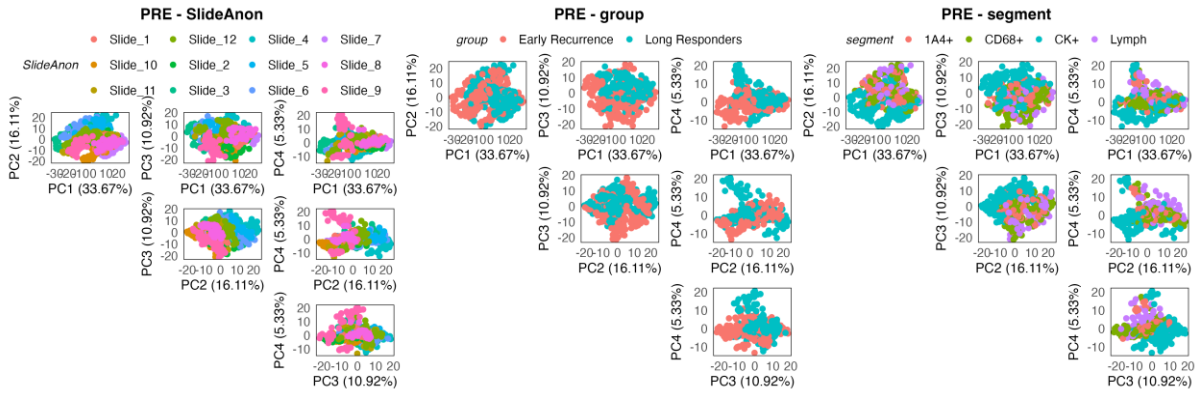


***Figure 40. PCA before data correction: It illustrates the distribution of the areas of interest across the first four principal components to identify the primary sources of variance within the dataset. The x-axis and y-axis represent the PCs that capture the highest percentage of transcriptomic variation. In the left panel, samples are colored by slide identity to highlight technical batches. The center panel classifies samples by relapse status, using red for early recurrence and light blue for long responders. The rightmost panel colors the areas of interest by morphological segment, with red for fibroblasts (1A4+), green for macrophages (CD68+), light blue for epithelial cells (CK+), and purple for the lymphoid compartment.***

**When AOIs are colored by slide, they form tight, non-overlapping clusters, indicating that slide identity is the dominant source of variance**. This is reflected in the principal components: PC1 explains 33.67% of the total variance, PC2 explains 16.11%, PC3 explains 10.92%, and PC4 explains 5.33%, with PC1 and PC2 together accounting for nearly half of the total variability.

When AOIs are colored by group (early recurrence vs long responders), a mild degree of separation becomes visible, suggesting that relapse status contributes to some biological variability. However, the two groups remain intermixed across the principal components.

Coloring AOIs by morphological segment reveals clearer biologically driven patterns, as epithelial ($CK^+$), macrophage ($CD68^+$), fibroblasts ($1A4^+$), and lymphoid regions tend to cluster according to their expected transcriptional identities. Nevertheless, this biological structure is a little bit overshadowed by the much stronger slide-specific clustering, confirming that technical variation introduced by slide processing remains the primary driver of variance before batch correction.

To determine the appropriate number of unwanted factors (k) for RUV4, the silhouette score was evaluated across values of k from 1 to 10. In this context, as it is shown in Figure 41, the silhouette score quantifies how strongly AOIs cluster according to slide identity: higher values indicate persistent slide-driven separation, whereas lower values reflect attenuation of this technical effect. The score decreased markedly at k = 2, indicating that two unwanted factors were sufficient to substantially reduce slide-specific variation.

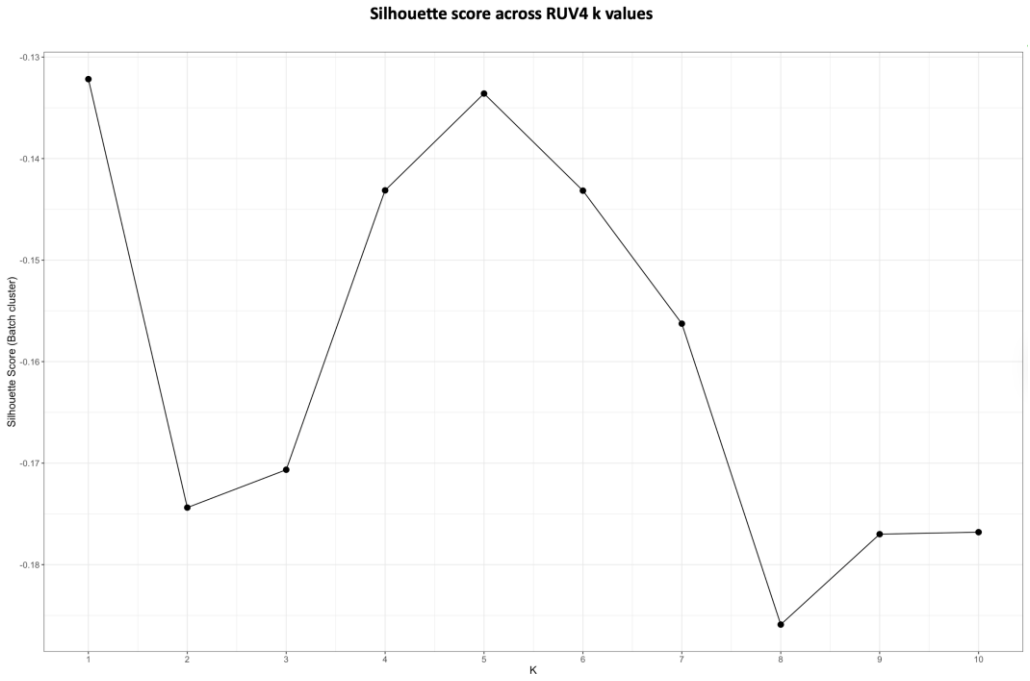


***Figure 41. Evaluation of the silhouette score to optimize the k parameter in RUV4: the graph displays the structural evaluation of clustering strength across different model settings. The x-axis represents the number of unwanted factors of variation, indicated as k, ranging from 1 to 10. The y-axis shows the silhouette score, a metric used to quantify the density and separation of sample clusters based on their slide of origin. Each point on the line plot represents the calculated score for a specific k value, providing a technical basis for selecting the most appropriate level of correction.***

Increasing k beyond this point did not further improve correction; at higher values, such as k = 8, the slide effect might also be minimized, but at the obvious risk of overcorrection

and the removal of genuine biological signal. For this reason, k = 2 was selected as the optimal compromise, effectively reducing artificial slide-driven transcriptional artifacts but preserving the underlying biological, morphological, and clinical patterns.

After applying RUV4 with *k = 2*, the overall variance structure of the dataset changes, reflecting the removal of a substantial portion of slide-related technical noise. This reduction is visible in the post-correction PCA (Figure 42), where the proportion of variance explained by the first components decreases (PC1 = 29%, PC2 = 11.07%, PC3 = 7.41%, PC4 = 6.38%).

