Hypoxia and the Tumour Microenvironment: An Analysis of Breast Cancer Heterogeneity

Thesis Submitted For
Degree of Doctor of Philosophy
By

Badran Elshenawy
Wolfson College,
Department of Oncology,
University of Oxford,
Trinity Term 2024

Supervisors

Professor Francesca Buffa **Professor Adrian Harris**

Statement of Originality

I hereby declare that this thesis is composed entirely by myself and has not been previously submitted, in whole or in part, for any degree at other institutions. All work presented, unless otherwise indicated by citations or acknowledgments, is solely my own. This assertion extends to the authorship of any related papers submitted alongside this thesis.

Research is intrinsically collaborative, and I wish to express my profound gratitude to all who have contributed to my development throughout my DPhil. For specific expressions of thanks, please refer to the Acknowledgments section.

SARS-CoV-2 Pandemic Impact Statement

The SARS-CoV-2 pandemic profoundly affected every facet of daily life, including the progress of my DPhil research. Due to disruptions caused by the pandemic, the collection of new bench-side data was significantly hindered. As a result, this thesis predominantly relies on data that was generated before the pandemic and obtained from publicly available sources.

Declaration

I hereby declare that this thesis is my own original work, undertaken after my registration for the degree of Doctor of Philosophy (DPhil) in Oncology at the University of Oxford. This work has not been submitted for a degree or any other qualification at this or any other institution.

Dedication

This thesis is dedicated to my family, whose unwavering support was crucial throughout my DPhil journey. Their presence and steadfast belief in my abilities were vital during the challenges posed by the COVID-19 pandemic, enabling me to persevere and succeed.

Acknowledgements

I extend my deepest gratitude to my supervisors, Professor Adrian Harris and Professor Francesca Meteora Buffa. Their mentorship transcended conventional boundaries, imbuing me with the principles of critical thinking, strategies for efficient execution, and skills for elegant presentation. Their profound influence has been pivotal in my development both as a scientist and as a person. I am profoundly grateful for their steadfast guidance and support.

Scientific research thrives on collaboration, and my thesis embodies this principle. I am particularly thankful to May Sin Ke, who conducted all the experiments and generated the vital hypoxia time series single cell data for chapter 4. My thanks also go to Salwa Lin for her work in producing the tumor micro-environment time series single cell data for chapter 5.

Funding

Mohamed Bin Rashid Al Maktoum Knowledge Foundation.

List of Submissions

Publications

Di Giovannantonio, M., Harris, B. H. L., Elshenawy, B., Harris, D. A., Barberis, A., Harris, A. L., & Buffa, F. M. (2024). Defining hypoxia through big data: A landmark study evaluating hypoxia gene expression signatures. *Nature Genetics*. (Submitted)

Elshenawy, B., Ke, M., Sheldon, H., Arora, A., Buffa, F. M. (2024). Chronic hypoxia in TNBC breast cancer cells drives resistance to immune cell-mediated killing. *Cancer Research*, 84(3_Supplement_1), B056. doi:https://doi.org/10.1158/1538-7445.ADVBC23-B056 (abstract)

Top of Form

Bottom of Form

Lin, S., Carroll, T., Elshenawy, B., Bridgeman, V., Malanchi, I., Banham, A. H., Harris, L. (2023). Transcriptome profiling of metastatic niche in a murine model of triple-negative breast cancer lung metastases. *Molecular Oncology*, 17(S1), 1-597. doi:https://doi.org/10.1002/1878-0261.13469 (abstract)

Elshenawy, B., Ke, M., Sheldon, H., Arora, A., Buffa, F. M. (2022). Single cell RNA-sequencing: A powerful yet still challenging technology to study cellular heterogeneity. *Bioessays*, 44(12), e2200084. doi:https://doi.org/10.1002/bies.202200084

Presentations

Ke, A. S., Elshenawy, B., Arora, A., Sheldon, H., Harris, A., & Buffa, F. (2023). Chronic hypoxia in TNBC breast cancer induces development of a novel population of CD24+/CD49b-high cells characterized by high expression of CST1. *American Association for Cancer Research (AACR) Special Conference in Cancer Research*, San Diego, USA.

Lin, S., Carroll, T., Elshenawy, B., Bridgeman, V., Malanchi, I., Banham, A. H., & Harris, A. L. (2023). Transcriptome profiling of metastatic niche in a murine model of triple-negative breast cancer lung metastases. *Annual Congress of the European Association for Cancer Research (EACR)*, Torino, Italy.

Lin, S., Carroll, T., Elshenawy, B., Bridgeman, V., Malanchi, I., De Val, S., Banham, A. H., & Harris, A. L. (2022). Spatial and temporal vascular remodelling in triple-negative breast cancer lung metastases. *22nd International Vascular Biology Meeting (IVBM)*, Oakland, California, United States of America.

Lin, S., Carroll, T., Elshenawy, B., AlMesned, F., Bridgeman, V., Malanchi, I., De Val, S., Banham, A. H., & Harris, A. L. (2022). Spatial and temporal single-cell transcriptional analysis of metastatic niche identifies distinct cell populations enriched in advanced lung metastases. *Weatherall Institute of Molecular Medicine (WIMM) Student Presentation Day 2022*, WIMM Research Centre, University of Oxford, Oxford, United Kingdom.

Abstract

Breast cancer, the most prevalent cancer worldwide, boasts a high 5-year survival rate exceeding 90% with early detection. Despite this, the disease's intrinsic heterogeneity and therapeutic resistance remain formidable challenges in clinical management, particularly in advanced stages. Among the four primary subtypes identified by hormone receptor expression—luminal A, luminal B, HER2-enriched, and triple-negative breast cancer (TNBC)—TNBC stands out for its aggressive metastatic behavior and high therapeutic resistance; this aggression is a direct byproduct of the dearth of targeted treatments as well as intrinsic heterogeneity. Furthermore, TNBC's diverse subtypes each exhibit unique molecular profiles that facilitate cancer adaptation and progression through different signaling pathways that are at best poorly characterized.

A differentiating factor in TNBC's aggression is its adaptation to hypoxia—a prevalent condition within the tumor micro-environment (TME). Hypoxia in tumors can be cyclical, acute, or chronic, each inducing distinct cellular expression programs. While the acute hypoxic response mediated by HIF1A is well-characterized through bulk transcriptomics, the complexities of cyclical and chronic hypoxia are less understood. Likewise, heterogeneity across different TNBC subtypes is very poorly characterized, thereby further complicating treatment and catalyzing aggression. bulk transcriptomics falls short in capturing the heterogeneity at the single-cell level, omitting crucial insights from the varied sub-populations of cancer cells responding to hypoxic stress.

This thesis opens with an examination of the acute hypoxic response in TNBC at the single-cell level, revealing previously unrecognized heterogeneity and functional specialization in response to hypoxia. Notably, a potential negative feedback loop between HIF2A and ESR1 was identified, implicating hypoxia in resistance to common estrogen-targeting therapies in breast cancer.

The subsequent chapter delves into both acute and chronic hypoxic responses in TNBC across multiple timepoints—1, 2, 7, and 14 days of hypoxia, compared to normoxia. This pioneering single-cell study on chronic hypoxia in TNBC unveiled a full recovery of the cell cycle under prolonged hypoxic conditions, supported by robust expression of proliferation markers and signature scoring with the hallmark G2M checkpoint. Advanced stages showed the emergence of epithelial-mesenchymal transition (EMT), angiogenesis, and senescence-associated secretory phenotypes by day 7, and a pronounced immune-inflammatory phenotype by day 14, featuring heightened interferon signaling. These findings were cross-validated against the TNBC subset of the Single Cell Breast Cancer Atlas (ScBrAtlas) using signature scoring, confirming the clinical relevance of the observed phenotypic adaptations.

The final chapter extends the investigation to the stromal micro-environment of TNBC metastases in mouse lungs, exploring compositional shifts over time and identifying key stromal cell types driving advanced carcinogenesis. Employing scRNA-Seq's novel cell-cell communication analysis, this work elucidates the intricate networks and ligand-receptor pairs that orchestrate angiogenesis and tumor progression, highlighting the transformative potential of single-cell technologies in understanding and combating breast cancer.

Table of Contents

List of Figures

List of Abbreviations

Chapter 1	
Abbreviation	**Definition**
ADCC	Antibody-Dependent Cellular Cytotoxicity: An immune response where antibodies bind to a target cell, leading to its destruction by immune cells.
AR	Androgen Receptor: A receptor that binds to androgens (male hormones) and is involved in the development of male characteristics and some types of cancer.
BCL2	B-cell CLL/lymphoma 2: A protein family involved in regulating cell death (apoptosis), often overexpressed in cancer cells to inhibit apoptosis.
BRCA1/BRCA2	Breast Cancer gene 1 and 2: Genes that produce tumor suppressor proteins, mutations in which significantly increase the risk of breast cancer.
BRCA	Breast Cancer gene: Genes involved in DNA repair, mutations of which are linked to a higher risk of breast cancer.
CDK4/6	Cyclin-Dependent Kinases 4 and 6: Enzymes that drive cell cycle progression, often targeted in cancer therapy to prevent tumor growth.
CEACAM5	Carcinoembryonic Antigen-Related Cell Adhesion Molecule 5: A protein involved in cell adhesion, often upregulated in cancers.
CEACAM6	Carcinoembryonic Antigen-Related Cell Adhesion Molecule 6: A protein similar to CEACAM5, involved in cell adhesion and frequently upregulated in cancers.
COVID-19	Coronavirus Disease 2019: A global pandemic caused by the SARS-CoV-2 virus that significantly impacted research and experimental workflows.
E2	Estradiol: A form of estrogen, a hormone that plays a crucial role in the development of female reproductive tissues and breast cancer.
E0771	A murine triple-negative breast cancer cell line used in metastasis research.
EDN2	Endothelin 2: A protein involved in vasoconstriction and cell proliferation, often implicated in cancer development and metastasis.
EMT	Epithelial-to-Mesenchymal Transition: A process by which epithelial cells gain migratory and invasive properties, often leading to metastasis.
ER-	Estrogen Receptor Negative: Refers to breast cancers that do not express estrogen receptors, often more aggressive and treated with non-hormonal therapies.
ER+	Estrogen Receptor Positive: Refers to breast cancers that express estrogen receptors, typically treated with hormone therapies.
ER	Estrogen Receptor: A nuclear hormone receptor that, when activated by estrogen, regulates gene expression related to breast cancer development.
FIH	Factor Inhibiting HIF: An enzyme that hydroxylates HIF1A, preventing it from activating hypoxia-responsive genes.
FSH	Follicle-Stimulating Hormone: A hormone involved in the regulation of reproductive processes, including the production of estrogen.
GLUT1	Glucose Transporter 1: A protein that facilitates the transport of glucose across the cell membrane, often upregulated in cancer cells.
GnRH	Gonadotropin-Releasing Hormone: A hormone responsible for the release of FSH and LH from the pituitary gland.
HER2-	Human Epidermal growth factor Receptor 2 Negative: Breast cancer

	subtype that does not overexpress HER2, often treated with hormone therapy if ER+.
HER2+	Human Epidermal growth factor Receptor 2 Positive: Breast cancer subtype characterized by overexpression of HER2, treated with targeted therapies.
HER2	Human Epidermal growth factor Receptor 2: A protein that promotes the growth of cancer cells, often overexpressed in certain breast cancers.
HIF1A	Hypoxia-Inducible Factor 1-Alpha: A transcription factor that regulates cellular response to low oxygen levels, playing a key role in tumor survival.
HIF1B	Hypoxia-Inducible Factor 1-Beta: A partner protein that dimerizes with HIF1A to activate hypoxia-responsive genes.
HREs	Hypoxia Response Elements: DNA sequences that HIF1A binds to in order to activate genes involved in the hypoxic response.
Ki-67	A marker used to measure cell proliferation, often used to assess the aggressiveness of breast cancer.
LDHA	Lactate Dehydrogenase A: An enzyme involved in converting pyruvate to lactate, playing a role in cancer cell metabolism.
LH	Luteinizing Hormone: A hormone that triggers ovulation and stimulates the production of other hormones like progesterone and estrogen.
LRP4	Low-density lipoprotein Receptor-related Protein 4: A receptor involved in cell signaling, sometimes implicated in cancer metastasis.
M	Mesenchymal: A subtype of TNBC associated with epithelial-to-mesenchymal transition and a high level of cell motility and invasiveness.
MCF7	A human breast cancer cell line often used as a model for estrogen receptor-positive breast cancer studies.
MDA231	MDA-MB-231: A triple-negative breast cancer cell line commonly used in cancer research.
MSL	Mesenchymal Stem-Like: A subtype of TNBC characterized by stem cell-like properties and high metastatic potential.
OER	Oxygen Enhancement Ratio: A measure of how much more effective radiation therapy is in the presence of oxygen compared to hypoxic conditions.
Pecam1	Platelet Endothelial Cell Adhesion Molecule 1: A protein involved in leukocyte transmigration and angiogenesis, commonly found on endothelial cells.
PGR	Progesterone Receptor: A receptor that, when activated by progesterone, regulates gene expression involved in reproductive health and breast cancer.
PGK1	Phosphoglycerate Kinase 1: An enzyme involved in glycolysis, often upregulated in cancer cells to support rapid growth.
PHDs	Prolyl Hydroxylase Domain proteins: Enzymes that regulate the stability of HIF1A by hydroxylating it, marking it for degradation.
PR-	Progesterone Receptor Negative: Refers to breast cancers that do not express progesterone receptors, often found in more aggressive subtypes like TNBC.
PR+	Progesterone Receptor Positive: Breast cancers that express progesterone receptors, often responsive to hormone therapy.
PR	Progesterone Receptor: A receptor that mediates the effects of progesterone, often linked to breast cancer.
R	Reactive Oxygen Species: Chemically reactive molecules containing

	oxygen, which can induce cellular damage and play a role in cancer progression.
ROR1	Receptor Tyrosine Kinase-like Orphan Receptor 1: A protein involved in cell signaling, particularly in cancer progression and metastasis.
ROS	Reactive Oxygen Species: Chemically reactive molecules containing oxygen, which can induce cellular damage and play a role in cancer progression.
SAC	Spindle Assembly Checkpoint: A cellular mechanism that ensures chromosomes are properly aligned before cell division proceeds.
SERM	Selective Estrogen Receptor Modulator: Drugs that modulate estrogen receptor activity with varying effects in different tissues.
SERD	Selective Estrogen Receptor Degrader: Drugs that target estrogen receptors for degradation, used in hormone therapy-resistant cancers.
SERDs	Selective Estrogen Receptor Degraders: Drugs that bind to and degrade the estrogen receptor, used in resistant ER+ breast cancer.
SERMs	Selective Estrogen Receptor Modulators: Drugs that act on the estrogen receptor with tissue-specific effects, used in breast cancer therapy.
SEMA5B	Semaphorin 5B: A protein involved in cell signaling and axon guidance, sometimes implicated in cancer.
SNAI1	Snail Family Transcriptional Repressor 1: A transcription factor that promotes EMT, contributing to cancer metastasis.
TGF-β	Transforming Growth Factor β: A cytokine that plays a role in cell growth, proliferation, and differentiation, often involved in cancer progression.
TME	Tumor MicroEnvironment: The environment surrounding a tumor, including non-cancerous cells, signaling molecules, and the extracellular matrix.
TNBC	Triple-Negative Breast Cancer: A subtype of breast cancer that lacks ER, PR, and HER2 receptors, making it difficult to treat.
TWIST1	TWIST1: A transcription factor that plays a role in promoting EMT, metastasis, and resistance to apoptosis in cancer.
VHL	Von Hippel-Lindau protein: An E3 ubiquitin ligase that targets HIF1A for degradation under normal oxygen levels.
VEGF	Vascular Endothelial Growth Factor: A signal protein that promotes the growth of new blood vessels, often upregulated in tumors.
ZEB1	Zinc-finger E-box Binding Homeobox 1: A transcription factor involved in EMT and metastasis in cancer.
Chapter 2	
AK4	Adenylate Kinase 4: An enzyme involved in cellular energy homeostasis.
ALDH	Aldehyde Dehydrogenase: A family of enzymes involved in the detoxification of aldehydes.
ATCC	American Type Culture Collection: A repository of cell lines, microorganisms, and related materials.
BCL	Base Call files: Raw data files generated by sequencing machines before conversion to FASTQ format.
CAIX/CA9	Carbonic Anhydrase IX (also known as CA9): A protein involved in maintaining pH balance in cells, particularly under hypoxic conditions.
C57BL/6	C57BL/6: A common inbred strain of laboratory mouse used in biomedical research.
cDNA	Complementary DNA: DNA synthesized from an mRNA template, used in gene expression studies.
CO2	Carbon Dioxide: A gas used to regulate cell culture environments.

CreERT2	CreERT2: A tamoxifen-inducible form of the Cre recombinase enzyme, used for controlled gene editing in genetically modified organisms.
DDIT4	DNA-Damage-Inducible Transcript 4: A gene regulated by HIF1A, involved in cell survival under hypoxia.
DEA	Differential Expression Analysis: A method to identify genes that are expressed differently across conditions or cell types.
DESeq2	A statistical method for analyzing RNA-Seq data, particularly used for differential expression analysis.
DMEM	Dulbecco's Modified Eagle's Medium: A widely used cell culture medium.
DMSO	Dimethyl Sulfoxide: A solvent commonly used in cryopreservation.
EGLN3	Egl Nine Homolog 3: A protein involved in the cellular response to oxygen levels, often regulated by HIF1A.
EPAS1	Endothelial PAS domain protein 1: Also known as HIF2A, a transcription factor involved in the response to low oxygen.
ESR1	Estrogen Receptor 1: A nuclear hormone receptor involved in the regulation of gene expression by estrogens.
FASTQ	FAST Quality Format: A text-based format for storing both biological sequence data and corresponding quality scores.
FastQC	Fast Quality Control: A software tool used to assess the quality of sequencing data.
FBS	Fetal Bovine Serum: A common supplement added to cell culture media.
FIH	Factor Inhibiting HIF: An enzyme that hydroxylates HIF1A, preventing it from activating hypoxia-responsive genes.
FOXM1	Forkhead Box M Protein 1: A transcription factor involved in cell cycle regulation and tumor development.
G2M	Gap 2/Mitosis: Phases of the cell cycle involved in DNA repair and cell division.
GAPDH	Glyceraldehyde-3-Phosphate Dehydrogenase: A glycolytic enzyme often used as a control in gene expression studies.
ggplot2	ggplot2: A data visualization package for the R programming language.
GPI	Glucose-6-Phosphate Isomerase: An enzyme involved in glycolysis and gluconeogenesis, also regulated by HIF1A.
GREB1	Gene Regulated by Estrogen in Breast Cancer 1: A gene involved in the estrogen receptor signaling pathway.
hCD2	Human Cluster of Differentiation 2: A protein marker used in immunology.
HIF1A	Hypoxia-Inducible Factor 1-Alpha: A key transcription factor regulating cellular response to low oxygen.
LDHA	Lactate Dehydrogenase A: An enzyme involved in converting pyruvate to lactate during anaerobic metabolism.
MAD	Median Absolute Deviation: A robust statistic for measuring the variability of a dataset.
MCF7	A human breast cancer cell line commonly used in research.
MDA231	MDA-MB-231: A triple-negative breast cancer cell line often used in cancer research.
Mlung	Mouse Lung cancer dataset: Refers to a dataset derived from lung cancer studies in mice.
MOI	Multiplicity Of Infection: The ratio of infectious agents (like viruses) to infection targets (like cells).
MSigDB	Molecular Signature Database: A comprehensive collection of gene sets used for gene set enrichment analysis.
MT-CYB	Mitochondrial Cytochrome B: A component of the mitochondrial electron

	transport chain involved in cellular respiration.
mRNA	Messenger RNA: A type of RNA that conveys genetic information from DNA to the ribosome, where it is used to produce proteins.
PBS	Phosphate-Buffered Saline: A buffer solution commonly used in biological research.
PDK1	Pyruvate Dehydrogenase Kinase 1: An enzyme that inhibits the pyruvate dehydrogenase complex, reducing oxidative phosphorylation.
PGK1	Phosphoglycerate Kinase 1: A glycolytic enzyme regulated by HIF1A, involved in ATP generation during glycolysis.
PCA	Principal Component Analysis: A technique for reducing the dimensionality of data while retaining most of the variance.
PVDF	Polyvinylidene Fluoride: A membrane used in Western blotting to transfer proteins from a gel.
QC	Quality Control: Procedures to ensure data accuracy and reliability.
RPMI 1640	Roswell Park Memorial Institute 1640 Medium: A type of cell culture medium often used in cancer research.
rt-qPCR	Real-Time Quantitative Polymerase Chain Reaction: A technique used to quantify mRNA levels in real-time.
SLC25A39	Solute Carrier Family 25 Member 39: A mitochondrial carrier protein involved in regulating glutathione levels.
STAR	Spliced Transcripts Alignment to a Reference: A software tool for aligning RNA-Seq reads to a reference genome.
tdTomato	tdTomato: A red fluorescent protein used as a reporter in cell biology experiments.
UMAP	Uniform Manifold Approximation and Projection: A technique for visualizing high-dimensional data.
UMI	Unique Molecular Identifier: A short sequence tag used to identify individual molecules in sequencing.
Chapter 3	
ADM	Adrenomedullin: A peptide involved in many physiological functions, including vasodilation, and is often upregulated in hypoxia.
ATAC-Seq	Assay for Transposase-Accessible Chromatin coupled with Sequencing: A technique used to study chromatin accessibility.
CA9	Carbonic Anhydrase 9: An enzyme that regulates pH in cells, particularly under hypoxic conditions.
CDK1	Cyclin-Dependent Kinase 1: A protein kinase crucial for the transition from G2 to M phase in the cell cycle.
ChIP-Seq	Chromatin Immunoprecipitation coupled with Sequencing: A method used to analyze protein interactions with DNA.
CYP1A1	Cytochrome P450 Family 1 Subfamily A Member 1: An enzyme involved in the metabolism of xenobiotics and is regulated by hypoxia.
CYP1B1	Cytochrome P450 Family 1 Subfamily B Member 1: An enzyme that metabolizes environmental chemicals and is implicated in cancer.
CYP4F22	Cytochrome P450 Family 4 Subfamily A Member 22: An enzyme involved in the metabolism of fatty acids.
DDIT4	DNA-Damage Inducible Transcript 4: A protein involved in cellular stress responses, including hypoxia.
DEA	Differential Expression Analysis: A method used to identify genes that are expressed differently across conditions or cell types.
EMT	Epithelial-Mesenchymal Transition: A process where epithelial cells transform into mesenchymal cells, gaining migratory and invasive

	properties.
EPAS1	Endothelial PAS domain-containing protein 1: Also known as HIF2A, a transcription factor involved in the response to low oxygen levels.
ESR1	Estrogen Receptor 1: A nuclear hormone receptor that regulates gene expression in response to estrogen.
FASTQ	FAST Quality Format: A text-based format for storing nucleotide sequences and their corresponding quality scores.
FastQC	Fast Quality Control: A tool to assess the quality of high-throughput sequencing data.
FOXM1	Forkhead Box M1: A transcription factor that plays a key role in cell cycle regulation, often upregulated in cancers.
GPI	Glucose Phosphate Isomerase: An enzyme that plays a critical role in glycolysis and gluconeogenesis.
GO	Gene Ontology: A framework for the model of biology that relates genes to known functions, processes, and cellular components.
GREB1	Growth Regulation by Estrogen in Breast cancer 1: A gene involved in estrogen-responsive growth regulation.
IFI44	InterFeron Induced 44: A protein involved in immune responses, particularly in interferon pathways.
FACS	Fluorescence-Activated Cell Sorting: A specialized type of flow cytometry that sorts cells based on fluorescence labeling.
HIF1A	Hypoxia-Inducible Factor 1-Alpha: A transcription factor that mediates cellular response to low oxygen levels.
INSYN1	Inhibitory Synapse-associated protein 1: A protein associated with inhibitory synapses in the nervous system.
KEGG	Kyoto Encyclopedia of Genes and Genomes: A database resource for understanding high-level functions and utilities of the biological system.
KRT23	Keratin 23: A protein that plays a role in cellular structure and has been implicated in cancer progression.
LASSO	Least Absolute Shrinkage and Selection Operator: A regression analysis method used for feature selection and regularization in machine learning.
LINC01016	Long INtergenic Coding RNA 1016: A non-coding RNA involved in gene regulation.
lncRNAs	Long Non-Coding RNAs: Long RNA molecules that do not code for proteins but regulate gene expression.
MKI67	Marker of Ki-67: A protein associated with cellular proliferation, used as a marker to gauge cell proliferation rates.
MMPED2	Metallophosphoesterase Domain-containing 2: A protein involved in cellular communication and signaling.
MSigDB	Molecular Signature Database: A collection of annotated gene sets for use with GSEA and other gene-set analyses.
MT-CYB	Mitochondrial Cytochrome B: A mitochondrial gene involved in the electron transport chain.
NDRG1	N-Myc Downstream-Regulated Gene 1: A stress-responsive protein involved in cellular differentiation and survival under stress conditions.
ORA	Over-Representation Analysis: A statistical method to determine whether specific genes or pathways are over-represented in a particular set of genes.
PCA	Principal Component Analysis: A statistical method used to reduce the dimensionality of large datasets while preserving most of the variance.
PGK1	Phosphoglycerate Kinase 1: An enzyme that plays a key role in

	glycolysis, especially under hypoxic conditions.
PLAC1	Placenta-Specific 1: A protein that is normally expressed in the placenta but can be aberrantly expressed in cancer.
QC	Quality Control: Procedures used to ensure the reliability and accuracy of data.
R	R: A programming language and environment for statistical computing and graphics.
Renoir	Renoir: An in-house machine learning suite used for data analysis and model optimization.
RNA-Seq	RNA Sequencing: A sequencing technique to study the complete set of RNA transcripts produced by the genome under specific circumstances.
RT-qPCR	Real-time Quantitative Polymerase Chain Reaction: A technique used to amplify and simultaneously quantify a targeted DNA molecule.
SC3	Single-cell Consensus Clustering: A method for identifying robust clusters in single-cell RNA sequencing data.
SLC25A39	Solute Carrier Family 25 Member 39: A mitochondrial carrier protein involved in glutathione transport.
SNAI2	Snail Family Transcriptional Repressor 2: A transcription factor involved in EMT and cancer metastasis.
SPOCK1	Sparc/Osteonectin, Cwcv, and Kazal-like domains proteoglycan 1: A protein involved in extracellular matrix interactions and EMT.
Star	Spliced Transcript Alignment to a Reference: A fast RNA-Seq read aligner used to map reads to a reference genome.
UMAP	Uniform Manifold Approximation and Projection: A technique for visualizing high-dimensional data by reducing it to two or three dimensions.
UBE2C	Ubiquitin-Conjugating Enzyme E2 C: An enzyme involved in the ubiquitination process, essential for cell cycle regulation.
VEGFA	Vascular Endothelial Growth Factor A: A protein that stimulates blood vessel formation, often upregulated in response to hypoxia.
Chapter 4	
AKT	AKt/Protein Kinase B: A key signaling protein involved in cell survival, growth, and metabolism, often activated in cancer.
BCL-2	B-cell CLL/lymphoma 2: A family of proteins involved in the regulation of apoptosis, often implicated in cancer cell survival.
BRCA1	Breast Cancer Associated Gene 1: A gene whose mutations are associated with a higher risk of breast and ovarian cancers.
CD44	Cluster of Differentiation 44: A cell surface glycoprotein involved in cell-cell interactions, migration, and adhesion, often overexpressed in cancer.
CDK1	Cyclin-Dependent Kinase 1: An essential kinase that regulates the cell cycle, particularly the transition from G2 phase to mitosis (M phase).
Clustree	Clustree: A tool used for visualizing clustering trees to optimize clustering parameters in single-cell data analysis.
c-Src	Cellular Sarcoma: A non-receptor tyrosine kinase involved in the regulation of cellular structure and signaling, linked to cancer progression.
CSC	Cancer Stem Cell: Cells within a tumor that possess the ability to self-renew and differentiate, driving tumor growth and recurrence.
DEG	Differentially Expressed Genes: Genes that exhibit significant changes in expression levels between different conditions or treatments.
EGFR	Epidermal Growth Factor Receptor: A receptor tyrosine kinase that,

	when activated, promotes cell proliferation, survival, and migration.
EMT	Epithelial-Mesenchymal Transition: A process by which epithelial cells lose their polarity and adhesion, gaining migratory and invasive properties.
ERK	Extracellular signal-Regulated Kinase: A kinase involved in the MAPK/ERK signaling pathway, regulating cell division, survival, and differentiation.
FAK	Focal Adhesion Kinase: A protein tyrosine kinase involved in cellular adhesion, migration, and survival, often implicated in cancer metastasis.
GSEA	Gene Set Enrichment Analysis: A computational method used to determine whether a set of genes shows statistically significant differences between two biological states.
HIF1A	Hypoxia-Inducible Factor 1Alpha: A key transcription factor subunit that regulates the cellular response to hypoxia.
HIFs	Hypoxia-Induced Factors: Transcription factors that mediate the cellular response to low oxygen levels.
HREs	Hypoxia-Response Elements: DNA sequences in gene promoters that HIFs bind to, activating transcription of hypoxia-responsive genes.
ID1	Inhibitor of DNA-binding 1: A protein that inhibits transcription factor binding, involved in cell differentiation and associated with tumorigenesis.
ID3	Inhibitor of DNA-binding 3: Similar to ID1, it inhibits transcription factor binding and is linked to tumorigenesis and cancer progression.
IFIT2	Interferon-Induced Tetratricopeptide 2: A protein involved in the immune response, particularly in antiviral defense.
IFIT3	Interferon-Induced Tetratricopeptide 3: Similar to IFIT2, involved in antiviral defense and immune response.
IL32	Interleukin 32: A pro-inflammatory cytokine involved in immune response and cancer, particularly in promoting tumorigenesis and metastasis.
IL6	Interleukin 6: A cytokine involved in inflammation and maturation of B cells, also implicated in cancer progression and immune evasion.
ISG15	Interferon-Stimulated Gene 15: A ubiquitin-like protein that is involved in the immune response, particularly in the defense against viral infections.
ITGA2	Integrin Alpha 2: A cell surface receptor involved in cell adhesion and signal transduction, important for cancer cell migration.
L1CAM	L1 Cell Adhesion Molecule: A molecule involved in cell adhesion and signaling, often linked to cancer metastasis and invasion.
LDHA	Lactate Dehydrogenase A: An enzyme involved in the conversion of pyruvate to lactate, linked to cancer metabolism and hypoxia.
MKI67	Marker of Ki-67: A protein that is a marker for cell proliferation, often used in cancer research to measure the growth fraction of a cell population.
MMP1	Matrix Metallopeptidase 1: An enzyme that breaks down extracellular matrix components, facilitating cancer cell invasion and metastasis.
MSigDb	Molecular Signature Database: A collection of gene sets used for gene set enrichment analysis, such as in GSEA.
Myc	Myc Oncogene: A gene that codes for a transcription factor involved in cell cycle progression, apoptosis, and cellular transformation.
NEAT1	Nuclear Paraspeckle Assembly Transcript 1: A long non-coding RNA involved in the regulation of gene expression and linked to cancer progression.

NF-κB	Nuclear Factor kappa-light-chain-enhancer of activated B cells: A protein complex that controls transcription of DNA, cytokine production, and cell survival.
PEA	Pathway Enrichment Analysis: A technique used to identify pathways that are significantly enriched in a list of genes, such as DEGs.
PCA	Principal Component Analysis: A statistical technique used to emphasize variation and bring out strong patterns in a dataset.
PI3K	Phosphoinositide 3-Kinase: A family of enzymes involved in cellular functions such as growth, proliferation, differentiation, and survival.
QC	Quality Control: Procedures ensuring the reliability and accuracy of experimental data.
RCAN1	Regulator of Calcineurin 1: A protein that modulates the activity of calcineurin, involved in various cellular processes including immune response.
SASP	Senescence-Associated Secretory Phenotype: A phenotype where senescent cells secrete inflammatory cytokines, growth factors, and proteases.
ScBrAtlas	Single-Cell Breast Cancer Atlas: A dataset that profiles the gene expression of individual cells in breast cancer, allowing for detailed cellular analysis.
SC3	Single-cell Consensus Clustering: A method for identifying robust clusters in single-cell RNA sequencing data.
SNN	Shared Nearest Neighbor: A method used in clustering algorithms to group cells with similar gene expression profiles.
TIMP1	Tissue Inhibitor of Metalloproteinases 1: A protein that inhibits metalloproteinases, enzymes that degrade the extracellular matrix.
TNBC	Triple-Negative Breast Cancer: A breast cancer subtype lacking estrogen receptor (ER), progesterone receptor (PR), and HER2 expression.
TNFSF10	Tumor Necrosis Factor Superfamily member 10: Also known as TRAIL, a cytokine involved in apoptosis and potentially cancer progression.
TRAIL	TNF-Related Apoptosis-Inducing Ligand: Another name for TNFSF10, involved in inducing apoptosis in cancer cells.
UBE2C	Ubiquitin-Conjugating Enzyme E2 C: An enzyme involved in the ubiquitination process, crucial for cell cycle regulation and progression.
VEGFA	Vascular Endothelial Growth Factor A: A protein that promotes the growth of new blood vessels (angiogenesis), crucial for tumor growth.
VHL	Von Hippel-Lindau: A tumor suppressor protein that targets HIF1A for proteasomal degradation by tagging it with ubiquitin.
ZEB1	Zinc finger E-box-Binding homeobox 1: A transcription factor that promotes EMT and is associated with cancer metastasis.
Chapter 5	
ACKR2	Atypical Chemokine Receptor 2: A receptor that modulates immune responses by scavenging chemokines.
Acta2	Actin Alpha 2: Also known as smooth muscle actin, a marker for smooth muscle cells and myofibroblasts.
Alox12	Arachidonate 12-Lipoxygenase: An enzyme that metabolizes arachidonic acid to produce signaling molecules involved in inflammation and cancer.
AMH	Anti-Müllerian Hormone: A hormone involved in the regulation of reproductive organ development and is also a marker of ovarian reserve.

AMHR2	Anti-Müllerian Hormone Receptor Type 2: A receptor that binds to AMH and mediates its effects, particularly in reproductive tissue development.
Amigo2	Adhesion Molecule with Ig Like Domain 2: A protein involved in cell adhesion and neuronal development, with potential roles in cancer.
Atp13a3	ATPase 13A3: A protein involved in cation transport across membranes, potentially implicated in neurological disorders.
Bgn	Biglycan: A small leucine-rich proteoglycan involved in the organization of the extracellular matrix and associated with fibrosis and cancer.
Ccl21a	Chemokine (C-C Motif) Ligand 21A: A chemokine involved in lymphocyte trafficking and immune responses, with implications in cancer metastasis.
Cdh1	Cadherin 1: Also known as E-cadherin, a cell adhesion molecule critical for maintaining epithelial cell integrity, often downregulated in cancer.
Clecl4a	C-Type Lectin Domain Family 4 Member A: A protein involved in immune responses, particularly in recognizing pathogen-associated molecular patterns.
Col13a1	Collagen Type XIII Alpha 1 Chain: A type of collagen involved in the structural integrity of tissues, particularly in muscle and skin.
Col1a1	Collagen Type I Alpha 1 Chain: A fibrillar collagen essential for the tensile strength of skin, tendons, and bones.
Col4a1	Collagen Type IV Alpha 1 Chain: A major component of the basement membrane, playing a role in maintaining tissue structure and function.
Col4a2	Collagen Type IV Alpha 2 Chain: Works alongside Col4a1 in forming the basement membrane, crucial for tissue integrity.
Cox4i2	Cytochrome C Oxidase Subunit 4 Isoform 2: A subunit of cytochrome c oxidase, which is involved in the electron transport chain and cellular respiration.
CXCL14	Chemokine (C-X-C motif) ligand 14: A chemokine involved in immune surveillance and possibly in cancer progression.
CypA	Cyclophilin A: A peptidyl-prolyl isomerase involved in protein folding and immune response, also implicated in cancer progression.
Dcn	Decorin: A proteoglycan that interacts with collagen and other matrix proteins, playing a role in tissue organization and wound healing.
DEA	Differential Expression Analysis: A statistical approach used to identify differences in gene expression levels between different conditions or cell types.
DEGs	Differentially Expressed Genes: Genes that show statistically significant changes in expression levels between different conditions or groups.
Dusp1	Dual Specificity Phosphatase 1: An enzyme that dephosphorylates MAPKs, playing a role in regulating cellular responses to stress and growth signals.
ECM	Extracellular Matrix: A complex network of proteins and other molecules that provide structural and biochemical support to surrounding cells.
ECs	Endothelial Cells: Cells that line the interior surface of blood vessels and are involved in various physiological functions, including angiogenesis.
E0771	E0771: A murine breast cancer cell line derived from lung metastases, commonly used in cancer research.
Emp2	Epithelial Membrane Protein 2: A protein involved in cell adhesion and signaling, with roles in tumor progression and immune response.
EMT	Epithelial-Mesenchymal Transition: A biological process where epithelial cells transform into mesenchymal stem cells, aiding in cancer metastasis.

Fabp4	Fatty Acid-Binding Protein 4: A protein involved in the transport of fatty acids and other lipids within cells, with roles in metabolism and inflammation.
Fabp5	Fatty Acid-Binding Protein 5: Similar to Fabp4, it is involved in intracellular lipid transport and is associated with metabolic diseases and cancer.
FACS	Fluorescence-Activated Cell Sorting: A specialized type of flow cytometry that sorts a mixture of cells into distinct populations based on fluorescent labeling.
FN1	Fibronectin 1: A high-molecular-weight glycoprotein of the extracellular matrix that plays a crucial role in cell adhesion, growth, and migration.
Gja5	Gap Junction Protein Alpha 5: Also known as connexin 40, a protein that forms gap junctions for cell-cell communication, particularly in the heart.
GPCR	G-Protein-Coupled Receptor: A large family of cell surface receptors that respond to various external signals and activate intracellular pathways.
GSEA	Gene Set Enrichment Analysis: A method used to determine whether a predefined set of genes shows statistically significant differences in expression between two biological states.
Hba-x	Hemoglobin Subunit Alpha-X: A variant of the hemoglobin alpha chain, expressed during embryonic development.
Hbb-bh1	Hemoglobin Subunit Beta-BH1: A variant of the hemoglobin beta chain, expressed during embryonic development.
Hbb-y	Hemoglobin Subunit Beta-Y: Another variant of the hemoglobin beta chain, also expressed during embryonic development.
HGF	Hepatocyte Growth Factor: A growth factor that stimulates cell growth, cell motility, and morphogenesis, playing a key role in angiogenesis and wound healing.
HIAR	Hypoxia-Induced Angiogenesis Regulator: A molecule involved in the regulation of angiogenesis in response to low oxygen levels, often associated with tumor growth.
Hmcn1	Hemicentin 1: An extracellular matrix protein involved in tissue integrity and repair, particularly in the skin and vascular systems.
HSP90	Heat Shock Protein 90: A molecular chaperone that assists in protein folding, stabilization, and degradation, with a critical role in cancer cell survival.
Ifitm3	Interferon-Induced Transmembrane Protein 3: A protein that plays a role in antiviral responses by inhibiting viral entry into cells.
IL-32	Interleukin 32: A cytokine involved in inflammatory responses and may play a role in cancer development.
Itga1	Integrin Subunit Alpha 1: A protein that pairs with β1 integrin to mediate cell adhesion and signal transduction between cells and the extracellular matrix.
Itgb1	Integrin Subunit Beta 1: A component of various integrins involved in cell adhesion, migration, and signal transduction.
Krt19	Keratin 19: A type I keratin involved in the structure of epithelial cells, often used as a marker in cancer diagnosis.
LR	Ligand-Receptor: The interaction between a ligand (a signaling molecule) and its receptor, which initiates a cellular response.
Ltbp4	Latent Transforming Growth Factor Beta Binding Protein 4: A protein that modulates the activity of TGF-β, playing a role in tissue repair and fibrosis.
Maf	MAF BZIP Transcription Factor: A transcription factor involved in cellular

	differentiation, particularly in the immune system and lens development.
MAPK	Mitogen-Activated Protein Kinase: A kinase enzyme that plays a critical role in cellular signaling, affecting cell division, differentiation, and stress responses.
mCherry	mCherry: A red fluorescent protein used in molecular biology to label and visualize cells or proteins.
MHC-II	Major Histocompatibility Complex Class II: Molecules found on antigen-presenting cells that are crucial for initiating immune responses by presenting antigens to T-cells.
Miz-1	Myc-Interacting Zinc Finger Protein 1: A transcription factor that interacts with c-Myc, involved in regulating cell proliferation and differentiation.
MMP	Matrix Metalloproteinase: A group of enzymes that break down extracellular matrix proteins, playing a role in tissue remodeling and cancer metastasis.
MsIn	Mesothelin: A cell surface protein overexpressed in several types of cancer, making it a target for immunotherapy.
PECAM1	Platelet Endothelial Cell Adhesion Molecule 1: A cell adhesion molecule critical for leukocyte transmigration and vascular integrity.
Pdgfra	Platelet-Derived Growth Factor Receptor Alpha: A receptor tyrosine kinase that binds to PDGF and plays a role in cell growth and development, especially in mesenchymal cells.
Plek	Pleckstrin: A protein involved in the regulation of cytoskeletal organization and signal transduction in platelets and immune cells.
Pltap	Phospholipid Transfer Protein: A protein involved in lipid metabolism, facilitating the transfer of phospholipids between lipoproteins.
Plvap	Plasmalemma Vesicle Associated Protein: A protein involved in the formation of diaphragms in fenestrated endothelia, playing a role in vascular permeability.
Postn	Periostin: A matricellular protein involved in tissue repair, fibrosis, and cancer, particularly in promoting tumor cell survival and metastasis.
Prox1	Prospero Homeobox Protein 1: A transcription factor critical for lymphatic development and also implicated in cancer progression.
QC	Quality Control: Procedures implemented to ensure the reliability and accuracy of experimental data.
Rarres2	Retinoic Acid Receptor Responder Protein 2: A protein involved in the regulation of inflammation and adipogenesis, also known as chemerin.
RGD	Arginine-Glycine-Aspartic Acid Motif: A peptide sequence found in many proteins involved in cell adhesion and signaling.
Rpl23a	Ribosomal Protein L23a: A component of the 60S ribosomal subunit involved in protein synthesis.
Rpl7a	Ribosomal Protein L7a: A component of the ribosome involved in protein synthesis, playing a role in cellular growth and proliferation.
SC3	Single-cell Consensus Clustering: A clustering technique for single-cell RNA-seq data that integrates multiple clustering results to improve accuracy.
SCMs	Single-cell Clustering Methods: Computational techniques used to classify single-cell RNA sequencing data into distinct cell populations.
Sec14l3	SEC14 Like Lipid Binding Protein 3: A protein that may play a role in lipid metabolism and intracellular transport.
Sema3c	Semaphorin 3C: A signaling molecule involved in neural development and cancer, particularly in guiding cell movement and angiogenesis.
Serping1	Serpin Family G Member 1: A protein involved in the regulation of the

	complement system, playing a role in immune responses and inflammation.
SFTPA1	Surfactant Protein A1: A protein that plays a role in innate immunity and maintaining lung stability by reducing surface tension.
Sftpb	Surfactant Protein B: Essential for lung function, facilitating the spread of surfactant on the lung surface.
Sftpc	Surfactant Protein C: A protein involved in lung function, reducing surface tension in the alveoli and preventing lung collapse.
SMA	Smooth Muscle Actin: An actin isoform found in smooth muscle cells, used as a marker for identifying these cells.
SMCs	Smooth Muscle Cells: Muscle cells that are found within the walls of blood vessels and organs, involved in involuntary contractions.
SNN	Shared Nearest Neighbor: A graph-based clustering method often used in single-cell analysis to identify cell populations based on their gene expression profiles.
Sparc	Secreted Protein Acidic and Rich in Cysteine: A protein that modulates cell-matrix interactions and is involved in tissue remodeling and cancer.
Sparcl1	Secreted Protein Acidic and Rich in Cysteine-Like 1: A protein involved in regulating cell adhesion and migration, particularly in the nervous system and cancer.
Tcf21	Transcription Factor 21: A transcription factor involved in the regulation of cell differentiation and organ development, with roles in both development and cancer.
Tm4sf1	Transmembrane 4 L Six Family Member 1: A cell surface protein involved in cell proliferation, migration, and cancer metastasis.
TME	Tumor Microenvironment: The environment surrounding a tumor, including surrounding blood vessels, immune cells, and other cells, which influences cancer progression.
Top2a	DNA Topoisomerase II Alpha: An enzyme that controls the topological states of DNA during transcription and replication, often targeted in cancer therapy.
Trp53i11	Tumor Protein P53 Inducible Protein 11: A protein that is regulated by p53 and may play a role in apoptosis and cancer biology.
Ucp2	Uncoupling Protein 2: A mitochondrial protein that regulates the production of reactive oxygen species and energy balance, linked to metabolic disorders and cancer.
UMAP	Uniform Manifold Approximation and Projection: A dimensionality reduction technique that helps visualize high-dimensional data by mapping it to two or three dimensions.
Upk3b	Uroplakin 3B: A protein involved in the structure of the urothelium, the tissue lining the bladder and other parts of the urinary tract.
VEGF	Vascular Endothelial Growth Factor: A signaling protein that promotes the growth of new blood vessels, often upregulated in tumors to support angiogenesis.
VEGFR	Vascular Endothelial Growth Factor Receptor: A receptor that binds VEGF, triggering signaling pathways that promote blood vessel formation.
Vwf	von Willebrand Factor: A blood glycoprotein involved in hemostasis, essential for platelet adhesion to sites of vascular injury.
ZEB1	Zinc finger E-box-Binding homeobox 1: A transcription factor that promotes EMT and is associated with cancer metastasis.

Notation

Due to the computational focus of this thesis, it is essential to clearly reference the software functions and packages used in the analyses. For R, the notation ***package::function()*** will be employed to indicate both the package name and the function it contains. This notation mirrors R's syntax, where the double colon differentiates the package from its function.

Table of Functions

Function	Description
Seurat	
Seurat::subset()	Subsets a Seurat object based on quality control metrics, typically used for initial data cleaning.
Seurat::RunPCA()	Performs principal component analysis on single-cell RNA sequencing data stored within a Seurat object.
Seurat::RunUMAP()	Executes Uniform Manifold Approximation and Projection (UMAP) for dimensionality reduction of scRNA-Seq data.
Seurat::FindNeighbors()	Constructs a nearest neighbor graph essential for subsequent community detection algorithms.
Seurat::FindClusters()	Conducts clustering analysis to identify groups of similar cells based on gene expression patterns.
Seurat::FindMarkers()	Identifies differentially expressed genes between two groups, aiding in the characterization of cell types.
Seurat::FindAllMarkers()	Detects differentially expressed genes across individual clusters to pinpoint unique cellular signatures.
Seurat::DimPlot()	Generates UMAP plots annotated by metadata columns or categorical variables, facilitating visual data exploration.
Seurat::FeaturePlot()	Creates plots annotated by gene expression levels or continuous variables, useful for gene expression profiling.
Seurat::ElbowPlot()	Produces elbow plots to help determine the optimal number of principal components to retain.
Seurat::VlnPlot()	Generates violin plots for visualizing the distribution of gene expression across cells.
Seurat::CellCycleScoring()	Scores cell cycle phases to classify cells into different stages, which is critical for studies of cell cycle effects.
Seurat::SCTransform()	Applies normalization, detects variable features, and scales expression data using Pearson residuals for improved analysis accuracy.
SCPubr	
SCpubr::do_FeaturePlot()	Offers enhanced aesthetics for feature plots over those provided by Seurat, improving the visual representation of data.
SCpubr::do_DimPlot()	Provides superior visual aesthetics for dimensional reduction plots compared to Seurat's implementation.
SCpubr::do_VlnPlot()	Delivers enhanced visual quality for violin plots, making it easier to discern expression distributions.
clustree	
clustree::clustree()	Generates clustering trees to visualize the relationships between clusters at different resolutions, aiding in the selection of optimal clustering parameters.
ShinyCell	
ShinyCell::CreateConfig()	Generates a configuration file for Shiny applications, streamlining

	the setup process for interactive data visualization.
ShinyCell::makeShinyApp()	Builds interactive Shiny applications tailored to display scRNA-Seq data, enhancing data exploration and presentation.
Monocle3	
Monocle3::cluster_cells()	Performs clustering analysis using either the Louvain or Leiden algorithm on scRNA-Seq data, helping identify distinct cellular populations.
Monocle3::plot_cells()	Primary plotting function that displays cells using PCA or UMAP coordinates, annotated with metadata or features for comprehensive visual insights.
Monocle3::learn_graph()	Learns a principal graph from dimensionality reduction outputs to trace cell developmental trajectories.
Monocle3::order_cells()	Orders cells in pseudotime based on their positions in a learned principal graph, critical for understanding cell development and differentiation pathways.
CellChat	
CellChat::identifyOverExpressedGenes()	Identifies genes that are over-expressed in each cluster, providing insights into the unique functional roles of different cell groups.
CellChat::identifyOverExpressedInteractions()	Detects interactions between ligand-receptor pairs that are significantly over-expressed across clusters, highlighting potential communication pathways.
CellChat::computeCommunProb()	Calculates communication probabilities within the network among potential interacting cell groups, offering a quantitative measure of intercellular communication.
CellChat::netVisual_circle()	Generates circular plots that visually represent the signaling network, aiding in the interpretation of complex interaction data.
CellChat::netVisual_chord()	Produces chord diagrams to illustrate the directional flow of signals within the cellular network, facilitating understanding of cell signaling dynamics.
CellChat::netAnalysis_signalingRole_heatmap()	Creates heatmaps that summarize the key signaling roles within the network, providing a clear overview of signaling patterns and interactions.
CellChat::netAnalysis_contribution()	Generates bar plots that detail the contribution of individual ligand-receptor pairs to specific signaling pathways, helping to pinpoint critical components of cellular communication.

1. Chapter 1: Introduction

1.1. Breast Cancer Overview: Epidemiology, Molecular Subtypes, and Current Therapeutic Strategies

Breast cancer remains the most diagnosed cancer worldwide, with an annual incidence of 2.3 million cases, a mortality rate of 685,000, and a five-year prevalence of 7.8 million (Coughlin, 2013). When detected early, it is also one of the most effectively treated cancers, boasting a five-year survival rate exceeding 90%.

Despite advances in treatment and detection, the incidence of breast cancer has surged by 200% over the past four decades across both developed and developing countries (Lidegaard & Kroman, 2005). This increase is attributed to a multifactorial aetiology, incorporating both environmental and genetic factors. Key environmental factors include family history, obesity, alcohol use, age, and reproductive history, such as contraceptive use, late menopause, early menarche, and delayed first childbirth. Genetically, mutations in BRCA1 and BRCA2 are significant predictors of increased cancer risk (Coughlin, 2013). Given its ongoing prevalence and impact, there is a pressing need for enhanced research into more effective detection techniques and therapeutic interventions.

Breast cancer is classified into several molecular subtypes based on the expression profiles of estrogen receptor (ER), progesterone receptor (PR), and human epidermal growth factor receptor 2 (HER2) **[Figure 1-1]** (Singh & Yadav, 2021). Breast cancer is categorized into several molecular subtypes, each with distinct characteristics and treatment approaches. Luminal A, characterized by the presence of estrogen receptor (ER), progesterone receptor (PR), and the absence of HER2 (ER+, PR+, HER2-), typically has the best prognosis and is primarily treated with hormonal therapy. Luminal B subtype (ER+, PR+, HER2+) has a worse prognosis compared to Luminal A and is treated with both hormonal therapy and HER2-targeted agents. The HER2-enriched subtype, which lacks ER and PR but expresses HER2 (ER-, PR-, HER2+), is known for its aggressiveness and poor prognosis, managed with HER2-targeted therapies.

Triple-negative breast cancer (TNBC), characterized by the absence of estrogen receptor (ER), progesterone receptor (PR), and human epidermal growth factor receptor 2 (HER2) expression (ER^-, PR^-, $HER2^-$), represents a highly aggressive subtype distinguished by extensive metastatic potential and intrinsic resistance to conventional therapeutic modalities, rendering it the most clinically challenging breast cancer variant to manage. While recent therapeutic advances have introduced targeted interventions including poly(ADP-ribose) polymerase (PARP) inhibitors for BRCA-mutated tumors and immune checkpoint inhibitors such as pembrolizumab in combination with chemotherapy, the clinical efficacy of these approaches remains constrained by the rapid emergence of acquired resistance mechanisms and limited patient response rates. Consequently, the heterogeneous molecular landscape and paucity of universally targetable oncogenic drivers in TNBC underscore the critical imperative for developing novel precision therapeutic strategies that can circumvent resistance pathways and address the underlying biological complexity of this aggressive malignancy.

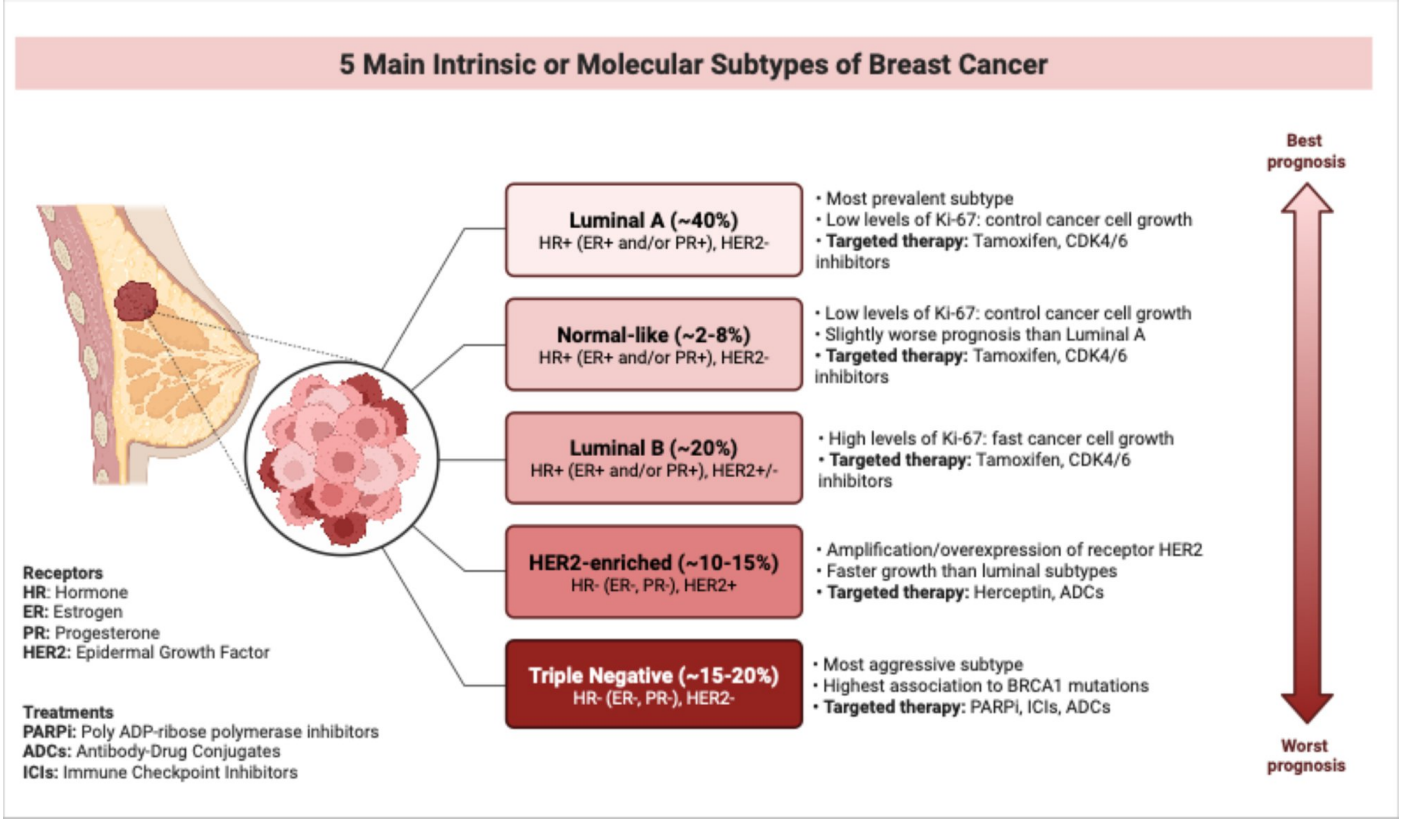


Figure 1-1: Main Intrinsic or Molecular Subtypes of Breast Cancer *– This diagram categorizes the five principal molecular subtypes of breast cancer based on receptor status: Luminal A, Luminal B, HER2-enriched, Triple Negative (TNBC), and Normal-like. Each subtype is depicted with corresponding characteristics and typical prognostic outcomes:* ***Luminal A (HR+, HER2-):*** *Representing approximately 40% of cases, this is the most prevalent subtype, characterized by low levels of the Ki-67 marker, indicating slower cancer cell growth. Common treatments include targeted therapies such as hormonal therapy.* ***Luminal B (HR+, HER2+/–):*** *Constituting about 20% of cases, this subtype generally exhibits higher Ki-67 levels than Luminal A, suggesting faster cell proliferation. Treatments often involve a combination of hormonal therapy and, depending on HER2 status, targeted HER2 therapies.* ***HER2-enriched (HR–, HER2+):*** *Making up about 10-15% of cases, these tumors typically show overexpression of the HER2 receptor and a more aggressive growth pattern than Luminal subtypes. Targeted HER2 therapies like Herceptin are commonly used.* ***Triple Negative (HR–, PR–, HER2–):*** *Accounting for 15-20% of cases, TNBC is noted for its aggressive behaviour and poor prognosis, particularly due to its high metastatic potential and resistance to conventional therapies. It frequently occurs in younger women and is associated with BRCA1 mutations. Figure adapted from (Singh & Yadav, 2021).*

For estrogen receptor-positive (ER+) tumours, treatment strategies predominantly involve the use of aromatase inhibitors and selective estrogen receptor modulators (SERMs). Aromatase inhibitors, as illustrated in **[Figure 1-2]**, inhibit the conversion of testosterone to oestradiol and androstenedione to oestrone. This reduction in oestradiol levels decreases ER activation in breast cancer cells, effectively impeding tumor progression. However, resistance to aromatase inhibitors is a significant challenge in aggressive breast cancer cases. This resistance can arise through various mechanisms, including mutations in the ER that lead to constitutive receptor signaling independent of oestradiol. Additional resistance mechanisms include the upregulation of cyclin D1 and CDK4/6, which drive cell cycle progression independent of ER-mediated signaling, overexpression of BCL2 family members to inhibit apoptosis, and activation of alternative growth factor signaling pathways that compensate for reduced ER signaling (C. X. Ma et al., 2015).

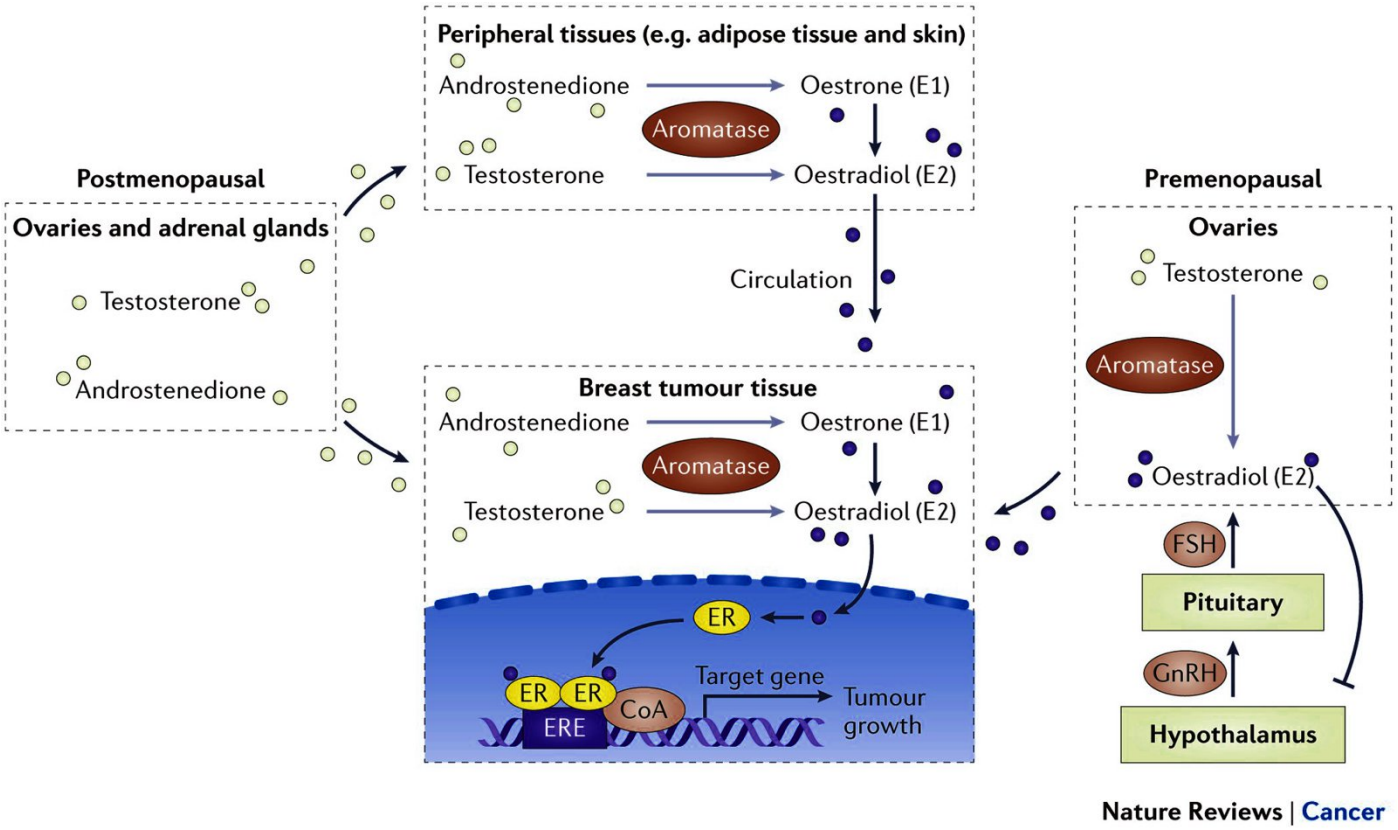


Figure 1-2: Aromatase Inhibitors Mechanism of Action – *Estrogen production occurs through a complex signaling cascade initiated by the hypothalamus releasing Gonadotropin-releasing hormone (GnRH), which stimulates the pituitary gland. The pituitary gland then secretes Follicle-stimulating hormone (FSH) and Luteinizing hormone (LH), which circulate to the ovaries and prompt the production of testosterone in the theca cells. Testosterone is converted to estradiol (E2) via the aromatase enzyme. Aromatase inhibitors target this critical conversion step, reducing the availability of estrogen receptors (ER) for ER+ tumours and consequently hindering tumor growth and progression. Figure adapted from (C. X. Ma et al., 2015).*

SERMs, such as tamoxifen, exhibit tissue-specific modes of action **[Figure 1-3]**. In breast tissues and breast cancers, they function as competitive antagonists to ER, inhibiting ER-mediated signaling. Conversely, in bone and cardiac tissues, SERMs act as partial agonists, thereby mitigating bone density loss associated with breast cancer and menopause and offering cardiovascular benefits. Nonetheless, resistance to SERMs frequently develops due to tumor adaptations, such as mutations in the ER that diminish SERM binding affinity while still permitting oestradiol binding. Tumours may also activate alternative signaling pathways that obviate the need for estrogen signaling, thereby promoting cancer progression despite SERM therapy (Patel & Bihani, 2018).

To address these challenges, selective estrogen receptor degraders (SERDs) have been developed as a more effective treatment for metastatic and resistant ER+ tumours **[Figure 1-3]**. SERDs bind to the ER and induce a conformational change that targets the receptor for degradation. Consequently, SERDs act as universal antagonists across all tissues, irrespective of physiological or pathological context. Despite these advances, resistance to hormonal therapies remains problematic. Nonetheless, the combined use of SERMs and SERDs in targeting ER continues to be a cornerstone in the pharmacological management of ER+ breast cancer (Bhatia et al., 2023).

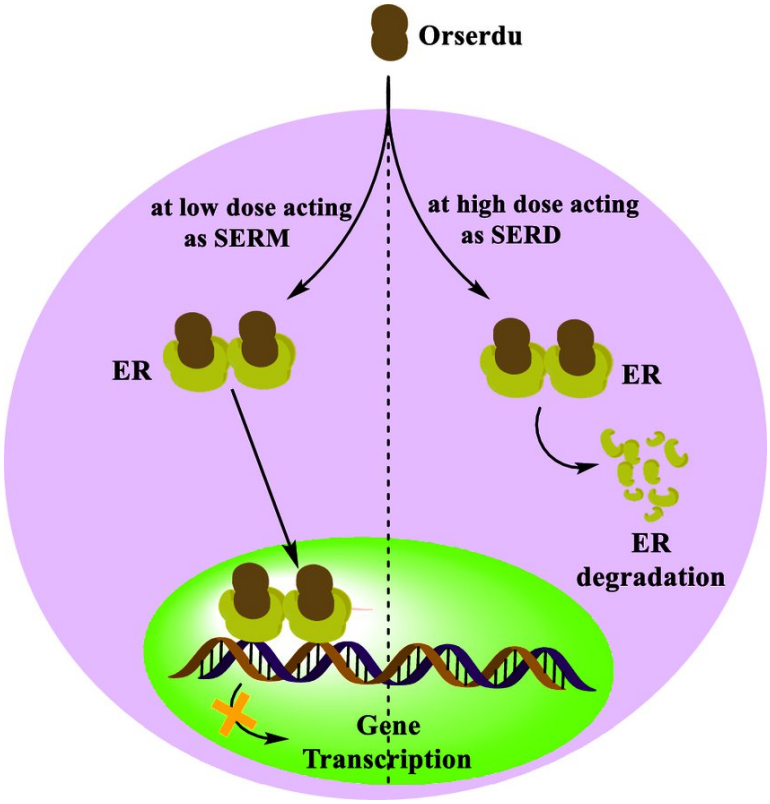


Figure 1-3: Mechanisms of Action of Selective Estrogen Receptor Modulators (SERMs) and Selective Estrogen Receptor Degraders (SERDs) — *This diagram illustrates how SERMs and SERDs modulate estrogen receptor (ER) activity in breast cancer treatment. SERMs function as competitive antagonists to ER in breast tissues, inhibiting estrogen-mediated signaling, while acting as partial agonists in other tissues, such as bone. In contrast, SERDs bind to the ER, inducing a conformational shift that leads to the receptor's degradation, uniformly suppressing ER signaling across all tissues. Orserdu, depicted in the figure, exemplifies a next-generation SERD that uniquely acts as a SERM at low doses and transitions to a SERD mechanism at higher concentrations, showcasing its dual functionality. Figure adapted from (Bhatia & Thareja, 2023).*

For HER2-positive tumours, such as the HER2-enriched and Luminal B subtypes of breast cancer, monoclonal antibodies like Herceptin are crucial in clinical settings to curb tumor progression **[Figure 1-4]**. Herceptin specifically targets the extracellular domain of the HER2 receptor, inhibiting its signaling pathway. By binding to this domain, Herceptin prevents the dimerization of HER2, blocks the cleavage of its extracellular domain—an alternative pathway for its constitutive activation—and tags HER2-positive tumor cells for destruction through antibody-dependent cellular cytotoxicity (ADCC) mediated by its Fc domain (Nami et al., 2018). However, treatment resistance remains a significant challenge, particularly in more aggressive tumours.

Contrastingly, these treatment modalities are ineffective against triple-negative breast cancer (TNBC) due to its lack of the necessary receptors, namely ER, PR, and HER2, which are targeted in other subtypes. The next section will explore the inherent heterogeneity and limited treatment options for TNBC, underscoring why it represents a primary focus of this research project due to its particularly aggressive nature.

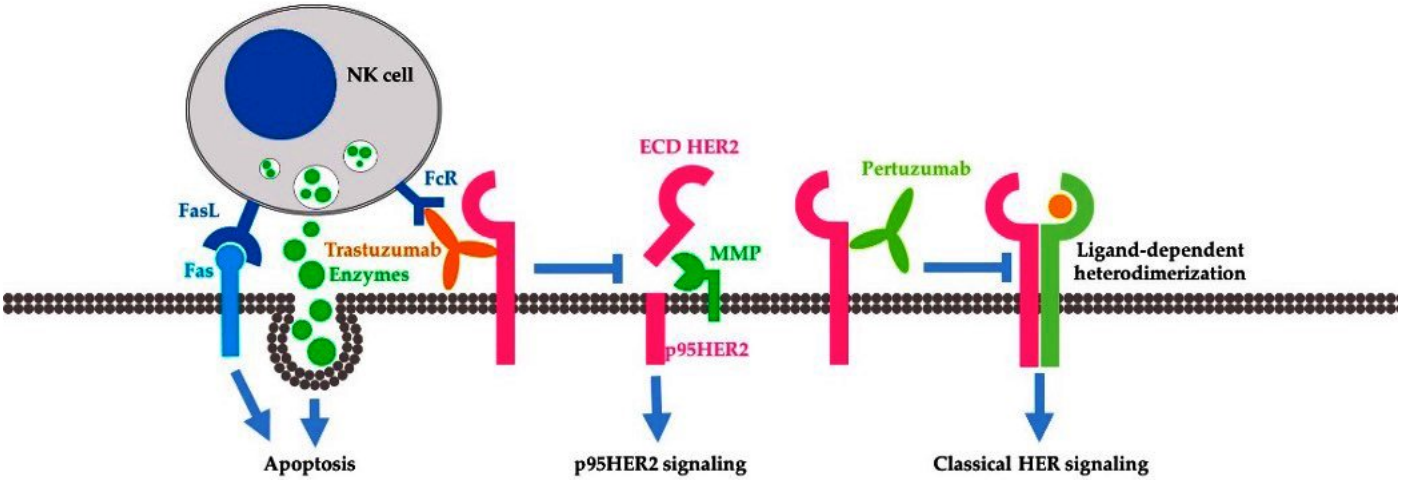


Figure 1-4: Mechanism of Action of Trastuzumab/Herceptin in HER2-Positive Breast Cancer – *Herceptin, a monoclonal antibody, specifically targets the HER2 receptor's extracellular domain to disrupt its signaling pathway. By binding to this domain, Herceptin inhibits HER2 dimerization, essential for receptor activation. It also facilitates antibody-dependent cellular cytotoxicity (ADCC) by marking the receptor for destruction by immune cells and prevents the proteolytic cleavage of HER2's extracellular domain, thus blocking constitutive ligand-independent signaling activation. This illustration highlights the multifaceted actions of Herceptin, including its interaction with NK cells and its impact on downstream HER2 signaling pathways. Figure adapted from (Nami et al., 2018).*

1.2. Current Therapeutic Strategies

Triple-negative breast cancer (TNBC), characterized by the absence of estrogen receptor (ER), progesterone receptor (PR), and human epidermal growth factor receptor 2 (HER2) expression (ER^-, PR^-, $HER2^-$), represents a highly aggressive subtype distinguished by extensive metastatic potential and intrinsic resistance to conventional therapeutic modalities, rendering it the most clinically challenging breast cancer variant to manage. TNBC demonstrates the poorest prognosis among breast cancer subtypes, particularly in metastatic cases, with five-year survival rates dropping to approximately 12% in advanced disease stages (Mustacchi et al., 2019).

The absence of well-characterized molecular targets that define other breast cancer subtypes necessitates reliance on systemic chemotherapy as the primary therapeutic intervention, though recent advances have introduced targeted approaches that exploit specific vulnerabilities within TNBC tumor biology.

1.2.1. Traditional Chemotherapeutic Approaches

Current standard-of-care treatment modalities for TNBC predominantly involve aggressive multi-agent chemotherapy regimens that, while effective in a subset of patients, are associated with substantial systemic toxicity and limited long-term efficacy.

The cornerstone of TNBC chemotherapy encompasses three primary drug classes: taxanes, anthracyclines, and platinum-based agents, each targeting distinct cellular processes essential for cancer cell survival and proliferation.

Taxane-Based Therapy

Taxane derivatives, particularly paclitaxel and docetaxel, represent first-line therapeutic agents in TNBC management, as detailed in **[Figure 1-5]**. Paclitaxel exerts its cytotoxic effects through high-affinity binding to β-tubulin subunits within microtubule structures, promoting microtubule polymerization while simultaneously inhibiting depolymerization. This stabilization disrupts normal microtubule dynamics essential for mitotic spindle formation, leading to prolonged mitotic arrest at the metaphase-anaphase transition through activation of the spindle assembly checkpoint (SAC).

The sustained checkpoint activation ultimately triggers apoptotic cell death pathways, preferentially affecting rapidly dividing cancer cells (Mandapati & Lukong, 2023).

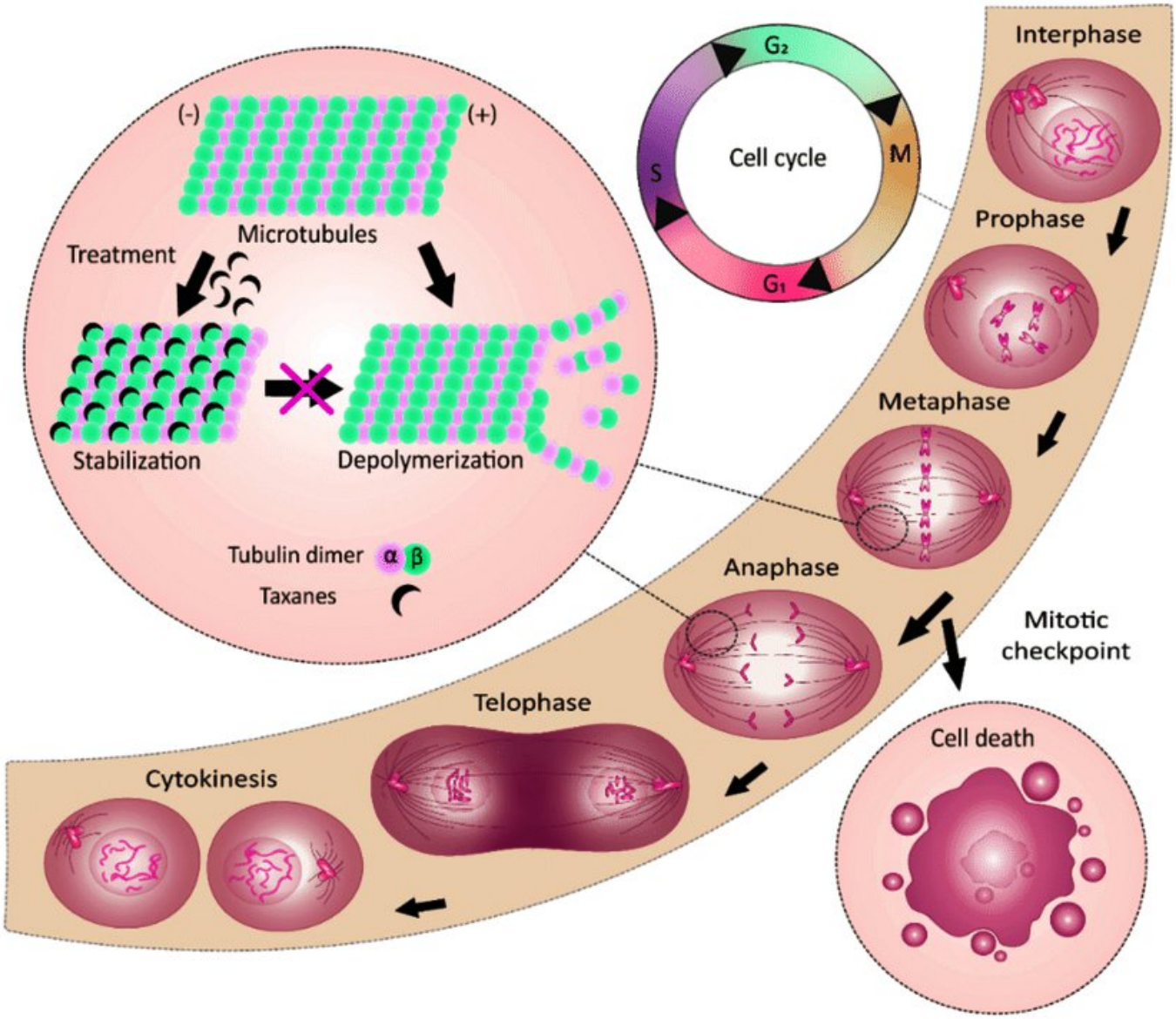


Figure 1-5: Paclitaxel Mechanism of Action in TNBC Therapy *– Paclitaxel, a taxane derivative widely used in treating TNBC, targets and binds to beta-tubulin within the microtubule structures of cancer cells. This interaction leads to the stabilization of microtubules, preventing their normal depolymerization which is crucial for cell division. This stabilization impedes the normal arrangement and function of mitotic spindles during metaphase and anaphase, thereby disrupting cell cycle progression. The interference with mitotic spindle functions triggers the spindle assembly checkpoint (SAC), leading to cell cycle arrest at the metaphase-anaphase transition. This arrest subsequently induces apoptosis as the cell can no longer proceed with division. Through these mechanisms, paclitaxel exerts its cytotoxic effects by both halting cell division and promoting the programmed cell death of TNBC cells. Figure adapted from (Škubník et al., 2021).*

Despite their clinical efficacy, taxane derivatives are associated with significant dose-limiting toxicities, including cumulative peripheral neuropathy, bone marrow suppression, and hypersensitivity reactions that can compromise treatment adherence and patient quality of life.

Anthracycline-Based Therapy

Anthracyclines, including doxorubicin and epirubicin, constitute another fundamental component of TNBC chemotherapy regimens. These agents primarily target topoisomerase II, an essential nuclear enzyme responsible for relieving topological stress during DNA replication and transcription. As illustrated in **[Figure 1-6]**, anthracyclines function through a dual mechanism involving DNA

intercalation and topoisomerase II inhibition. The drugs form stable ternary complexes with DNA and topoisomerase II, preventing enzyme-mediated re-ligation of transiently created double-strand breaks. This results in the accumulation of DNA lesions that overwhelm cellular repair mechanisms, ultimately triggering apoptotic cell death (Venkatesh & Kasi, 2024).

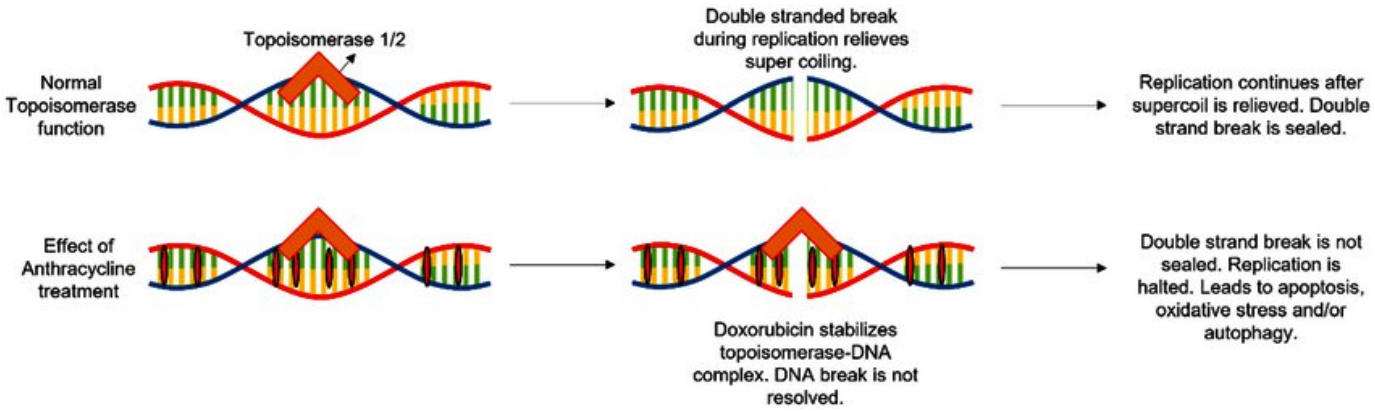


Figure 1-6: Anthracycline Mechanism of Action – *Anthracyclines, a cornerstone of TNBC treatment, exert their therapeutic effects by targeting topoisomerase II. They function by inducing double-stranded DNA breaks, subsequently attracting topoisomerase II to these breaks, where the enzyme is stabilized and effectively trapped within the DNA-enzyme complex. This action sequesters available topoisomerase II enzymes, preventing their participation in further DNA replication processes. The accumulation of unresolved DNA breaks leads to the disruption of DNA replication, resulting in cell cycle arrest and ultimately apoptotic cell death. Figure adapted from (Mandapati & Lukong, 2022).*

The clinical utility of anthracyclines is significantly limited by their propensity to cause dose-dependent cardiotoxicity, attributed to oxidative stress and mitochondrial dysfunction in cardiac myocytes, necessitating careful cardiac monitoring and dose optimization strategies.

Platinum-Based Therapy

Platinum compounds, particularly carboplatin and cisplatin, demonstrate enhanced efficacy in TNBC, especially in tumors harboring BRCA1/2 mutations or other homologous recombination deficiencies. These agents function as DNA cross-linking agents, forming both intra- and inter-strand DNA adducts that impede replication fork progression, as depicted in **[Figure 1-13]**. The cytotoxic efficacy of platinum agents is particularly pronounced in homologous recombination-deficient tumors, where the inability to repair DNA cross-links through high-fidelity repair pathways creates a state of synthetic lethality (Dasari & Tchounwou, 2014).

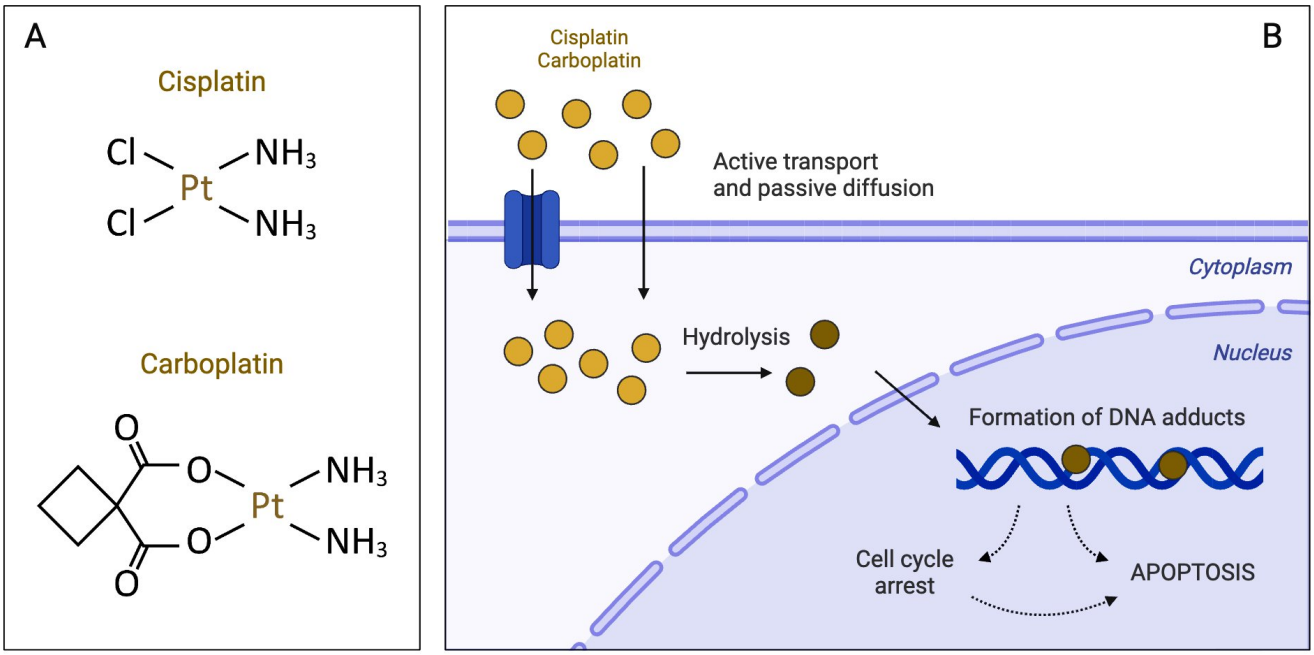


Figure 1-7: Cisplatin Mechanism of Action – *Cisplatin is an example of platinum-based agents used in cancer treatment. Cisplatin is transported inside the cells through a combination of passive diffusion through the cell membrane as well as active transport, and it crosses the nuclear membrane to reach the DNA and induce DNA damage in the form of strand breaks. This is especially effective in BRCA-mutated cells due to the attenuated functionality of the homologous recombination DNA repair pathway, thereby facilitating the persistence of DNA damage that is then responsible for cell cycle arrest and apoptosis. Figure was developed by Tinkara Kreft in Biorender based on this publication (Cheng et al., 2021).*

Clinical application of platinum agents is constrained by their associated nephrotoxicity, neurotoxicity, and ototoxicity, requiring careful patient selection and monitoring protocols.