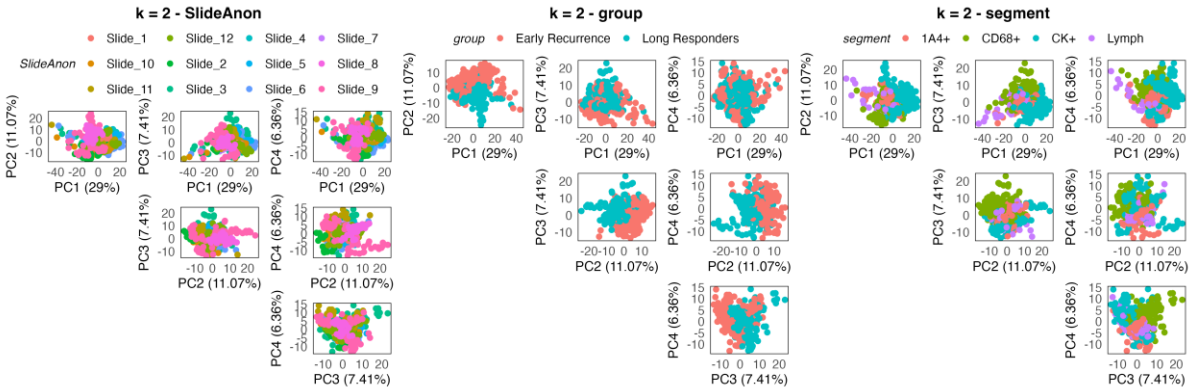


***Figure 42. PCA after RUV4 correction: this figure shows the reorganized variance structure of the dataset following the application of the RUV4 algorithm with k = 2. Like the previous PCA, the axes represent the principal components, but with updated percentages of explained variance. The panels maintain the same coloring scheme for comparison.***

The lower percentages indicate that the technical variance previously dominating the dataset has been attenuated, leaving a residual variance that is more evenly distributed across components. **When AOIs are colored by slide, the distinct clusters observed before correction are no longer present, and AOIs from different slides appear more intermixed, indicating that slide identity no longer drives the global variance structure**. Coloring AOIs by relapse state shows a distribution that remains partially overlapping, but without the distortion introduced by slide-specific effects. Importantly, the expected biological structure is preserved: epithelial, macrophage, stromal, and lymphoid segments still cluster according to their transcriptional identities. This confirms that using k = 2 reduces technical variation without compromising the underlying biological patterns. If a much higher number of unwanted factors had been used, the explained variance of the first PCs would have dropped far more dramatically, suggesting that not only technical but also biological signal was being removed. To verify that RUV4 effectively attenuated the slide-associated technical variation without compromising the biological signal, a new hierarchical agglomerative clustering was performed on the normalized data. This post-correction analysis, visualized in the heatmap in

Figure 43, provides a direct comparison with the pre-correction clustering (Figure 39) and allows for an assessment of the corrected data structure.

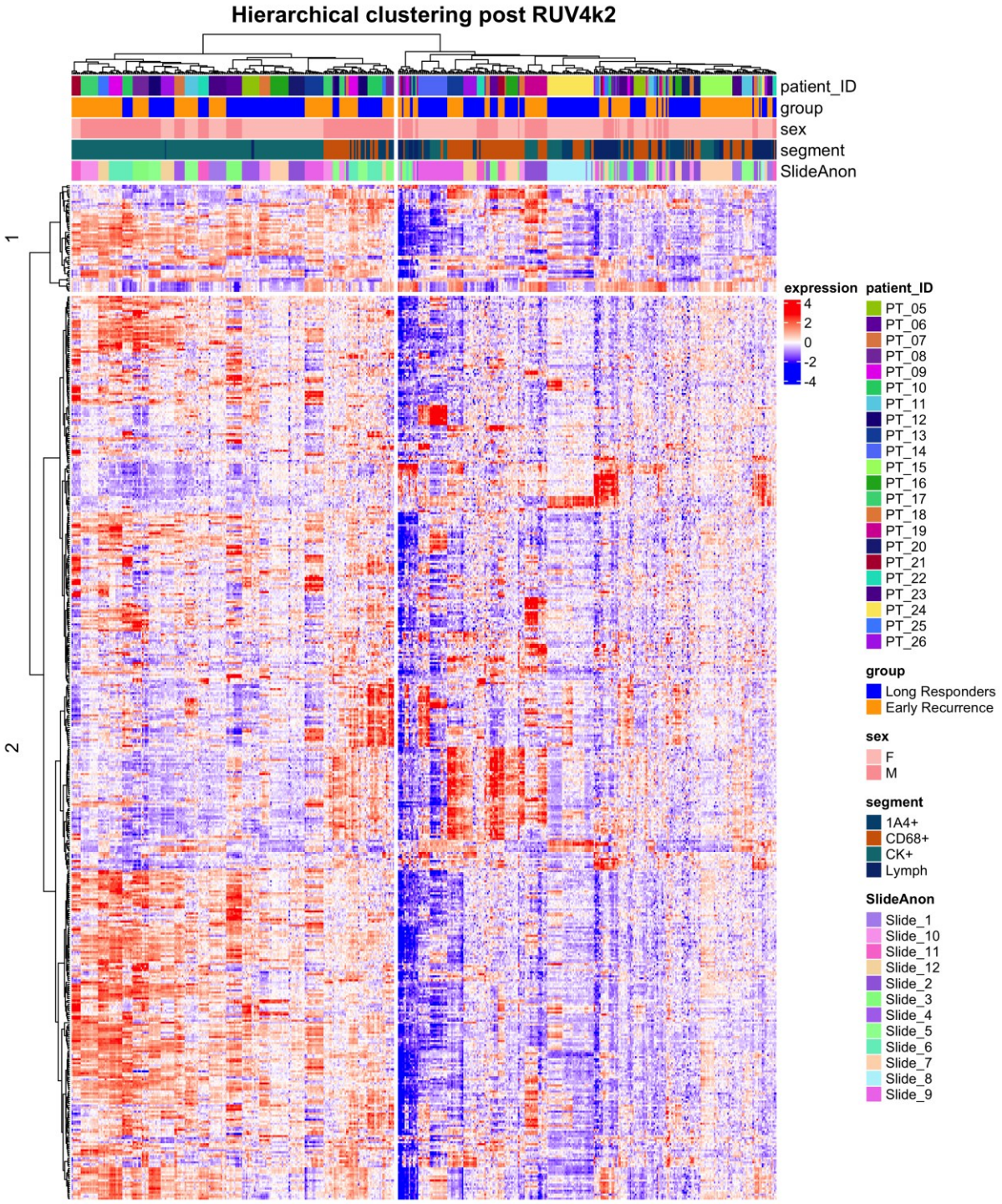


***Figure 43. Unsupervised hierarchical clustering post RUV4 normalization with k = 2. This heatmap maintains the same technical structure and legend as Figure 39 to allow for a direct comparison. The x-axis represents the areas of interest (AOIs), and the y-axis shows the highly variable genes. The metadata bars at the top identify the patient ID, clinical group (Early Recurrence in orange, Long Responders in blue), sex, morphological segment (CK+, CD68+, 1A4+, Lymph), and slide identity. The color scale represents standardized gene expression levels (Z-scores).***

AOIs from the same slide don't form cohesive blocks anymore; instead, the SlideAnon bar shows a clear intermixing of colors across the entire dendrogram. This confirms the successful removal of technical noise. **The benefit of the RUV4 correction is most visible in the epithelial (CK+) samples**. After correction, **these samples group together into one clear and solid cluster**, which didn't happen before. This shows that the unique genetic signature of the epithelial cells, which was previously hidden by technical noise, has been re-established.

## Preliminary Insights from Spatially Resolved DEA

The GeoMx DSP platform enabled the isolation of transcriptional signals from specific cellular compartments, bypassing the inherent resolution limits of bulk RNA-seq. The technical integrity of the dataset was confirmed by a strong correlation between the physical parameters of the areas of illumination and sequencing performance metrics. Morphological assessment revealed a highly heterogeneous tissue architecture. The selected AOIs spanned surface areas from ~4,000 $\mu m^2$ to ~117,000 $\mu m^2$, with nuclei distributions reflecting the distinct biological densities of each segment: (i) lymphoid segments, represented the most densely populated regions, with nuclei counts reaching peaks of $> 2,600$ in organized structures and maintaining a robust average above 600 nuclei; (ii) tumor segments, exhibited variable cellularity, typically ranging between 200 and 800 nuclei, depending on the growth pattern and compactness of the epithelioid mesothelioma; (iii) macrophages and fibroblasts segments, were characterized by lower cellular density.

To understand whether any transcriptional pattern might be evidence in association with recurrence, a differential expression analysis was performed on the different cell types in a spatial context. The $CK^+$ tumor compartment was evaluated to identify intrinsic molecular drivers of aggressiveness, distinguishing ER from LR cells. Within the lymphoid niches, the investigation focused on whether the higher TIL levels in ER patients reflected a functional immune response or a state of chronic, ineffective inflammation. $CD68^+$ myeloid cells were isolated to assess if specific macrophage polarizations contribute to defending the tumor from immune attack in ER cases, while the $1A4^+$ fibroblasts compartment was included to investigate whether a more reactive and denser extracellular matrix in ER cases promotes a permissive environment for tumor infiltration.