1.2.2. Targeted Therapeutic Approaches

PARP Inhibitor Therapy

The recognition that a significant subset of TNBC cases exhibit deficiencies in homologous recombination repair, either through germline BRCA1/2 mutations (~15-20% of cases) or other mechanisms affecting the BRCA pathway, has led to the development of poly(ADP-ribose) polymerase (PARP) inhibitors as targeted therapeutic agents. PARP enzymes play crucial roles in single-strand DNA break repair through the base excision repair pathway. PARP inhibition in homologous recombination-deficient cells creates a synthetic lethal interaction, as the accumulation of unrepaired single-strand breaks leads to replication fork collapse and double-strand break formation that cannot be adequately repaired **[Figure 1-8]** (Lord & Ashworth, 2017).

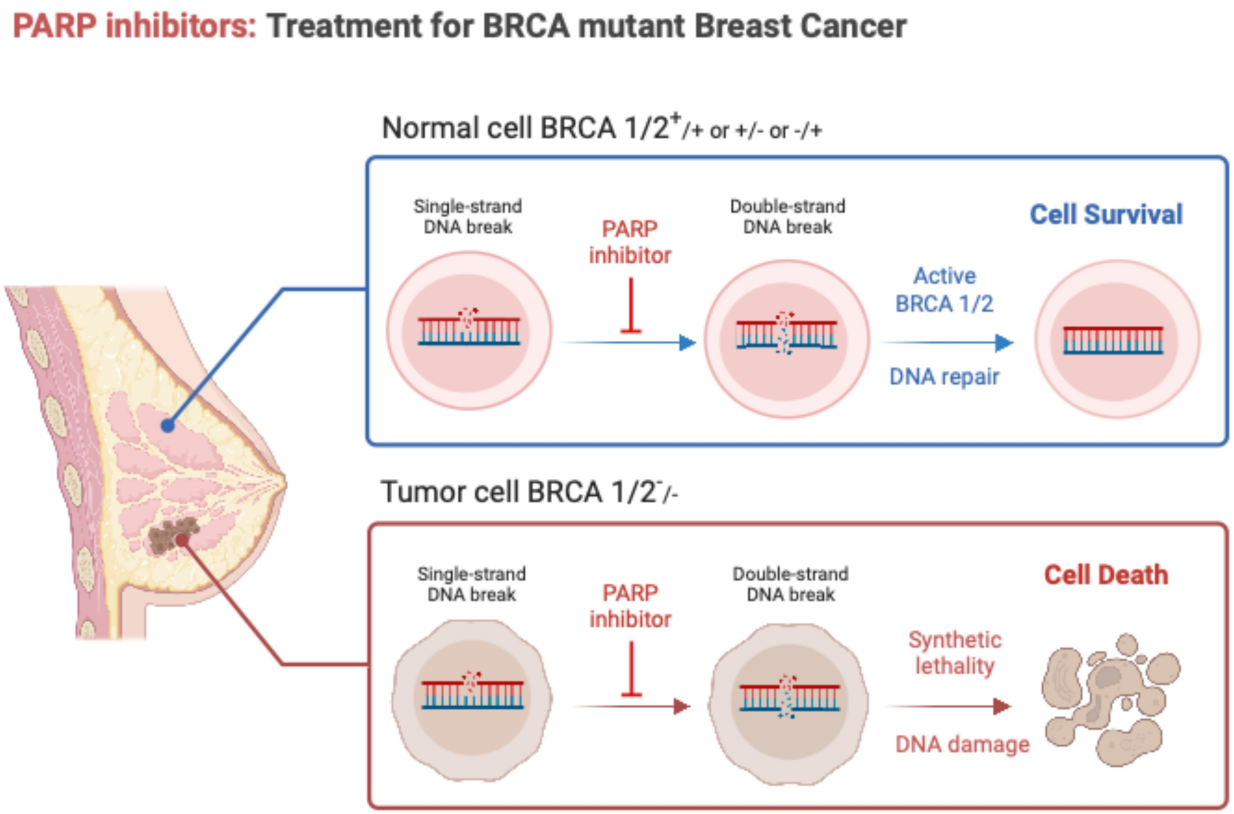


Figure 1-8: Mechanism of action of PARP inhibitors in BRCA-mutant breast cancer — *In normal cells with functional BRCA1/2 proteins (top panel), single-strand DNA breaks are initially processed by PARP enzymes. When PARP inhibitors block this pathway, the resulting double-strand breaks are efficiently repaired by homologous recombination via active BRCA1/2, allowing cell survival. In tumor cells with BRCA1/2 mutations (bottom panel), PARP inhibition leads to accumulation of double-strand breaks that cannot be repaired due to deficient homologous recombination. This creates synthetic lethality, where the combination of BRCA deficiency and PARP inhibition results in excessive DNA damage and tumor cell death. This selective targeting of BRCA-deficient cancer cells while sparing normal cells forms the therapeutic basis for PARP inhibitor treatment in hereditary breast cancers. Adapted from (Polyak & Garber, 2011).*

Olaparib, the first FDA-approved PARP inhibitor for TNBC, demonstrated significant clinical benefit in the OlympiAD trial, showing improved progression-free survival in patients with germline BRCA1/2 mutations compared to standard chemotherapy (Robson et al., 2017). Subsequently, talazoparib received approval based on the EMBRACA trial results, which demonstrated superior efficacy with reduced toxicity compared to physician's choice chemotherapy in the same patient population (Litton et al., 2018).

Despite these advances, resistance to PARP inhibitors emerges through multiple mechanisms, including restoration of homologous recombination capacity through secondary BRCA mutations, upregulation of alternative DNA repair pathways, and drug efflux pump activation. Additionally, the clinical benefit of PARP inhibitors remains largely restricted to the subset of patients with confirmed homologous recombination deficiency, highlighting the need for improved patient selection biomarkers (Pettitt et al., 2018).

Immunotherapeutic Interventions

The recognition that TNBC tumors often exhibit higher mutational burdens and increased tumor-infiltrating lymphocyte (TIL) density compared to other breast cancer subtypes has catalyzed investigation into immunotherapeutic approaches. The relatively immunogenic nature of TNBC, characterized by increased neoantigen presentation and immune cell infiltration, provides a rational basis for immune checkpoint inhibition strategies (Yi et al., 2021).

Pembrolizumab, a monoclonal antibody targeting the programmed death-1 (PD-1) receptor, received FDA approval for TNBC treatment based on results from the KEYNOTE-355 trial. In this study, the combination of pembrolizumab with nab-paclitaxel demonstrated improved progression-free survival compared to chemotherapy alone in patients with PD-L1-positive tumors (combined positive score ≥10) (Cortes et al., 2022). The mechanism of action involves blocking the PD-1/PD-L1 immune checkpoint axis, thereby preventing T-cell exhaustion and restoring anti-tumor immune responses **[Figure 1-9]**.

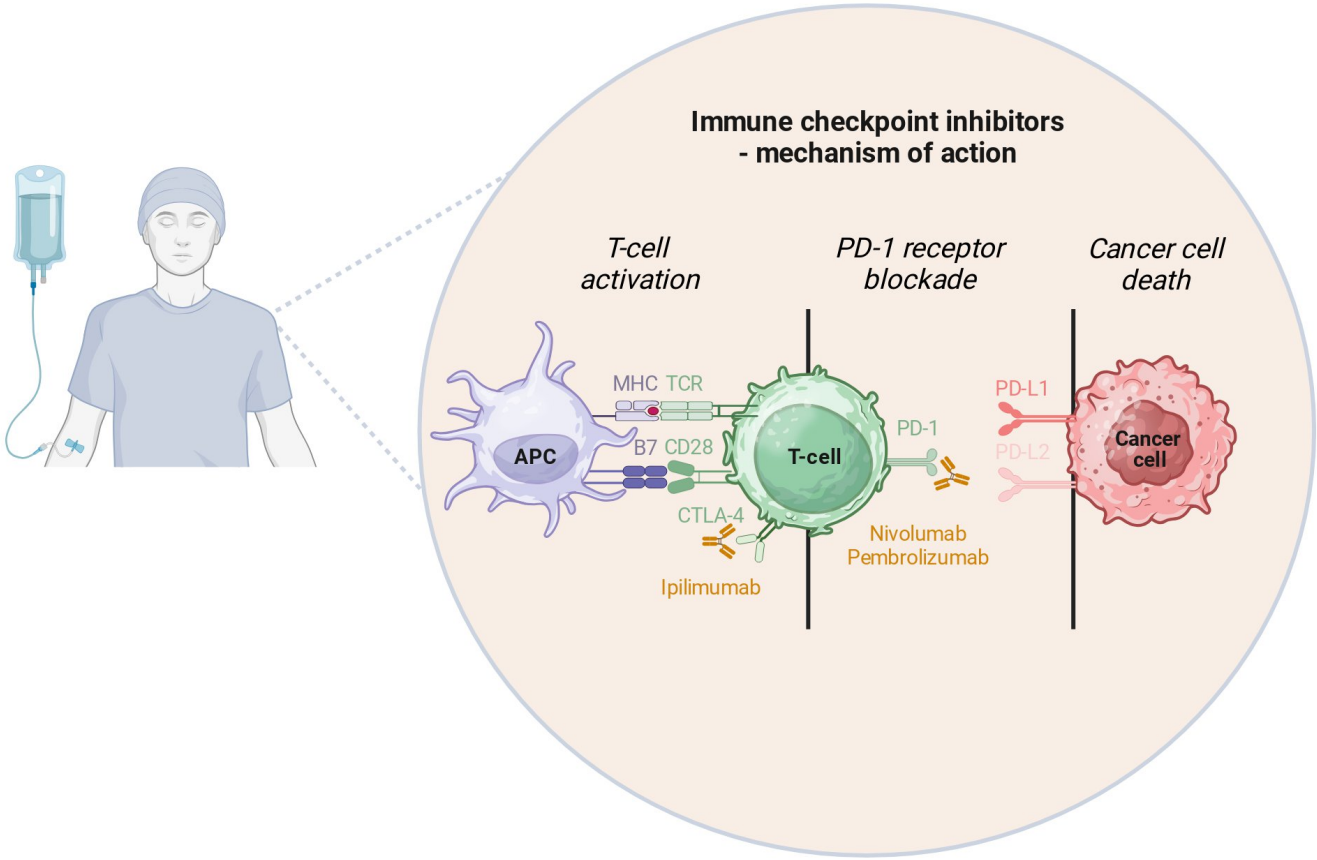


Figure 1-9: Mechanism of action of immune checkpoint inhibitors, specifically pembrolizumab — *The diagram illustrates the three-stage process of immune checkpoint inhibition in cancer treatment. In stage 1 (T-cell activation), antigen-presenting cells (APCs) activate T-cells through MHC-TCR interaction and co-stimulatory signals including B7-CD28 binding. In stage 2 (PD-1 receptor blockade), activated T-cells express PD-1 receptors that would normally be inhibited by PD-L1 on cancer cells, leading to immune suppression. Pembrolizumab and nivolumab block this PD-1/PD-L1 interaction, while ipilimumab targets the CTLA-4 checkpoint pathway. In stage 3 (cancer cell death), checkpoint inhibition prevents immune suppression, allowing T-cells to maintain their cytotoxic activity and effectively target and destroy cancer cells. This mechanism restores the immune system's ability to recognize and eliminate malignant cells by removing inhibitory signals that tumors use to evade immune surveillance. Created in BioRender. Ekheden, I. (2024) https://app.biorender.com/citation/66dec6a0110bdbb20bedfaba.*

Atezolizumab, targeting the PD-L1 receptor, also showed clinical efficacy in combination with nab-paclitaxel in the IMpassion130 trial, though subsequent studies have yielded mixed results, emphasizing the importance of patient selection based on PD-L1 expression levels (Schmid et al., 2020).

Despite these therapeutic advances, response rates to immune checkpoint inhibitors in TNBC remain modest (approximately 20-30%), and resistance mechanisms including immune evasion, altered antigen presentation, and immunosuppressive tumor microenvironment factors limit long-term efficacy. Furthermore, the identification of optimal predictive biomarkers beyond PD-L1 expression remains an active area of investigation (Y. Liu et al., 2023).

1.2.3. Emerging Therapeutic Strategies

Antibody-Drug Conjugates

Sacituzumab govitecan, an antibody-drug conjugate targeting Trop-2, represents a novel therapeutic approach that has demonstrated significant clinical activity in heavily pre-treated metastatic TNBC. The ASCENT trial showed improved overall survival and progression-free survival compared to standard chemotherapy, leading to FDA approval for this indication (Bardia et al., 2021).

CDK4/6 Inhibitor Combinations

While CDK4/6 inhibitors have shown limited single-agent activity in TNBC due to frequent loss of retinoblastoma protein function, combination strategies with other targeted agents or immunotherapy are under active investigation to overcome resistance mechanisms (Hu et al., 2021).

1.2.4. Therapeutic Challenges and Resistance Mechanisms

The clinical management of TNBC remains significantly hampered by several interconnected challenges that limit therapeutic efficacy and contribute to poor patient outcomes. The extensive cytotoxicity associated with current chemotherapeutic regimens results from their non-selective mechanism of action, causing significant collateral damage to rapidly dividing normal tissues. This approach is predicated on the differential DNA repair capacity between tumor and normal cells, yet the therapeutic window remains narrow, resulting in substantial treatment-related morbidity.

Moreover, TNBC exhibits remarkable intertumoral and intratumoral heterogeneity, which significantly complicates therapeutic approaches. Molecular profiling studies have identified distinct TNBC subtypes, including basal-like 1 and 2, immunomodulatory, mesenchymal, mesenchymal stem-like, and luminal androgen receptor subtypes, each characterized by unique biological behaviors and therapeutic vulnerabilities **[Figure 1-10]**. This molecular diversity, combined with the limited availability of universally targetable oncogenic drivers, underscores the critical need for precision medicine approaches that can address the underlying biological complexity of this heterogeneous disease.

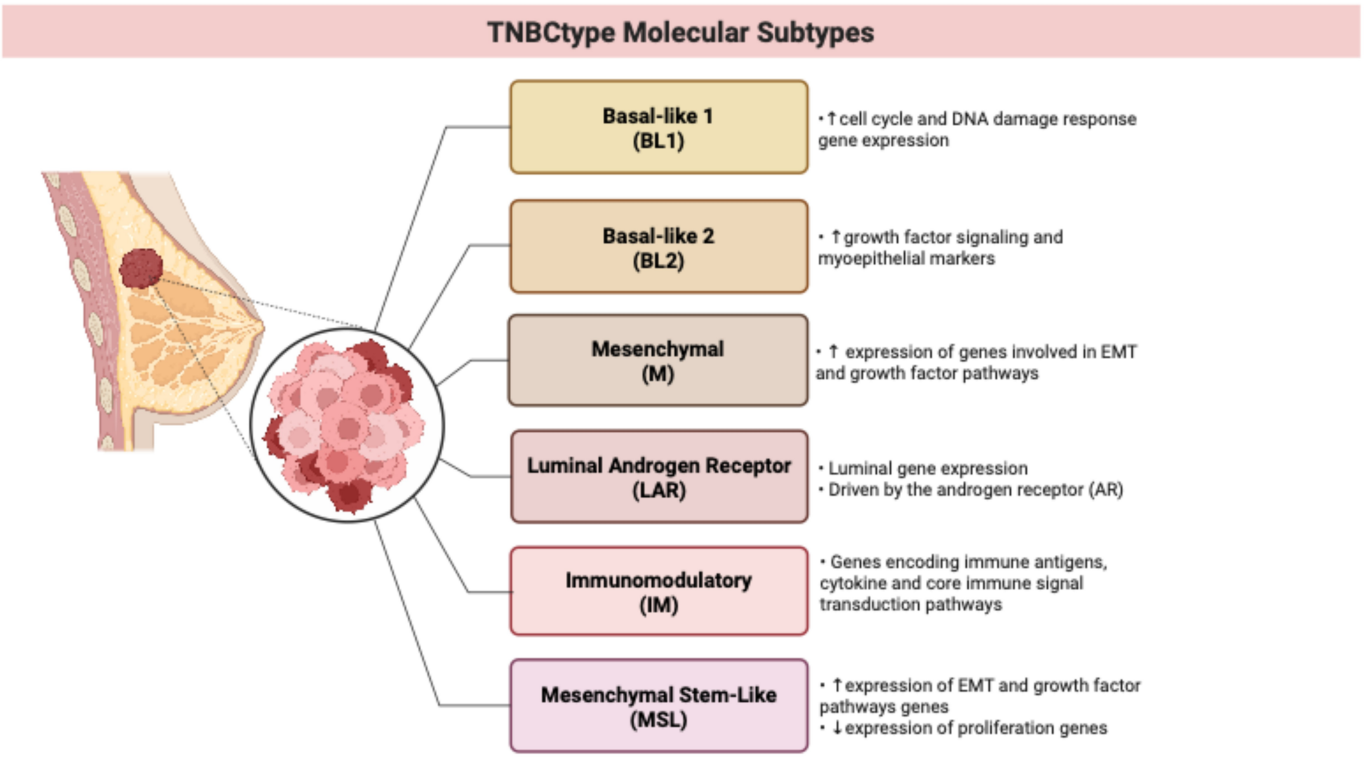


Figure 1-10: TNBC Subtype Heterogeneity – *The management and treatment of Triple-Negative Breast Cancer (TNBC) is notably complex due to its inherent subtype heterogeneity. There are several distinct subtypes of TNBC, each characterized by unique biological behaviours and responses to treatment. These include Basal-like 1 (BL1), Basal-like 2 (BL2), Immunomodulatory (IM), Mesenchymal (M), Mesenchymal Stem-like (MSL), and Luminal Androgen Receptor (LAR) subtypes. Each subtype exhibits specific traits such as differential growth factors expression, immune response modulation, and varying degrees of epithelial-to-mesenchymal transition, which influence their interactions with the surrounding microenvironment. This variability not only complicates the therapeutic approach but also contributes to genomic instability, enhanced proliferation, therapeutic resistance, and progression through the invasion-metastasis cascade. Figure adapted from (C.-J. Li et al., 2021).*

The development of acquired resistance represents another major therapeutic challenge, occurring through diverse mechanisms including enhanced DNA repair capacity, activation of alternative survival pathways, immune evasion, and selection of resistant tumor cell clones. These resistance mechanisms often emerge rapidly, limiting the durability of therapeutic responses and necessitating the development of combination strategies and next-generation targeted agents.

1.2.5. Future Therapeutic Directions

The heterogeneous molecular landscape of TNBC necessitates continued research focused on identifying novel therapeutic targets and developing precision medicine strategies. Areas of active investigation include targeting metabolic vulnerabilities, exploiting synthetic lethal interactions beyond BRCA deficiency, enhancing immunotherapeutic responses through combination approaches, and developing strategies to overcome treatment resistance. The integration of advanced genomic profiling, biomarker discovery, and novel drug development approaches holds promise for improving therapeutic outcomes in this challenging malignancy.

The following section will examine the critical influence of hypoxia on TNBC biology, exploring its impact on tumor progression, therapeutic resistance, and the tumor microenvironment, as understanding these hypoxic responses may reveal additional therapeutic vulnerabilities and inform the development of more effective treatment strategies.

1.3. Hypoxia and The Hypoxic Response

Hypoxia within the tumor microenvironment (TME) significantly contributes to the aggressiveness and heterogeneity observed in Triple-Negative Breast Cancer (TNBC). Characterized by rapid tumor

proliferation, TNBC frequently results in more pronounced hypoxia compared to other breast cancer subtypes. This severe hypoxia stems from an imbalance between the rapid tumor growth and the available blood supply, leading to areas within the tumor that suffer from critically low oxygen and nutrient levels (Srivastava et al., 2023).

In such low oxygen conditions, TNBC cells activate the hypoxic response, a physiological adaptation mechanism. Under typical conditions, this response helps adjust the stromal environment to restore oxygen levels to tissue-specific normal ranges, known as physoxia, which for most tissues varies between 3-8%. For breast tissue specifically, physoxia is typically around 6-7%, but in TNBC, median oxygen concentrations can drop below 2%, triggering an intense hypoxic response (McKeown, 2014).

The regulatory mechanism of this response centres on Hypoxia-Induced Factor 1 Alpha (HIF1A), which is constantly synthesized and regulated post-translationally, rather than through gene transcription or protein translation **[Figure 1-11]**. The control of HIF1A involves its hydroxylation by prolyl hydroxylases (PHDs) and Factor-Inhibiting HIF (FIH), processes that are highly oxygen-dependent. In the absence of adequate oxygen, these enzymes fail to function effectively, leading to the stabilization and accumulation of HIF1A. Once stabilized, HIF1A forms a complex with HIF1B, and this complex then moves to the nucleus. Within the nucleus, it binds to hypoxic response elements (HREs) on the DNA, activating genes that enable the cell to adapt to and survive under hypoxic conditions (Masoud & Li, 2015).

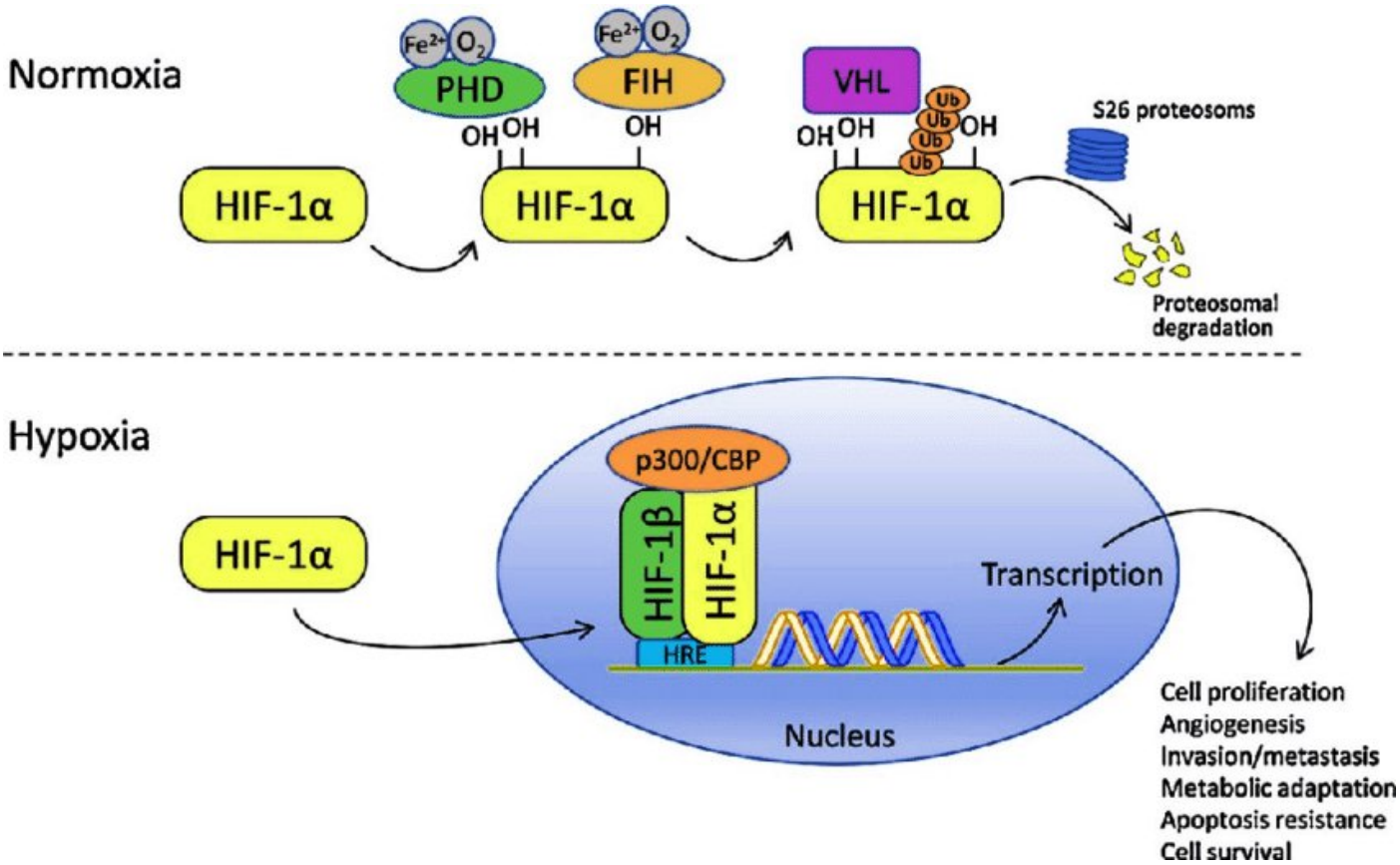


Figure 1-11: HIF1A Signaling Cascade *— HIF1A stabilization and activity are tightly regulated in cells under normoxic conditions by a cyclical process of production and degradation. HIF1A undergoes hydroxylation by HIF-prolyl hydroxylases (PHDs) and factor-inhibiting HIF (FIH), targeting it for proteasomal degradation. PHDs hydroxylate prolyl residues, while FIH hydroxylates asparaginyl residues, both enhancing the binding affinity of HIF1A for the Von Hippel-Lindau (VHL) protein, an E3 ubiquitin ligase that polyubiquitinates HIF1A, marking it for degradation. Under normoxic conditions, this process prevents HIF1A accumulation. In hypoxia, decreased oxygen levels inhibit PHD and FIH activity, leading to HIF1A stabilization. Elevated HIF1A then dimerizes with HIF1B, forming a complex that translocate to the nucleus. This complex binds to hypoxia response elements (HREs) in the DNA, activating transcription of genes that drive key tumorigenic processes such as proliferation, angiogenesis, metabolic reprogramming, resistance to apoptosis, and promotion of invasion and metastasis. Figure adapted from (Ali, 2018).*

Several studies have demonstrated the critical role of HIF1A in the hypoxic response of triple-negative breast cancer (TNBC). Research utilizing a lentiviral vector with HIF1A siRNA has effectively suppressed the hypoxic response in TNBC, leading to decreased metastatic capability, reversal of resistance to chemotherapy, and reduced cell proliferation (S. Li et al., 2015). Another study introduced HIF1A siRNA through a multifunctional cationic porphyrin-grafted lipid (CPGL) microbubble system, which also doubles as a photodynamic therapy (PDT) agent. The major finding from this research indicates that combining HIF1A inhibition with PDT significantly curtails TNBC growth under hypoxic conditions and boosts the effectiveness of PDT (S. Sun et al., 2018).

The hypoxic response significantly influences several critical aspects of tumor progression and malignancy **[Figure 1-12]**. It markedly enhances vascular endothelial growth factor (VEGF) signaling, which is essential for new blood vessel formation and thereby supports angiogenesis (Krock et al., 2011). Hypoxia also leads to the upregulation of transcription factors such as SNAI1, ZEB1, and TWIST1, which are crucial for epithelial-mesenchymal transition (EMT) and subsequent metastasis. Additionally, there is an increase in the expression of genes related to glycolysis, including LDHA, GLUT1, and PGK1 (Hapke & Haake, 2020). This metabolic shift, part of the Warburg effect, allows cancer cells to rely on glycolysis for energy production and biomass generation even when oxygen is sufficient, supporting rapid proliferation (Liberti & Locasale, 2016). Furthermore, hypoxia can significantly reduce the effectiveness of radiotherapy by up to 50%, as tumor cells in low oxygen conditions are less susceptible to radiation-induced DNA damage **[Figure 1-13]**.

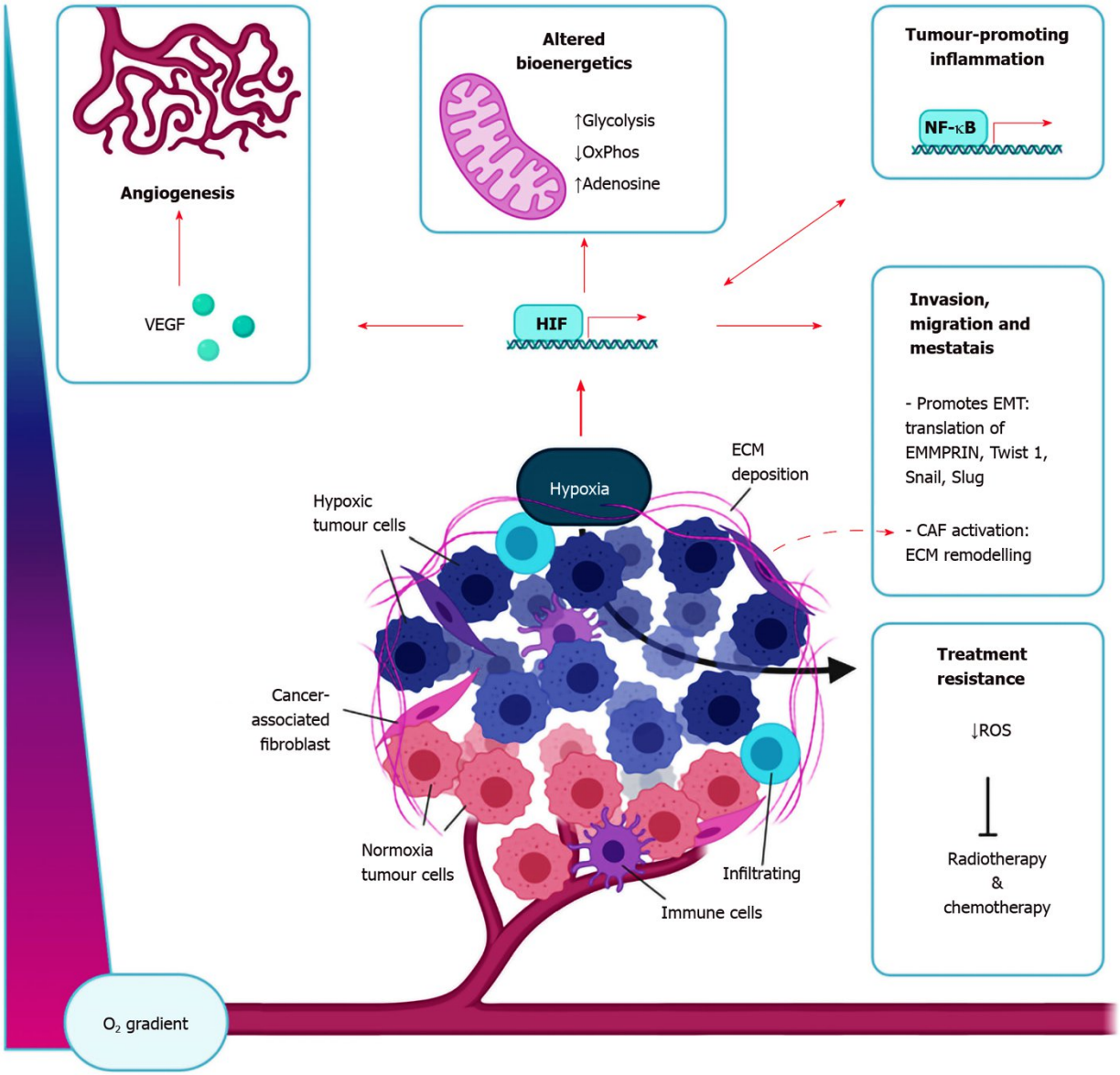


Figure 1-12: Hypoxic Tumor Microenvironment (TME) Interactions and Enhanced Carcinogenesis — *Hypoxia within the tumor microenvironment significantly contributes to advanced carcinogenesis by driving complex biological processes. Central to this is the activation of hypoxia-inducible factors, which lead to increased expression and release of vascular endothelial growth factor (VEGF), promoting angiogenesis. Hypoxia profoundly influences cellular metabolism, notably by upregulating glycolytic enzymes such as LDHA, PGK1, and GLUT1, while simultaneously reducing oxidative phosphorylation. This metabolic shift supports tumor survival under low oxygen conditions. Additionally, hypoxia impairs the effectiveness of radiotherapy and promotes resistance to chemotherapy, further complicating treatment outcomes. Critically, hypoxia fosters tumor invasion and metastasis by inducing epithelial-to-mesenchymal transition (EMT), evidenced by upregulation of key transcription factors such as TWIST1, SNAI1, and ZEB1. These factors orchestrate cellular changes that enhance tumor cell mobility and invasion, pivotal in the progression of carcinogenesis. Figure adapted from (King et al., 2021).*

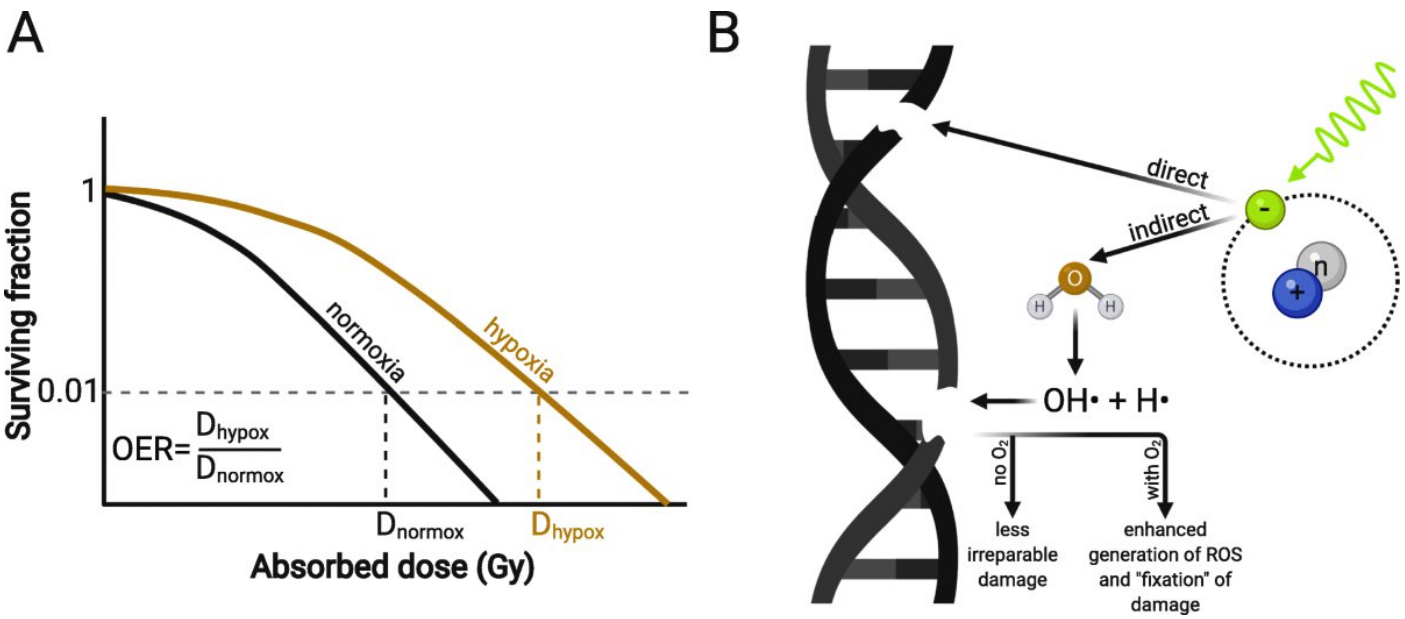


Figure 1-13: Radiotherapeutic Resistance Under Hypoxia - *The efficacy of radiotherapy is markedly reduced under hypoxic conditions, requiring increased doses of radiation to achieve equivalent cancer cell mortality observed under normoxic conditions. The graph illustrates the relationship between the absorbed dose of radiation (Gy) and the surviving fraction of cells, highlighting the dose dependency on oxygen availability. Under hypoxia, the surviving fraction curve (red) shifts to the right compared to normoxia (black), indicating a higher radiation dose (D_hypox) is needed to reach similar levels of cell kill. The oxygen enhancement ratio (OER) represents the increased effectiveness of radiation under oxygenated versus hypoxic conditions, with a lower OER indicating diminished radiation-induced DNA damage in hypoxia. This resistance arises from a reduction in reactive oxygen species (ROS) production, critical for radiation-induced DNA damage, as well as hypoxia-induced cell cycle arrest that lowers cellular susceptibility to radiation and the upregulation of genes promoting tumor adaptability and therapeutic resistance. Figure adapted from (Telarovic et al., 2021).*

Hypoxia within the tumor microenvironment (TME) is not uniform and manifests in various forms. Acute hypoxia refers to short-term oxygen deprivation lasting up to 72 hours. Chronic hypoxia denotes prolonged periods of reduced oxygen availability that can extend from weeks to years within tumors. Cyclical hypoxia involves rapid and repeated fluctuations between hypoxic and normoxic states due to dynamic changes in the tumour's blood supply, reflecting the complex nature of oxygen availability in the TME (Saxena & Jolly, 2019).

While acute hypoxia has been well-documented through bulk transcriptomic studies, cyclical and chronic hypoxia remain less understood due to challenges in replicating these conditions in vitro. This thesis focuses on these underexplored areas, aiming to provide deeper insights into how TNBC adapts to varying levels of hypoxia, utilizing advanced single-cell analysis techniques to uncover the nuanced interplay of cellular responses within the hypoxic TME.

Another aspect not yet fully characterized in the literature is the hypoxic heterogeneity across different TNBC subtypes. Utilizing TCGA data and the Buffa hypoxia signature, we conducted a signature scoring analysis (median score) on bulk RNA-Seq data across various TNBC subtypes. The boxplot **[Figure 1-14A]** compares TNBC to other hormone-sensitive breast cancer types, distinctly showing higher hypoxia levels in TNBC compared to other subtypes.

The heatmap **[Figure 1-14B]** illustrates DEA markers for hypoxia versus normoxia across TNBC subtypes, emphasizing canonical hypoxia markers shared across all subtypes, such as VEGFA and NDRG1, alongside unique markers specific to each subtype. This represents the first evidence of hypoxic heterogeneity within TNBC subtypes, underscoring the clinical relevance of these findings, derived from TCGA data.

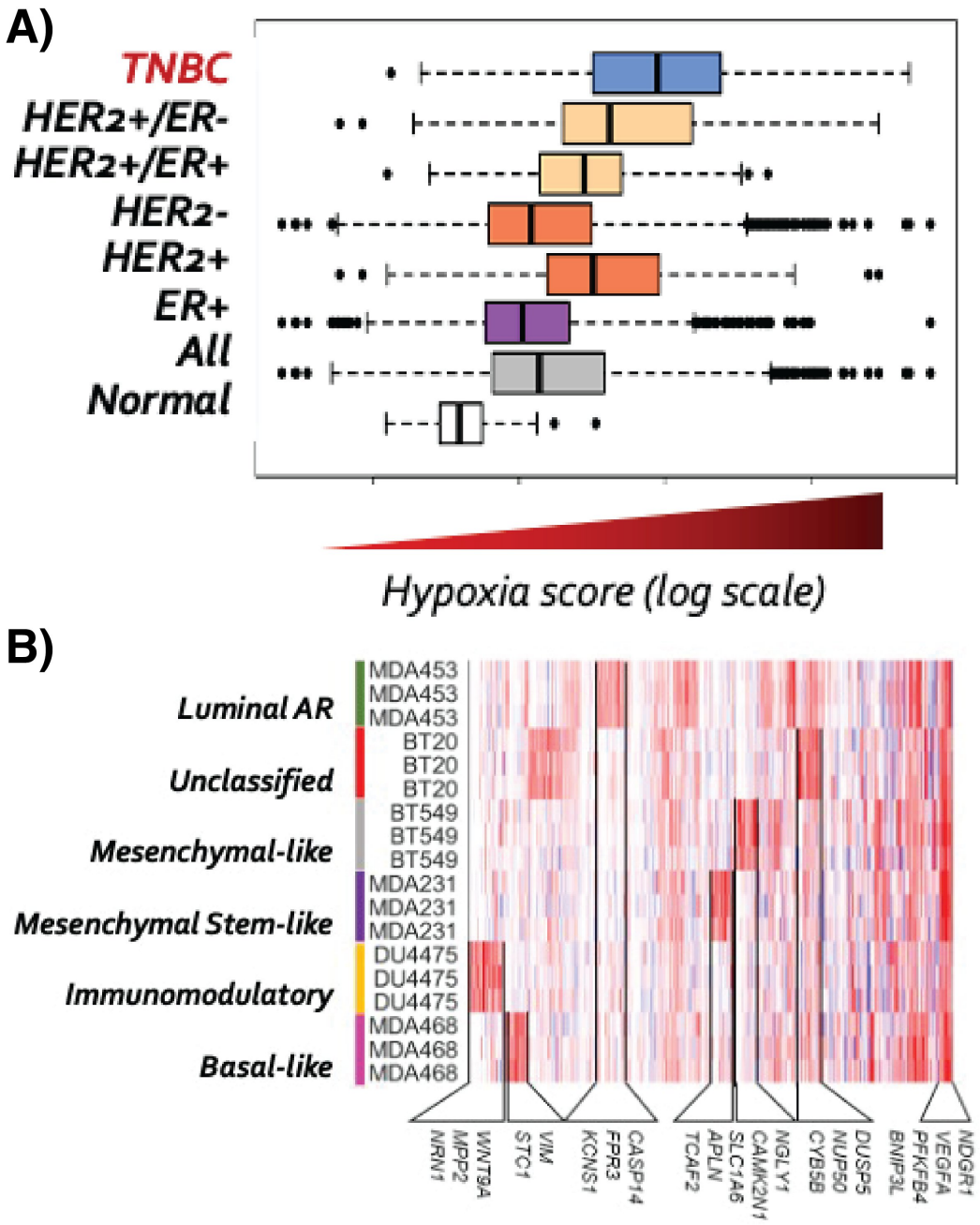


Figure 1-14: Hypoxic Response Heterogeneity Across Different TNBC Subtypes – *A) Boxplot displays the Buffa hypoxia signature score across TNBC and other breast cancer subtypes, illustrating that TNBC generally exhibits a higher hypoxia score compared to other subtypes. B) Heatmap of DEA markers delineating canonical hypoxia markers shared among TNBC subtypes alongside unique markers specific to each TNBC subtype within the TCGA dataset. This analysis presents preliminary evidence of hypoxic heterogeneity within TNBC subtypes, highlighting the need for further investigations to enhance pharmacological strategies and therapeutic precision.* ***(Unpublished data) (Communications with Professor Francesca Buffa)***

1.4. Thesis Goals & Aims

This thesis aimed to elucidate the heterogeneity of the hypoxic response across different breast cancer subtypes at the single-cell level, particularly focusing on the distinctions between acute and chronic hypoxia. The investigation extends to characterize the stromal micro-environment of breast cancer across different stages of development, aiming to elucidate the roles physical and cellular components of the Tumor Microenvironment (TME) play in driving tumor progression and development at the single-cell level. This exploration into single-cell analysis arose as a natural extension of previous work conducted with bulk transcriptomics, aiming to uncover uncharacterized heterogeneity within cell populations previously assumed homogeneous.

The data compiled in this thesis reveal that hypoxic responses are varied and heterogeneous, leading to the formation of sub-populations of cells whose functional specialization appears to be a key driver of tumor adaptation. Another significant finding is the identification of multiple cell subtypes within the TME that are potent drivers of invasion and angiogenesis, crucial components of the invasion-metastasis cascade.

The specific objectives were:

1. To characterize the heterogeneity of the hypoxic response across various breast cancer subtypes at the single-cell level.