In the following sections, Figure 44 (A, B, C) presents the volcano plots, which utilize a dual threshold, padj and log2FC. This restrictive approach identifies the driver genes with the most pronounced and statistically significant expression changes: only 13 genes in the tumor

compartment, 4 in the macrophages, and 3 in the lymphoid areas reached this high-stringency cut-off, 0 DEGs in fibroblasts comparison analysis.

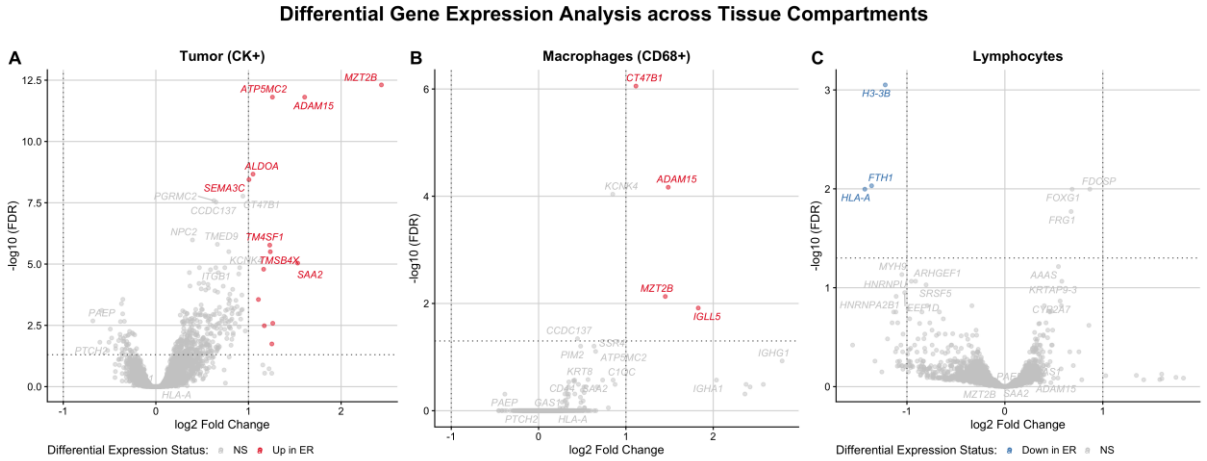


***Figure 44. Volcano plots of differential expression across compartments: the vertical dashed line marks the $|log_2FC| = 1$ threshold, and the horizontal dashed line indicates FDR = 0.05. Axes represent $log_2$ fold change (ER vs LR) on the x-axis and $-log_{10}$(FDR) on the y-axis. Genes significantly upregulated in ER are shown in red, whereas genes downregulated in ER are shown in blue; non-significant genes are displayed in grey.***

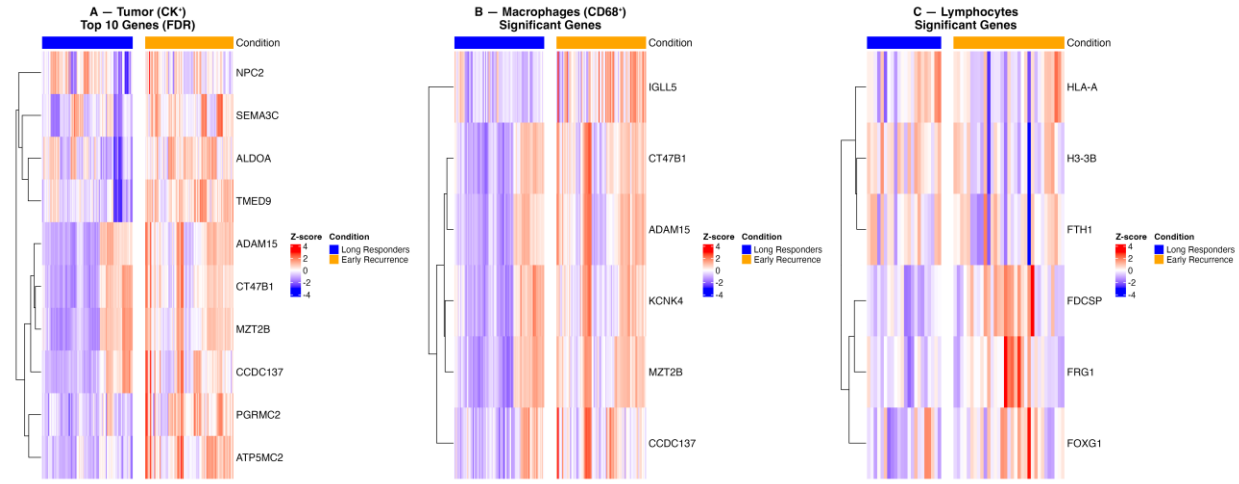


***Figure 45. Heatmaps of compartment-specific transcriptional signatures in ER and LR patients: (A) Top 10 genes by significance (FDR < 0.05) in $CK^+$ tumors. (B) Six significantly associated genes (FDR < 0.05) in $CD68^+$ macrophages. (C) Six significantly associated genes (FDR < 0.05) in lymphoid regions. Rows are genes; columns are AOIs. Expression values are Z-scores, scaled across all samples, to show relative differences. The condition bar shows the relapse status: ER (orange) and LR (blue).***

In contrast, Figure 45 (A, B, C) features heatmaps displaying the top genes ranked solely by their statistical significance (FDR < 0.05), regardless of their absolute fold change. This visualization includes the top 10 of 364 significant genes for the tumor section, and 6 genes each for the $CD68^+$ and lymphoid segments. Visual inspection of the heatmaps reveals a

sharp chromatic separation of Z-scores between ER and LR groups across all analyzed compartments.

In the $CK^+$ **tumor compartment**, the transcriptional landscape is **dominated by an aggressive state specific to ER patients**. As shown in Table 2, these tumors **activate genes for rapid growth and metabolic reprogramming.** Key findings include ***MZT2B*** upregulation, which regulates the **mitotic spindle and indicates increased replication**. *MZT2B* also boosts malignant traits by enhancing **mitochondrial bioenergetics** via *COX5B*, providing **fuel for faster mitosis** (Ding et al., 2025). This is supported by a Warburg-like metabolic shift, with increased glycolytic and mitochondrial subunits **(*ALDOA, ATP5MC2*). *ALDOA*'s upregulation sustains glycolysis** and **helps create an immunosuppressive, macrophage-driven niche**, matching the aggressive ER profile (Ji et al., 2025).

| Compartment | Gene | Full Name | Direction | log2FC |
|---|---|---|---|---|
| Tumor | MZT2B | Mitotic Spindle Organizing Protein 2B | Up in ER | 2.435 |
| | ADAM15 | ADAM Metallopeptidase Domain 15 | Up in ER | 1.606 |
| | ALDOA | Aldolase, Fructose-Bisphosphate A | Up in ER | 1.049 |
| | ATP5MC2 | ATP Synthase Membrane Subunit C2 | Up in ER | 1.259 |
| | SAA2 | Serum Amyloid A2 | Up in ER | 1.531 |
| | SEMA3C | Semaphorin 3C | Up in ER | 1.005 |
| | TMSB4X | Thymosin Beta-4 X-linked | Up in ER | 1.236 |

***Table 2: Genes upregulated in the ER within tumor compartment.***

| Compartment | Gene | Full Name | Direction | log2FC |
|---|---|---|---|---|
| Macrophages | ADAM15 | ADAM Metallopeptidase Domain 15 | Up in ER | 1.485 |
| | IGLL5 | Immunoglobulin Lambda-Like Polypeptide 5 | Up in ER | 1.828 |
| | MZT2B | Mitotic Spindle Organizing Protein 2B | Up in ER | 1.451 |
| | CT47B1 | Cancer/Testis Antigen 47B1 | Up in ER | 1.115 |

***Table 3. Genes upregulated in the ER within the macrophages segment.***

**Crucially, this aggressive program is not confined to the neoplastic cells but is mirrored in the tumor microenvironment**. In the $CD68^+$ macrophage compartment (Table 3), all significant DEGs are uniquely associated with the ER phenotype. Most notably, the **simultaneous upregulation of *ADAM15* and *MZT2B* in both tumor and myeloid cells** identifies a synchronized remodeling axis.

The shared expression of *ADAM15*, a **metalloproteinase known to facilitate cell-cell interactions and matrix degradation, suggests that ER-associated macrophages are functionally "reprogrammed" to act as facilitators of tumor invasion** rather than mediators of an anti-tumor response.

In contrast, **the lymphoid compartment** highlights the **immunological advantage of LR patients** (Table 4). The **high expression of *HLA-A* in LR pts** underscores a **preserved antigen-presentation machinery**, allowing the **adaptive immune system to maintain control** over residual disease (Dhatchinamoorthy et al., 2021). Conversely, the **absence of these activation markers in ER lymphoid niches** suggests a state of **immune ignorance or dysfunction, leaving the microenvironment permissive to rapid progression.** In addition to *HLA-A*, **LR lymphoid niches** also showed **increased expression of ferritin genes, particularly *FTH1***. This result is consistent with an immunologically active microenvironment; in fact, ferritin upregulation in lymphocytes is a well-described marker of metabolic engagement and oxidative-stress handling during immune activation (Sacco et al., 2021). Notably, this pattern contrasts with the bulk LR *GAS1*-high tumors, where *FTH1* and *FTL* were markedly downregulated, reflecting a metabolically quiescent, iron-depleted state. Together, these observations indicate that LR patients retain metabolically competent and functionally responsive lymphocytes. In contrast, **ER tumors' microenvironment adopts a low-iron, low-metabolism program that may facilitate immune evasion.** To summarize, these results reveal that **early recurrence** is probably driven by a multifaceted program: **an aggressive tumor-intrinsic state, a supportive macrophage niche, and a potentially dysfunctional lymphoid environment.**

| Compartment | Gene | Full Name | Direction | log2FC |
|---|---|---|---|---|
| **Lymphocytes** | **HLA-A** | Major Histocompatibility Complex, Class I, A | Up in LR | -1.430 |
| | **FTH1** | Ferritin Heavy Chain 1 | Up in LR | -1.361 |
| | **H3-3B** | H3.3 Histone Variant B | Up in LR | -1.222 |

***Table 4. Genes upregulated in the LR within the lymphocytes compartment.***

# 5. Conclusion

This study aimed to identify the characteristics that distinguish patients with early recurrence from those with a prolonged response following CRS-HIPEC, and to determine whether intrinsic tumor or microenvironmental factors could contribute to this heterogeneity.

## The Dual Transcriptional Profile of e-DMPeM

The results of this study reveal that the clinical trajectory of epithelioid-DMPeM is governed by a deep molecular dichotomy, where the balance between metabolic drive and homeostatic restraint determines the timing of recurrence. Bulk RNA-seq data show that ER is fueled by intense biosynthetic activation. In these patients, the tumor orchestrates rapid growth (driven by MYC, E2F, and mTORC1) together with marked structural plasticity, as reflected by the activation of EMT and angiogenesis. This phenotype extends beyond the cancer cells to the surrounding microenvironment, in which fibroblasts acquire an activated, matrix-remodeling myCAF state. Through spatial transcriptomics, it was possible to observe how this aggressiveness manifests as a collaboration between different cell types. At the tumor core ($CK^+$), genes such as *ALDOA* and *MZT2B* sustain accelerated metabolism and cell division. Simultaneously, within the macrophage niches ($CD68^+$), the expression of *ADAM15* suggests these immune cells facilitate tissue invasion rather than oppose it, clearing the path for disease progression. On the opposite, LR is rooted in deep quiescence, where *GAS1* acts as a guardian of tumor indolence. The regulatory network of *GAS1*-high tumors is characterized by a coordinated lockdown of pro-tumorigenic pathways, centered on the inhibition of EIF2 signaling. This central suppression, coupled with the downregulation of master regulators like *MYC*, *EGFR*, and *HIF1A*, drives a shift toward a non-proliferative state and reduced neo-angiogenesis.

*GAS1* expression in LR patients could impose a metabolic block: tumor cells markedly reduce protein synthesis and iron utilization (*FTH1*/*FTL*), entering a metabolically "cold" and stable state. Within the lymphoid niches of LR patients, lymphocytes appear to preserve their antigen-recognition capacity and to manage oxidative stress more effectively, possibly through a more balanced regulation of ferritin genes. This combination suggests that, while the tumor becomes metabolically quiescent, the immune compartment remains functionally competent. In short, epithelioid mesothelioma oscillates between a proliferative state and a *GAS1*-mediated quiescent state. It is the ability to maintain this metabolic and immunological balance that ultimately could determine long-term clinical stability.

## *GAS1* as a Candidate Regulator of Tumor Aggressiveness: Future Investigations

The most important result that emerged from DE analysis and IHC validation, is the overexpression and production of growth arrest-specific protein 1 in long-responder patients. It is reduced in ER tumors and strongly influences transcriptional programs, even within the LR group. It appeared as a potential regulator of tumor aggressiveness and is certainly a relevant candidate as a prognostic marker. Evidence from other tumor types supports this interpretation. In colorectal cancer, for example, *GAS1* has been reported as a downstream target of *TCF7L1*, and its repression appears necessary to sustain migratory and invasive programs. Restoring *GAS1* expression in this context tends to attenuate these aggressive phenotypes, suggesting that *GAS1* may contribute to restraining tumor progression (King et al., 2024).

Future investigations should explore the functional role of *GAS1* through in vitro studies, leveraging the experimental tools already available for this gene. In parallel, emerging nanotechnology-based approaches further expand the potential applicability of *GAS1*. Recent studies have shown that *GAS1* can be delivered using Atosiban-conjugated three-way junction RNA nanoparticles (3WJ-pRNA), enabling targeted release of *GAS1*-enriched extracellular vesicles (Y.-X. Li et al., 2025). This platform appears stable and biocompatible, supporting the feasibility of RNA-based delivery strategies. Although developed outside oncology, these findings suggest that similar nanoparticle-mediated approaches could be explored in peritoneal mesothelioma.

The discrepancy between bulk RNA-seq and spatial transcriptomics for *GAS1*, whereby it does not emerge as a DEG in the spatial analysis, does not invalidate the bulk findings; rather, it underscores the complementary nature of the two approaches. Bulk RNA-seq captures global, tissue-level differences that may be driven by changes in cellular composition, whereas spatial profiling reveals that *GAS1* is not differentially regulated within any specific compartment. This indicates that the bulk signal likely reflects macro-architectural differences across the tissue, rather than compartment-intrinsic transcriptional modulation.

## Limitations and Future Perspective of Bulk Transcriptomics in DMPeM

This work has some limitations that should be recognized. First, the sample size is inherently constrained by the rarity of DMPM and by the strict selection criteria required for CRS-HIPEC, which limits statistical power, particularly for stratified analyses. Second, both bulk and spatial transcriptomic data were generated from FFPE material, which, while clinically valuable, presents technical challenges regarding RNA integrity and sequencing depth.

Bulk RNA-seq, despite its power to capture global trends, averages signals across heterogeneous cell populations. In a disease like epithelioid-DMPeM, where epithelial, stromal, and immune components are tightly interwoven, this approach can mask cell-type-specific programs or rare, spatially restricted states that influence recurrence.