2. To characterize the heterogeneity of the hypoxic response between acute and chronic hypoxia at the single-cell level.

3. To characterize the role hypoxia plays in altering and shaping the stromal tumor microenvironment (TME) of breast cancer at the single-cell level during its early development.

However, the onset of the COVID-19 pandemic significantly disrupted experimental workflows and data generation, necessitating revisions to the initial goals. Achievements under these constrained circumstances included:

1. A comprehensive characterization of hypoxic response heterogeneity at the single-cell level for acute hypoxia using the MCF7 cell line.

2. A detailed examination of the differences between acute and chronic hypoxia using the MDA231 cell line, providing insights into the temporal dynamics of hypoxia.

3. An in-depth characterization of the breast cancer stromal TME over time using the E0771 cell line in mouse lung metastases, elucidating its evolution during early tumor development.

These accomplishments reflect the scope of progress made under challenging conditions and lay the groundwork for further in-depth studies in this field.

2. Chapter 2: Materials & Methods

2.1. Experimental Methodology

The experimental procedures detailed in the sections concerning the MDA231 and Mlung datasets were conducted by collaborators during their doctoral research. May Sin Ke is credited with generating the MDA231 dataset, while Salwa Lin is responsible for the Mlung dataset. The descriptions provided in the following sections represent paraphrased accounts of their methodologies and are not derived from my personal experimentation or original development.

2.1.1. MCF7 Dataset

Cell Culture and Growth Conditions: The cell lines were grown in an environment regulated at 37ºC and 5% CO2 with a humidity control. The medium used was DMEM, enriched with a 10% addition of FBS. For experiments under low oxygen levels, the cultures were placed in a hypoxic chamber, specifically an INVIVO2 400 (Pro-Lab Diagnostics), with the oxygen concentration finely controlled at 0.1% and maintained alongside 5% CO2.

Protein Detection via Immunoblotting: Proteomic analysis involved resolving cellular proteins by electrophoresis on a 10% polyacrylamide gel followed by their transfer onto a PVDF membrane. Various primary antibodies were employed typically at a dilution of 1:1000, including those targeting HIF1α sourced from BD Transduction Laboratories, USA; CAIX provided by J. Pastrorek from the Institute of Virology, Slovak Republic; and others like LDHA, PDK1, AK4, ALDH from Santa-Cruz and β actin from Sigma. Densitometric analysis of the blots was performed using the ImageJ software.

SmartSeq2 sequencing procedures, not detailed here, were performed in accordance with the methodologies established in (Picelli et al., 2013a). These summarized descriptions reframe previously reported protocols by (Ledaki et al., 2015), retaining essential methodological steps while ensuring the text remains original and non-replicative.

2.1.2. MDA231 Time Series Dataset

Procurement of Cell Lines: The MDA-MB-231 breast cancer cell lines were obtained from ATCC in Manassas, Virginia. They were cryopreserved immediately upon arrival and verified for absence of mycoplasma contamination.

Culture and Maintenance Conditions: The cells were grown in DMEM supplemented with 10% fetal bovine serum and 1% penicillin antibiotic mixture. The cultures were housed in sterilized, non-toxic plasticware approved for tissue culture from Corning. For normoxic conditions, the cultures were kept in an incubator providing 5% CO2 at 37°C. Hypoxic conditions were simulated using a Ruskin InvivO2 400 workstation set to maintain an oxygen concentration of 1%. Cells were given 12 hours to adapt post-plating before exposure to hypoxia. Routine cell passaging occurred every 5 to 7 days at about 80% confluency, involving the removal of media, washing with PBS, and detaching cells using 0.25% trypsin-EDTA for three minutes at 37°C. Cells were then diluted in fresh media at a 1:8 ratio for replating. Media changes were performed bi-weekly.

Cryopreservation Procedures: For long-term storage, cells were trypsinized, suspended in complete medium, and centrifuged at 1,000 xg for 5 minutes. The pellet was resuspended in a cryoprotective solution containing 90% DMEM and 10% DMSO, distributed into cryovials at 1ml per vial, and initially frozen at -80°C using a NALGENE freezing container. After 24 hours, vials were transferred to liquid nitrogen for permanent storage.

Single-cell RNA Sequencing Techniques: The BD Rhapsody system was employed for comprehensive profiling of the TNBC MDA-MB-231 cell line, due to its ability to handle large cell

volumes and its enhanced multiplexing capabilities. The system arranges individual cells alongside magnetic beads with unique molecular identifiers within a microfluidic apparatus. This setup supports efficient gene expression profiling by capturing cDNA with cell-specific identifiers, which aids in precise gene expression analysis. The multiplexing technique used allows for concurrent analysis of eight different samples, tagged distinctly using the system's antibody library.

Hypoxia Experimentation Techniques: Cells were plated strategically, starting with 0.75 million cells, and scaled up to 2.0 million cells on the 7th and 10th days, with cultures maintained and sub-cultured every third day over 14 days. For short-term hypoxia studies, the cells were not sub-cultured after the initial seeding. Media changes were conducted every other day with media pre-conditioned to the specific oxygen levels of the experiments. Normoxic controls followed a similar protocol without hypoxic preconditioning.

mRNA Library Preparation: Following the protocols of the BD Rhapsody™ system, mRNA libraries were created by first releasing Sample Tags from the Cell Capture Beads. This was followed by amplifying the total mRNA directly on the beads using a random priming strategy, concluded by index PCR amplification. Cleanup of the resulting cDNA was carried out using SPRIselect beads as per the manufacturer's instructions.

Next-Generation Sequencing: After library preparation, sequencing of both the whole transcriptome and Sample Tag cDNA libraries was performed on an Illumina NovaSeq 6000 S4 sequencer, fully utilizing the sequencer's capacity to ensure extensive coverage of the samples.

2.1.3. Mlung Time Series Dataset

E0771 Parental Cancer Cell Origin: The foundational E0771 cell line, characterized by its triple-negative medullary breast adenocarcinoma properties, was isolated from a naturally occurring tumor in C57BL/6 mice. Despite minimal ERα presence confined to the cytoplasm, this line is categorized as triple-negative. We are grateful to Professor Ruth Muschel at the University of Oxford for supplying this cell line.

Generation of tdTomato-Expressing E0771 Cells: The primary E0771 line was genetically modified using a LeGO-iT2 lentivirus, which encodes the tdTomato reporter. During transduction, cells were seeded in twelve-well plates at a density of 100,000 cells/well and exposed to the lentivirus at an MOI of 10 in a medium blended with RPMI 1640, polybrene, and FBS. After spinfection and a 24-hour incubation, the medium was refreshed to remove any residual virus. Fluorescent microscopy confirmed successful transduction, and fluorescence-activated cell sorting isolated cells with high tdTomato expression for further expansion and cryopreservation.

Engineering mCherry-Secreting hCD2 E0771 Cells: Our colleagues, led by Professor Illaria Malanchi at The Francis Crick Institute, provided the mCherry-hCD2 labelled E0771 line. This variant was crafted by introducing a vector encoding for a secreted, cell-uptake optimized mCherry protein, enabling the tracing of local cell interactions. The cultured lines—including the unaltered E0771, tdTomato, and mCherry-hCD2 versions—were all maintained in DMEM fortified with FBS.

Inducible tdTomato Reporter Mice: Th conditional Wt1-CreERT2; tdTomato mouse model was sourced from Professor Nicola Smart's team at Oxford. These mice carry alleles for both inducible Cre recombinase and the tdTomato reporter, allowing for controlled gene editing and cellular tracking. Cross-breeding strategies were employed to produce the desired genotypes for experimental and control cohorts, facilitating subsequent tumor modeling studies.

Single-cell Preparation and Sequencing: Single-cell samples, isolated via cell sorting, were processed for sequencing at the MRC WIMM Single Cell Genomics Facility, directed by Dr. Neil Ashley. Utilizing the 10X Genomics Chromium platform and reagents, we generated a library

capturing single-cell transcriptomes. Quality assessments were made post-amplification, and sequencing was conducted through Novogene's services. The data, including sequences and comprehensive quality reports, were received. For accurate alignment and expression analysis, a customized reference was constructed by Dr Thomas Caroll, and the Cell Ranger pipeline was employed without normalization to compile the feature-barcode matrices.

2.2. Computational Methodology

2.2.1. Pre-Processing Pipeline

2.2.1.1. MCF7 Dataset

The pre-processing pipeline for the sequencing data involves several key steps, each employing specialized software tools to ensure data quality and reliability. Initially, the quality of raw sequencing reads was assessed using FastQC (Andrews et al., 2012), which provides a comprehensive overview of data quality across various metrics. Following quality assessment, the reads were trimmed to remove adapters and low-quality sections using Cutadapt, thereby improving the quality of the subsequent alignment.

For aligning the processed reads to the reference genome, we employed STAR (Dobin et al., 2013a), a widely-used tool that efficiently handles large amounts of RNA sequencing data with high accuracy. After alignment, FeatureCounts was utilized to quantify gene-level features from the aligned reads, enabling the precise measurement of gene expression.

To consolidate and review quality control metrics from each of these stages, MultiQC (Ewels et al., 2016a) was used. This tool aggregates results from various processing steps into a single report, facilitating an easy and comprehensive evaluation of the entire dataset. To consolidate and review quality control metrics from each of these stages, MultiQC (Ewels et al., 2016a) was used. This tool aggregates results from various processing steps into a single report, facilitating an easy and comprehensive evaluation of the entire dataset.

The entire pre-processing workflow was automated and managed using Snakemake (Köster & Rahmann, 2012), a workflow management system that ensures reproducibility and scalability. Snakemake orchestrates the execution of each step in the pipeline, handling dependencies and allowing for efficient use of computational resources.

2.2.1.2. MDA231 Dataset

The MDA231 dataset was analysed using the BD Rhapsody analysis pipeline (W. Li et al., 2024) available on the Seven Bridges Platform. This platform hosts the pre-processing of the data, which includes several critical steps to ensure the quality and usability of the sequencing data for downstream analysis.

2.2.1.3. Mlung Data

Pre-processing of the single-cell RNA-Seq data was performed using the Cell Ranger pipeline, developed by 10X Genomics, as detailed by (Zheng et al., 2017). This command-line software suite is operated on the CCB cluster and is designed to handle a range of crucial tasks. The process begins with the **cellranger mkfastq** command, which demultiplexes raw base call (BCL) files into FASTQ files. Following this, the **cellranger count** command takes these FASTQ files and executes several functions including alignment, filtering, barcode counting, and UMI counting, ensuring comprehensive data preparation for downstream analysis.

- Bottom of Form

2.2.2. scRNA-Seq Analysis Pipeline

Following completion of pre-processing and the generation of the final count matrix for the data, the full analysis pipeline is executed in Seurat (Hao et al., 2021a) in the R programming language. The steps of the analysis pipeline are the following:

2.2.2.1. Data Import

Data is imported into a seurat object using ***CreateSeuratObject()***.

2.2.2.2. Quality Control

The initial step in the analysis pipeline involves comprehensive quality control (QC) to filter out low-quality cells and genes with low expression levels, thereby ensuring the integrity of the data. QC is divided into sub-sections, with Cell QC targeting the removal of low-quality cells that may skew data distribution. These include dying cells exhibiting apoptotic or stressed phenotypes and doublets. Key metrics used in Cell QC include:

- **Library Size**, which represents the total counts per cell. High counts suggest doublets, while low counts indicate dying cells.
- **Number of Detected Features**, which measures the total unique genes detected per cell. A high number of features may indicate doublets, and a low number reflects dying cells.
- **Percentage of Mitochondrial Genes**, where the detection of mitochondrial genes suggests mitochondrial permeation due to cell stress or damage, often seen in cells undergoing apoptosis or necrosis.

Removal of these cells is crucial to prevent distortion of the analysis results. The thresholds for QC metrics are determined using data visualization tools; histograms are generated with **ggplot2::geom_histogram()** and violin plots with **Seurat::VlnPlot()** to help visualize the distribution of QC metrics. An alternative, more mathematical method of setting thresholds involves detecting outliers, defined as data points lying more than three median absolute deviations (MADs) from the median on either side of the distribution, using the **scater::IsOutlier()** function.

Gene QC focuses on filtering out features that show very low expression levels across the dataset, which could affect the variability analysis and dimensionality reduction stages of the pipeline. This thorough QC process ensures that only high-quality, representative data is used for subsequent analyses, maintaining the accuracy and reliability of the findings.

2.2.2.3. Normalisation & Variable Feature Detection

Following quality control, the next crucial step in the analysis pipeline involves normalization and variable feature detection. This stage adjusts the count matrix to account for differences in library size among the cells. It also identifies genes exhibiting high levels of variation relative to their mean expression, as these hypervariable features are likely to represent the primary sources of both biological and technical variation within the data. Proper identification of these features is essential for effective dimensionality reduction and clustering in the single-cell analysis pipeline. The normalization, variable feature selection, and data scaling processes are all executed using the **Seurat::SCTransform()** function (Saelens et al., 2019), which comprehensively manages these tasks to ensure the data is prepared for downstream analyses.

2.2.2.4. Cell Cycle Scoring

Cell cycle scoring is a critical step in the analysis pipeline due to the significant variation that cell cycle processes contribute to transcriptomic data. The goal of this step is to delineate the impact of cell cycle variations on the biological diversity observed within the dataset. This process is carried out using the ***Seurat::CellCycleScoring()*** function, which utilizes specific gene signatures for the G2M and S phases of the cell cycle, as outlined by (Tirosh et al., 2016). The function calculates an aggregate score for each cell based on the expression levels of these signature genes, effectively quantifying cell cycle effects. This scoring is implemented in the backend through the ***Seurat::AddModuleScore()*** function, enabling a nuanced analysis of cell cycle influences on gene expression profiles.

2.2.2.5. Dimensionality Reduction & Visualisation

Dimensionality reduction is crucial for simplifying the complexity of transcriptomic data, facilitating effective visualization and exploration of the underlying sources of variation. The primary techniques employed in this analysis are PCA and UMAP, which are executed using the ***Seurat::runPCA()*** and ***Seurat::runUMAP()*** functions, respectively, as described by (McInnes et al., 2020).

To effectively use UMAP for projection, it is necessary to determine the appropriate number of principal components (PCs) derived from PCA. This selection is guided by elbow plots, generated via the ***Seurat::ElbowPlot()*** function, which display the standard deviation of each PC. The 'elbow' point in these plots indicates the number of PCs that capture the most significant variability in the data, balancing the inclusion of biological signals while minimizing technical noise. This approach assumes that the data's technical variation is minimal, allowing the major PCs to predominantly reflect biologically relevant variation.

2.2.2.6. Clustering Analysis

Clustering analysis is fundamental in identifying distinct sub-populations of cells within the dataset that may represent unique biological phenotypes or different cell types. The Seurat pipeline incorporates community detection algorithms for this purpose, particularly the Leiden and Louvain methods as highlighted by (Traag et al., 2019). These algorithms are executed using the ***Seurat::FindNeighbors()*** and ***Seurat::FindClusters()*** functions. For ***Seurat::FindNeighbors()***, the number of principal components (PCs) to be used is determined using elbow plots. For ***Seurat::FindClusters()***, a critical step is setting the resolution parameter, which dictates the granularity of the clustering; too high a resolution can lead to over-clustering, while too low may overlook significant cell populations.

Additionally, clustering trees, a tool developed to optimize the resolution parameter, play a crucial role in this process. Implemented via the clustree package, as described in (Zappia & Oshlack, 2018), clustering trees help visualize how changes in resolution affect the number and stability of the resulting clusters, ensuring that the analysis is finely tuned to capture the most meaningful biological variations within the data.

2.2.2.7. Differential Expression Analysis (DEA) and Pathway Enrichment Analysis (PEA)

Differential Expression Analysis (DEA) is essential for identifying genes that are differentially expressed across different experimental groups, clusters, cell types, or biological phenotypes. Seurat supports various methods for DEA, including the Wilcoxon signed-rank test, which is the default non-parametric test used for identifying genes with differential expression patterns between two groups. Another option is DESeq2, which employs a model based on the negative binomial distribution, typically used for bulk RNA-Seq data analysis, as detailed by (Love et al., 2014). These tests are implemented in Seurat through the ***Seurat::FindMarkers()*** and ***Seurat::FindAllMarkers()*** functions.

Pathway Enrichment Analysis (PEA) builds on DEA by identifying biological pathways that are significantly enriched with differentially expressed genes. PEA is instrumental in pinpointing key pathways affected under various conditions or within specific clusters. The primary methods used in PEA are Over-Representation Analysis (ORA) and Gene Set Enrichment Analysis (GSEA) as described by (Subramanian et al., 2005). ORA assesses whether certain pathways are over-represented among differentially expressed genes compared to their general frequency in the genome and is particularly effective when analyzing gene lists from studies with clear differential expression. This analysis is facilitated by tools like clusterProfiler. GSEA, on the other hand, does not require a cutoff for gene inclusion; it evaluates the presence of genes from predefined sets across the entire ranked list of genes from the dataset, identifying even subtle changes in gene expression. GSEA is conducted using the fgsea package, noted by (Korotkevich et al., 2021).

Both ORA and GSEA necessitate reference gene signatures that represent biological pathways. These signatures are often sourced from comprehensive databases like the Molecular Signature Database (MSigDB), curated by (Liberzon et al., 2015), which houses an extensive collection of gene sets for pathway analysis. In the context of R programming, these gene signatures are accessible through the msigdbr package, facilitating their integration into PEA workflows.

2.2.2.8. Pseudotime Analysis

Pseudotime methods in the context of scRNA-Seq are indeed powerful for deciphering the complex nature of cellular differentiation and development. Many tools are available for trajectory inference, and they differ in their assumptions and functionality. The main tool used for pseudotime analysis in this project was **Monocle3** (Cao et al., 2019). In a recent benchmarking review (Saelens et al., 2019), it was one of the top-performing tools with adequate integration with the Seurat system and R language.

2.2.2.9. Shiny Apps

Shiny Apps, integrated within the R programming language, are an interactive visualization tool that is invaluable for data presentation and exploration. They are particularly effective for engaging with complex datasets interactively, facilitating a dynamic analysis experience.

For scRNA-Seq datasets, Shiny apps have been developed using the **ShinyCell** package (Ouyang et al., 2021) to create tailored visualizations. Additionally, custom Shiny apps for Chapter 6 were designed from scratch using the Shiny package to meet specific analytical needs.

Key features of Shiny apps enhance data exploration by allowing collaborators to interact directly with the data. These functionalities include:

- **Gene Expression Inspection:** Users can check the expression of individual genes through interactive UMAP plots that visually represent data clustering based on gene expression levels.
- **Data Grouping Exploration:** Shiny apps allow for the examination of how data groups according to various metadata columns, also visualized through UMAP plots.
- **Distribution Analysis:** Users can analyze the distribution of gene expression values or metadata characteristics using interactive box plots and violin plots.

These interactive features make Shiny apps a robust tool for detailed data analysis and presentation, thereby allowing for effective communication and exploration of computational findings by bioinformaticians and experimental biologists alike.

2.2.2.10. Cell-Cell Communication Analysis Overview

Cell-cell communication analysis (S. Jin et al., 2021) focuses on elucidating the mechanisms through which different clusters or sub-populations of cells interact via specific signaling pathways. This analysis is pivotal for understanding cellular communications and adaptations within their biological environments, providing insights into cellular functions and interactions in both healthy and diseased states.

The analysis is conducted using the **CellChat** package in R (S. Jin et al., 2021), leveraging a normalized expression matrix that details gene expression levels across various cells. CellChat also utilizes the CellChat database (CellChatDB), which houses comprehensive data on known ligand-receptor pairs and biological signaling pathways.

The methodology includes several key steps:

1. **Identification of Over-expressed Genes:** This step uses the ***CellChat::identifyOverExpressedGenes()*** function to pinpoint genes that are over-

expressed within specific cell sub-populations, preparing the groundwork for further analysis of ligand-receptor (LR) pairs.

2. **Identification of Over-expressed Interactions:** Following gene identification, the analysis progresses with the ***CellChat::identifyOverExpressedInteractions()*** function, which detects potential over-expressed interactions.

3. **Compute Communication Probability:** The ***CellChat::computeCommunProb()*** function calculates communication probabilities, employing a trimean approach for averaging gene expression per cell group, thus focusing on fewer but stronger interactions.

4. **Network Aggregation:** This step aggregates individual communication probabilities per LR pair into broader communication probabilities per signaling pathway using the ***CellChat::aggregateNet()*** function. This reflects interactions across multiple LR pairs and provides a comprehensive view of the communication networks.

Visualization techniques are integral to presenting the results, utilizing unique visual tools like hierarchy plots, chord plots, and circle plots. Each tool offers a different perspective on the communication networks identified, aiding in the interpretation and dissemination of the complex data.

Further details about the statistical methods used by CellChat to estimate communication probabilities and the meticulous curation of the CellChatDB are available in the package's publications and online vignettes.

3. Chapter 3: Single-Cell RNA-Seq Exploration of Hypoxic Heterogeneity in ER+ Breast Cancer

3.1. Introduction

Hypoxia, defined as a state of low oxygen levels relative to the adjacent normal tissues, is a hallmark of many solid tumors and has been implicated in inducing resistance to various therapeutic interventions (Rofstad, 2000). In the tumor microenvironment, hypoxia arises as rapidly proliferating cancer cells exceed the capacity of their blood supply, resulting in a mismatch between the high metabolic demands of these cells and the available oxygen supply. This discrepancy leads to a reduction in tissue oxygenation.

The presence of hypoxia triggers an increase in the levels of Hypoxia-Inducible Factor 1-alpha (HIF1A), initiating a robust transcriptional response that enables cells to survive under oxygen-deprived conditions. Under physiological conditions, the HIF1A-dependent hypoxic response serves a protective function, helping cells endure low oxygen levels and work towards restoring normal oxygenation. However, in the context of cancer, this adaptive mechanism is co-opted to promote tumor aggressiveness, enhance angiogenesis, and facilitate the metastatic spread, integral components of the invasion-metastasis cascade (Eslami-S et al., 2020; Kakkad et al., 2019).

Approximately 8% of the genome is regulated by the HIF1A-mediated hypoxic response, encompassing genes involved in various metabolic processes, cell proliferation, and survival pathways. Additionally, this response regulates a multitude of microRNAs (miRNAs) and long non-coding RNAs (lncRNAs), adding a layer of post-transcriptional control that shapes cellular behavior under hypoxia (Z. Ma et al., 2022). Despite the well-characterized heterogeneity of the hypoxic response across different cell types within the tumor microenvironment (TME), variations within specific cancer cell populations remain underexplored. Conventional transcriptomic techniques, such as microarrays and bulk RNA sequencing, typically measure an averaged response across cell populations, often missing crucial genes active in subpopulations. This limitation prevents these methods from detecting subtle but potentially important heterogeneities that could influence the development of distinct cancer cell subpopulations with either synergistic or antagonistic effects on tumor survival (X. Yu et al., 2021).

Clinical studies have consistently shown that tumor hypoxia, as determined by the analysis of hypoxia-regulated genes, serves as an independent marker of poor prognosis (Dhani et al., 2015; Forker et al., 2018; Gee et al., 2010; Kronblad et al., 2006; Van den Eynden et al., 2007). Numerous biomarkers and gene profiles have been identified to clinically characterize hypoxia (Abu-Jamous et al., 2017; Buffa et al., 2010a; Favaro et al., 2011; Neumeister et al., 2012), but the implications of cellular heterogeneity in hypoxic responses have been less frequently addressed.

A particularly responsive gene in the transcriptional hypoxic profile is carbonic anhydrase 9 (CA9) (Abu-Jamous et al., 2017; Buffa et al., 2010a; Lal et al., 2001; Loncaster et al., 2001). CA9 encodes a transmembrane glycoprotein catalyzing the reversible hydration of carbon dioxide to bicarbonate (Opavský et al., 1996). This bicarbonate is then transported back into cells, assisting in the maintenance of intracellular pH (McIntyre et al., 2016; Swietach et al., 2008), thereby allowing CA9 to serve as a crucial regulator of pH homeostasis (Swietach et al., 2010). While CA9 is scarcely detectable in most normal tissues, its expression is markedly increased in hypoxic regions of tumours. Recent meta-analyses encompassing over 24,000 patients from 129 studies have demonstrated that elevated CA9 levels significantly correlate with enhanced metastasis-free, disease-free, progression-free, and overall survival rates, across various tumor types (with the exception of renal cell carcinoma) (Chia et al., 2001; Span et al., 2003; van Kuijk et al., 2016). In basal-like breast tumours, however, high CA9 expression has been associated with increased resistance to chemotherapy and a generally poorer prognosis.

Our recent investigations have identified two distinct populations within hypoxia-induced cells differentiated by their CA9 expression in 2D cultures (Ledaki et al., 2015). This study demonstrated

that CA9-positive hypoxic cells could give rise to both CA9-positive and CA9-negative populations, whereas CA9-negative cells were only capable of generating CA9-negative offspring. This finding prompts further examination of hypoxia-induced heterogeneity at the single-cell level to ascertain whether observed variations are due to pre-existing programmed differences or if they result from hypoxia-induced differentiation and the formation of specialized subpopulations. To this end, we have conducted detailed studies on the response to hypoxia in an estrogen receptor-positive breast cancer cell line, MCF7, extensively used in hypoxia research (Comşa et al., 2015). This approach allows us to dissect the cellular dynamics under hypoxic conditions, providing insights into the adaptative strategies that cells employ and their impact on disease progression and therapeutic resistance.

Through these studies, we aim to illuminate the complexities of hypoxia-driven cellular behaviors, particularly focusing on how differential expression of CA9 can influence tumor phenotype and response to treatments. This could potentially lead to the identification of novel therapeutic targets that could disrupt the hypoxic adaptations of tumor cells, offering new avenues for treatment in hypoxia-impacted cancers. Our findings may also enhance our understanding of tumor microenvironment dynamics, contributing to more precise and effective cancer management strategies.

3.2. Methods

3.2.1. Experimental

3.2.1.1. Hypoxic Timepoint Justification

MCF7 cells were cultured following the protocol previously described (Ledaki et al., 2015) and summarized in Chapter 2 of this thesis. Hypoxia was established at a defined oxygen concentration of 0.1%. The duration of hypoxic exposure was determined based on the expression levels of CA9, which were quantitatively assessed through Western blotting **[Figure 3-1A]** and FACS analysis **[Figure 3-1B]**. These assessments were performed after 24 and 72 hours of hypoxia, with the maximal induction of CA9 observed at 72 hours. Consequently, MCF7 cells were subjected to hypoxia for this duration to ensure significant CA9 expression, aligning with the aims of our study.

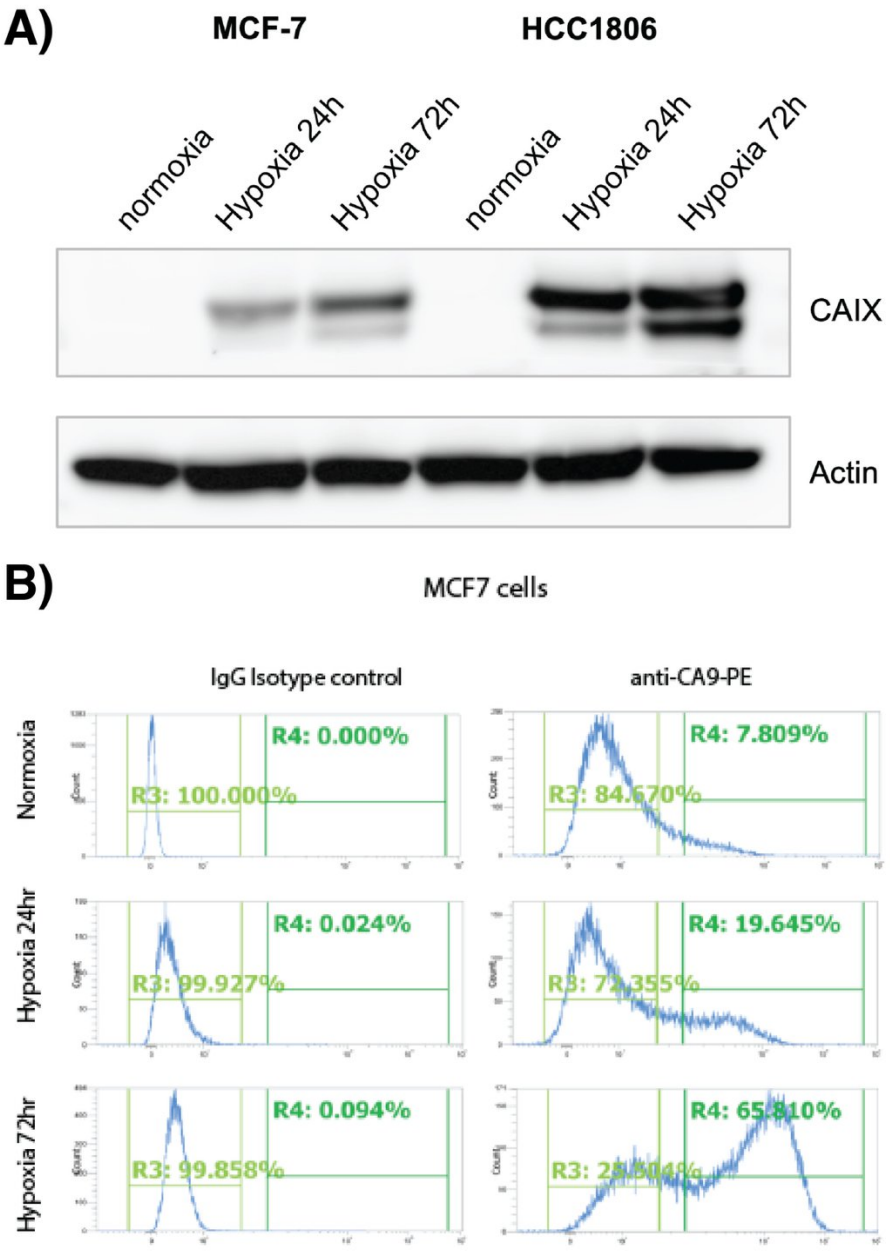


***Figure 3-1**: **MCF7 Immunoblotting and FACS Sorting during Normoxia, 24H Hypoxia, and 72H Hypoxia -** A) Immunoblotting results demonstrating maximal CA9 induction after 72 hours of hypoxia. B) FACS sorting analysis corroborates the maximal induction of CA9 under 72 hours of hypoxia. Notably, at 72 hours of hypoxia, there is a clear emergence of CA9 expression heterogeneity, with one population expressing low levels and another exhibiting high levels of CA9, an indication of the cellular heterogeneity that will be characterized soon.*

3.2.1.2. Single Cell Sequencing

SmartSeq single cell sequencing was used as described previously (Picelli et al., 2013b).

3.2.2. Computational

All have been extensively characterized in Chapter 2 and are also included here briefly for completion.

3.2.2.1. Pre-Processing Pipeline

Quality control of FASTQ files was performed using FastQC (Andrews et al., 2012) to assess the quality of the raw sequencing data. Subsequent read trimming was carried out with Cutadapt (Martin, 2011) to remove adapter sequences and low-quality bases, ensuring the accuracy of the following analyses. The trimmed reads were then aligned to the reference genome using STAR (Dobin et al.,

2013b), a rapid alignment tool known for its efficiency in handling large datasets. Feature counting was conducted with FeatureCounts (Liao et al., 2014) to quantify gene expression levels from the aligned reads.

To streamline the management of these data processing steps and ensure reproducibility, the entire preprocessing workflow was integrated and automated using Snakemake (Mölder et al., 2021). This workflow management system efficiently coordinated the execution of each step and facilitated the aggregation of all quality control metrics through MultiQC (Ewels et al., 2016b), providing a comprehensive overview of the pipeline's performance and output quality at each stage.

3.2.2.2. Primary Analysis Pipeline

The analysis pipeline was developed and executed using the R programming language, with the Seurat package (Hao et al., 2021b) employed for single-cell RNA sequencing (scRNA-Seq) data analysis. The primary aim was to enable effective visualization and identification of cellular subpopulations within the dataset.

Dimensionality reduction was a key component of the analysis, implemented in two sequential steps to enhance data interpretability. The initial step involved Principal Component Analysis (PCA), which served to reduce the complexity of the dataset by transforming it into a set of orthogonal variables that capture the most variance. Following PCA, Uniform Manifold Approximation and Projection (UMAP) (McInnes et al., 2020) was applied. UMAP further distilled the data into a two-dimensional representation, facilitating easier visualization and cluster detection.

Cluster analysis was conducted using the Louvain algorithm, a method renowned for its efficacy in detecting community structures within large networks. The selection of the number of principal components to retain and the resolution parameter for clustering was systematically determined using elbow plots and clustering trees, tools provided by the clustree package (Zappia & Oshlack, 2018). These visual aids helped identify optimal parameters by illustrating variance retention and cluster stability across different resolutions.

Concerning batch effects, which are prevalent challenges in scRNA-Seq datasets, integration methods are often utilized to mitigate these discrepancies across different samples. However, in this analysis, we decided against employing such integration techniques. Our rationale was based on observations that integration methods can sometimes obliterate biological variability, leading to a homogenization of the data that masks genuine biological differences (Tran et al., 2020). By avoiding these methods, we aimed to preserve the intrinsic biological diversity inherent in the dataset, accepting the trade-off of potential batch influences for a more authentic representation of the cellular phenotypes present.

3.2.2.3. Pathway Enrichment Analysis

Pathway analysis was conducted using the **clusterProfiler** package (G. Yu et al., 2012), a powerful tool for statistical analysis and visualization of functional profiles for genes and gene clusters. The enriched pathways were identified by testing gene sets against databases such as Gene Ontology (GO) and Kyoto Encyclopaedia of Genes and Genomes (KEGG) through over-representation analysis (ORA).

ORA is a method that identifies biological pathways significantly enriched in a given list of genes compared to a background gene set. This technique helps ascertain whether certain biological functions or processes are over-represented in the context of the dataset being analysed. By applying ORA, we were able to extract meaningful insights from the GO and KEGG databases, providing a deeper understanding of the biological processes and pathways that are potentially involved in the cellular phenotypes observed in our study.

The integration of **clusterProfiler** with these databases enables a comprehensive exploration of the functional landscape, facilitating a robust interpretation of gene expression patterns and their biological implications. This approach not only enhances the biological relevance of the data but also aids in pinpointing key pathways that may be crucial in the underlying mechanisms of the condition under study.

3.2.2.4. Pseudotime Analysis

Pseudotime analysis was conducted using Monocle3 (Cao et al., 2019), a robust tool specifically designed for analysing single-cell RNA sequencing data to uncover developmental trajectories and dynamic cellular processes. Monocle3 is particularly advantageous for its ability to handle multiple separate trajectories and accommodate non-linear trajectory structures within complex datasets (Saelens et al., 2019).

3.3. Results

3.3.1. Acute Hypoxic Response Is The Primary Source of Variation At The Single Cell Level

In this study, 384 MCF7 cells were sequenced, including 192 in normoxia and 191 in hypoxia. The distribution of quality control (QC) metrics was analysed using histograms, which facilitated the establishment of thresholds for data filtration **[Figure 3-2]**. Specifically, both upper and lower bound thresholds were defined for library size and the number of detected features **[Figure 3-2A&B]**, while only an upper threshold was set for the percentage of mitochondrial genes **[Figure 3-2C]**. A comprehensive scatterplot illustrating all these metrics, alongside their respective thresholds, was also generated **[Figure 3-3]**. This visualization not only confirmed the appropriateness of the thresholds but also ensured the flexibility of the metrics to minimize potential data loss. For further information on these QC metrics and the process involved in setting thresholds, please refer to Chapter 2.

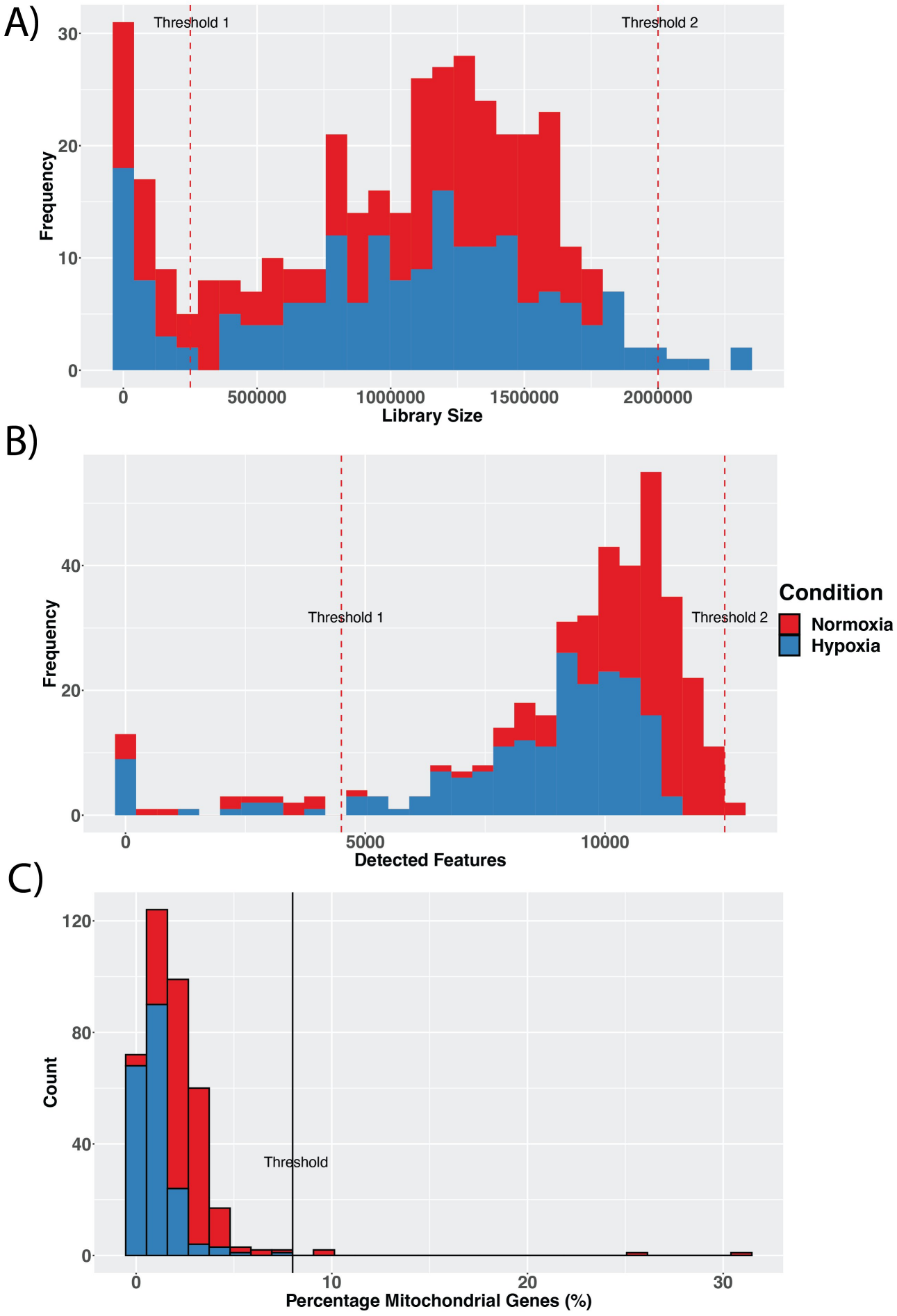


Figure 3-2: Quality Control Parameters for MCF7 Dataset, Segregated by Condition - A) Library Size Distribution: *This plot displays the range of library sizes observed. Cells with smaller library sizes may be under cellular stress or represent incomplete sequencing data, potentially skewing the analysis. Conversely, unusually large library sizes often indicate doublets or multiplets, which can also distort results. Thresholds are set at a minimum of 250,000 and a maximum of 2,000,000 to exclude these outliers.*

B) Distribution of Detected Features: *Shown here are the total number of features detected per cell. Cells with a low count of detected features are likely experiencing cellular stress or insufficient capture, affecting data integrity. Cells at the higher end of the spectrum are typically doublets. A lower threshold of 4,500 and an upper threshold of 12,500 are applied to manage these extremes.*
C) Mitochondrial Gene Percentage: *This graph illustrates the percentage of mitochondrial genes, with higher percentages indicating significant cellular stress or mitochondrial dysfunction, which necessitates removal from the dataset. An upper limit of 8% has been established for mitochondrial content. Each panel distinguishes between normoxic and hypoxic conditions, enabling a direct comparison of cellular states under different experimental conditions.*

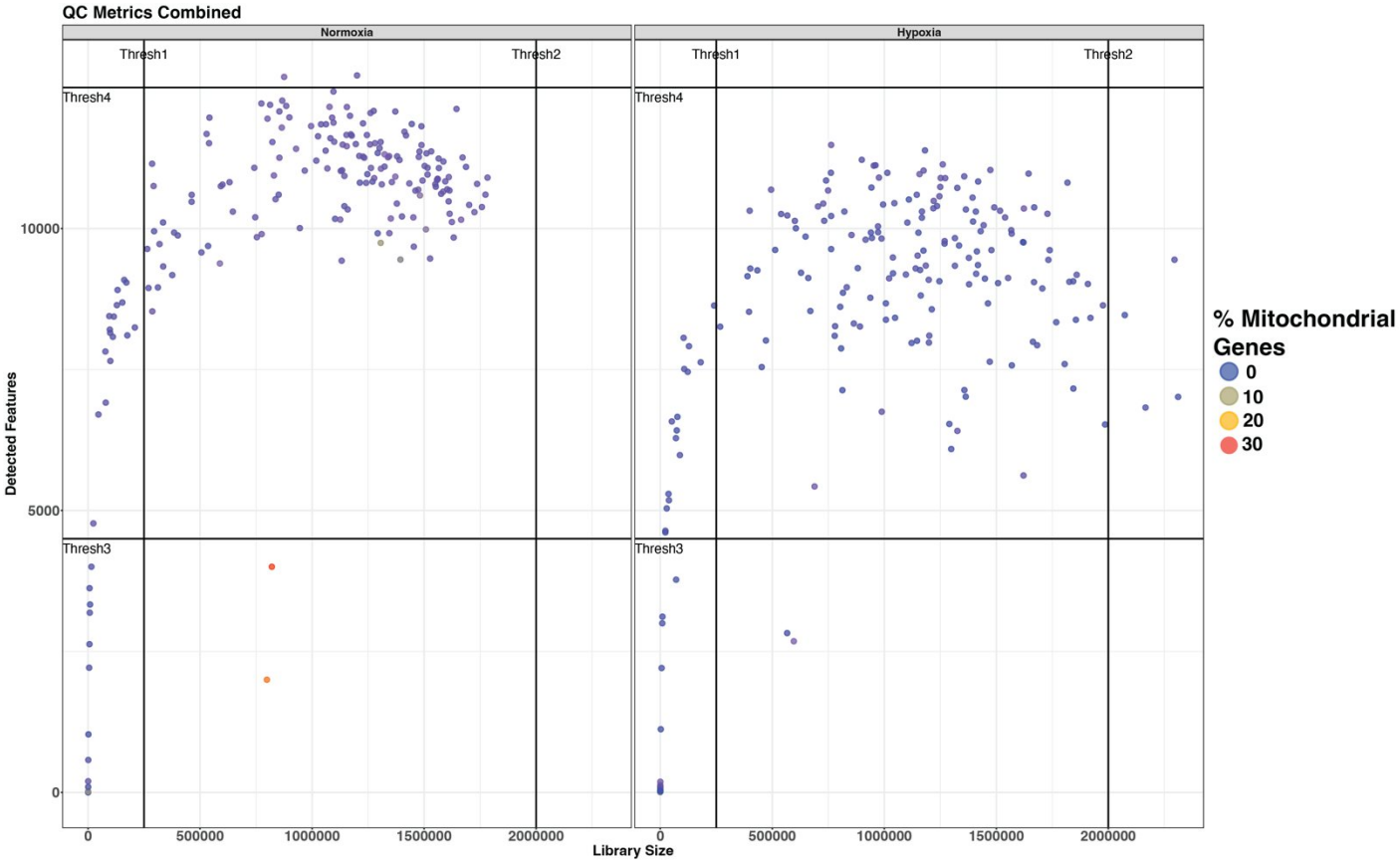


Figure 3-3: QC Metrics Scatterplot Across Conditions - *This scatterplot maps the quality control (QC) metrics for the dataset, with the x-axis representing library size and the y-axis indicating the number of detected features. Data points are colored based on the percentage of mitochondrial genes, providing a visual cue of mitochondrial content in each cell. Horizontal lines mark the thresholds for the number of detected features, while vertical lines denote the library size thresholds. These thresholds are critical for filtering out cells that do not meet the quality criteria. The plot is divided into panels for each condition—normoxia and hypoxia—allowing for a comparative view of how these thresholds impact data filtration in different cellular states. This facet design ensures that the viewer can assess the consistency and efficacy of the filtration process across varying experimental conditions.*

Following the application of these QC thresholds, a total of 71 cells (18% of the initial dataset) were excluded from further analysis. The subsequent stages of the processing pipeline included normalization, variable feature selection, and data. The processed data resulted in a normalized count matrix, which was then subjected to Principal Component Analysis (PCA) and Uniform Manifold Approximation and Projection (UMAP) for dimensionality reduction and to aid in the visualization of the data structure.