Furthermore, short-read sequencing offers limited resolution of transcript isoforms. Given that pleural mesothelioma frequently exhibits complex alternative splicing and chromosomal rearrangements, DMPeM likely shares a similar genomic complexity that standard RNA-seq cannot fully resolve. This limitation is particularly relevant in mesothelioma biology; large genomic studies of malignant pleural mesothelioma have shown extensive alternative splicing, splice-altering mutations, and fusion transcripts (Bueno et al., 2016). In parallel, structural genomic analyses have revealed that pleural mesothelioma frequently exhibits complex chromosomal rearrangements such as chromothripsis and chromoplexy, which generate aberrant transcripts and neoantigenic peptides that are difficult to detect with standard short-read RNA-seq (Mansfield et al., 2019). Although equivalent analyses are lacking for diffuse malignant peritoneal mesothelioma, the strong molecular similarity between the two diseases suggests that comparable splicing complexity and structural variation are likely to be present. Future studies integrating long-read RNA sequencing, whole-genome structural profiling, and isoform-aware approaches will therefore be essential to fully characterize transcript diversity, fusion events, and their contribution to clinical heterogeneity in DMPeM.

## Future Directions for Spatial Transcriptomics in Peritoneal Mesothelioma

Spatial transcriptomics represents a transformative opportunity for peritoneal mesothelioma research, allowing the dissection of the complex cellular architecture where malignant, immune, and stromal cells interact. Unlike bulk RNA-seq, these technologies anchor gene expression to precise tissue coordinates, mapping molecular profiles back to their anatomical context (X. Li & Wang, 2021). While standard clustering and dimensionality reduction can be applied, the true power of these datasets lies in integrating spatial proximity with expression data to infer communication networks and microenvironmental niches shaped by tissue architecture (B. Liu et al., 2022). However, significant limitations remain, including high costs, complex experimental workflows, and a general lack of standardized bioinformatic pipelines for data integration (Merritt et al., 2020). In addition, many platforms still struggle to achieve true single-cell resolution, complicating deconvolution in dense tissue areas. Despite these hurdles, this approach is essential for identifying the spatial determinants of immune dysfunction and therapeutic resistance. While previous studies have focused almost exclusively on pleural mesothelioma (Torricelli et al., 2023), this work is among the first to apply whole-transcriptome spatial profiling to diffuse malignant peritoneal mesothelioma. The compartmental metadata generated in this study provides a powerful foundation for high-resolution spatial analyses. A key strength of this dataset lies in the inclusion of mixed-population AOIs. The integration of functional enrichment analyses will represent a crucial next step for the future; such analyses would allow for a transition from a gene-centric view to a broader understanding of the biological networks governing peritoneal mesothelioma progression. Moreover, a central objective for future research, will be to determine whether the immunological differences observed are local or systemic. By comparing the profiles of lymphocytes within peritumoral niches (in direct contact with $CK^+$ cells) against those in isolated, distant lymphoid segments, we can clarify whether the "metabolic fitness" of LR patients is an intrinsic systemic trait or if, in ER cases, lymphocytes are functionally "switched off" only upon physical contact with the aggressive tumor mass. Similarly, these mixed AOIs allow for mapping macrophage diversity based on their "neighborhood." Future studies could verify if proximity to CAF-rich regions induces a more immunosuppressive or matrix-remodeling phenotype, acting as a biochemical barrier that facilitates early recurrence. This thesis offers a new interpretative model for DMPeM. It suggests that clinical stability depends on the ability to preserve a homeostatic balance between the tumor, the stroma, and

the immune system. Validating these spatial interactions remains the crucial next step in translating these biological observations into concrete clinical strategies.

# 6. Bibliography


- Adusumalli, R., & Banala, R. R. (2025). Cancer-associated fibroblasts at the crossroads of tumor progression and therapy resistance: From heterogeneity to precision reprogramming. *Journal of the Egyptian National Cancer Institute*, *37*(1), 81. https://doi.org/10.1186/s43046-025-00334-7
- Ambrosetti, M. C., Bariani, M., Zamboni, G. A., Valletta, R., & Bonatti, M. (2025). The Many Faces of the Angry Peritoneum. *Diagnostics*, *15*(9), 1163. https://doi.org/10.3390/diagnostics15091163
- Anders, S., Pyl, P. T., & Huber, W. (2015). HTSeq—A Python framework to work with high-throughput sequencing data. *Bioinformatics*, *31*(2), 166–169. https://doi.org/10.1093/bioinformatics/btu638
- Belfiore, A., Busico, A., Bozzi, F., Brich, S., Dallera, E., Conca, E., Capone, I., Gloghini, A., Volpi, C. C., Cabras, A. D., Pilotti, S., Baratti, D., Guaglio, M., Deraco, M., Kusamura, S., & Perrone, F. (2019). Molecular Signatures for Combined Targeted Treatments in Diffuse Malignant Peritoneal Mesothelioma. *International Journal of Molecular Sciences*, *20*(22), 5817. https://doi.org/10.3390/ijms20225817
- Bhatt, H., Vakil, D. P., Quintero, L., Motta, M., Vashchenko, M., Rodriguez, B., Samuels, S. K., Jayant, K., Magloire, J., Gannon, C. J., & Llaguna, O. H. (2026). Characteristics of Long-Term Survivors With Peritoneal Mesothelioma—Insights From the National Cancer Database. *The American Surgeon™*, *92*(1), 53–61. https://doi.org/10.1177/00031348251350994
- Boffetta, P. (2007). Epidemiology of peritoneal mesothelioma: A review. *Annals of Oncology*, *18*(6), 985–990. https://doi.org/10.1093/annonc/mdl345
- Bolger, A. M., Lohse, M., & Usadel, B. (2014). Trimmomatic: A flexible trimmer for Illumina sequence data. *Bioinformatics*, *30*(15), 2114–2120. https://doi.org/10.1093/bioinformatics/btu170
- Bonde, A., Singh, R., Prasad, S. R., Kamireddy, D., Aggarwal, A., Ramani, N., Saboo, S., Shanbhogue, K., Dasyam, A. K., & Katabathina, V. S. (2023). Mesotheliomas and Benign Mesothelial Tumors: Update on Pathologic and Imaging Findings. *RadioGraphics*, *43*(3), e220128. https://doi.org/10.1148/rg.220128
- Bootsma, S., Bijlsma, M. F., & Vermeulen, L. (2023). The molecular biology of peritoneal metastatic disease. *EMBO Molecular Medicine*, *15*(3), e15914. https://doi.org/10.15252/emmm.202215914
- Bozzi, F., Brich, S., Dagrada, G. P., Negri, T., Conca, E., Cortelazzi, B., Belfiore, A., Perrone, F., Gualeni, A. V., Gloghini, A., Cabras, A., Brenca, M., Maestro, R., Zaffaroni, N., Casali, P., Bertulli, R., Deraco, M., & Pilotti, S. (2016). Epithelioid peritoneal

mesothelioma: A hybrid phenotype within a mesenchymal-epithelial/epithelial-mesenchymal transition framework. *Oncotarget*, *7*(46), 75503–75517. https://doi.org/10.18632/oncotarget.12262