The PCA and UMAP visualizations provided critical insights into the dataset. PCA revealed that the primary source of variation along the first principal component (PC1) was attributed to the hypoxic response (responsible for 35.2% of the variation in the data), while the second principal component (PC2) primarily captured the variation due to the clustering of the cells (to be discussed later) **[Figure 3-4A&B]**. Correspondingly, the UMAP results aligned with those of the PCA, clearly displaying two distinct cell populations, indicative of the separate conditions of hypoxia and normoxia **[Figure 3-4C&D]**. It is important to note that distances within UMAP visualizations do not hold a valid metric for interpretation, so the observed heterogeneity was rigorously validated through subsequent downstream analyses.

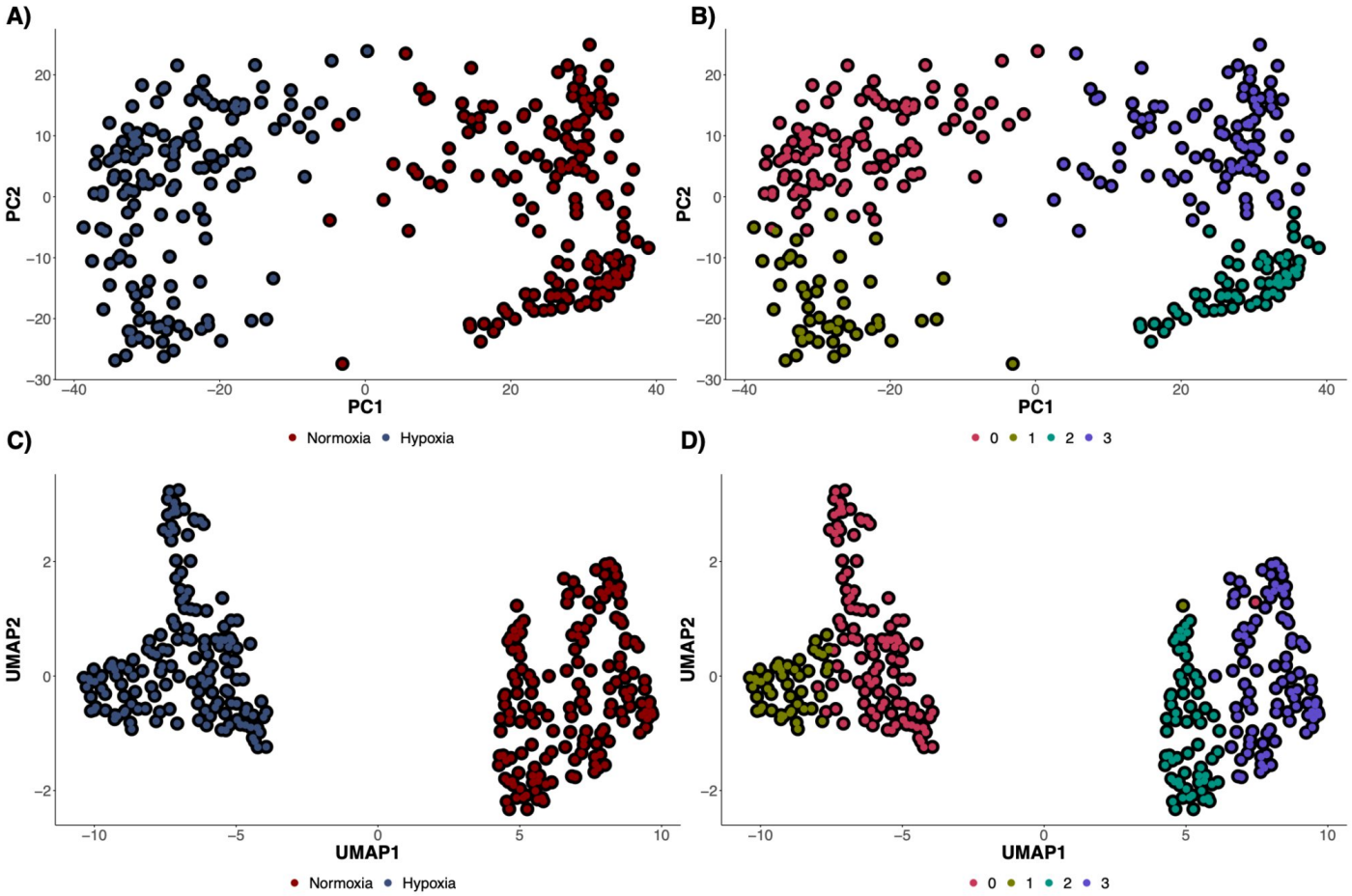


Figure 3-4: Structural Analysis of MCF7 Dataset Under Different Conditions - A) PCA Plot by Hypoxic Condition: *This plot colors data points based on whether cells were under normoxic or hypoxic conditions, illustrating that the primary variation along PC1 is driven by the hypoxic response. This observation underscores the strong impact of hypoxia on cellular behaviour.* ***B) PCA Plot by Identified Clusters:*** *This plot shows data points coloured according to distinct clusters identified within the dataset. It highlights that the secondary variation along PC2 corresponds to the formation of subpopulations, indicating differential cellular states within both normoxic and hypoxic conditions.* ***C) and D) UMAP Plots:*** *These plots reinforce the patterns observed in the PCA plots, providing a refined visualization that supports the separation of cell states and clusters. The UMAP plots further validate the presence of distinct cellular adaptations in response to hypoxia, as well as the emergence of subpopulations.*

3.3.2. HIF1A-Dependent Hypoxic Targets Detected at the Single Level

The first objective of this analysis was to validate the fidelity of the HIF1A-dependent hypoxic response. As part of the pre-processing pipeline, after normalizing the data, variance analysis was conducted for this scRNA-Seq dataset and visualised using a mean-variance plot **[Figure 3-5]**. This analysis revealed significant variability in the expression of core hypoxic response genes, including ADM, NDRG1, DDIT4, CA9, and VEGFA. Each of these genes is a recognized target of the HIF1A transcription factor, underscoring their roles in cellular responses to hypoxia. Furthermore, these genes are integral components of both the Buffa hypoxia signature (Buffa et al., 2010a) and the broader hallmark hypoxia signature (Liberzon et al., 2015), confirming their critical involvement in hypoxia-related processes.

Other key features identified in this variance analysis include CYP1A1, CYP1B1 and TK1 **[Figure 3-5]**. CYP1B1 is Cytochrome P450 Family 1 Subfamily B Member 1. CYP1B1 is a strong promoter of carcinogenesis that is significantly upregulated in breast cancer, and its upregulation in ER-positive breast cancers has been shown to be driven by HIF1A binding to ERalpha, nuclear translocation of the receptor complex, and promotion of CYP1B1 expression by binding to the ER response element on the CYP1B1 promoter. An alternative mechanism of HIF1A-dependent ERalpha activation to promote CYP1B1 expression involves MAP kinases and their potentiation of ERalpha signalling through post-translation modifications (Min et al., 2022).

Thymidine kinase 1 (TK1) is a crucial enzyme involved in DNA replication and repair. Its expression is tightly regulated by the cell cycle, with peak levels occurring during the S phase. This phase-specific expression makes TK1 serum levels particularly valuable as biomarkers for assessing the therapeutic efficacy of CDK 4/6 inhibitors in ER-positive breast cancers (Bagegni et al., 2017). Additionally, the significant variability of TK1 expression in response to hypoxia is noteworthy. This variability is likely attributable to the extensive alterations in the cell cycle dynamics that hypoxia triggers, underscoring the potential impact of environmental stress on cellular proliferation processes.

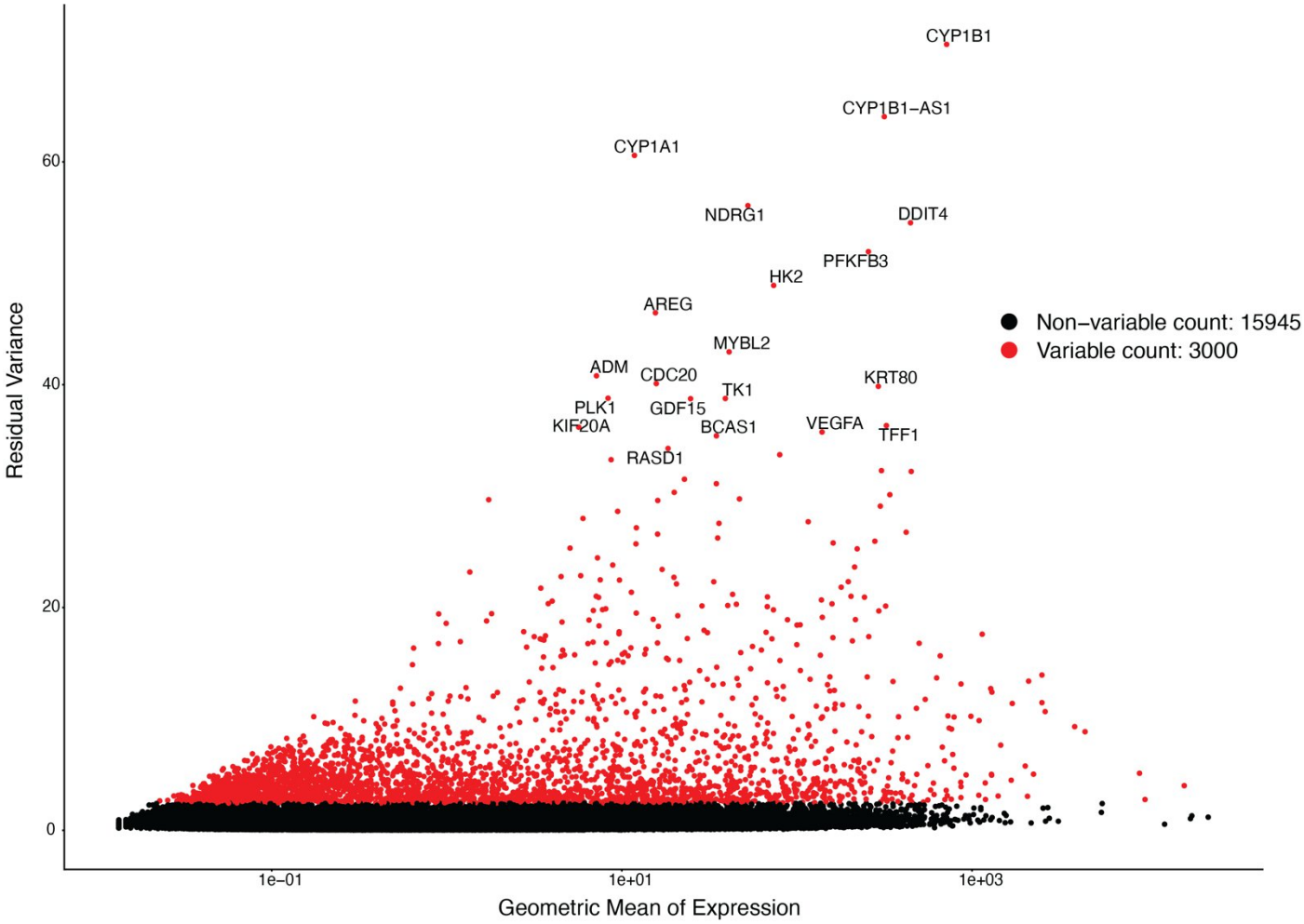


Figure 3-5: Mean-Variance Plot for MCF7 Dataset - *This plot depicts the relationship between the geometric mean of gene expression (x-axis) and the residual variance (y-axis). Each point represents a feature (gene), with the colour indicating whether the feature is classified as variable (red) or non-variable (black). The variance analysis is configured to automatically select the top 3000 features exhibiting the highest variability, which are critical for further biological interpretations. Prominent among the top variable features are canonical markers of hypoxia, such as DDIT4, VEGFA, and NDRG1. These genes are known targets of the HIF1A pathway and play crucial roles in the cellular response to hypoxia, making their high variability particularly significant.*

To further validate the level of activation of the transcriptional response to hypoxia, expression levels of CA9 and ADM were visualised using violin plots **[Figure 3-6A]**; these features showed high leves of activation in the hypoxic state. The canonical controls were RPL11 and ACTB **[Figure 3-6B]**, both common housekeeping genes, did not change in hypoxia versus normoxia in MCF7 cells.

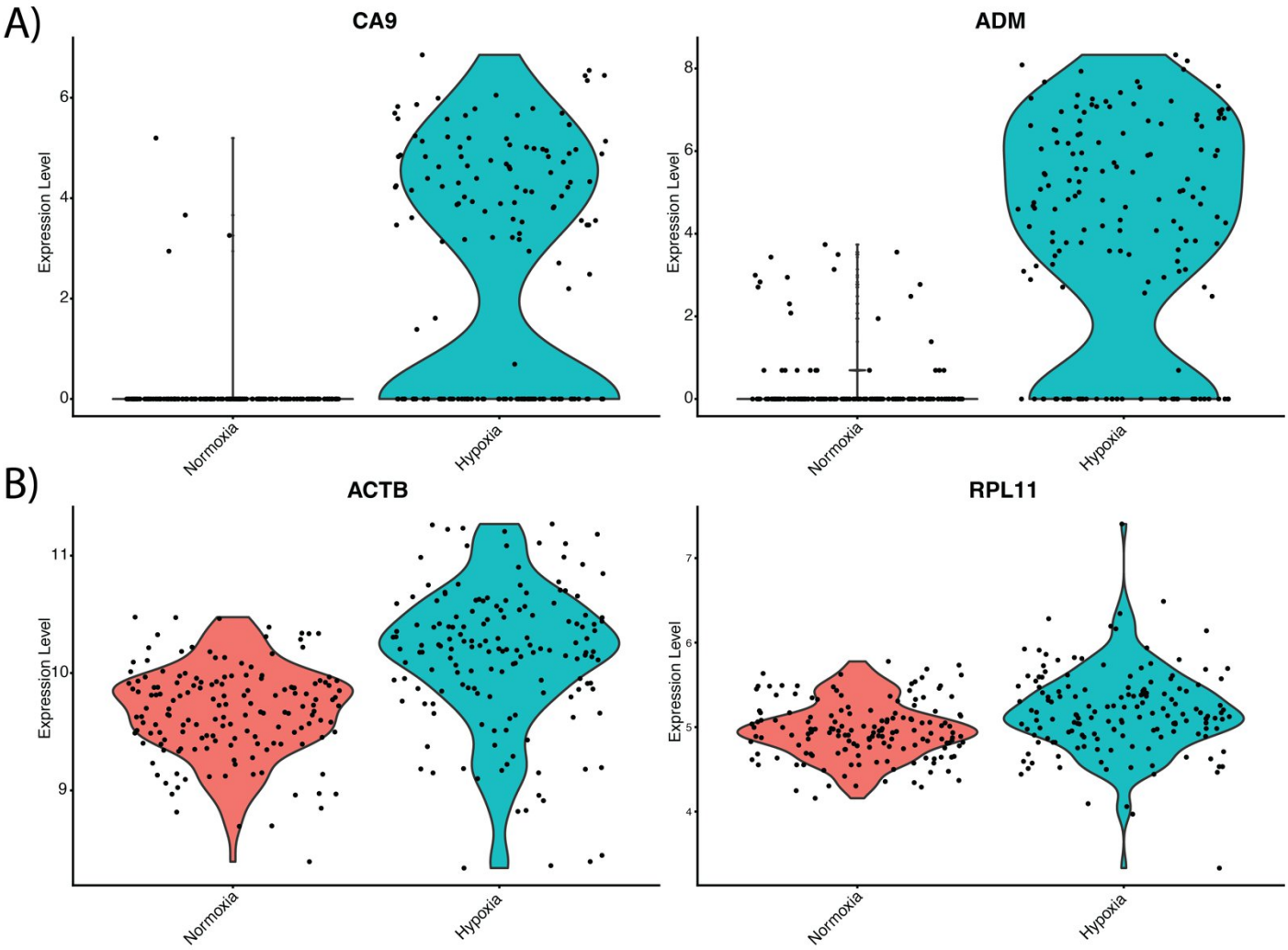


Figure 3-6: Expression Variability of Hypoxia-Related and Housekeeping Genes - *This figure illustrates the variability and expression patterns of key genes under normoxic and hypoxic conditions.* ***A) Hypoxia Response Genes:*** *Panels show the expression levels of CA9 and ADM, which are significantly elevated in hypoxic conditions compared to normoxia, underscoring their roles as canonical markers of the hypoxic response.* ***B) Housekeeping Genes:*** *Expression levels of ACTB and RPL11 are depicted, demonstrating their stability across both normoxic and hypoxic states. These housekeeping genes serve as controls, maintaining consistent expression irrespective of experimental changes.*

Subsequent analyses utilized the feature loadings from PC1 to perform Gene Set Enrichment Analysis (GSEA) with Gene Ontology Biological Process (GO BP) terms **[Figure 3-7]**. The results identified "response to oxygen levels" and "epithelial cell differentiation" as the top enriched pathways. The enhancement of the former underscores a robust hypoxic response, while the latter suggests the potential development of an epithelial-to-mesenchymal transition (EMT) phenotype. Conversely, pathways such as "cell cycle" and "cell division" were among the most downregulated, indicating a hypoxia-induced cell cycle arrest as a response to cellular stress.

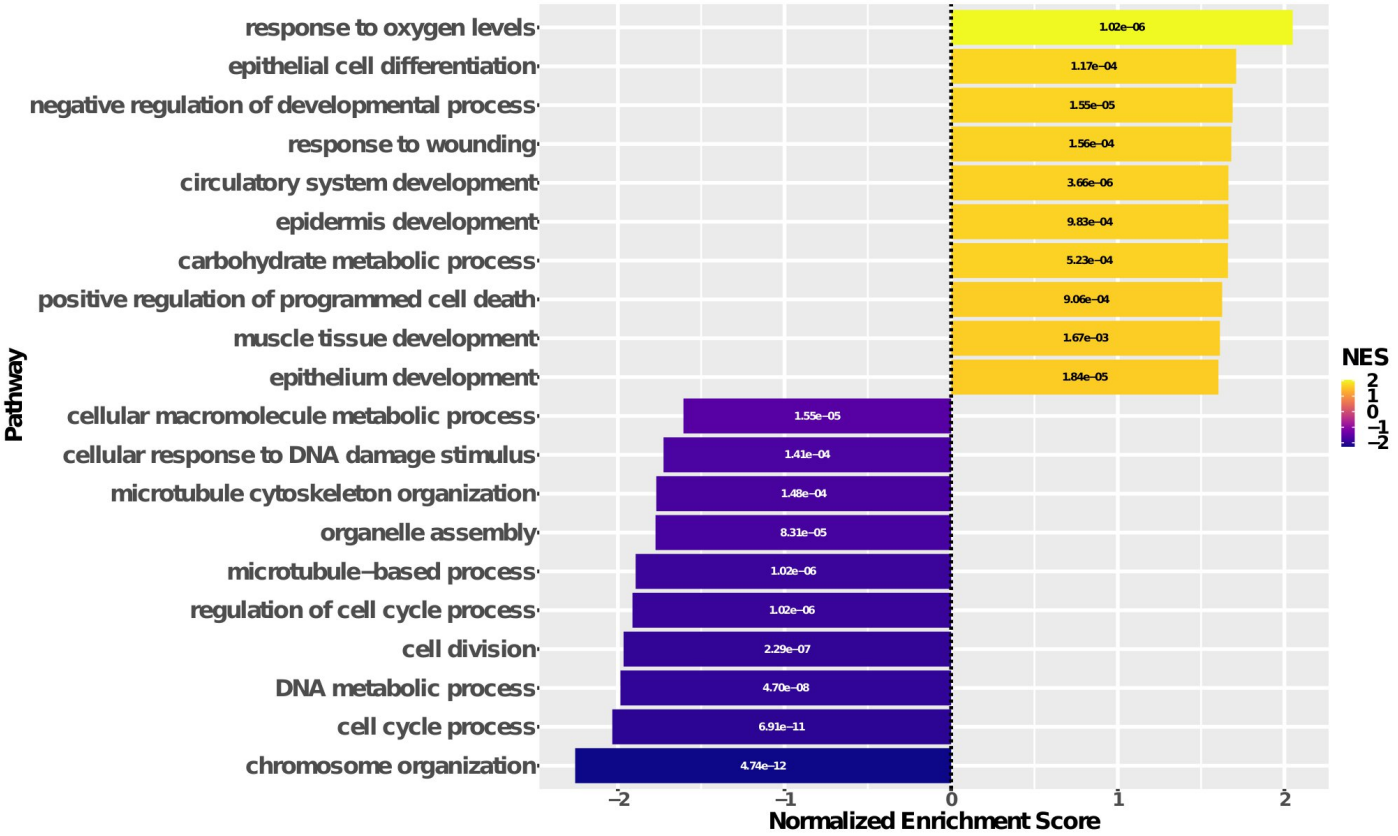


Figure 3-7: Gene Set Enrichment Analysis (GSEA) Based on PC1 Feature Loadings *- This figure presents the GSEA results using PC1 feature loadings derived from PCA, which targeted the primary source of variability in the dataset associated with hypoxic response. The bar chart illustrates the top 10 enriched and bottom 10 depleted pathways, where the x-axis represents the normalized enrichment score, and the y-axis lists the pathways. Numbers inside the bars indicate adjusted p-values. Significantly upregulated pathways include 'response to oxygen levels' and 'epithelial cell differentiation,' underscoring the influence of hypoxic signaling and suggesting the induction of epithelial-to-mesenchymal transition (EMT). In contrast, pathways such as 'cell division' and 'cell cycle' are markedly downregulated, indicative of a hypoxia-induced cell cycle arrest. These results highlight the complex dual impact of hypoxia on cellular behaviour—fostering adaptations to low oxygen conditions with activation of the HIF1A-dependent hypoxic response while concurrently restricting proliferative activities to prevent cell death.*

3.3.3. Hypoxia-Driven Cell Cycle Arrest

Hypoxia acts a predominant form of cellular stress, and it is well-documented to have substantial effects on the cell cycle, which the preceding GSEA analysis of PC1 feature loadings hints at. Using Seurat's cell cycle scoring functionality (Tirosh et al., 2016), MCF7 cells were classified into three phases of the cell cycle: G1, G2/M, and S phase. G2M and S phases are representative of cellular proliferation whereas G1 is representative of cell cycle arrest and/or senescence. Cell cycle classification was overlayed on MCF7 UMAP plots **[Figure 3-8A]**. There was a clear shift towards the G1 phase with a 2.5-fold increase in the percentage of G1 cells in the hypoxic state **[Figure 3-8B].** These results were further validated by quantification and visualization of MKI67 expression levels in this dataset. MKI67, widely recognized as a marker for cell proliferation, is routinely used in both real-time quantitative PCR (rt-qPCR) and immunohistochemistry (X. Sun & Kaufman, 2018). MKI67 expression levels showed marked reduction in MKI67 *($W = 5511.5$, $p\text{-value} < 2.2e\text{-}16$)* **[Figure 3-8C]**.

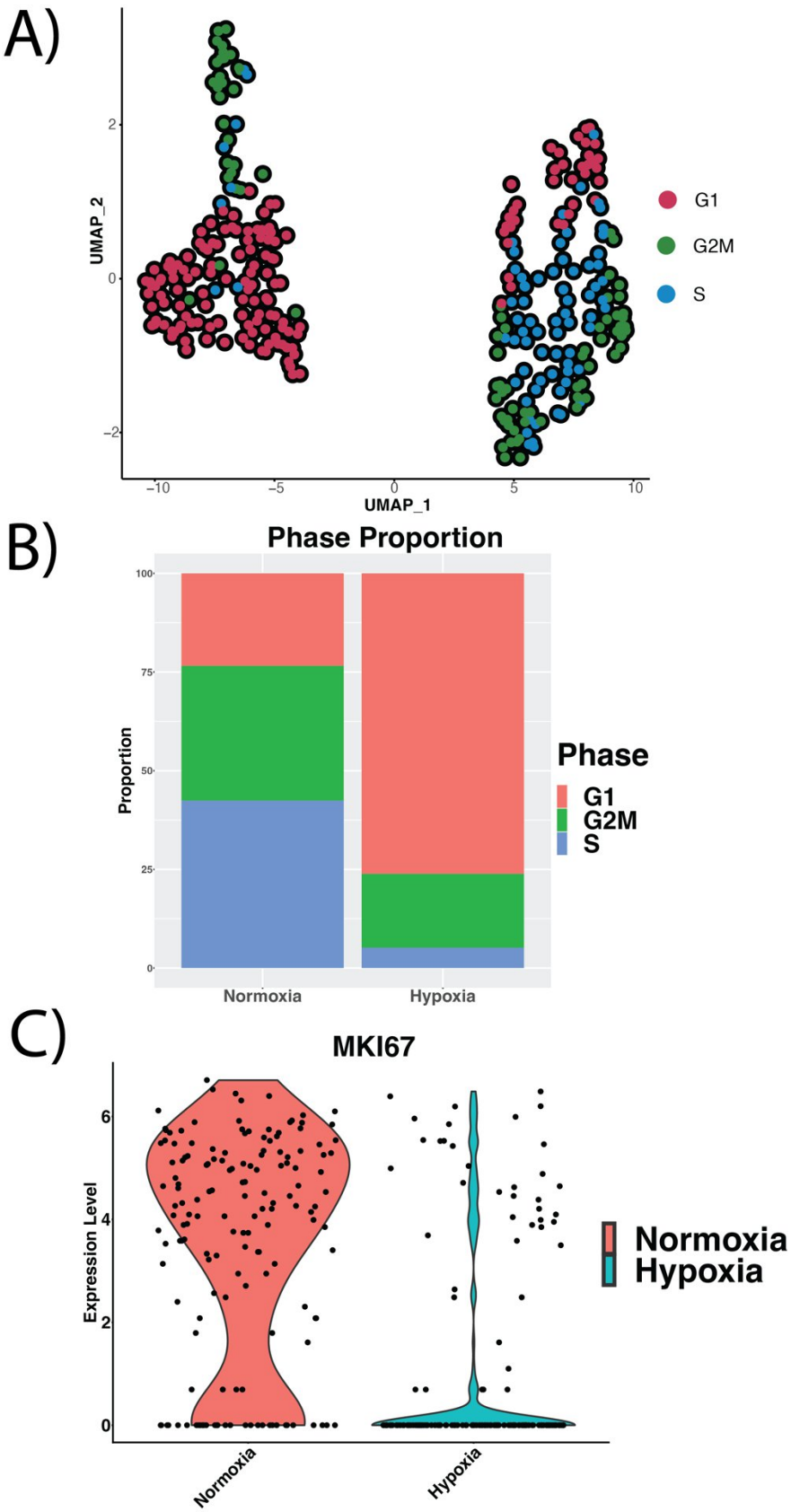


Figure 3-8: Analysis of Cell Cycle Dynamics and MKI67 Expression Under Hypoxic Conditions - A) UMAP Plot*: This visualization depicts the distribution of cells across the G2M, S, and G1 phases of the cell cycle, color-coded for clarity. Notably, there is a pronounced shift towards the G1 phase in the hypoxic state, as seen in the left cluster of the UMAP, indicating a hypoxia-induced slowdown or arrest in the cell cycle progression.* ***B) Cell Cycle Phase Barplots****: These barplots display the proportion of cells in each phase of the cell cycle for normoxic and hypoxic conditions separately. The data reiterate the significant increase in the G1 population under hypoxia, further validating the UMAP findings.* ***C) MKI67 Expression Analysis****: Shows the expression levels of MKI67, a key marker of cell proliferation, under both normoxic and hypoxic conditions. A decrease in MKI67 expression under hypoxia corroborates the observed cell cycle arrest, illustrating the suppression of proliferation in low oxygen environments. MKI67 expression is quantified here using reverse transcription quantitative PCR (RT-qPCR), a sensitive technique for measuring mRNA levels.* Top of FormBottom of Form

The exploration of cell cycle changes between normoxic and hypoxic conditions necessitated additional investigations. We assessed the expression levels of key cell cycle regulators and conducted signature scoring for the Hallmark G2M checkpoint pathway to understand these dynamics thoroughly.

The primary markers evaluated were CDK1 and UBE2C **[Figure 3-9A&B]**. CDK1, a vital enzyme in cell cycle regulation, associates with cyclins A and B to drive the mitotic process (Qian et al., 2015). Ubiquitin-conjugating enzyme E2 C (UBE2C) plays a pivotal role in the anaphase-promoting complex/cyclosome (APC/C), essential for the metaphase-anaphase transition through its role in protein ubiquitination and degradation. Notably, research indicates that in esophageal squamous cell carcinoma, UBE2C's activity is enhanced by FOXM1, which promotes cell cycle progression (Nicolau-Neto et al., 2018). Additionally, recent analyses have underscored the importance of both CDK1 and UBE2C as critical indicators of cell cycle activity, identified through advanced computational approaches in single-cell RNA sequencing (Huang et al., 2022).

The Hallmark pathways from the Molecular Signature Database (MSigDB) are meticulously curated to epitomize key biological phenotypes, having been derived from a broad collection of gene sets within the MSigDB to minimize redundancy and enhance the portrayal of biological processes (Liberzon et al., 2015). Particularly relevant to this analysis are the Hallmark G2M checkpoint and Hallmark E2F targets pathways **[Figure 3-9C&D]**. The G2M checkpoint pathway encompasses critical genes necessary for the cell's transition into mitosis, while the E2F targets pathway focuses on genes regulated by E2F, a pivotal transcription factor governing the transition from the G1 to the S phase of the cell cycle.

These pathways serve as robust indicators of cell cycle activity, reflecting the dynamic cellular responses to environmental cues. For this study, the pathway signatures were imported using the msigdbr package and analysed quantitatively through the ***Seurat::AddModuleScore()*** function in Seurat.

Visual assessments of these pivotal cell cycle markers **[Figure 3-9A&B]** alongside the Hallmark G2M checkpoint and E2F targets pathways **[Figure 3-9C&D]** consistently demonstrate a significant reduction in cell cycle processes during acute hypoxia. This evidence supports earlier findings from cell cycle scoring analysis, highlighting the influence of hypoxic stress in moderating cell cycle progression to promote cellular adaptation and survival.

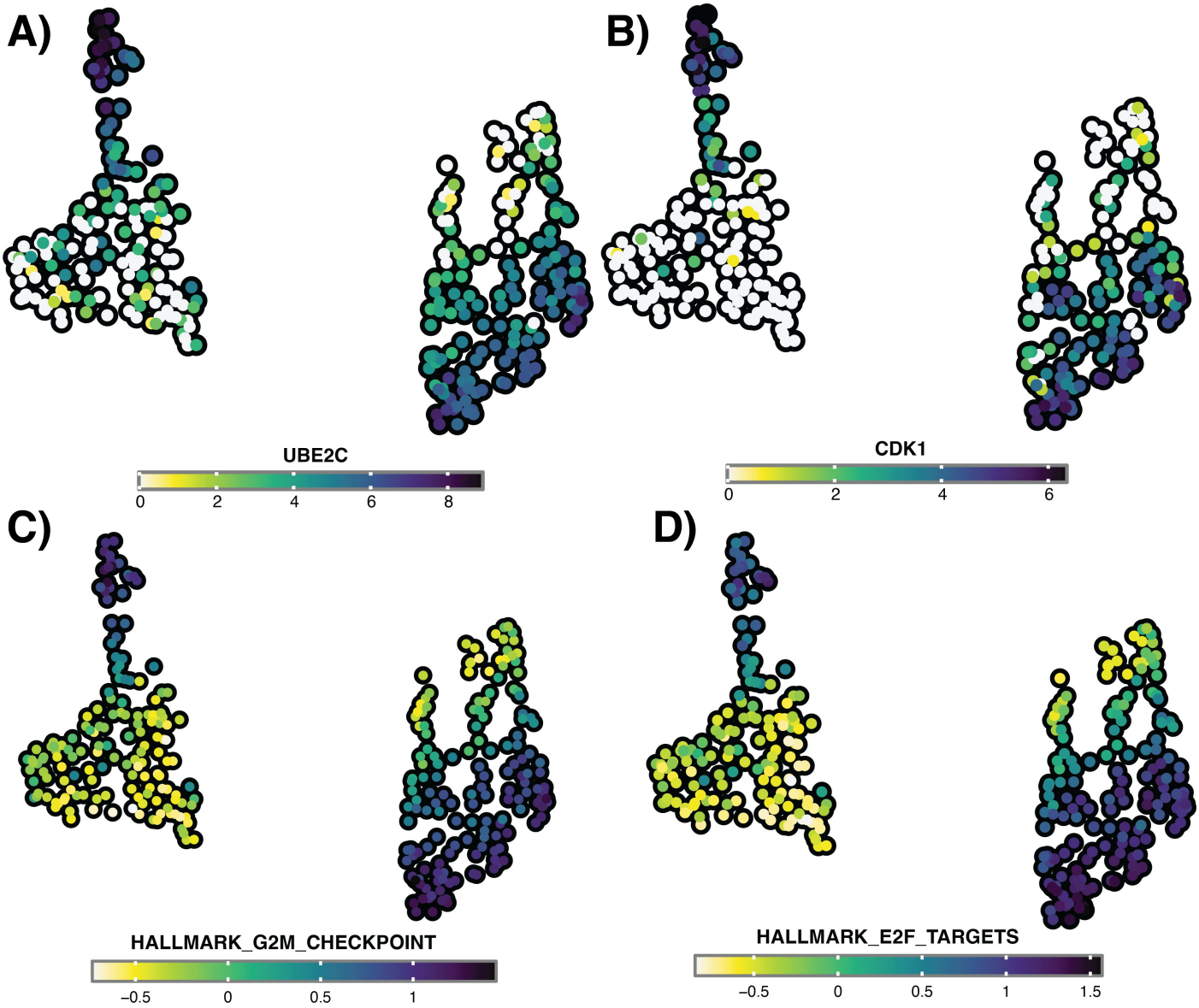


Figure 3-9: Visualization of Cell Cycle Arrest in MCF7 Cells Under Hypoxia - A) UBE2C Expression UMAP Plot*: This plot shows the expression levels of UBE2C, significantly downregulated in most hypoxic cells, particularly at the bottom of the UMAP plot. Notable is a minor upregulation in a subset of hypoxic cells within the G2M phase.* ***B) CDK1 Expression UMAP Plot****: Similarly, CDK1 expression, essential for mitotic control, is largely reduced under hypoxia, especially in the lower regions of the plot, highlighting hypoxia-induced disruptions in cell cycle progression.* ***C) Hallmark G2M Checkpoint Signature Scores UMAP Plot****: Illustrates the distribution of signature scores for the Hallmark G2M checkpoint pathway. The scores are predominantly lower across the hypoxic cells, which aligns with the downregulation patterns observed in the UBE2C and CDK1 expression plots, indicating suppressed cell cycle activity.* ***D) E2F Targets Signature Scores UMAP Plot****: This plot shows the signature scores for E2F targets, critical for cell proliferation and DNA synthesis. Similar to the G2M checkpoint, there is a marked decrease in activity across most of the hypoxic population, reinforcing the evidence of cell cycle arrest.*

Given the substantial influence of hypoxia on cell cycle regulation, it was equally interesting to consider the other side of the equation: the effect of cell cycle variations on the induction of known hypoxia markers under conditions of acute hypoxia. Selected for their prominent roles in both the Buffa and Hallmark hypoxia signatures, ADM, BNIP3, VEGFA, and NDRG1 served as the canonical markers of hypoxia . Analysis of their expression across various cell cycle phases—including S and G2M—revealed a consistent induction, highlighting the pervasive and intense nature of the hypoxic response, even in cells that are actively progressing through the cycle **[Figure 3-10]**. This observation underscores the profound interplay between cellular oxygen sensing and cell cycle mechanics, illuminating how hypoxia can impose its effects across the cellular landscape.

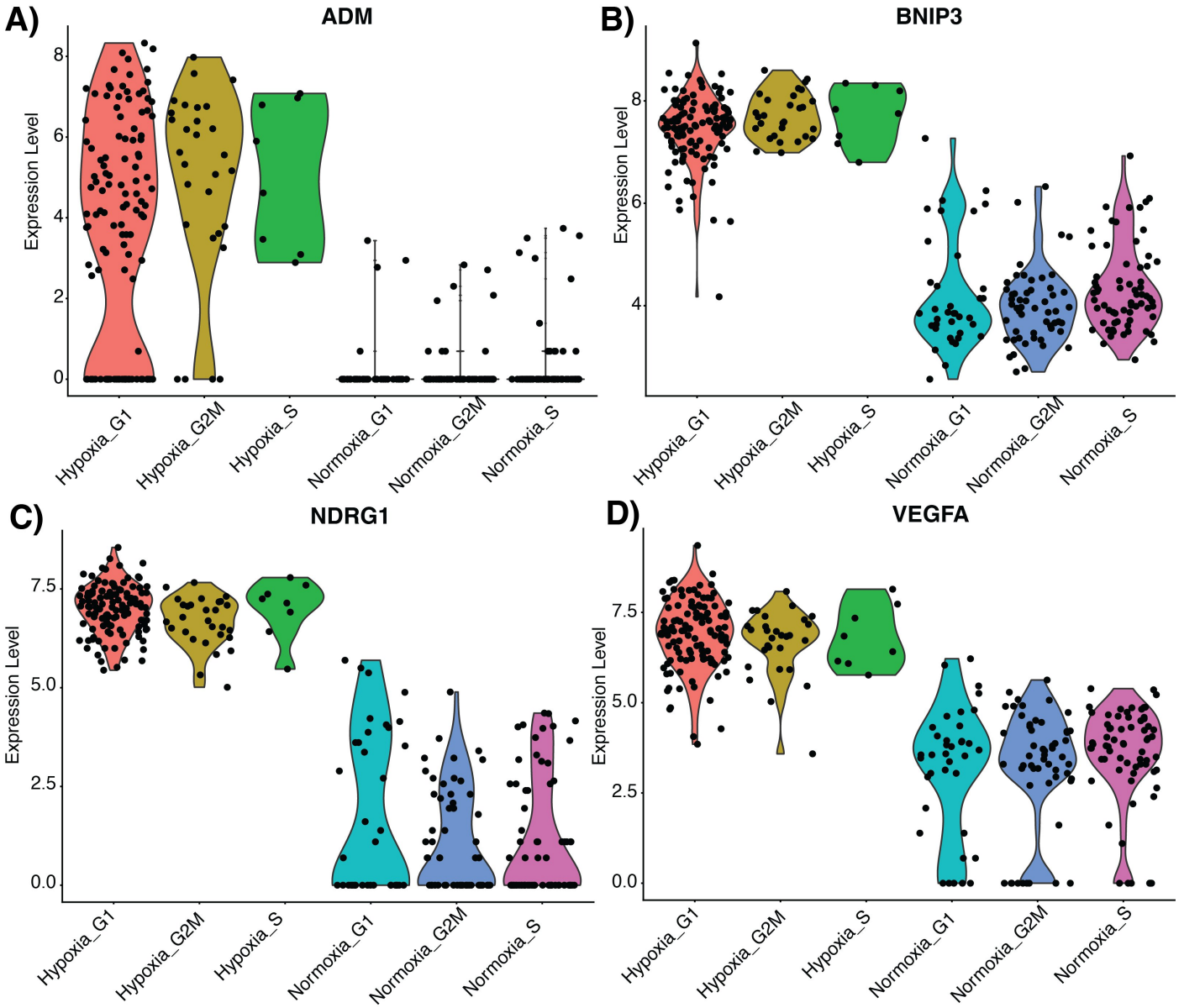


Figure 3-10: Induction of Hypoxia Markers Across Different Cell Cycle Phases in Acute Hypoxia - *This figure displays violin plots illustrating the expression levels of canonical hypoxia markers ADM, BNIP3, NDRG1, and VEGFA, stratified by unique combinations of cell cycle phases and hypoxic or normoxic conditions. Each plot examines the effects of the cell cycle on the induction of these hypoxia markers. The plots are segmented by cell cycle phases (G1, S, G2M) and condition (hypoxia or normoxia) to explore potential differential effects on the expression of these markers. The analysis aims to identify any variances in hypoxia marker induction that are influenced by the cell cycle, providing insights into how cellular proliferation dynamics may impact cellular responses to hypoxic stress.*

3.3.4. Hypoxic Response Heterogeneity and Functional Specialisation of Cell Subpopulations

Differential Expression Analysis (DEA) was performed to pinpoint the primary markers defining the hypoxic state at the single-cell level. The initial hypothesis anticipated that the core targets would be the well-known canonical hypoxic markers, which are typically characterized by their significant induction detectable through RT-qPCR, microarrays, and bulk RNA sequencing techniques. Additionally, we expected to identify variant features potentially indicative of functional specialization within cells under hypoxic conditions.

The DEA results were visualized using a heatmap that displayed the top 20 upregulated and top 20 downregulated genes **[Figure 3-11]**. Among the upregulated genes were established hypoxia markers such as VEGFA, BNIP3, and NDRG1. These markers not only showed substantial variability in the mean-variance analysis but also emerged as critical effectors of hypoxic signaling.

Their expression profiles exhibited highly consistent patterns of elevated expression across the hypoxic cells, emphasizing their essential roles in mediating the hypoxic response.

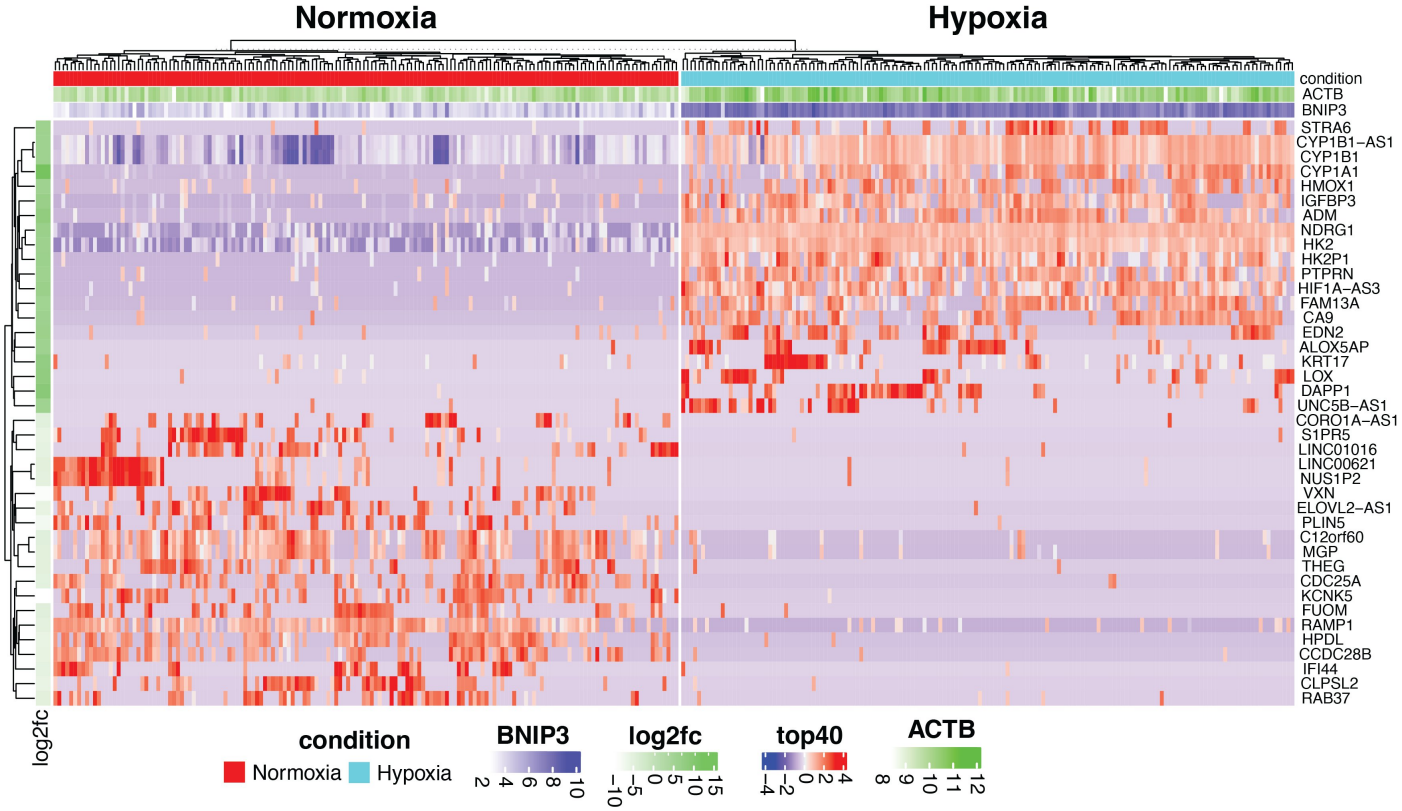


Figure 3-11: Differential Expression Analysis (DEA) Heatmap Comparing Hypoxia and Normoxia *- This heatmap presents the results from a Differential Expression Analysis (DEA) comparing hypoxic and normoxic conditions, specifically focusing on the top 20 upregulated and downregulated genes in response to hypoxia. Each row within the heatmap represents a distinct gene, while columns correspond to individual cells, illustrating the gene expression variations under each condition. The column annotations display the expression levels of ACTB and BNIP3, selected for their relevance as markers: BNIP3 highlights the hypoxic response due to its upregulation in low oxygen environments, and ACTB serves as a baseline transcriptional marker because of its role as a housekeeping gene. Additionally, the heatmap includes annotations for condition labels (hypoxia and normoxia) and row annotations for average log2 fold change levels, which were utilized to order the DEA features, thereby identifying the top and bottom 20 genes.*

However, the analysis also uncovered a distinct group of genes — LOX, KRT17, and EDN2 — which were more strongly induced yet displayed more heterogeneity in expression. This variability suggests these genes might play roles in driving functional differentiation within the hypoxic environment. Their diverse expression patterns across the hypoxic cell population highlight their potential importance in the nuanced adaptations to hypoxia.

Among the top downregulated features in the DEA were CDC25A, KCNK5, and SLC27A4. CDC25A, a critical tyrosine protein phosphatase, plays a pivotal role in cell cycle regulation. It primarily facilitates cell cycle progression by activating CDK1 through the dephosphorylation of its inhibitory sites, thereby enabling CDK1's catalytic functions. The observed downregulation of CDC25A under hypoxic conditions aligns with its role in cell cycle arrest, a common response to hypoxia as the cell adapts to reduced oxygen availability.

KCNK5, a potassium channel, and SLC27A4, a fatty acid transporter, also exhibited decreased expression in response to hypoxia. The downregulation of these genes hints at broader metabolic and ionic changes that occur as cells adjust to hypoxic stress. Although intriguing, the specific implications of reduced KCNK5 and SLC27A4 expression under these conditions were not the focus of this study and warrant further investigation to fully understand their roles and impacts in hypoxic adaptation.

To explore potential functional specialization within the hypoxic population, clustering analysis was conducted using the Louvain community detection algorithm. The resolution parameter was

optimized through clustering trees to accurately delineate the groups, resulting in the identification of four distinct clusters—two under normoxic conditions and two under hypoxic conditions.

The results of the clustering analysis were effectively visualized using PCA and UMAP plots **[Figure 3-12]**. The PCA plot distinctly demonstrates that the primary source of variation along PC2 is attributed to cluster formation **[Figure 3-12A]**. Complementarily, the UMAP plot provides a clear visual differentiation, segmenting hypoxic and normoxic cells into separate subpopulations within this multidimensional space **[Figure 3-12B]**.

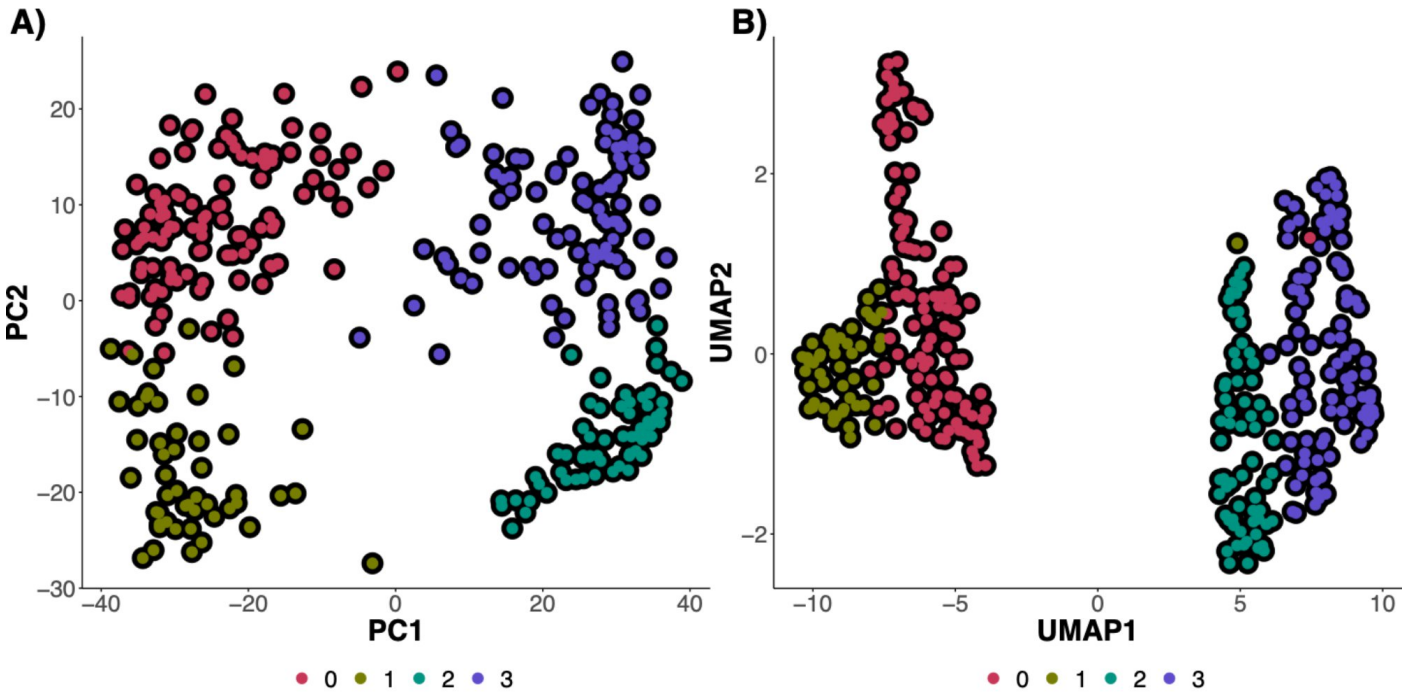


Figure 3-12: PCA and UMAP Plots Demonstrating Cellular Clustering in Response to Hypoxia - A) PCA Plot*: This plot highlights the significant clustering of cells, with the primary source of variation along the second principal component (PC2) being attributed to cluster formation. This suggests distinct phenotypic subgroups within the hypoxic an normoxic cell populations.* ***B) UMAP Plot****: Complementing the PCA findings, this UMAP visualization further delineates the division of cells into discrete clusters, showcasing how hypoxic and normoxic conditions contribute to the formation of distinct cellular subpopulations. Each cluster likely represents a unique adaptive or phenotypic response to hypoxic stress.*

Further characterization of the clusters was achieved through bar plots, which detail the composition of each cluster. This analysis confirms that Clusters 0 and 1 are associated with hypoxic conditions, while Clusters 2 and 3 correspond to normoxic conditions **[Figure 3-13A&B]**. Notably, Clusters 0 and 1 predominantly consist of cells in the G1 phase of the cell cycle (indicative of cell cycle arrest and stunted proliferation), whereas Clusters 2 and 3 are primarily composed of cells in the G2M and S phases **[Figure 3-13C&D]**.

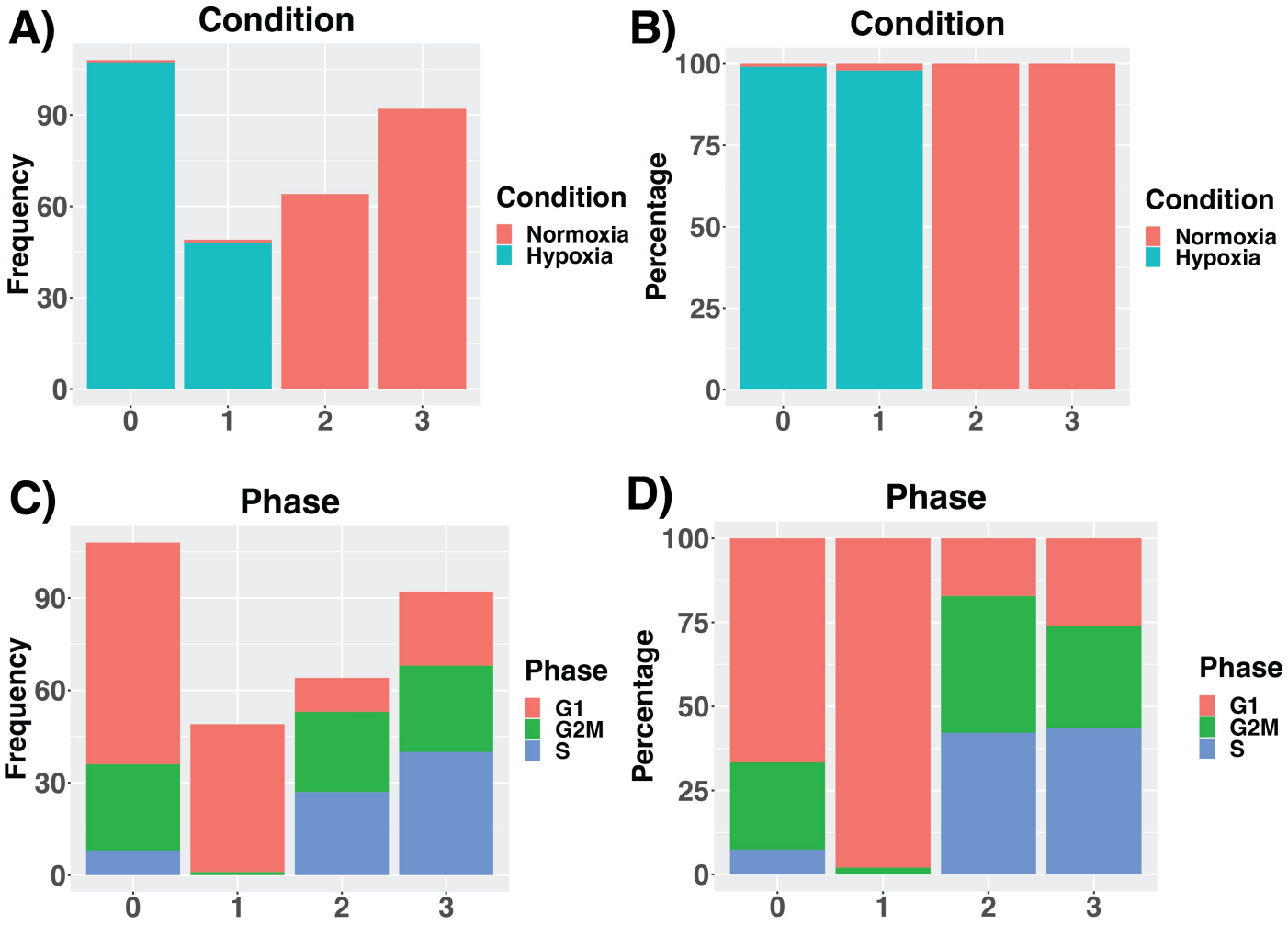


Figure 3-13: Cluster Composition Analysis - A) and B) Condition Composition: *These bar plots illustrate the count and proportion of cells within each cluster categorized into either normoxia or hypoxia.* ***C) and D) Cell Cycle Phase Distribution****: Corresponding to the cell cycle, these plots display the number and percentage of cells in each cluster across different phases of the cell cycle (G1, S, G2M). This clearly demonstrates the dominance of cellular arrest and reduced proliferation in the hypoxic clusters.*

Given the importance of these clusters and the biological variation they likely represent, further investigation was warranted. DEA analysis was carried out to identify the top marker genes for each cluster. The top 10 markers for each cluster (ordered by average log2fold change) were visualised in a heatmap **[**Figure 3-14**]**.

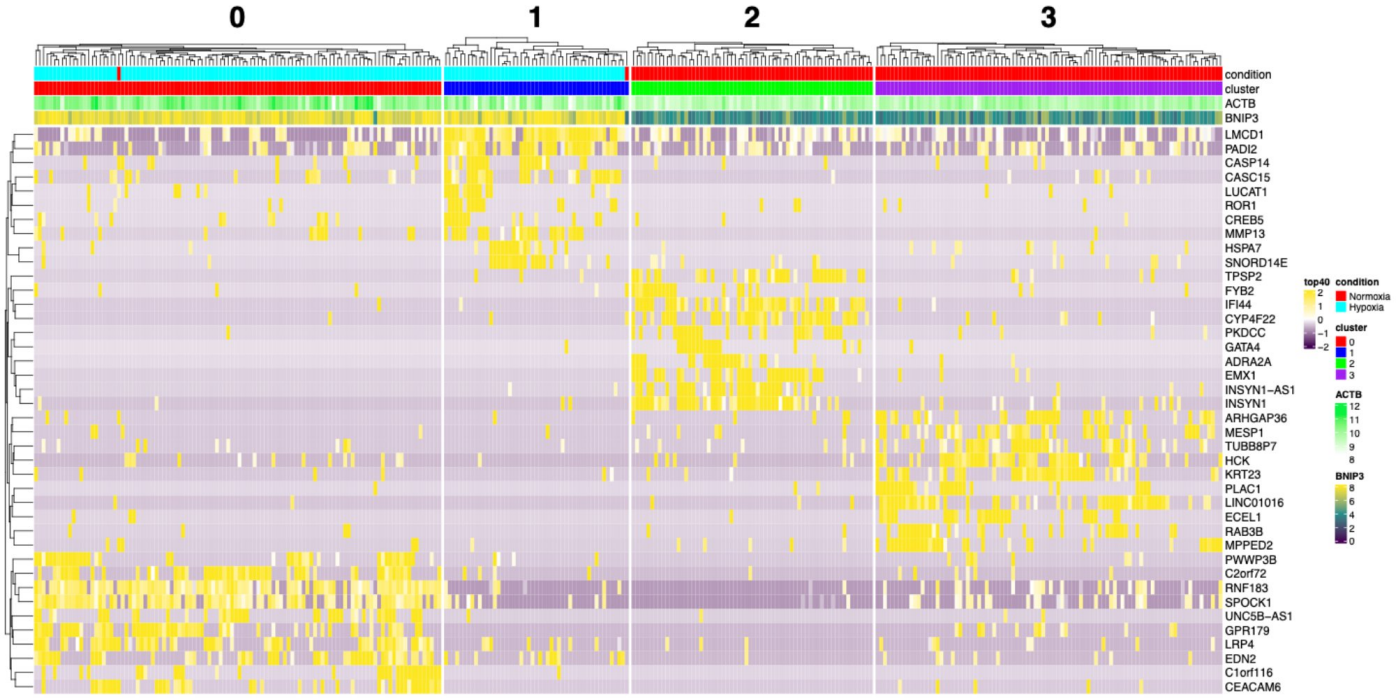


Figure 3-14: Differential Expression Analysis for Cluster-Specific Markers - *This heatmap displays the results of a differential expression analysis, identifying the top 10 genes for each cluster to delineate distinct molecular signatures. Each row within the heatmap represents a specific gene, and columns represent individual cells, showcasing the variability in gene expression across the clusters. Column annotations include the expression levels of ACTB and BNIP3, which are crucial for interpreting the data: BNIP3, upregulated in low oxygen conditions, serves as an indicator of the hypoxic response, while ACTB, a housekeeping gene, provides a baseline for transcriptional activity. Additionally, this heatmap is annotated with condition labels—distinguishing between hypoxic (clusters 0 and 1) and normoxic (clusters 2 and 3) states—to offer insights into the experimental conditions of the gene expression patterns observed.*

In the normoxic cluster 2 of breast cancer cells, several key genes such as IFI44, CYP4F22, EMX1, and INSYN1, along with its antisense mRNA, emerged as top regulated features. IFI44, an interferon-induced protein, plays a crucial role in immune response mechanisms by acting as a downstream target in the interferon alpha pathway. Despite the incomplete understanding of its full biological functions, IFI44 has been linked to poor clinical outcomes in head and neck squamous cell carcinoma (Liberzon et al., 2015; Pan et al., 2020). Its paralogue, IFI44L, contrasts by promoting cancer cell apoptosis through interactions with STAT1, which triggers apoptosis via the JAK-STAT1 signaling pathway (Zeng et al., 2023).

INSYN1, typically found in inhibitory synapses within the central nervous system, showed unexpected expression in normoxic MCF7 breast cancer cells, hinting at novel roles beyond neuronal functions. Similarly, EMX1, a transcription factor known for its role in neuronal development, was expressed in these cells. In sarcomas, EMX1 suppresses the Wnt/beta-catenin signaling pathway, which is linked to the development of aggressive cancer traits, suggesting that its downregulation under hypoxic conditions might facilitate cancer progression by enabling stem cell-like characteristics (Jimenez-García et al., 2021).

In the normoxic cluster 3, proteins such as KRT23, PLAC1, MMPED2, and LINC01016 were highlighted as top regulators. KRT23, a member of the keratin family, is notably implicated in various aspects of advanced cancer progression including cellular immortalization and metastasis. It directly influences the activity of human telomerase reverse transcriptase (hTERT), crucial for telomere elongation and tumor stabilization during tumorigenesis (Zhang et al., 2017). In ovarian cancer, KRT23 enhances the cancer's invasive traits by affecting the TGF-β/Smad signaling pathway, essential for acquiring migratory and invasive capabilities (Ren et al., 2020).

PLAC1, traditionally associated with placental tissues, displayed an unusual expression in the MCF7 line, indicating a possible role in cancer beyond its known function. Its expression in ovarian cancer is particularly influenced by p53 mutations, with wild-type p53 suppressing PLAC1 while mutations enhance its expression, correlating with poorer survival rates (Devor et al., 2017). Similarly, in

prostate cancer, high levels of PLAC1 have been linked with adverse outcomes, making it a potential target for immunotherapy (Mahmoudian et al., 2019).

Lastly, MPPED2, a cyclic nucleotide phosphodiesterase, regulates important cellular communication pathways by degrading cyclic mononucleotides like cAMP and cGMP. Its expression, often reduced in breast cancer through epigenetic silencing, when reactivated, shows a significant reduction in cancer progression, underscoring its potential as a therapeutic target. This suggests that hypoxia may exacerbate cancer progression by downregulating tumor suppressor genes such as MPPED2, adding complexity to the landscape of cancer progression (Pellecchia et al., 2019).

In the hypoxic cluster 0, prominent marker genes included SPOCK1, IGFBP5, and EDN2, each contributing significantly to the tumour's pathophysiology in unique ways. SPOCK1, a membrane proteoglycan involved in the epithelial-mesenchymal transition (EMT), has been shown to promote carcinogenesis by enhancing cancer cell survival and EMT processes. In pancreatic ductal adenocarcinoma, SPOCK1 reduction leads to increased apoptosis and cellular senescence, implicating it in cancer progression and poor patient outcomes (J. Li et al., 2020). Similarly, in breast cancer, SPOCK1's overexpression correlates with enhanced invasion and metastasis (Fan et al., 2016), and its role in hypoxia-induced EMT further underscores its influence in aggressive cancer dynamics.

IGFBP5, another critical player, modulates the activity of insulin-like growth factors, crucial for cellular growth and survival. It regulates these factors by binding and either promoting or inhibiting the activation of insulin-like growth factor receptors, which are involved in several oncogenic processes (J. Wang et al., 2015). In melanoma, IGFBP5 acts as a tumor suppressor by inhibiting epithelial-mesenchymal transition and reducing the activation of critical signaling pathways such as the TGF-β/Smad pathway. Conversely, in breast cancer, IGFBP5 can enhance cancer progression by supporting survival and proliferation pathways through its interactions with IGF1R and estrogen receptors (Dittmer, 2022). Interestingly, IGFBP5 has also been identified as a marker of cellular senescence in adipose tissues, suggesting its potential role in stress response and tumor suppression (Saul et al., 2022).

Hypoxic cluster 1 features genes like SNAI2 and ROR1, which play vital roles in tumor aggressiveness and metastasis. SNAI2, a transcription factor, is crucial for EMT by regulating the expression of E-cadherin and promoting metastatic capabilities through TGF-β mediated pathways (Ding et al., 2013). It also regulates other EMT markers such as ZEB1 and ZEB2, enhancing their expression and thus facilitating more aggressive cancer traits (Guaita et al., 2002). ROR1, an orphan receptor, interacts with Wnt5a to promote non-canonical Wnt signaling, influencing cell proliferation, apoptosis, and EMT. This receptor's activity highlights the complex regulatory mechanisms that cancer cells exploit to thrive under hypoxic conditions and resist therapeutic interventions (Menck et al., 2021; Zhou et al., 2020).

In our study, Differential Expression Analysis (DEA) genes were utilized to conduct a GSEA pathway enrichment analysis, employing GO BP terms to elucidate the biological processes involved under different oxygen conditions. The analysis was divided into two main categories: normoxic and hypoxic conditions.

In normoxic conditions, particularly in clusters 2 and 3 as shown in [**Figure 3-15B**], there is a significant enrichment of DNA replication and proliferative pathways, consistent with the S phase fractions noted in these clusters. Cluster 2 is distinctly involved in nuclear chromosome segregation and DNA replication, highlighting its strong link to cell division and genetic stability. Conversely, Cluster 3 shows a marked enrichment in ion transport processes, indicating an active role in cellular metabolism and homeostasis maintenance, crucial for cell survival and function.

Under hypoxic conditions, as illustrated in [**Figure 3-15A**], clusters 0 and 1 exhibit diverse pathway activations. Cluster 0 is significantly enriched in pathways related to skin, muscle, and epithelial

development, as well as cell adhesion, suggesting roles in structural development and integrity that could influence tissue architecture and intercellular interactions. Cluster 1, on the other hand, is primarily focused on cell motility, chemotaxis, and blood vessel development, underscoring mechanisms that may facilitate metastatic spread and angiogenesis, crucial for adapting to hypoxic stress.

This analysis highlights how oxygen availability influences cellular behavior through differential pathway activation. In hypoxic clusters, there is a noticeable shift towards mechanisms that enhance survival under low oxygen conditions, such as angiogenesis and cell motility, potentially contributing to cancer progression. In contrast, normoxic clusters [**Figure 3-15B**] prioritize cell division and metabolic stability, fostering growth and proliferation under more favorable conditions. The distinct pathway engagement in different oxygen environments not only underscores the adaptability of cells to their microenvironment but also pinpoints potential therapeutic targets, especially those active in hypoxic tumor microenvironments, suggesting avenues for targeted interventions.

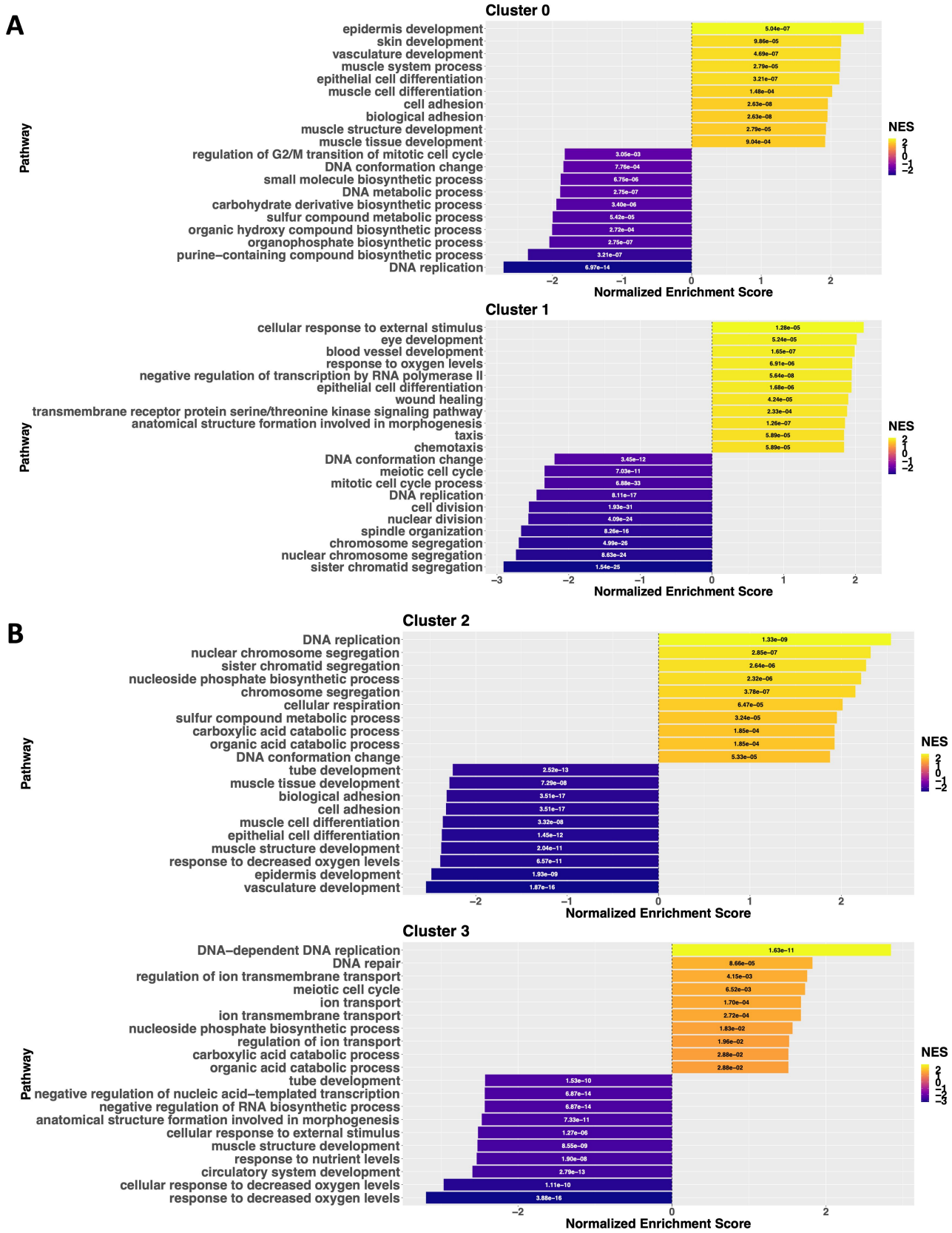


Figure 3-15: Gene Set Enrichment Analysis (GSEA) Results for Normoxic and Hypoxic Clusters - Visualization Structure: *Each plot displays pathways along the y-axis and normalized enrichment scores along the x-axis. Bars extending to the right indicate positive enrichment and upregulation of pathways, whereas bars extending to the left denote negative enrichment and downregulation. The numbers inside the bars represent the adjusted p-values for each pathway, providing a statistical measure of significance.* ***A) Hypoxic Clusters:*** *This plot illustrates significant pathway enrichment in processes related to angiogenesis, epithelial-to-mesenchymal transition (EMT), and cellular motility, highlighting adaptive mechanisms triggered by low oxygen conditions. Conversely, pathways involved in the cell cycle and mitosis are notably downregulated, suggesting a shift away from proliferation towards survival and migration under hypoxia.* ***B) Normoxic Clusters:*** *Cluster 2 displays strong enrichment in pathways associated with DNA replication and chromosome segregation, indicating active cell division and genetic material management crucial for sustained cellular proliferation. Cluster 3: shows enrichment in ion transport and metabolic processes, essential for*

maintaining cellular function and homeostasis in oxygen-rich conditions. Additionally, there's a marked downregulation of pathways associated with hypoxia and angiogenesis, reflecting the normoxic stability of the environment.

3.3.5. EPAS1-Driven Negative Regulation of ER Signalling As a Potential Mechanism of Therapeutic Resistance to Hormonal Therapy in ER+ Cancers

Emerging research highlights a critical interaction between estrogen receptor (ESR1) signaling and hypoxia-inducible factor 1-alpha (HIF1A), which could significantly influence the dynamics of cancer progression. Initial investigations revealed an ESR1-binding element within the HIF1A gene sequence, establishing a direct regulatory link between these two critical factors. This study employs chromatin immunoprecipitation sequencing (ChIP-Seq) to further delineate the landscape of genes that are co-regulated by both HIF1A and ESR1 (J. Yang et al., 2015). The presence of dual response elements in their promoters suggests a sophisticated and integrated regulatory mechanism where ESR1 and HIF1A not only coexist but collaboratively influence the expression of key molecular pathways that drive crucial oncogenic processes.

To probe deeper into the regulatory mechanisms at play, this study utilized ridge plots to examine the expression patterns of HIF1A, ESR1, and a curated set of 148 genes known to respond to both regulators under varying conditions of oxygen availability. The analysis targeted genes such as EGLN3, DDIT4, GAPDH, and GPI, observing their expression under normoxic and hypoxic conditions **[Figure 3-16]**. A particularly notable observation was made in hypoxic cluster 1, where there was a marked decrease in ESR1 expression. This prompted a targeted analysis of ESR1's canonical targets, including PGR, GREB1, and FOXM1, which similarly showed diminished expression levels **[Figure 3-17]**, suggesting a systemic downregulation of ESR1 activity under hypoxia.

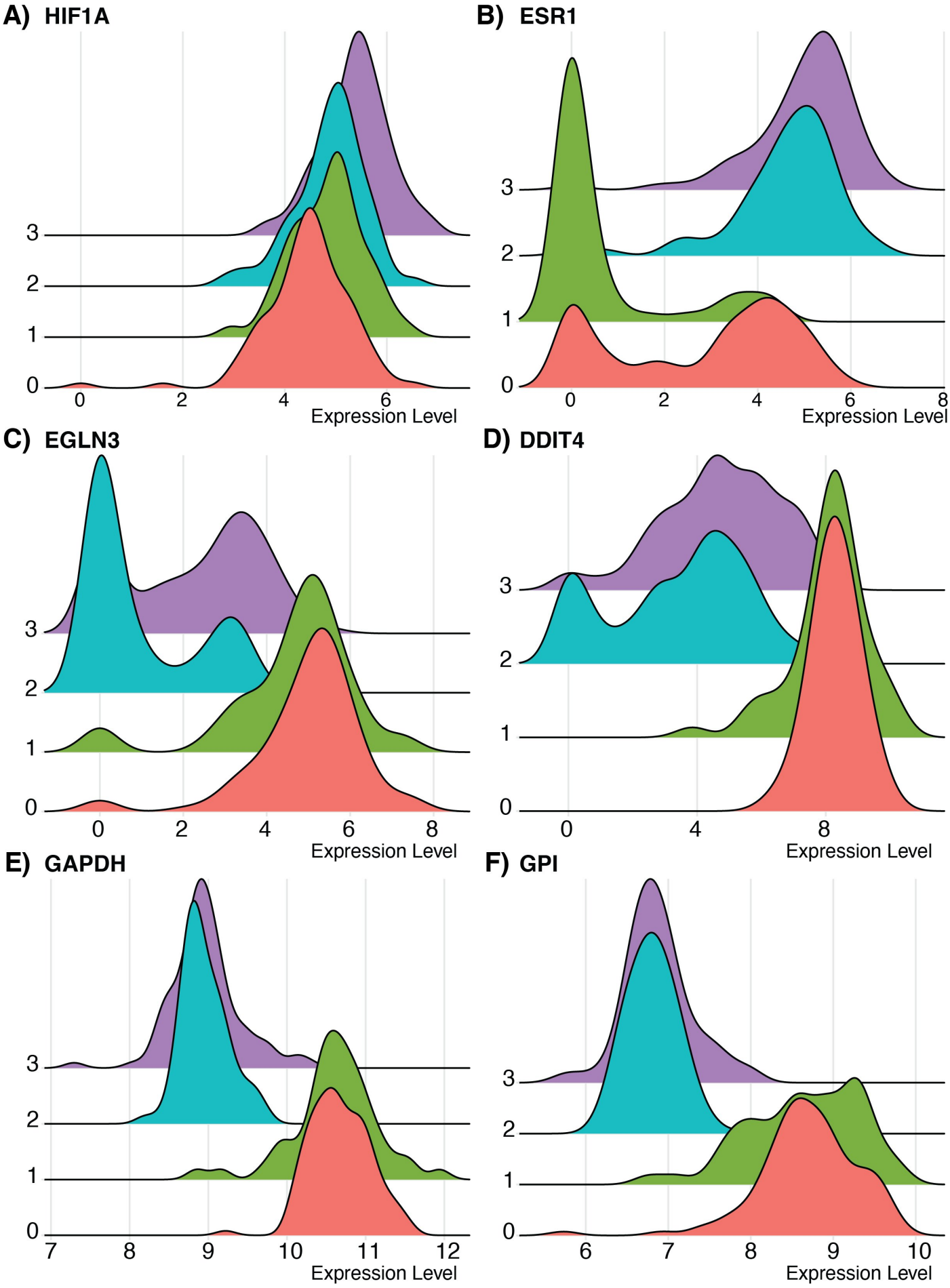


Figure 3-16: Ridge Plots Illustrating Expression Patterns of Dually Responsive Genes - *These ridge plots display the expression levels of key genes regulated by both hypoxia-inducible factor 1-alpha (HIF1A) and estrogen receptor (ESR1) across various cell clusters, highlighting how these genes respond under different cellular conditions. The plots include HIF1A and ESR1 themselves,*

as well as a selection of genes that are dually responsive to these transcription factors, such as EGLN3, DDIT4, GAPDH, and GPI. Notably, HIF1A shows consistent expression across all clusters, reflecting its regulation primarily at the post-translational level rather than through changes in mRNA levels. In contrast, ESR1 exhibits a significant decrease in expression within cluster 1, indicating a potential regulatory suppression in this cluster.

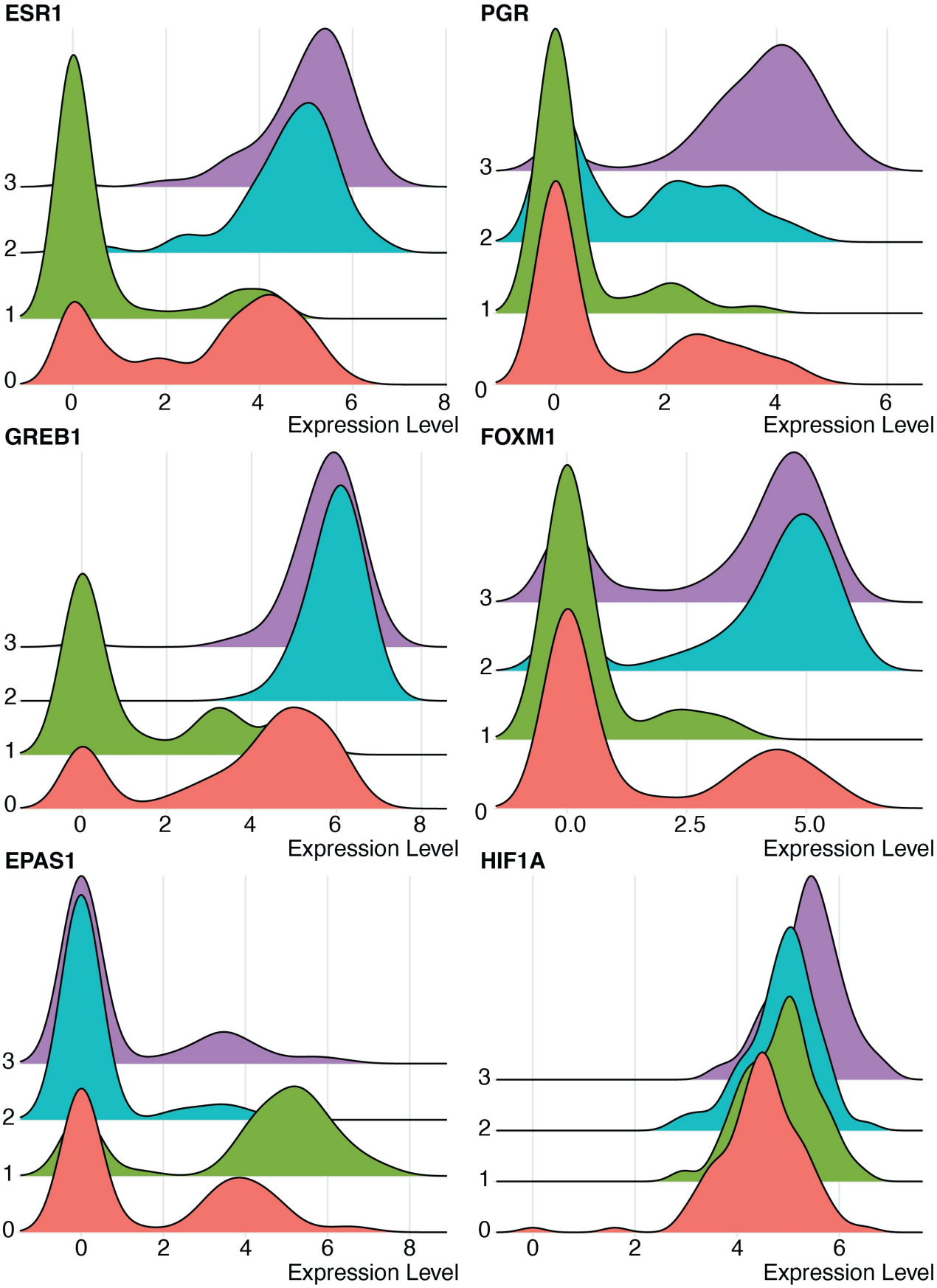


Figure 3-17: Expression Dynamics of ESR1 and Its Canonical Targets Across Cell Clusters - *This figure presents ridge plots depicting the expression levels of estrogen receptor (ESR1) and its canonical targets, including PGR, GREB1, and FOXM1, across different cell clusters identified in the study. Additionally, the plots show the expression of EPAS1 (HIF2A) and HIF1A to compare their expression patterns with those of ESR1 and its targets. Notably, ESR1 and its targets exhibit higher expression levels in cluster*

0 but significantly lower levels in cluster 1, closely mirroring the expression pattern of ESR1 itself. This observation suggests a coordinated regulation of ESR1 and its downstream targets within this hypoxic cluster. Intriguingly, EPAS1 expression is elevated in cluster 1, the same cluster where ESR1 and its targets are downregulated, suggesting a potential inverse relationship or regulatory interference between EPAS1 and ESR1 signaling pathways.

Intriguingly, in the same cluster where ESR1 levels were decreased, EPAS1 levels were found to be elevated **[Figure 3-17]**. This inverse relationship sparked the hypothesis that EPAS1 and ESR1 might be involved in a negative feedback loop. Such a loop could suggest that under extended hypoxic conditions, an increase in EPAS1 suppresses ESR1 expression and its downstream signaling pathways. This mechanism could potentially serve as a short-term tumor-suppressive response by reducing ER-associated proliferative signaling. However, paradoxically, this could promote long-term cancer progression by potentially enhancing resistance to therapies targeting these pathways.

To validate the observed patterns, signature scoring analysis was conducted using HIF2A (EPAS1) and ESR1 gene signatures derived from the comprehensive CollecTRI database (Müller-Dott et al., 2023), which compiles transcriptional regulatory interactions backed by experimental evidence; ***Seurat::AddModuleScore()*** was the signature scoring method of choice here. The results were visualized using UMAP and ridge plots, demonstrating high EPAS1 signature scores predominantly in hypoxic clusters, whereas ESR1 signaling appeared consistent across clusters **[Figure 3-18]**.