- Brcic, L., & Kern, I. (2020). Clinical significance of histologic subtyping of malignant pleural mesothelioma. *Translational Lung Cancer Research*, *9*(3), 924–933. https://doi.org/10.21037/tlcr.2020.03.38
- Bueno, R., Stawiski, E. W., Goldstein, L. D., Durinck, S., De Rienzo, A., Modrusan, Z., Gnad, F., Nguyen, T. T., Jaiswal, B. S., Chirieac, L. R., Sciaranghella, D., Dao, N., Gustafson, C. E., Munir, K. J., Hackney, J. A., Chaudhuri, A., Gupta, R., Guillory, J., Toy, K., … Seshagiri, S. (2016). Comprehensive genomic analysis of malignant pleural mesothelioma identifies recurrent mutations, gene fusions and splicing alterations. *Nature Genetics*, *48*(4), 407–416. https://doi.org/10.1038/ng.3520
- Carbone, M., Ly, B. H., Dodson, R. F., Pagano, I., Morris, P. T., Dogan, U. A., Gazdar, A. F., Pass, H. I., & Yang, H. (2012). Malignant mesothelioma: Facts, Myths, and Hypotheses. *Journal of Cellular Physiology*, *227*(1), 44–58. https://doi.org/10.1002/jcp.22724
- Carlson, B., Harmath, C., Turaga, K., Kindler, H. L., Armato, S. G., & Straus, C. (2022). The role of imaging in diagnosis and management of malignant peritoneal mesothelioma: A systematic review. *Abdominal Radiology*, *47*(5), 1725–1740. https://doi.org/10.1007/s00261-022-03464-x
- Chen, B., Khodadoust, M. S., Liu, C. L., Newman, A. M., & Alizadeh, A. A. (2018). Profiling Tumor Infiltrating Immune Cells with CIBERSORT. In L. Von Stechow (Ed.), *Cancer Systems Biology* (Vol. 1711, pp. 243–259). Springer New York. https://doi.org/10.1007/978-1-4939-7493-1_12
- Consonni, D., Migliore, E., Gangemi, M., Cavone, D., Vimercati, L., Piro, S., Giovannetti, L., Zabeo, V., Comiati, V., Curti, S., Mattioli, S., Landi, M. T., Gioscia, C., Eccher, S., Murano, S., D'Agostin, F., Genova, C., Perduri, R., Grappasonni, I., … Mensi, C. (2025). Peritoneal mesothelioma and asbestos exposure: A population-based case–control study in Italy, 2000–2021. *Occupational and Environmental Medicine*, *82*(10), 495–503. https://doi.org/10.1136/oemed-2025-110414
- Cooper, J., & Giancotti, F. G. (2019). Integrin Signaling in Cancer: Mechanotransduction, Stemness, Epithelial Plasticity, and Therapeutic Resistance. *Cancer Cell*, *35*(3), 347–367. https://doi.org/10.1016/j.ccell.2019.01.007
- Cummins, K. C., Sears, O., Bitar, E. R., Shen, C., Tsung, A., & Ruff, S. M. (2026). Hormonal Status May Contribute to Sex-Based Survival Differences in Epithelioid Peritoneal Mesothelioma. *Annals of Surgical Oncology*. https://doi.org/10.1245/s10434-026-19455-x

- Dacic, S. (2022). Pleural mesothelioma classification—Update and challenges. *Modern Pathology*, *35*, 51–56. https://doi.org/10.1038/s41379-021-00895-7
- Dhatchinamoorthy, K., Colbert, J. D., & Rock, K. L. (2021). Cancer Immune Evasion Through Loss of MHC Class I Antigen Presentation. *Frontiers in Immunology*, *12*, 636568. https://doi.org/10.3389/fimmu.2021.636568
- Dobin, A., Davis, C. A., Schlesinger, F., Drenkow, J., Zaleski, C., Jha, S., Batut, P., Chaisson, M., & Gingeras, T. R. (2013). STAR: Ultrafast universal RNA-seq aligner. *Bioinformatics*, *29*(1), 15–21. https://doi.org/10.1093/bioinformatics/bts635
- Dubreuil, J., Giammarile, F., Rousset, P., Rubello, D., Bakrin, N., Passot, G., Isaac, S., Glehen, O., & Skanjeti, A. (2017). The role of 18F-FDG-PET/ceCT in peritoneal mesothelioma. *Nuclear Medicine Communications*, *38*(4). https://journals.lww.com/nuclearmedicinecomm/fulltext/2017/04000/the_role_of_18f_fdg_pet_cect_in_peritoneal.6.aspx
- Ewels, P., Magnusson, M., Lundin, S., & Käller, M. (2016). MultiQC: Summarize analysis results for multiple tools and samples in a single report. *Bioinformatics*, *32*(19), 3047–3048. https://doi.org/10.1093/bioinformatics/btw354
- Gabay, M., Li, Y., & Felsher, D. W. (2014). MYC Activation Is a Hallmark of Cancer Initiation and Maintenance. *Cold Spring Harbor Perspectives in Medicine*, *4*(6), a014241–a014241. https://doi.org/10.1101/cshperspect.a014241
- Gagnon-Bartsch, J. A., & Speed, T. P. (2012). Using control genes to correct for unwanted variation in microarray data. *Biostatistics*, *13*(3), 539–552. https://doi.org/10.1093/biostatistics/kxr034
- Gilani, S. N. S., Mehta, A., Garcia-Fadrique, A., Rowaiye, B., Jenei, V., Dayal, S., Chandrakumaran, K., Carr, N., Mohamed, F., Cecil, T., & Moran, B. (2018). Outcomes of cytoreductive surgery with hyperthermic intraperitoneal chemotherapy for peritoneal mesothelioma and predictors of survival. *International Journal of Hyperthermia*, *34*(5), 578–584. https://doi.org/10.1080/02656736.2018.1434902
- Goodwin, S., McPherson, J. D., & McCombie, W. R. (2016). Coming of age: Ten years of next-generation sequencing technologies. *Nature Reviews Genetics*, *17*(6), 333–351. https://doi.org/10.1038/nrg.2016.49
- Greenbaum, A., & Alexander, H. R. (2020). Peritoneal mesothelioma. *Translational Lung Cancer Research*, *9*(S1), S120–S132. https://doi.org/10.21037/tlcr.2019.12.15
- Han, L., Li, X., Cao, M., Cao, Y., & Zhou, L. (2018). Development and validation of an individualized diagnostic signature in thyroid cancer. *Cancer Medicine*, *7*(4), 1135–1140. https://doi.org/10.1002/cam4.1397

- Heljasvaara, R., Aikio, M., Ruotsalainen, H., & Pihlajaniemi, T. (2017). Collagen XVIII in tissue homeostasis and dysregulation—Lessons learned from model organisms and human patients. *Matrix Biology*, *57–58*, 55–75. https://doi.org/10.1016/j.matbio.2016.10.002
- Hernandez, S., Lazcano, R., Serrano, A., Powell, S., Kostousov, L., Mehta, J., Khan, K., Lu, W., & Solis, L. M. (2022). Challenges and Opportunities for Immunoprofiling Using a Spatial High-Plex Technology: The NanoString GeoMx® Digital Spatial Profiler. *Frontiers in Oncology*, *12*, 890410. https://doi.org/10.3389/fonc.2022.890410
- Hiltbrunner, S., Fleischmann, Z., Sokol, E. S., Zoche, M., Felley-Bosco, E., & Curioni-Fontecedro, A. (2022). Genomic landscape of pleural and peritoneal mesothelioma tumours. *British Journal of Cancer*, *127*(11), 1997–2005. https://doi.org/10.1038/s41416-022-01979-0
- Hong, M., Tao, S., Zhang, L., Diao, L.-T., Huang, X., Huang, S., Xie, S.-J., Xiao, Z.-D., & Zhang, H. (2020). RNA sequencing: New technologies and applications in cancer research. *Journal of Hematology & Oncology*, *13*(1), 166. https://doi.org/10.1186/s13045-020-01005-x
- Husain, A. N., Chapel, D. B., Attanoos, R., Beasley, M. B., Brcic, L., Butnor, K., Chirieac, L. R., Churg, A., Dacic, S., Galateau-Salle, F., Hiroshima, K., Hung, Y. P., Klebe, S., Krausz, T., Khoor, A., Litzky, L., Marchevsky, A., Nabeshima, K., Nicholson, A. G., … Colby, T. V. (2024). Guidelines for Pathologic Diagnosis of Mesothelioma: 2023 Update of the Consensus Statement From the International Mesothelioma Interest Group. *Archives of Pathology & Laboratory Medicine*, *148*(11), 1251–1271. https://doi.org/10.5858/arpa.2023-0304-RA
- Joseph, N. M., Chen, Y.-Y., Nasr, A., Yeh, I., Talevich, E., Onodera, C., Bastian, B. C., Rabban, J. T., Garg, K., Zaloudek, C., & Solomon, D. A. (2017). Genomic profiling of malignant peritoneal mesothelioma reveals recurrent alterations in epigenetic regulatory genes BAP1, SETD2, and DDX3X. *Modern Pathology*, *30*(2), 246–254. https://doi.org/10.1038/modpathol.2016.188
- Karunakaran, K. B., & Ganapathiraju, M. K. (2024). Malignant peritoneal mesothelioma interactome with 417 novel protein–protein interactions. *BJC Reports*, *2*(1), 42. https://doi.org/10.1038/s44276-024-00062-w
- Kastenhuber, E. R., & Lowe, S. W. (2017). Putting p53 in Context. *Cell*, *170*(6), 1062–1078. https://doi.org/10.1016/j.cell.2017.08.028
- Kieffer, Y., Hocine, H. R., Gentric, G., Pelon, F., Bernard, C., Bourachot, B., Lameiras, S., Albergante, L., Bonneau, C., Guyard, A., Tarte, K., Zinovyev, A., Baulande, S., Zalcman, G., Vincent-Salomon, A., & Mechta-Grigoriou, F. (2020). Single-Cell Analysis Reveals Fibroblast Clusters Linked to Immunotherapy Resistance in Cancer. *Cancer Discovery*, *10*(9), 1330–1351. https://doi.org/10.1158/2159-8290.CD-19-1384