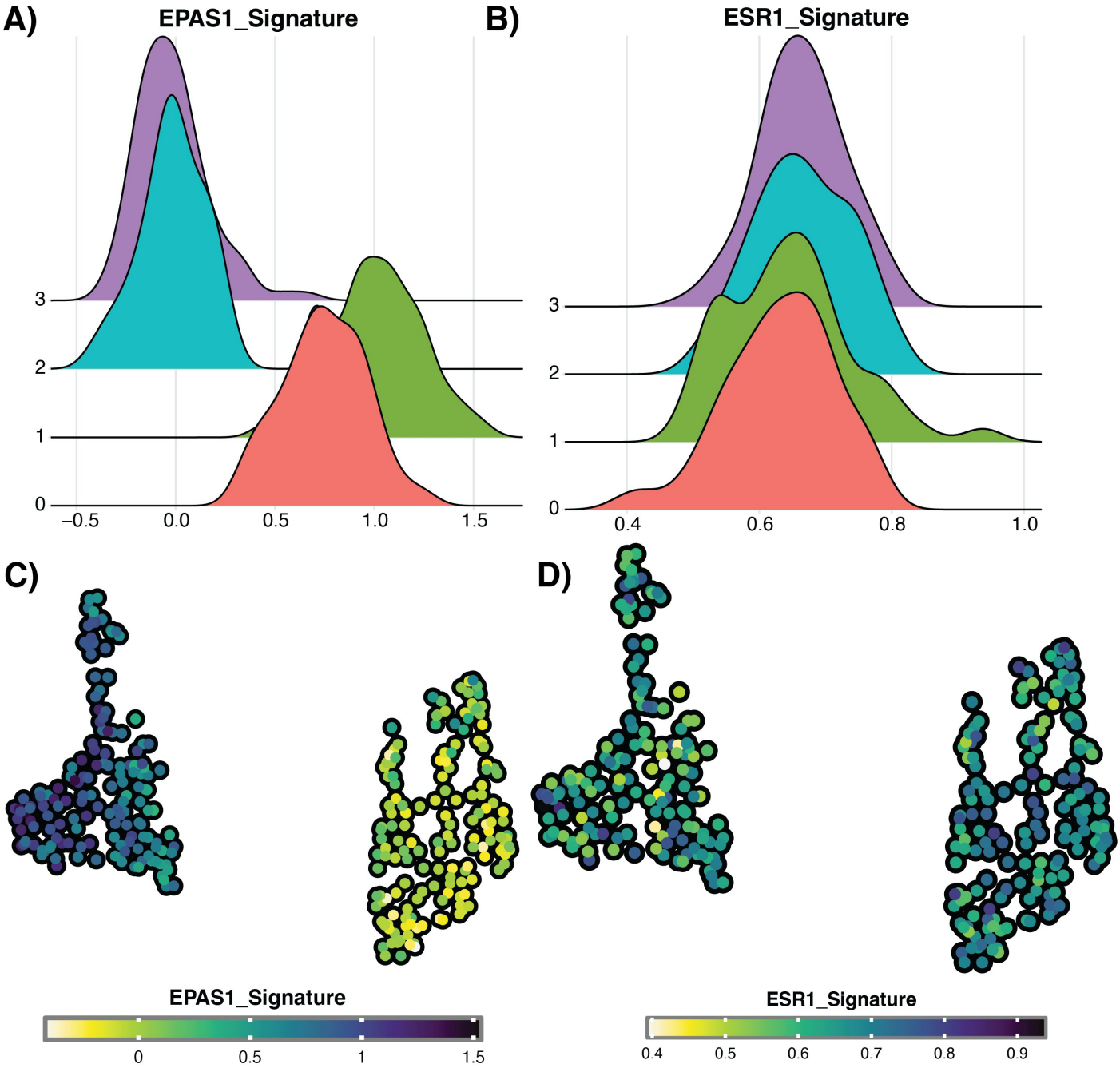


Figure 3-18: Signature Scoring for EPAS1 and ESR1 Across Hypoxic and Normoxic Clusters - *This figure comprises both UMAP and violin plots, illustrating the dynamics of EPAS1 and ESR1 signatures across clusters within the dataset. Panel A displays UMAP plots where the colour intensity indicates the strength of the EPAS1 and ESR1 signature score within each cluster. Notably, EPAS1 signaling is markedly elevated in hypoxic clusters, suggesting active hypoxia signaling in these regions. Conversely, ESR1 signaling remains relatively stable and uniform across all clusters, indicating a consistent level of estrogen receptor activity independent of hypoxic status. Panel B shows corresponding violin plots for these signatures. The plots for EPAS1 highlight its predominant expression in hypoxic conditions, mirroring the UMAP data, which suggests a robust activation of hypoxia-driven pathways in these clusters. For ESR1, the violin plots demonstrate a consistent expression across all conditions, reinforcing the UMAP findings and suggesting that ESR1's role is maintained across varying levels of oxygen availability.*

Given the potential for regulatory overlap and the influence of other factors such as HIF1A, further validation experiments are crucial. Currently, real-time quantitative PCR (rt-qPCR) and Western blot analyses are underway, focusing specifically on EPAS1 and ESR1-specific features. These experiments aim to provide a more robust understanding of the molecular interactions at play.

3.3.6. Machine Learning-Derived Hypoxia Signature Identifies Mitochondrial Genes as Key Predictors of the Hypoxic State In An ER+ Cancer Cell Line

Following the comprehensive analysis, a key area of interest emerged: the application of classification machine learning algorithms to this dataset. Traditionally, scRNA-Seq data have posed significant challenges for the use of classical machine learning models, particularly linear ones, due

to the inherent data sparsity. More complex models that exhibit high non-linearity, as well as deep neural networks, are technically feasible; however, these approaches generally prioritize high classification accuracy over interpretability, which is not particularly useful for this specific application.

In response to these challenges, a logistic regression model incorporating a lasso penalty—referred to here as a lasso classifier—was employed. The primary objective of this classifier was to refine the feature space effectively, selecting only those features that contribute most significantly to the binary classification of the hypoxic state. This method of feature selection is especially valuable in scenarios where reducing dimensionality without losing predictive power is crucial.

The process of training size selection, hyperparameter tuning, and model optimization was meticulously carried out using our specialized in-house machine learning suite, Renoir (Barberis et al., 2024). The result of this rigorous analytical approach was the identification of a set of key features, hereafter referred to as the 'lasso signature'. This signature comprises the features that the lasso classifier identified as having the highest impact on determining the hypoxic state of the cells.

The best performing model (> 99% accuracy) identified a refined signature of 3 genes: **MT-CYB [Figure 3-19A]**, **PGK1 [Figure 3-19B]**, and **SLC25A39 [Figure 3-19C]**.

PGK1 (Phosphoglycerate Kinase 1) is a well-characterized glycolytic enzyme regulated by HIF1A. It is also a member of the Buffa (Buffa et al., 2010a) and the hallmark (Liberzon et al., 2015) hypoxia signatures. As a HIF1A-regulated gene, PGK1 plays a pivotal role in glycolysis, facilitating the transition of 1,3-bisphosphoglycerate to 3-phosphoglycerate. This enzyme is crucial for cellular adaptation under low oxygen conditions, effectively enhancing glycolytic flux to compensate for diminished mitochondrial oxidative phosphorylation.

MT-CYB (Mitochondrial Cytochrome B), despite being a mitochondrial gene, is notably downregulated in hypoxia. Typically considered an artifact when detected in single-cell RNA sequencing data, MT-CYB in this dataset is distinctly associated with the hypoxic state. Recent literature supports that the downregulation of mitochondrial activity mediated by HIF1A is a defensive mechanism against oxidative stress-induced apoptosis, highlighting the adaptive response of cells to hypoxic stress (H.-S. Li et al., 2019).

SLC25A39 (Solute Carrier Family 25 Member 39), a mitochondrial membrane carrier, is instrumental in regulating glutathione levels within mitochondria to support oxidative phosphorylation. The observed downregulation of SLC25A39 suggests a strategic reduction in mitochondrial activity during hypoxia. This reduction is a survival strategy aimed at minimizing mitochondrial-derived reactive oxygen species and evading apoptosis under oxygen-limited conditions (M.-C. Kim et al., 2021).

The discovery that mitochondrial genes serve as powerful negative predictors of hypoxic states is particularly striking and suggests a need for a paradigm shift in single-cell analysis. Traditionally regarded as mere artifacts of mitochondrial permeation, these mitochondrial genes may represent critical sources of biological variation. This insight challenges the conventional interpretation and underscores the potential of these genes to provide valuable information about cellular responses and adaptations. Recognizing and harnessing this potential could significantly enhance our understanding of hypoxic signaling and adaptation. It is important to emphasize that this analysis is highly correlative rather than causal; establishing causality will require further experimental validation, such as targeted knockout or rescue studies. Recognizing and harnessing this potential could significantly enhance our understanding of hypoxic signaling and adaptation.

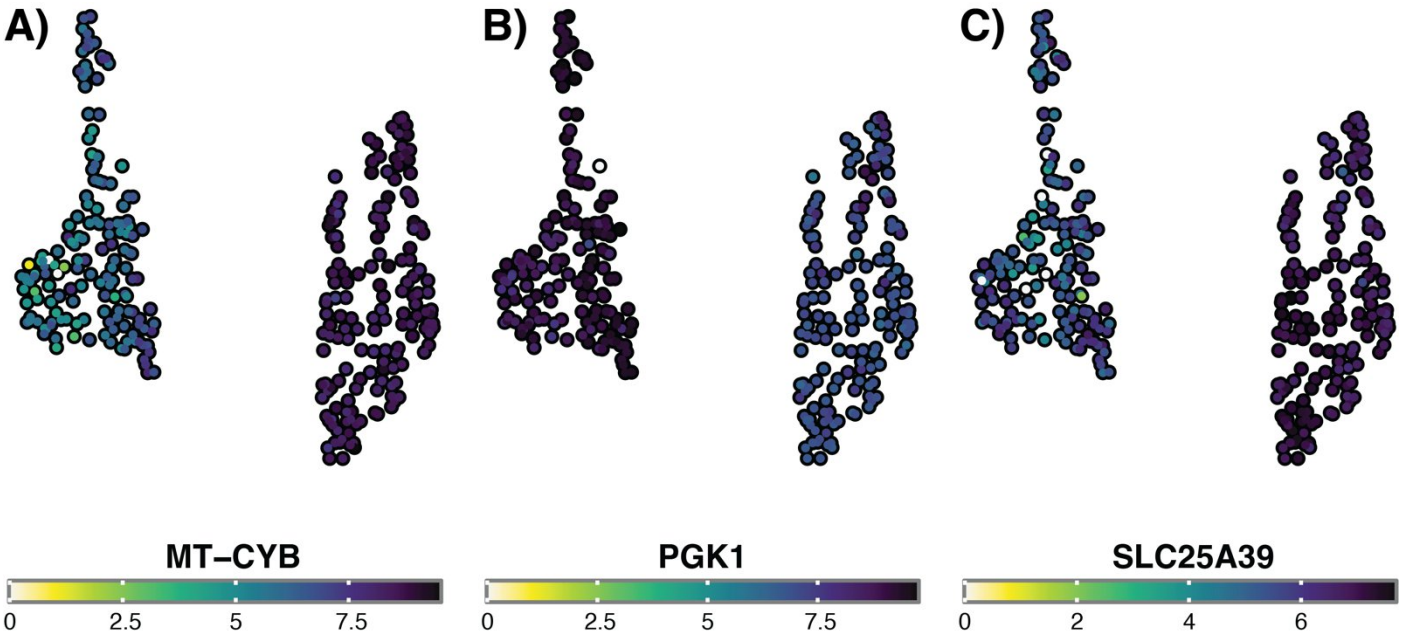


Figure 3-19: Expression UMAP Plots of Lasso Signature - A) MT-CYB: *This panel displays the expression of MT-CYB, a mitochondrial protein essential to oxidative metabolism. Notably, its expression is inversely correlated with hypoxia, identifying it as a negative predictor of this condition.* ***B) PGK1:*** *Depicts the expression of PGK1, a critical enzyme in glycolytic metabolism, which is important for understanding cellular energy processes in various conditions including hypoxia, though it does not serve as a negative predictor.* ***C) SLC25A39:*** *This panel shows the expression of SLC25A39, a mitochondrial carrier involved in glutathione transport. Similar to MT-CYB, SLC25A39 is inversely proportional to hypoxic conditions, thus acting as another negative predictor.*

3.4. Discussion

This study was designed to delve into the heterogeneity of hypoxia responses within the MCF7 breast cancer cell line, a model well-regarded for its relevance to both hypoxia and hormone response studies. Previous investigations had identified two distinct populations within cancer cell lines: one that induced carbonic anhydrase IX (CA9) and another that did not. This heterogeneity prompted further examination at the single-cell level to understand the implications on gene expression. Interestingly, while CA9 heterogeneity was evident, those cells with the highest levels of CA9 under hypoxia did not exhibit a broader range of hypoxia-induced pathways compared to CA9-negative cells, suggesting that CA9 heterogeneity might not extend to other hypoxia-regulated genes.

3.4.1. Hypoxia-Induced Heterogeneity and Cell Cycle Dynamics

The investigation delineated two principal clusters of genes activated by hypoxia: one primarily encompassing established metabolic genes and targets of hypoxia-inducible factors (HIFs), and another exhibiting more pronounced heterogeneity and enhanced fold induction. A notable alteration in cell cycle dynamics was observed under hypoxic conditions, with a significant transition from the S phase to the G1 phase. Despite these shifts, the activation of hypoxia-responsive genes appeared consistent across different cell cycle phases, indicating a disconnection between cell cycle changes and hypoxia-driven gene expression.

The identified gene clusters exemplify functional diversification within a population previously considered homogeneous in its response to hypoxia. It is hypothesized that these clusters collaborate to bolster tumor survival and adaptation. One cluster demonstrates high levels of EPAS1 and low levels of ESR1, correlating with a potential chemotherapeutic resistance phenotype. Conversely, the other cluster is markedly enriched for pathways related to muscle, skin, and vascular development, suggesting its involvement in modifying tissue architecture and influencing intercellular communication. This bifurcation underscores the complex, multifaceted nature of cellular responses to hypoxic stress within tumors.

3.4.2. Chronic Hypoxia and Clinical Implications

Most studies typically employ short durations of hypoxia (e.g., 24 hours); however, our findings suggest that extending this to 72 hours provides a more robust induction of CA9 and clearer delineation of hypoxic responses, especially for ER+ tumours. This has significant implications for understanding tumor physiology under chronic hypoxic conditions and could inform the development of treatments that more effectively target these adaptive responses.

3.4.3. Mitochondrial Metabolism and Patient Classification

The use of a logistic regression classifier with a lasso penalty in this study was pivotal for identifying minimal gene sets that can accurately distinguish between normoxic and hypoxic conditions.

Notably, two of the three key genes identified in the lasso signature are involved in mitochondrial metabolism. This classification underscores the substantial impact of hypoxia on mitochondrial metabolism, emphasizing how critical mitochondrial reprogramming is in response to hypoxic signaling.

3.4.4. Implications for Drug Resistance and Hormone Therapy

Finally, the distinct expression patterns of hypoxia-induced genes point toward potential differential pathways of adaptation among cells, which could influence drug resistance and hormone responsiveness.

While preliminary observations hinted at a possible inverse relationship between EPAS1 and ESR1, this negative correlation was not confirmed in our analysis. Nonetheless, the concept of hypoxic heterogeneity contributing to reduced efficacy of SERMs and SERDs remains clinically relevant for ER^+ tumors.

Ongoing work is focused on exploring the HIF2A–ESR1 axis in greater depth, with the goal of informing potential dual therapy strategies targeting both HIF2A and ESR1 to enhance SERM and SERD effectiveness in resistant cases.

3.4.5. Future Directions: Chromatin Accessibility and Cellular Signaling

To probe the underlying mechanisms of observed behaviours, it will be crucial to utilize assays like ATAC sequencing to assess chromatin accessibility in single cells. Additionally, identifying specific surface markers such as CEACAM5, CEACAM6, EDN2, LRP4, ROR1, and SEMA5B for these clusters could provide insights into their biological functions and response to treatments like fulvestrant.

4. Chapter 4: Temporal Dynamics of Hypoxic Responses: Unveiling Variations Over Time.

4.1. Introduction

Hypoxia, defined as oxygen levels below physiological norms, plays a critical role in the tumor microenvironment. This condition often arises as cancer cells proliferate rapidly and exceed their available supply of oxygen and nutrients. In response, these cells initiate a hypoxic response, primarily mediated through the stabilization and subsequent activation of Hypoxia-Induced Factors (HIFs) (Z. Chen et al., 2023).

Regulation of HIF1A levels in the cytoplasm is precisely controlled by the hydroxylation of its prolyl residues through prolyl hydroxylases (PHDs). This post-translational modification enhances the binding affinity between HIF1A and von Hippel-Lindau (VHL), an E3 ubiquitin ligase, leading to the poly-ubiquitination and proteasomal degradation of HIF1A. This oxygen-sensing system's efficiency relies on the oxygen-dependency of PHDs; a reduction in oxygen levels decreases PHD activity, allowing HIF1A stabilization, its dimerization with HIF1B, and subsequent nuclear translocation. Inside the nucleus, the HIF complex triggers transcriptional changes by binding to hypoxia-response elements (HREs) within the DNA, regulating genes essential for adaptation to hypoxic conditions (Masoud & Li, 2015).

Hypoxia significantly impacts cancer development and progression. In the short term, it induces cellular stress and leads to the formation of a necrotic core within tumours, with a gradient of hypoxia extending to the periphery of the blood supply. Over the long term, hypoxic tumours harness the hypoxic response to develop traits that promote angiogenesis, invasion, and metastasis, which are critical steps in the invasion-metastasis cascade of advanced carcinogenesis.

The progression of tumours under hypoxic conditions is facilitated through various mechanisms. While hypoxia does not directly cause DNA damage, it promotes strand breaks, hypermutation, and overall genomic instability through indirect mechanisms (Kaplan & Glazer, 2020). Genomic instability is a hallmark of cancer, enabling rapid mutation and selection of aggressive cancer phenotypes under adverse conditions such as hypoxia (Y. Yao & Dai, 2014). This forms the basis of the clonal expansion and evolution theory of tumor progression (Greaves & Maley, 2012). Another significant impact of hypoxia is the promotion of a cancer stem-cell phenotype (Heddleston et al., 2010). Hypoxic signaling induces cell cycle arrest, supports epithelial-mesenchymal transition (EMT), and facilitates the transition of cells into a senescence-associated secretory phenotype (Welford & Giaccia, 2011). This phenotype, characterized by high plasticity and strong therapeutic resistance, acts as a reservoir for the tumor and a primary cause of relapse in patients post-treatment.

Hypoxia presents in various forms: acute, cyclical, and chronic, each with distinct implications for cancer progression. Acute hypoxia, which involves transcriptomic changes occurring within 24 to 72 hours, has been extensively studied. Cyclical hypoxia, characterized by intermittent periods of low oxygen followed by re-oxygenation, is clinically relevant but difficult to simulate in vitro and complex to interpret. Chronic hypoxia, defined by sustained low oxygen levels extending beyond 72 hours, is poorly understood at both bulk and single-cell levels, primarily due to challenges in maintaining prolonged hypoxic conditions without compromising cell growth and viability. Despite these challenges, chronic hypoxia is the most clinically relevant, and its under-characterization represents a substantial gap in our understanding (Saxena & Jolly, 2019).

This chapter seeks to bridge this gap by exploring the heterogeneity of the hypoxic response over time, specifically between acute and chronic phases, within the MDA231 cell line, which is categorized under the MSL subtype of Triple-Negative Breast Cancer (TNBC). Originally, the study

aimed to characterize hypoxic responses across various TNBC subtypes over time. However, due to experimental delays caused by the COVID-19 pandemic, the scope of these studies was limited.

4.2. Methods

Experimental and computational methods were as described in chapter 2.

This experiment has a fully crossed experimental design **[Table 1]** with two factors: condition (Hypoxia, Normoxia) and timepoint (Day 1, Day 2, Day 7, Day 14). Each condition was systematically sampled across all timepoints to investigate the effects of hypoxia relative to normoxic controls.Top of Form

Bottom of Form

Table 1: Experimental Design for Hypoxia/Normoxia Time-Series Analysis

	Timepoint			
Condition	**Day 1**	**Day 2**	**Day 7**	**Day 14**
Hypoxia	Hypoxia Day 1 (HD1)	Hypoxia Day 2 (HD2)	Hypoxia Day 7 (HD7)	Hypoxia Day 14 (HD14)
Normoxia	Normoxia Day 1 (ND1)	Normoxia Day 2 (ND2)	Normoxia Day 7 (ND7)	Normoxia Day 14 (ND14)

4.3. Results

4.3.1. Clustering Analysis Identifies 6 Clusters Across Acute & Chronic Hypoxia

The histograms and combined scatterplots **[Figure 4-1]** provide a visualization of the quality control (QC) metrics for our samples, presenting library size, number of detected genes, and mitochondrial gene content. Consistent shapes in the histograms across all sample tags suggest uniform quality among the various groups, each representing distinct timepoints and hypoxic states.

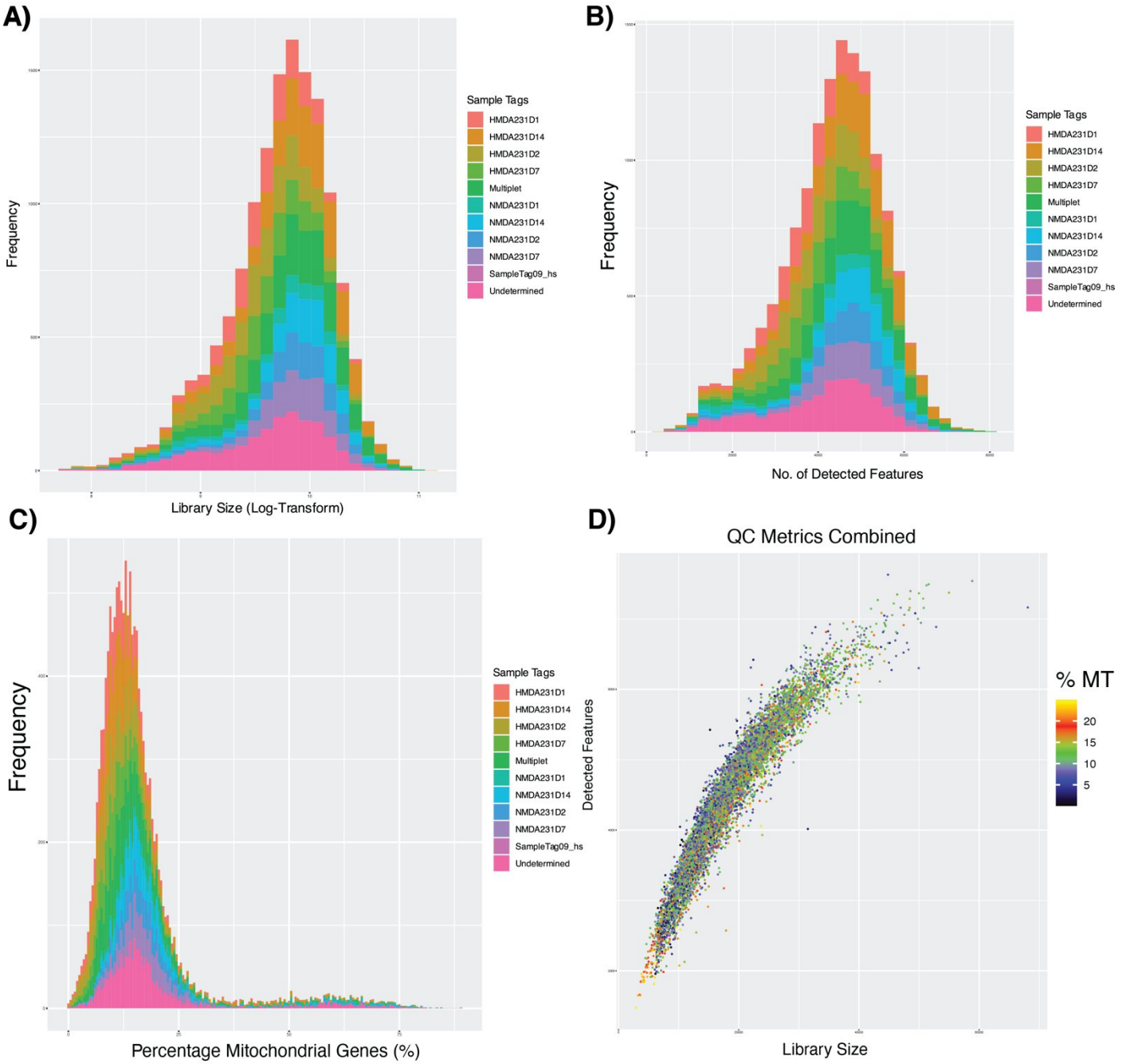


***Figure 4-1: Integrated Quality Control (QC) Metrics Analysis**. **A) Library Size Distribution:** Displays log-transformed library sizes across sample tags to highlight sequencing depth variation. **B) Unique Feature Counts:** Depicts the count of unique features detected for each sample tag, indicating gene expression profile richness. **C) Mitochondrial Gene Proportion:** Illustrates the percentage of mitochondrial genes per sample tag, with a threshold set at 25% to exclude stressed or dying cells. **D) Combined QC Scatter Plot:** This scatter plot merges all QC metrics, plotting detected features against library size and color-coding by the percentage of mitochondrial genes. The visualization confirms consistency across QC metrics, with data points grouped to indicate high-quality sequencing (dense central trend) and to flag potential outliers (isolated points with higher mitochondrial gene percentages or lower feature counts). Cells with 'undetermined' or 'multiplet' tags from the BD Rhapsody pipeline, suggestive of doublets or untagged cells, have been excluded to maintain data integrity.*

In our analysis, we did not set thresholds based on histograms for the majority of QC metrics. The BD Rhapsody platform has already implemented filtration criteria, as indicated by the tags 'multiplets' and 'undetermined' in the visualizations. 'Multiplets' denote potential doublets or multiplet events characterized by high library sizes and gene detection numbers, while 'undetermined' tags highlight cells that lack clear sample tag assignment and require exclusion.

An exception to this pre-set filtration is the application of an upper threshold for the percentage of mitochondrial genes, which we have determined to be 25%. This cutoff aims to exclude cells likely undergoing apoptosis. Although the mitochondrial gene content observed in the BD Rhapsody data appeared unusually high, particularly when contrasted with the MCF7 dataset discussed in Chapter

3, further review of comparable datasets from the literature confirmed that such distributions, while elevated, are not indicative of underlying issues with our data. Clustering of the cells was performed using the Louvain algorithm. The selection of an appropriate clustering resolution was informed by the clustering tree depicted **[Figure 4-2]**, leading to the adoption of a resolution parameter of 0.5 for the final clustering analysis. This resolution was chosen to allow for identification of unique biological variation whilst avoiding over-clustering and false biological variation.Top of FormBottom of Form

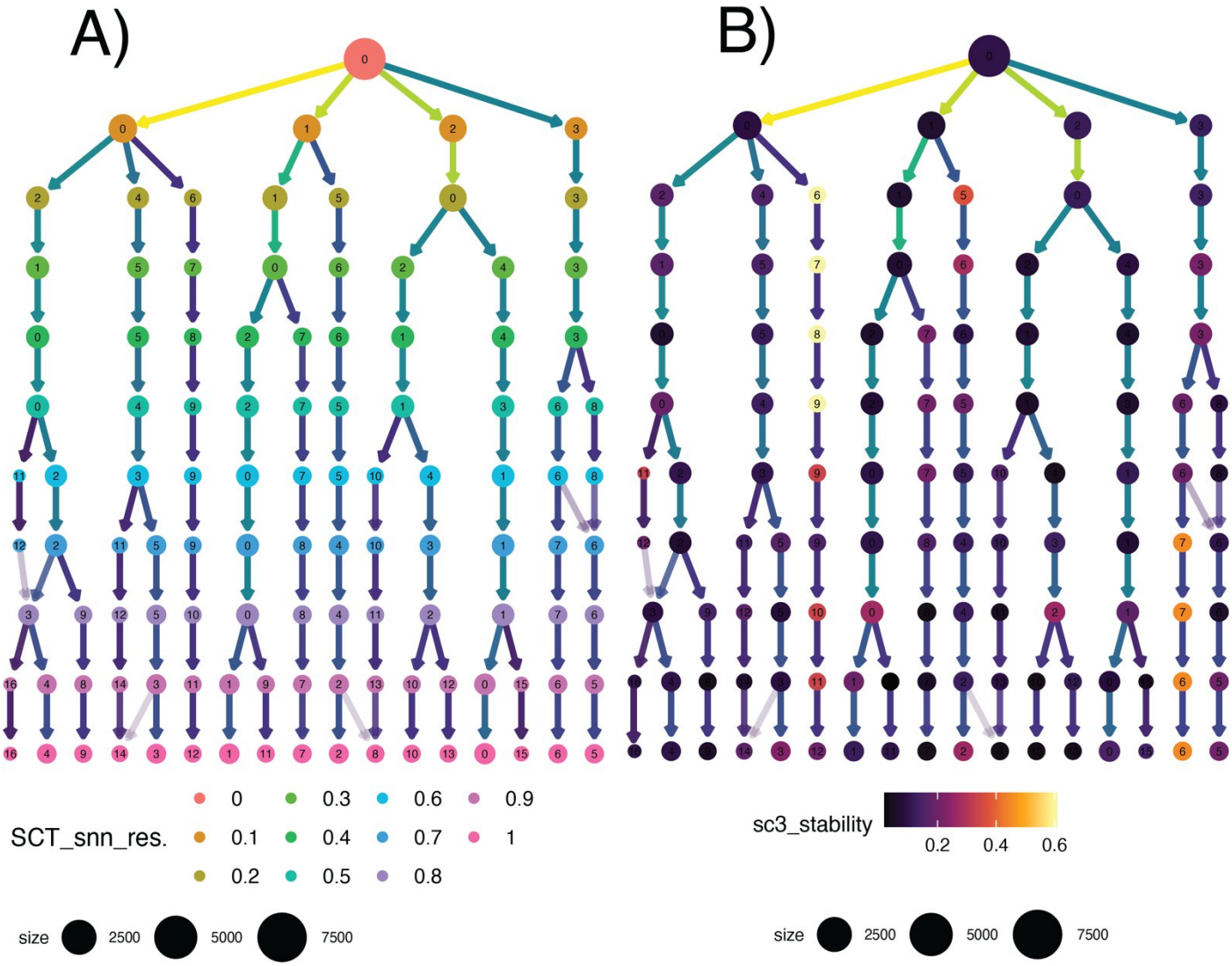


Figure 4-2: Clustering Trees from Louvain Algorithm Analysis - A) *Clustering Tree Coloured by Cluster Assignment: This dendrogram represents the hierarchical clustering structure determined by the Louvain algorithm, with the colour scheme reflecting the cluster identity assigned at various resolutions of shared nearest neighbor (SNN) modularity. The node size correlates with the number of cells in each cluster, and branch colors indicate different resolution parameters, highlighting how the granularity of clustering is influenced by the chosen resolution.* ***B)*** *Clustering Tree coloured by Cluster Stability: Here, the same dendrogram is coloured based on cluster stability scores derived from the SC3 (Single-cell Consensus Clustering) index, providing a visual assessment of cluster robustness. Warmer colors indicate higher stability, suggesting that clusters are consistent and reliable across different resolution parameters. The node size remains representative of cluster population, and the varying hues illustrate the differential stability of clusters, with the most stable clusters being potential targets for further characterization. The resolution parameter plays a crucial role in the Louvain algorithm's success, as it determines the depth of community detection within the data, affecting the sensitivity and specificity of the identified clusters. Clustree, a tool for visualizing clustering dendrograms across resolutions, provides insights into the optimal resolution setting, balancing between over-clustering and under-clustering, which is vital for accurate downstream analyses. Based on these figures, a final clustering resolution of 0.5 was chosen.****Top of Form***

After applying QC filters (including BD rhapsody built-in QC metrics and percentage mitochondrial genes), a total of 5,991 hypoxic cells and 3,243 normoxic cells were retained for further analysis. The breakdown of these cells across different conditions and timepoints was as follows: ND1 with 375 cells, ND2 with 848 cells, ND7 with 1,007 cells, ND14 with 1,013 cells for normoxic conditions; and HD1 with 1,365 cells, HD2 with 1,548 cells, HD7 with 1,252 cells, HD14 with 1,826 cells for hypoxic conditions **[Figure 4-3A]**.

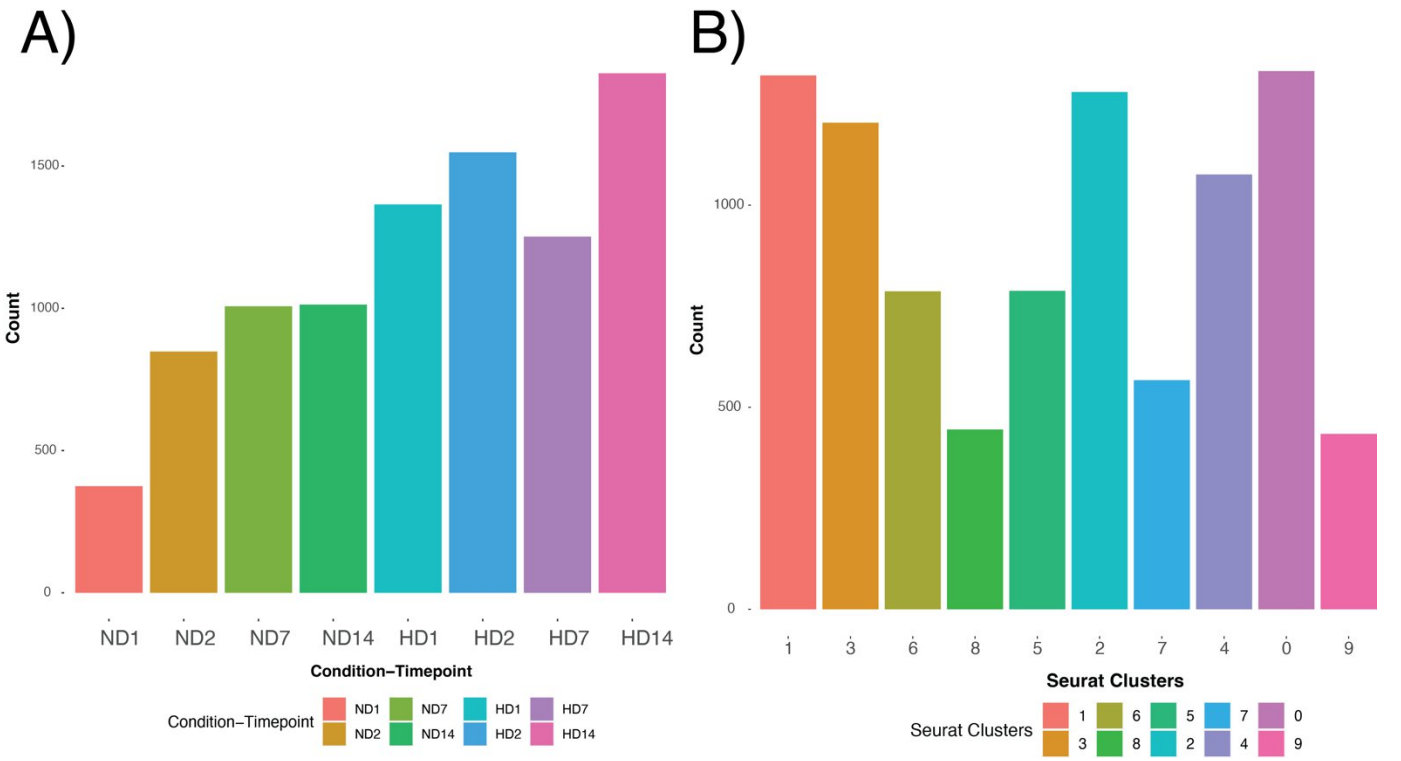
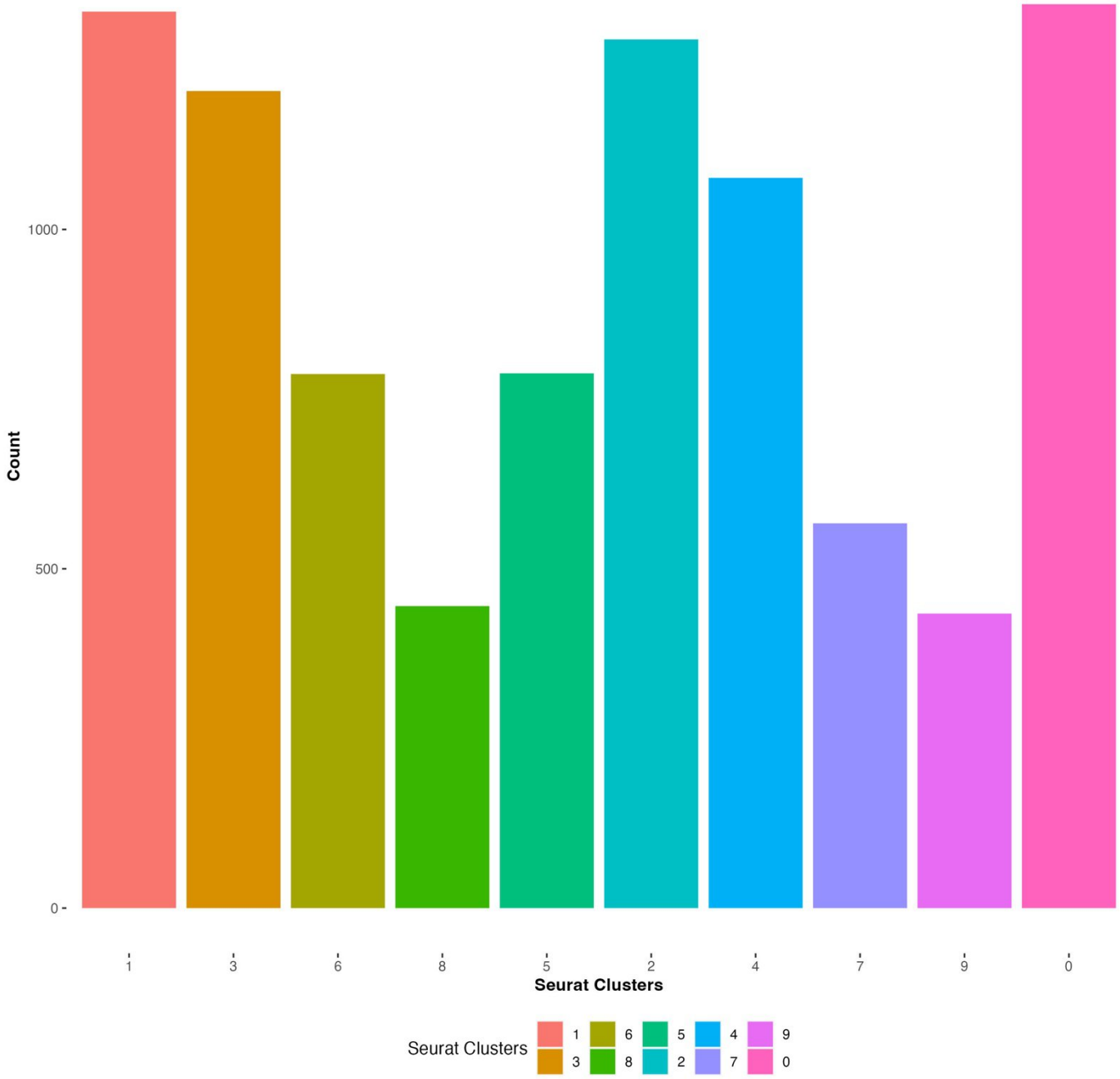
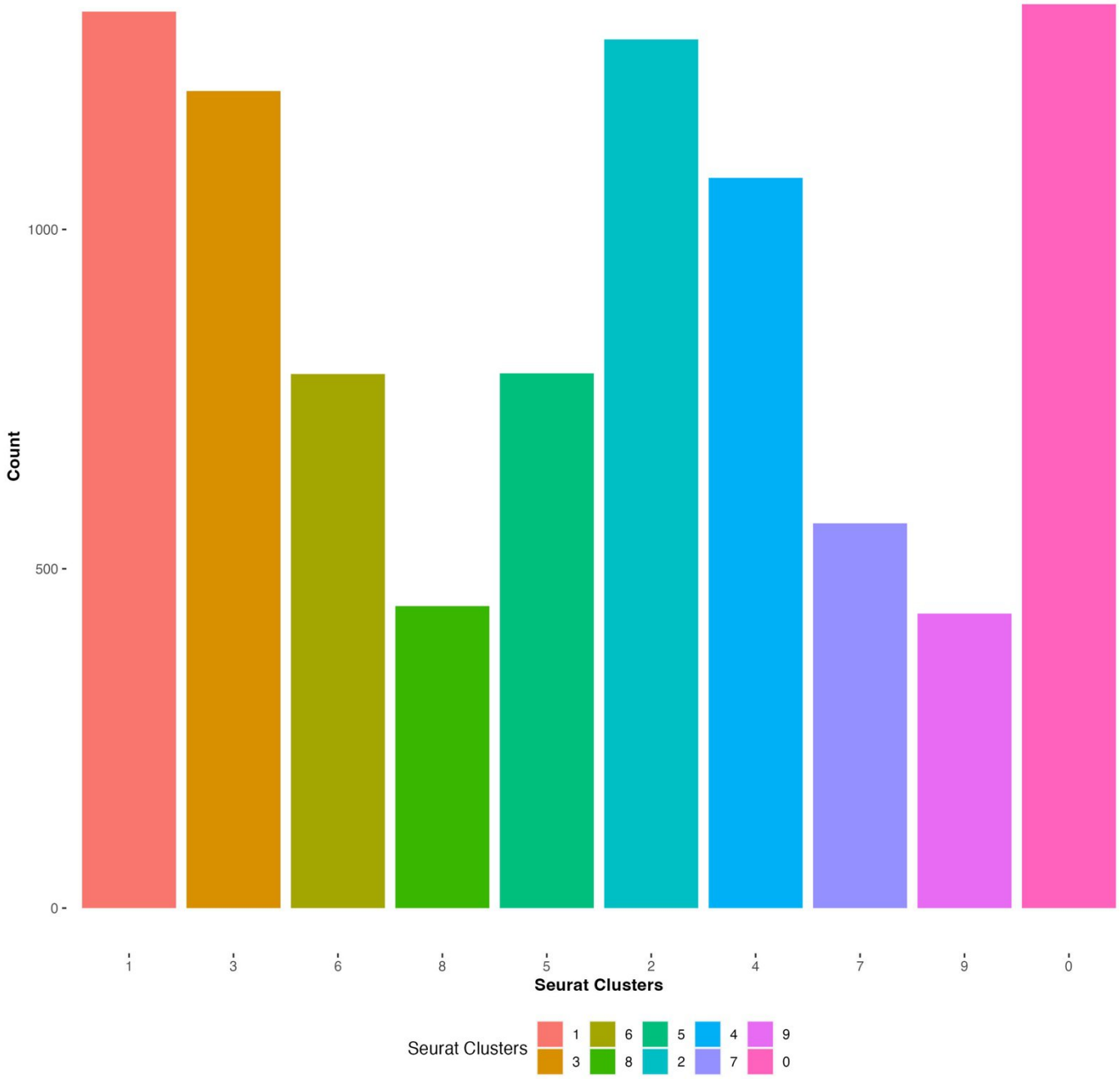


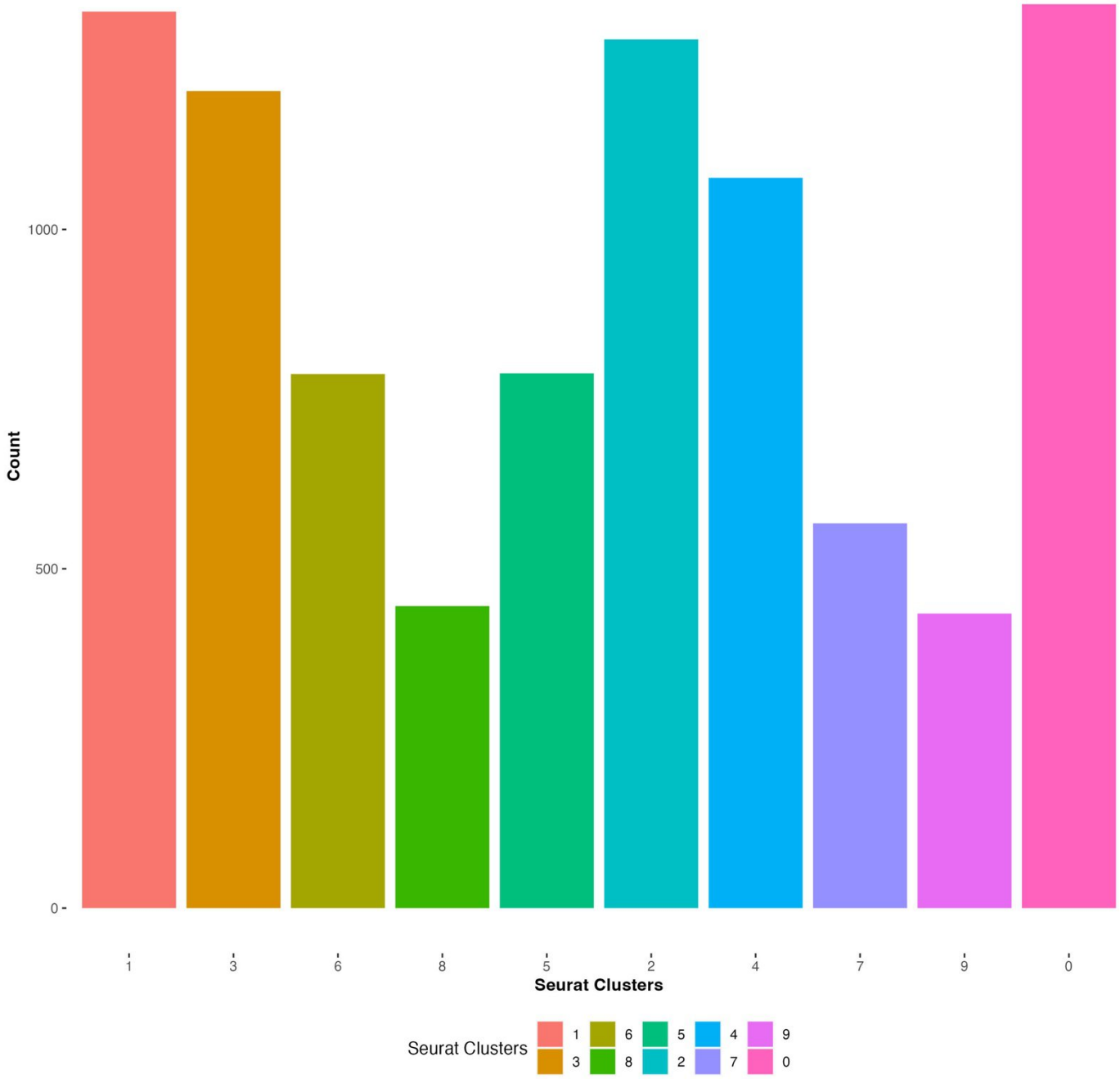
Figure 4-3: Count Plots for Cell Distribution Across Conditions, Timepoints, and Clusters - A) Distribution by Condition and Timepoint: *This plot displays the number of cells per each condition-timepoint combination, with labels such as ND1 for normoxia day 1, ND2 for normoxia day 2, and similarly for hypoxia days, showing a clear comparison of cell counts across different experimental groups.* ***B) Distribution by Seurat Clusters:*** *This graph illustrates the cell count distribution across different Seurat clusters. Clusters 1, 3, 6, and 8 predominantly consist of cells from normoxic conditions. Other clusters primarily represent hypoxic conditions, reflecting the dynamic response of cells to low oxygen environments.*

Subsequent analysis involved dimensionality reduction techniques such as PCA and UMAP to explore the data structure. The UMAP plots clearly segmented the hypoxic from the normoxic cells, each forming distinct groups **[Figure 4-4A]**. Specifically, while normoxic cells tended to cluster tightly without much variation across different timepoints, the hypoxic cells displayed a more dynamic arrangement. This semi-trajectory pattern observed in hypoxic cells—beginning with HD1 and extending through HD2, HD7, and concluding at HD14—suggests a progressive adaptation to hypoxic conditions over time **[Figure 4-4C]**.