- King, C. M., Ding, W., Eshelman, M. A., & Yochum, G. S. (2024). TCF7L1 regulates colorectal cancer cell migration by repressing GAS1 expression. *Scientific Reports*, *14*(1), 12477. https://doi.org/10.1038/s41598-024-63346-8
- Kipps, E., Tan, D. S. P., & Kaye, S. B. (2013). Meeting the challenge of ascites in ovarian cancer: New avenues for therapy and research. *Nature Reviews Cancer*, *13*(4), 273–282. https://doi.org/10.1038/nrc3432
- Kusamura, S., Baratti, D., De Simone, M., Pasqual, E. M., Ansaloni, L., Marrelli, D., Robella, M., Accarpio, F., Valle, M., Scaringi, S., Biacchi, D., Palopoli, C., Gazzanelli, S., Guaglio, M., & Deraco, M. (2023). Diagnostic and Therapeutic Pathway in Diffuse Malignant Peritoneal Mesothelioma. *Cancers*, *15*(3), 662. https://doi.org/10.3390/cancers15030662
- Kusamura, S., Kepenekian, V., Villeneuve, L., Lurvink, R. J., Govaerts, K., De Hingh, I. H. J. T., Moran, B. J., Van Der Speeten, K., Deraco, M., Glehen, O., Barrios-Sanchez, P., Baumgartner, J. M., Ben-Yaacov, A., Bertulli, R., Cashin, P., Cecil, T., Dayal, S., De Simone, M., Foster, J. M., … Yu, Y. (2021). Peritoneal mesothelioma: PSOGI/EURACAN clinical practice guidelines for diagnosis, treatment and follow-up. *European Journal of Surgical Oncology*, *47*(1), 36–59. https://doi.org/10.1016/j.ejso.2020.02.011
- Leek, J. T., Johnson, W. E., Parker, H. S., Jaffe, A. E., & Storey, J. D. (2012). The sva package for removing batch effects and other unwanted variation in high-throughput experiments. *Bioinformatics*, *28*(6), 882–883. https://doi.org/10.1093/bioinformatics/bts034
- Leggett, R. M., Ramirez-Gonzalez, R. H., Clavijo, B. J., Waite, D., & Davey, R. P. (2013). Sequencing quality assessment tools to enable data-driven informatics for high throughput genomics. *Frontiers in Genetics*, *4*. https://doi.org/10.3389/fgene.2013.00288
- Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., Durbin, R., & 1000 Genome Project Data Processing Subgroup. (2009). The Sequence Alignment/Map format and SAMtools. *Bioinformatics*, *25*(16), 2078–2079. https://doi.org/10.1093/bioinformatics/btp352
- Li, Y.-X., Wei, S.-Q., & Zheng, P.-S. (2025). Atosiban-conjugated 3WJ-pRNA nanoparticles delivering GAS1-enhanced extracellular vesicles: Targeting the decidua to combat recurrent miscarriage. *Journal of Nanobiotechnology*, *23*(1), 780. https://doi.org/10.1186/s12951-025-03776-2
- Lim, H. J., Wang, Y., Buzdin, A., & Li, X. (2025). A practical guide for choosing an optimal spatial transcriptomics technology from seven major commercially available options. *BMC Genomics*, *26*(1), 47. https://doi.org/10.1186/s12864-025-11235-3
- Liu, B., Li, Y., & Zhang, L. (2022). Analysis and Visualization of Spatial Transcriptomic Data. *Frontiers in Genetics*, *12*, 785290. https://doi.org/10.3389/fgene.2021.785290

- Liu, Y., Yang, Y., An, S., Wan, H., Zhang, M., Yin, R., Zhao, W., Huang, L., Zhao, Y., & Li, C. (2025). A novel malignant mesothelioma organoids-T cell co-culture platform for personalized immunochemotherapy testing. *Scientific Reports*, *15*(1), 41305. https://doi.org/10.1038/s41598-025-25087-0
- Love, M. I., Huber, W., & Anders, S. (2014). Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biology*, *15*(12), 550. https://doi.org/10.1186/s13059-014-0550-8
- Low, R. N., Barone, R. M., & Lucero, J. (2015). Comparison of MRI and CT for Predicting the Peritoneal Cancer Index (PCI) Preoperatively in Patients Being Considered for Cytoreductive Surgical Procedures. *Annals of Surgical Oncology*, *22*(5), 1708–1715. https://doi.org/10.1245/s10434-014-4041-7
- Low, R. N., Barone, R. M., & Rousset, P. (2021). Peritoneal MRI in patients undergoing cytoreductive surgery and HIPEC: History, clinical applications, and implementation. *European Journal of Surgical Oncology*, *47*(1), 65–74. https://doi.org/10.1016/j.ejso.2019.02.030
- Mansfield, A. S., Peikert, T., Smadbeck, J. B., Udell, J. B. M., Garcia-Rivera, E., Elsbernd, L., Erskine, C. L., Van Keulen, V. P., Kosari, F., Murphy, S. J., Ren, H., Serla, V. V., Schaefer Klein, J. L., Karagouga, G., Harris, F. R., Sosa, C., Johnson, S. H., Nevala, W., Markovic, S. N., … Vasmatzis, G. (2019). Neoantigenic Potential of Complex Chromosomal Rearrangements in Mesothelioma. *Journal of Thoracic Oncology*, *14*(2), 276–287. https://doi.org/10.1016/j.jtho.2018.10.001
- Manzini, V. D. P., Recchia, L., Cafferata, M., Porta, C., Siena, S., Giannetta, L., Morelli, F., Oniga, F., Bearz, A., Torri, V., & Cinquini, M. (2010). Malignant peritoneal mesothelioma: A multicenter study on 81 cases. *Annals of Oncology*, *21*(2), 348–353. https://doi.org/10.1093/annonc/mdp307
- Merritt, C. R., Ong, G. T., Church, S. E., Barker, K., Danaher, P., Geiss, G., Hoang, M., Jung, J., Liang, Y., McKay-Fleisch, J., Nguyen, K., Norgaard, Z., Sorg, K., Sprague, I., Warren, C., Warren, S., Webster, P. J., Zhou, Z., Zollinger, D. R., … Beechem, J. M. (2020). Multiplex digital spatial profiling of proteins and RNA in fixed tissue. *Nature Biotechnology*, *38*(5), 586–599. https://doi.org/10.1038/s41587-020-0472-9
- Modarresi, F., Faghihi, M. A., Lopez-Toledano, M. A., Fatemi, R. P., Magistri, M., Brothers, S. P., Van Der Brug, M. P., & Wahlestedt, C. (2012). Inhibition of natural antisense transcripts in vivo results in gene-specific transcriptional upregulation. *Nature Biotechnology*, *30*(5), 453–459. https://doi.org/10.1038/nbt.2158
- Möhrmann, L., Werner, M., Oleś, M., Knol, L., Arnold, J. S., Mundt, T., Paramasivam, N., Richter, D., Fröhlich, M., Hutter, B., Hüllein, J., Jahn, A., Scheffold, C., Möhrmann, E. E., Hanf, D., Kreutzfeldt, S., Heilig, C. E., Teleanu, M.-V., Lipka, D. B., … Glimm, H. (2025).

Comprehensive genomic and transcriptomic analysis enables molecularly guided therapy options in peritoneal and pleural mesothelioma. *ESMO Open*, *10*(4), 104532. https://doi.org/10.1016/j.esmoop.2025.104532

- Moolgavkar, S. H., Meza, R., & Turim, J. (2009). Pleural and peritoneal mesotheliomas in SEER: Age effects and temporal trends, 1973–2005. *Cancer Causes & Control*, *20*(6), 935–944. https://doi.org/10.1007/s10552-009-9328-9
- Moore, A. J., Parker, R. J., & Wiggins, J. (2008). Malignant mesothelioma. *Orphanet Journal of Rare Diseases*, *3*(1), 34. https://doi.org/10.1186/1750-1172-3-34
- Muller, P., Hrstka, R., Coomber, D., Lane, D. P., & Vojtesek, B. (2008). Chaperone-dependent stabilization and degradation of p53 mutants. *Oncogene*, *27*(24), 3371–3383. https://doi.org/10.1038/sj.onc.1211010
- Nair, N. U., Jiang, Q., Wei, J. S., Misra, V. A., Morrow, B., Kesserwan, C., Hermida, L. C., Lee, J. S., Mian, I., Zhang, J., Lebensohn, A., Miettinen, M., Sengupta, M., Khan, J., Ruppin, E., & Hassan, R. (2023). Genomic and transcriptomic analyses identify a prognostic gene signature and predict response to therapy in pleural and peritoneal mesothelioma. *Cell Reports Medicine*, *4*(2), 100938. https://doi.org/10.1016/j.xcrm.2023.100938
- Neumann, V., Löseke, S., Nowak, D., Herth, F. J. F., & Tannapfel, A. (2013). Malignant Pleural Mesothelioma. *Deutsches Ärzteblatt International*. https://doi.org/10.3238/arztebl.2013.0319
- Newman, A. M., Liu, C. L., Green, M. R., Gentles, A. J., Feng, W., Xu, Y., Hoang, C. D., Diehn, M., & Alizadeh, A. A. (2015). Robust enumeration of cell subsets from tissue expression profiles. *Nature Methods*, *12*(5), 453–457. https://doi.org/10.1038/nmeth.3337
- Reed, D., Gannon, D., & Dongarra, J. (2022). *Reinventing High Performance Computing: Challenges and Opportunities* (arXiv:2203.02544). arXiv. https://doi.org/10.48550/arXiv.2203.02544
- Ritchie, M. E., Phipson, B., Wu, D., Hu, Y., Law, C. W., Shi, W., & Smyth, G. K. (2015). Limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic Acids Research*, *43*(7), e47–e47. https://doi.org/10.1093/nar/gkv007
- Sacco, A., Battaglia, A. M., Botta, C., Aversa, I., Mancuso, S., Costanzo, F., & Biamonte, F. (2021). Iron Metabolism in the Tumor Microenvironment—Implications for Anti-Cancer Immune Response. *Cells*, *10*(2), 303. https://doi.org/10.3390/cells10020303
- Semenza, G. L. (2013). HIF-1 mediates metabolic responses to intratumoral hypoxia and oncogenic mutations. *Journal of Clinical Investigation*, *123*(9), 3664–3671. https://doi.org/10.1172/JCI67230
- Shao, C., Li, G., Huang, L., Pruitt, S., Castellanos, E., Frampton, G., Carson, K. R., Snow, T., Singal, G., Fabrizio, D., Alexander, B. M., Jin, F., & Zhou, W. (2020). Prevalence of

High Tumor Mutational Burden and Association With Survival in Patients With Less Common Solid Tumors. *JAMA Network Open*, *3*(10), e2025109. https://doi.org/10.1001/jamanetworkopen.2020.25109

- Sheng, R., Yin, Y., & Wang, X. (2025). Mesothelial and immune cells interplay in the tumor microenvironment. *Trends in Molecular Medicine*, *31*(10), 895–908. https://doi.org/10.1016/j.molmed.2025.03.014
- Sigismund, S., Avanzato, D., & Lanzetti, L. (2018). Emerging functions of the EGFR in cancer. *Molecular Oncology*, *12*(1), 3–20. https://doi.org/10.1002/1878-0261.12155
- Stamou, K., Tsamis, D., Pallas, N., Samanta, E., Courcoutsakis, N., Prassopoulos, P., & Tentes, A.-A. (2015). Treating peritoneal mesothelioma with cytoreductive surgery and hyperthermic intraperitoneal chemotherapy. A case series and review of the literature. *International Journal of Hyperthermia*, *31*(8), 850–856. https://doi.org/10.3109/02656736.2015.1075071
- Subramanian, A., Tamayo, P., Mootha, V. K., Mukherjee, S., Ebert, B. L., Gillette, M. A., Paulovich, A., Pomeroy, S. L., Golub, T. R., Lander, E. S., & Mesirov, J. P. (2005). Gene set enrichment analysis: A knowledge-based approach for interpreting genome-wide expression profiles. *Proceedings of the National Academy of Sciences*, *102*(43), 15545–15550. https://doi.org/10.1073/pnas.0506580102
- Sugarbaker, P. H. (1999). Successful management of microscopic residual disease in large bowel cancer. *Cancer Chemotherapy and Pharmacology*, *43*(7), S15–S25. https://doi.org/10.1007/s002800051093
- Sun, L., Li, C., & Gao, S. (2023). Diffuse malignant peritoneal mesothelioma: A review. *Frontiers in Surgery*, *9*, 1015884. https://doi.org/10.3389/fsurg.2022.1015884
- Tazzari, M., Brich, S., Tuccitto, A., Bozzi, F., Beretta, V., Spagnuolo, R. D., Negri, T., Stacchiotti, S., Deraco, M., Baratti, D., Camisaschi, C., Villa, A., Vergani, B., Rivoltini, L., Pilotti, S., & Castelli, C. (2018). Complex Immune Contextures Characterise Malignant Peritoneal Mesothelioma: Loss of Adaptive Immunological Signature in the More Aggressive Histological Types. *Journal of Immunology Research*, *2018*, 1–13. https://doi.org/10.1155/2018/5804230
- Torricelli, F., Donati, B., Reggiani, F., Manicardi, V., Piana, S., Valli, R., Lococo, F., & Ciarrocchi, A. (2023). Spatially resolved, high-dimensional transcriptomics sorts out the evolution of biphasic malignant pleural mesothelioma: New paradigms for immunotherapy. *Molecular Cancer*, *22*(1), 114. https://doi.org/10.1186/s12943-023-01816-9
- Winslow, D. J., & Taylor, H. B. (1960). Malignant peritoneal mesotheliomas.A clinicopathological analysis of 12 fatal cases. *Cancer*, *13*(1), 127–136. https://doi.org/10.1002/1097-0142(196001/02)13:1%253C127::AID-CNCR2820130124%253E3.0.CO;2-6

- Yu, H., Pardoll, D., & Jove, R. (2009). STATs in cancer inflammation and immunity: A leading role for STAT3. *Nature Reviews Cancer*, *9*(11), 798–809. https://doi.org/10.1038/nrc2734
- Zhai, L., Huang, D., Lin, L., Zheng, R., He, Q., Zhu, X., Liu, R., Fei, J., Zhang, Z., Zhang, X., & Zhou, J. (2025). Integrative analysis and experiments to explore GAS1 as a prognostic target for ovarian cancer based on angiogenesis-related genes. *Journal of Ovarian Research*, *19*(1), 19. https://doi.org/10.1186/s13048-025-01883-0
- Zhan, T., Rindtorff, N., & Boutros, M. (2017). Wnt signaling in cancer. *Oncogene*, *36*(11), 1461–1473. https://doi.org/10.1038/onc.2016.304


**Use of Generative AI Tools**

Generative artificial intelligence tools were used to support linguistic rephrasing. No scientific content or experimental results were generated by AI.