The clustering analysis using the Louvain algorithm, with a chosen resolution of 0.5 (based on clustering tree analysis, see **Figure 4-2** for reference), identified 10 distinct clusters. This analysis revealed that Clusters 1, 3, 6, and 8 were primarily composed of normoxic cells. On the other hand, specific clusters were clearly associated with certain days within the hypoxic condition: Cluster 5 with HD1, Clusters 2 and 7 with HD7, and Clusters 0 and 9 with HD14, which are shown in the UMAP plots **[Figure 4-4D]**.


Count
1000
500
0
1
3
6
8
5
2
4
7
9
0
Seurat Clusters
Seurat Clusters
1
3
6
8
5
2
4
7
9
0


Count
1000
500
0
1
3
6
8
5
2
4
7
9
0
Seurat Clusters
Seurat Clusters
1
3
6
8
5
2
4
7
9
0



Bottom of Form

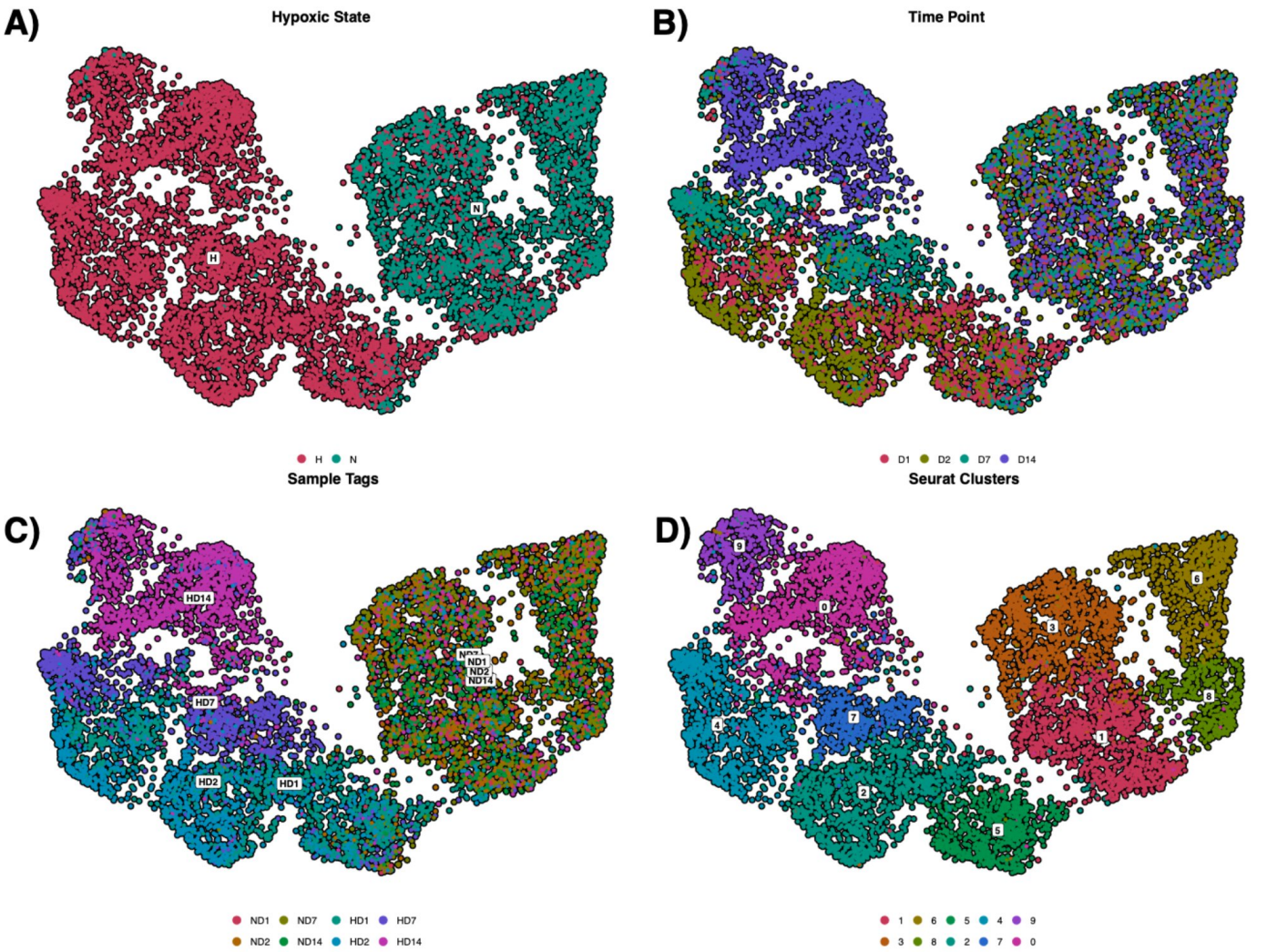


Figure 4-4: UMAP Plots of Data Structure - A) *Hypoxic State Annotation: This UMAP plot displays the distribution of cells annotated by hypoxic state, distinguishing between hypoxic and normoxic cells with distinct colors to indicate their respective locations within the data set.* ***B)*** *Timepoint Annotation: Cells are coloured according to their collection timepoints, illustrating the distribution of cells across different days, with each timepoint containing a mix of hypoxic and normoxic cells.* ***C)*** *Combined Hypoxic State and Timepoint Annotation: This visualization merges the hypoxic state and timepoint metadata, providing a more nuanced view of the dataset where the temporal progression of hypoxia is mapped alongside the oxygen conditions.* ***D)*** *Seurat Cluster Annotation: The final plot shows cells grouped into Seurat clusters determined by the Louvain clustering algorithm using an optimal resolution parameter of 0.5. The selection of this parameter was informed by analysis using clustree clustering trees. Clusters are visualized to show the segmentation of the data based on the combined influence of hypoxic state, timepoint, and inherent data structure identified through unsupervised clustering.*

4.3.2. Effect of Hypoxia on Cell Cycle Progression: Acute vs Chronic Adaptation

Hypoxia significantly impacts cancer cells by activating the HIF1A-dependent hypoxic response, leading to cell cycle arrest with cells predominantly remaining in the G1 phase. This response involves the HIF1A-mediated upregulation of p21 and p27, both of which are cyclin-dependent kinase (CDK) inhibitors (Druker et al., 2021). HIF1A influences these genes by partially displacing Myc from the promoters of p21 and p27, thereby enhancing their expression (Koshiji et al., 2004). Since Myc actively promotes cell cycle progression by naturally suppressing p21 and p27, its displacement allows these inhibitors to function, halting the cell cycle.

HIF1A also impedes cell cycle progression through a non-transcriptional mechanism that involves binding to cdc6. This interaction facilitates the nuclear localization of the MCM complex and inhibits the cdc7-dependent phosphorylation of the MCM complex. Consequently, this prevents the recruitment of Cdc45 and DNA polymerase A, effectively blocking replication fork firing and further stalling the cell cycle (Hubbi et al., 2013).

In normoxic conditions, cell cycle variation emerges as the predominant source of data variation, a conclusion supported by cell cycle regression analysis **[Figure 4-5A]**. This insight has guided the

decision to utilize only a single normoxic timepoint in future time series experiments, allowing for an increase in the sample size of hypoxic cells while reducing experimental costs.

In early hypoxia (day 1 and day 2), there is a notable predominance of cells in the G1 phase, suggesting cell cycle arrest **[Figure 4-5A]**. As the exposure extends to day 7 and day 14, the cells exhibit an adaptation to hypoxic stress, with a visible recovery of the cell cycle as more cells transition to the G2M and S phases. This shift during prolonged hypoxia is significant and is postulated to be driven by a 'HIF switch'—a transition from a HIF1A-dependent to a HIF2A-dependent hypoxic response (Slawski et al., 2024) **[Figure 4-5B]**. Although HIF1A and HIF2A target similar genes, HIF2A distinctly promotes cell cycle progression. In renal clear cell carcinomas (RCC), this is primarily seen in the induction of cyclin D1 expression through direct enhancer binding and its interaction with Myc (Gordan et al., 2007; Schödel et al., 2012).

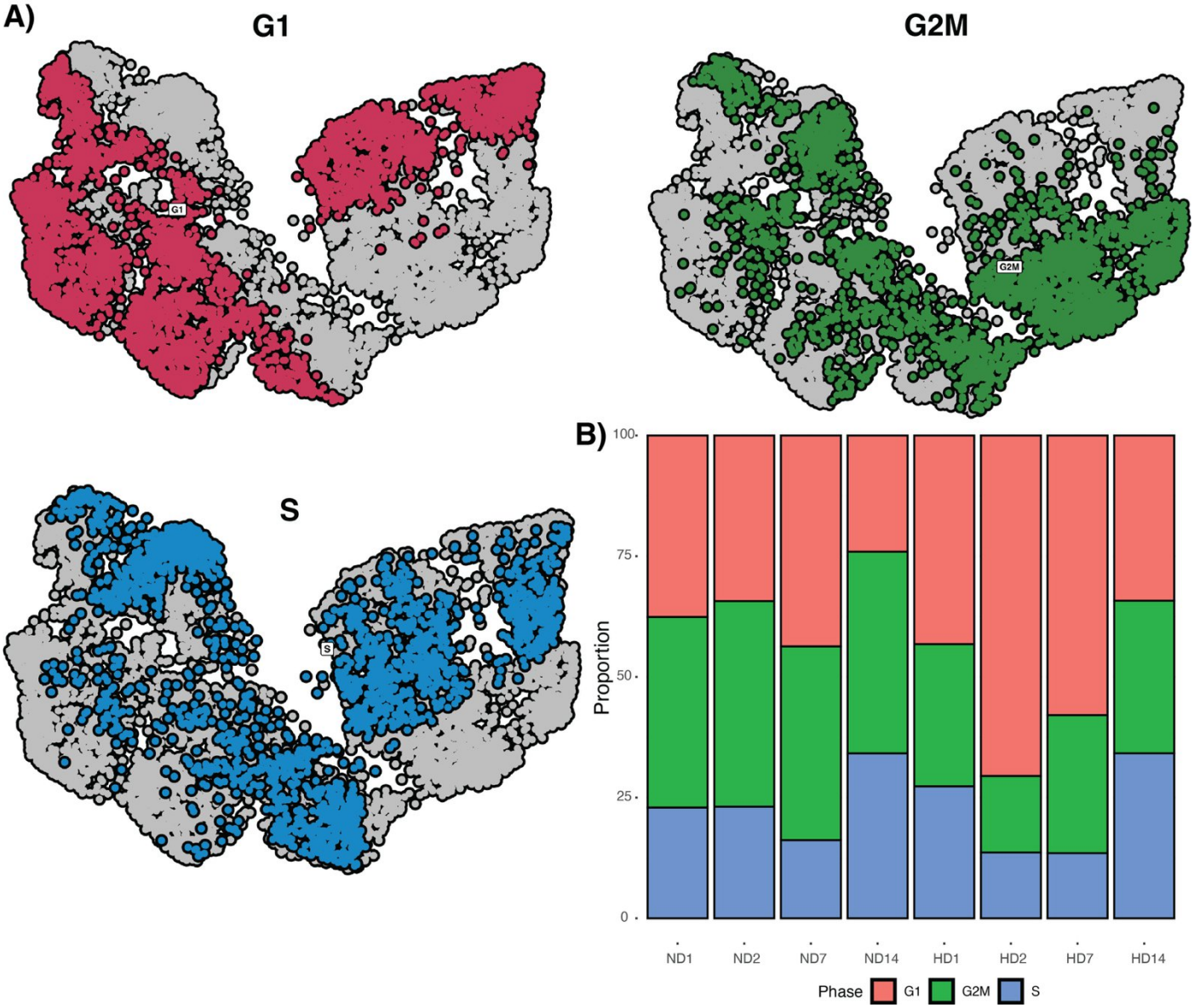


Figure 4-5: Cell Cycle Dynamics in Normoxic and Hypoxic Conditions Visualized by UMAP and Quantified by Barplots - A) UMAP Visualization of Cell Cycle Phases: *This panel illustrates the distribution of cells across the cell cycle phases—G1 (red), G2M (green), and S (blue). During normoxic conditions, variations in the cell cycle predominantly account for biological differences, supported by regression analysis detailed in supplementary figures. In contrast, acute hypoxia (HD1 and HD2) shows a pronounced shift towards the G1 phase, indicative of cell cycle arrest. However, in prolonged hypoxic conditions (HD7 and HD14), there is a noticeable normalization of the cell cycle distribution and prominence of the G2M and S phases, suggesting a significant cellular adaptation to extended low-oxygen environments.* ***B) Barplot Analysis of Cell Cycle Composition:*** *This barplot details the proportion of cells in each phase of the cell cycle across different conditions and timepoints. For cells in normoxia, the distribution across timepoints remains relatively consistent. Conversely, hypoxic cells demonstrate a marked increase in the G1 phase during the early stages of hypoxia (HD1 and HD2), with this trend becoming even more pronounced at HD2. Notably, by HD7 and HD14, there is a recovery to a more typical cell cycle profile and a reduction in G1 dominance, which underscores the cells' adaptation to hypoxic stress and their ability to resume normal cell cycle progression under chronic hypoxic conditions.*

Recognizing the pivotal changes in cell cycle dynamics between acute and chronic hypoxic conditions, further validation was essential. We examined the expression of crucial cell cycle progression markers and analysed signature scoring results for the Hallmark G2M checkpoint pathway.

The key markers evaluated included UBE2C **[Figure 4-6A]**, MKI67 **[Figure 4-6B]**, and CDK1 **[Figure 4-6C]**. MKI67, widely recognized as a marker for cell proliferation, is routinely used in both real-time quantitative PCR (rt-qPCR) and immunohistochemistry (X. Sun & Kaufman, 2018). Cyclin-dependent kinase 1 (CDK1) plays a critical role in cell cycle control, partnering with cyclins A and B to facilitate mitosis (Qian et al., 2015). Ubiquitin-conjugating enzyme E2 C (UBE2C) functions within the anaphase-promoting complex/cyclosome (APC/C), promoting the metaphase-anaphase transition by mediating the ubiquitination and subsequent degradation of target proteins. Notably, in oesophageal squamous cell carcinoma (ESCC), UBE2C is directly activated by FOXM1, enhancing its role in driving the cell cycle forward (Nicolau-Neto et al., 2018). Furthermore, both CDK1 and UBE2C have been highlighted as significant markers of cell cycle progression in a recent study that employed classification and feature ranking algorithms to identify key cell cycle predictors in scRNA-Seq (Huang et al., 2022).

The Hallmark G2M checkpoint pathway **[Figure 4-6D]**, a crucial gene signature from the Molecular Signature Database (MSigDb), epitomizes the efforts to streamline and unify gene sets that reflect specific biological states or processes. This pathway, along with other hallmark gene signatures, was curated to minimize redundancy and enhance the representational accuracy of biological processes in the database (Liberzon et al., 2015). Importation of these signatures was facilitated using the ***msigdbr*** package, and their scoring was accomplished through the ***Seurat::AddModuleScore()*** function.

The visual analyses of both key cell cycle markers and the Hallmark G2M checkpoint pathway present a consistent narrative **[Figure 4-6]**: cell cycle activity is notably suppressed during acute hypoxia phases but shows significant recovery in chronic hypoxia conditions. This pattern reaffirms observations from initial cell cycle scoring analyses, illustrating cancer cell adaptation to rising durations of hypoxic stress.Top of FormBottom of Form

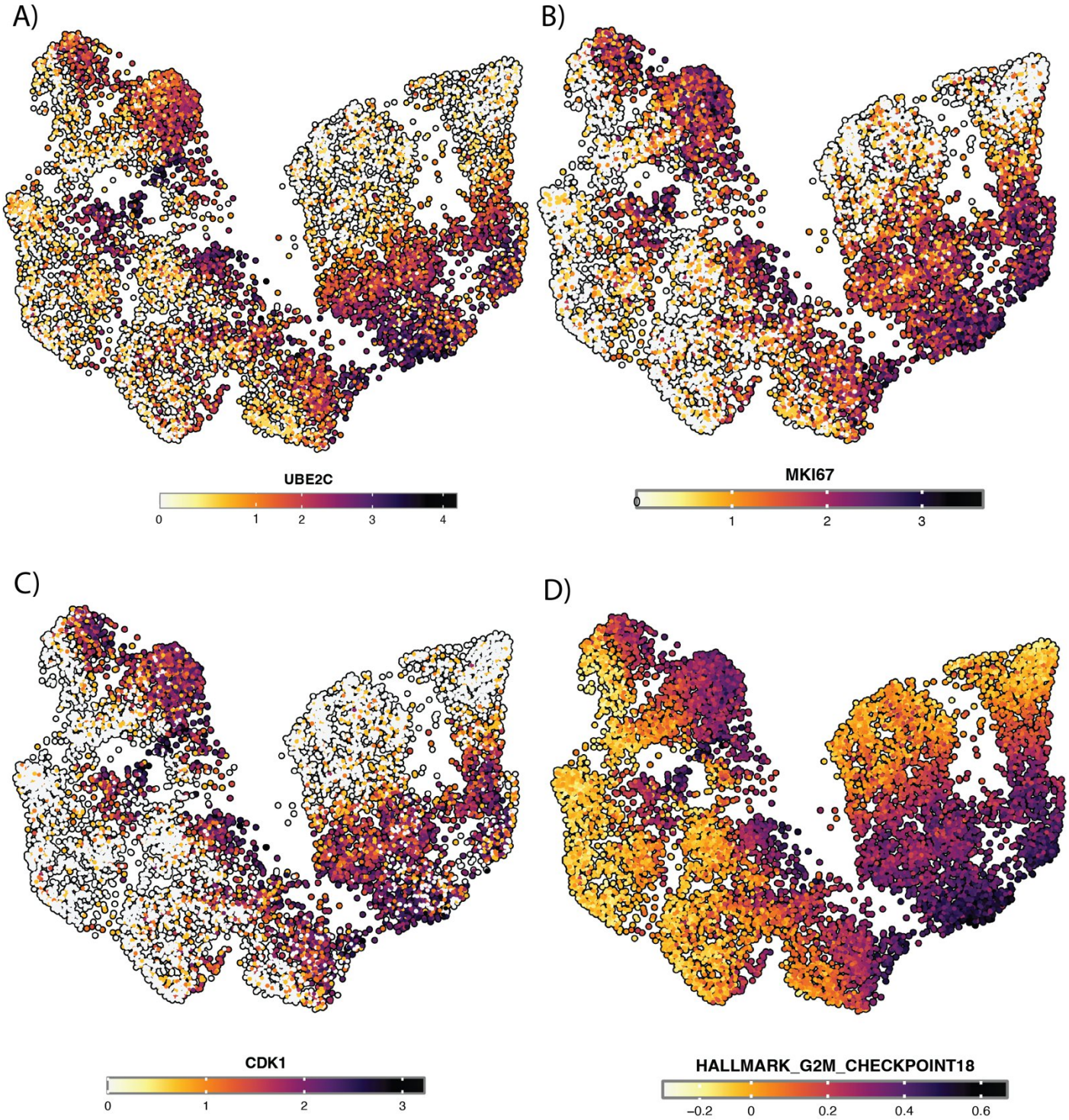


Figure 4-6: UMAP Feature Plots Highlighting Cell Cycle Dynamics and Hypoxic Response*: This figure displays UMAP feature plots illustrating the expression levels of key cell cycle markers and the Hallmark G2M Checkpoint pathway across different stages of hypoxia.* ***Top Panels (MKI67, CDK1, UBE2C)****: These plots reveal the expression patterns of MKI67, CDK1, and UBE2C. There is a notable predominance of cell cycle arrest markers during the acute hypoxic conditions of day 1 and day 2, followed by a significant recovery by day 14. This pattern underscores a cellular adaptation to prolonged hypoxia, potentially mediated by a HIF regulatory switch and Myc-dependent proliferation mechanisms.* ***Bottom Panel (Hallmark G2M Checkpoint Pathway)****: This plot, generated using the Seurat::AddModuleScore() method, visualizes the activity of the G2M Checkpoint pathway, echoing the observed trends in cell cycle dynamics through both acute and chronic hypoxic stages.Bottom of Form*

4.3.3. Continuous Activation of Hypoxic Response and Coordinated Upregulation of Adaptation Pathways Across Matched Hypoxic Timepoints

We focused on identifying key differentially expressed genes (DEGs) for each hypoxic timepoint using the ***Seurat::FindMarkers()*** function. This function utilizes the Wilcoxon signed-rank test to detect the most significant features for specific comparisons between groups, specifically contrasting hypoxic and normoxic cells at each timepoint.

We chose ***Seurat::FindMarkers()*** over ***Seurat::FindAllMarkers()*** because the latter compares a target group (e.g., HD1) against all other cells in the dataset, including both hypoxic and normoxic conditions. This approach can confound the results, as it does not provide a clean comparison between matched hypoxic and normoxic states. By using FindMarkers, we ensure that each hypoxic timepoint is directly compared with its normoxic counterpart, allowing for a more accurate and biologically relevant analysis.

The DEA results were visualized using the ***ComplexHeatmap*** package, as shown in **[Figure 4-7]**, highlighting the top 10 differentially expressed genes (DEGs) for each matched timepoint in hypoxia versus normoxia contrasts. Notably, the top DEGs at all timepoints include well-recognized hypoxia markers such as NDRG1, SLC2A3, BNIP3, and IGFBP3. These genes are prominent components of various hypoxia signatures, particularly the widely referenced Buffa hypoxia signature (Buffa et al., 2010b). The uniform appearance of these canonical markers across all contrasts underscores their consistent upregulation in response to hypoxia.

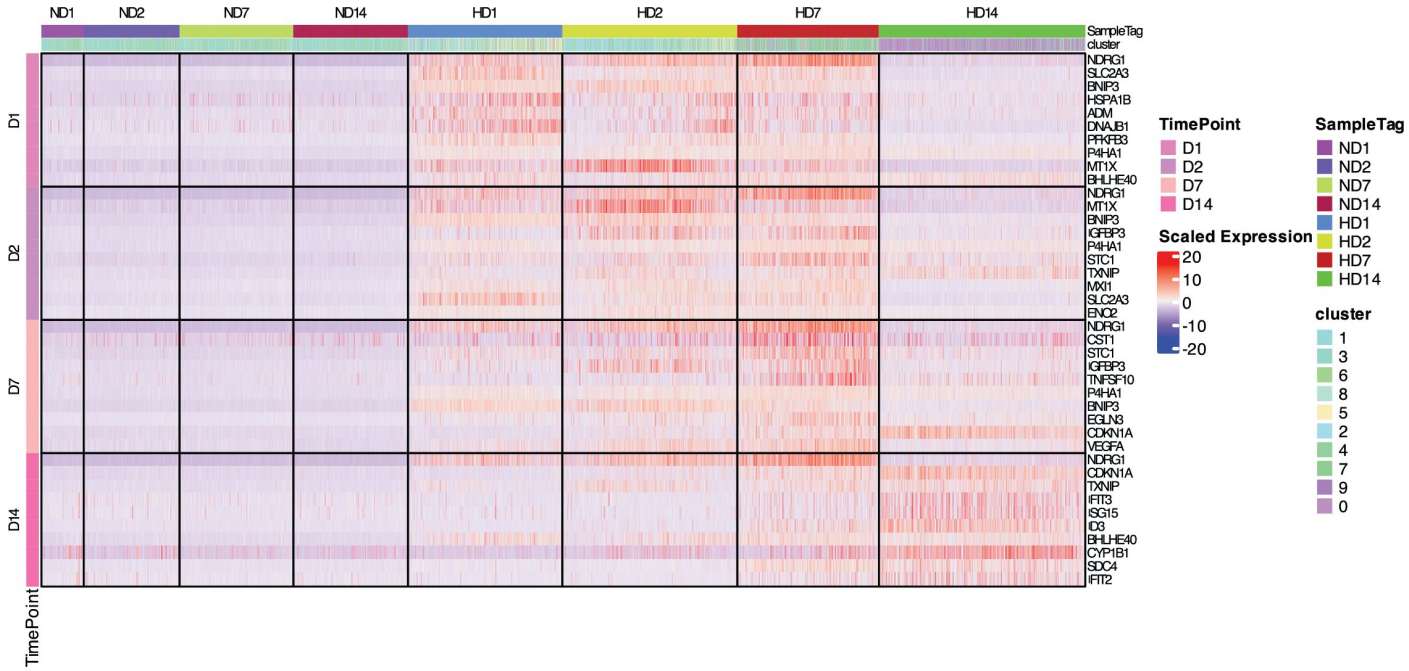


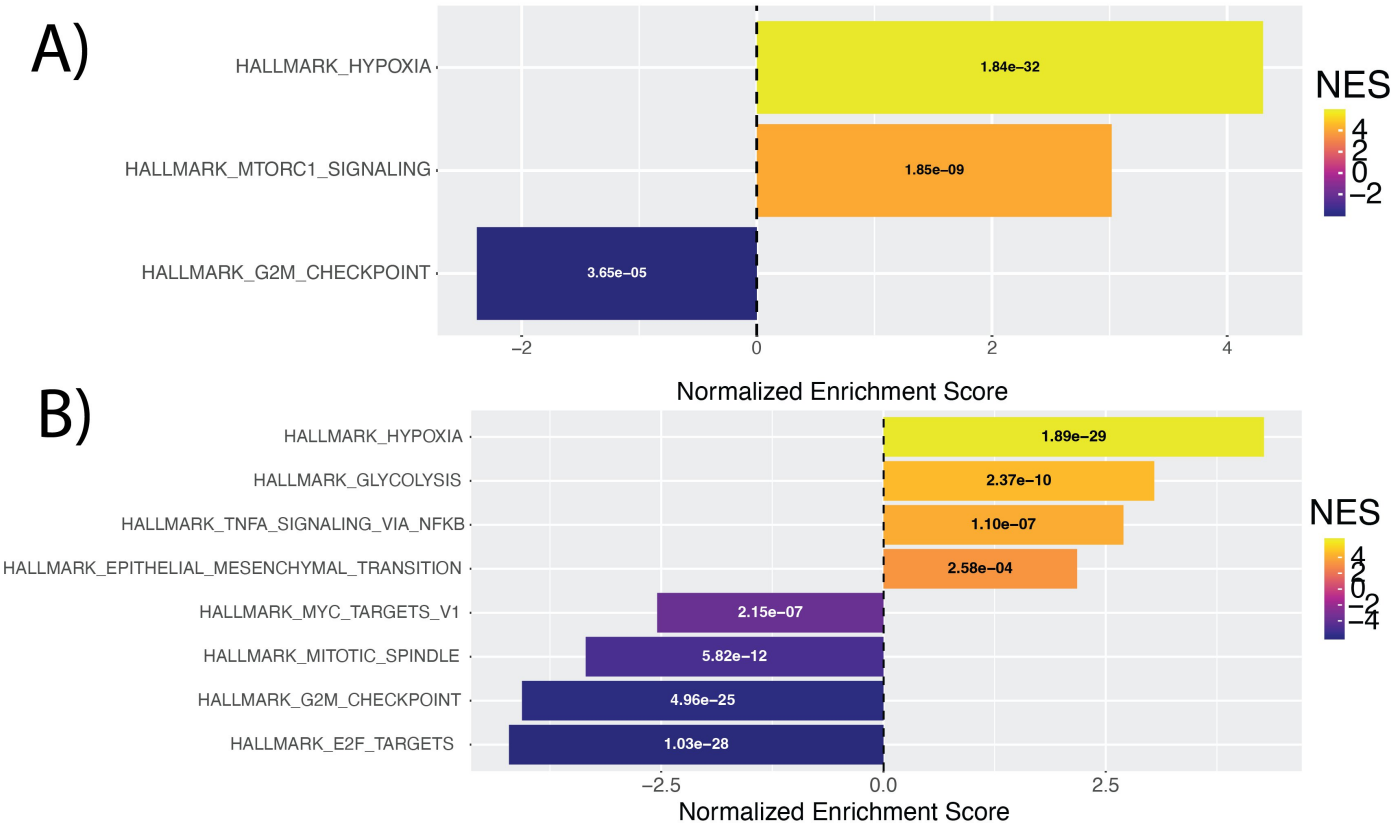
Figure 4-7: Differential Expression Analysis for Matched Hypoxic vs. Normoxic Timepoints: *This heatmap displays the top 10 differentially expressed genes (DEGs), ordered by average log2 fold change, across four matched timepoints of hypoxia (HD1, HD2, HD7, and HD14) compared to their respective normoxic conditions (ND1, ND2, ND7, and ND14). Each column in the heatmap is annotated with a SampleTag, indicating the specific combination of timepoint and condition, and cluster, indicating the Seurat clusters identified during the analysis. The row annotations specify the timepoints at which each DEG comparison was conducted. The selected top hypoxic markers are consistently lowly expressed in all normoxic conditions and show high expression across all corresponding hypoxic conditions, illustrating a clear differential response to hypoxia versus normoxia within the cellular populations.*

The analysis does not reveal unique features specific to chronic hypoxia adaptation, primarily because the comparisons are made between different hypoxic timepoints and their respective normoxic conditions, rather than among hypoxic states themselves. While this approach is excellent for confirming the overall hypoxic response's reliability, it limits the ability to pinpoint DEGs unique to prolonged hypoxic exposure. This methodological focus is optimal for validating the general response to hypoxia but less so for identifying distinct genetic expressions that emerge during sustained hypoxic conditions.

Following this, the results were utilized for pathway enrichment analysis conducted via Gene Set Enrichment Analysis (GSEA) with the hallmark gene sets. DEGs were ordered by average log2 fold change. The results of PEA are visualised using barplots **[Figure 4-8]**, and they form a cohesive picture with DEA analysis findings.

In the analysis comparing Day 1 hypoxia versus normoxia **[Figure 4-8A]**, the upregulation of the hallmark hypoxia pathway validates the initial cellular response to low oxygen levels, while the downregulation of the hallmark G2M checkpoint indicates the beginning of cell cycle arrest,

characteristic of acute hypoxia exposure. Moving to Day 2 **[Figure 4-8B]**, the persistence of upregulation in both the hallmark hypoxia and glycolysis pathways confirms a continued response to hypoxia, with the addition of an upregulated epithelial-mesenchymal transition (EMT) pathway suggesting the onset of more aggressive cellular behaviours due to extended hypoxic conditions. The downregulation of the hallmark G2M checkpoint and Myc targets further supports ongoing cell cycle arrest and reduced proliferative signaling. By Day 7 **[Figure 4-8C]**, the pattern remains consistent with continued upregulation of the EMT and hypoxia pathways, indicating sustained cellular adaptations to hypoxia. The hallmark Myc targets and G2M checkpoint continue to be suppressed, reflecting a prolonged inhibition of proliferation and cell cycle activity. By Day 14 **[Figure 4-8D]**, the analysis reveals new developments with the upregulation of TNFA signaling and interferon signaling pathways, pointing to the emergence of an immune-inflammatory response. The continuous suppression of Myc targets highlights a long-term inhibition of proliferative pathways as the cells adapt to chronic hypoxic conditions.



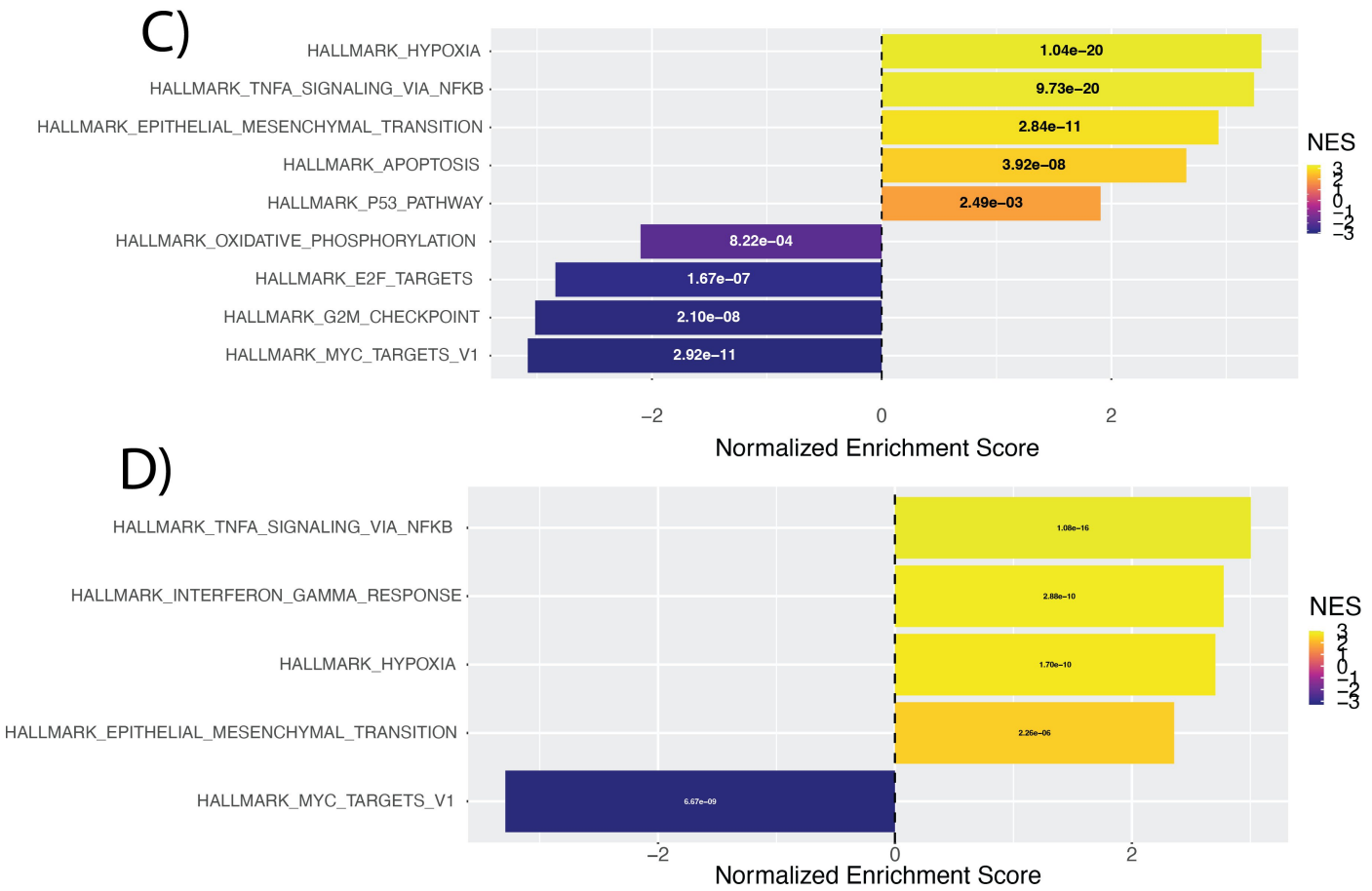


Figure 4-8: Pathway Enrichment Analysis (PEA) Results Using GSEA: *This figure illustrates the results of Pathway Enrichment Analysis conducted using Gene Set Enrichment Analysis (GSEA) with hallmark gene signatures as gene sets. The x-axis represents the normalized enrichment score, while the y-axis lists the pathways. Numbers inside the bars show the adjusted p-values, and only statistically significant pathways are included in this analysis.* ***A) Hypoxia vs. Normoxia Contrast for Day 1:*** *This panel shows upregulation of the hallmark hypoxia pathway and downregulation of the hallmark G2M checkpoint, reflecting the initial cellular response to hypoxic stress and associated cell cycle arrest, typical of acute hypoxia.* ***B) Hypoxia vs. Normoxia for Day 2:*** *Notable pathways include the continued upregulation of hallmark hypoxia and hallmark glycolysis, both characteristic of ongoing hypoxic response. Additionally, the hallmark epithelial-mesenchymal transition (EMT) is upregulated, suggesting the onset of more aggressive phenotypic changes. In contrast, hallmark G2M checkpoint and hallmark Myc targets are downregulated, indicating sustained cell cycle arrest and suppression of proliferative signaling.* ***C) Hypoxia vs. Normoxia for Day 7:*** *The upregulation of hallmark EMT and hallmark hypoxia persists, maintaining the patterns observed in earlier stages. Downregulation continues for hallmark Myc targets and G2M checkpoint, supporting the previous interpretations of prolonged hypoxic response and its effects.* ***D) Hypoxia vs. Normoxia for Day 14****: This panel shows upregulation in hallmark TNFA signaling and hallmark interferon signaling, pointing towards the development of an immune-inflammatory phenotype. Hallmark Myc targets remain downregulated, consistent with the ongoing suppression of proliferative pathways.*

These insights from the pathway enrichment analysis demonstrate a clear progression of cellular responses over the course of exposure to hypoxia, illustrating both immediate impacts and adaptations to prolonged hypoxic stress that are consistent with the DEA findings.

4.3.4. Analysing Temporal Hypoxic Adaptations in TNBC: Cluster-Specific Phenotypic Developments

The subsequent analysis focused on understanding the biological changes at different hypoxic timepoints relative to hypoxia day 1 (HD1), the first hypoxic timepoint in the dataset. The rationale for this approach is that contrasts against normoxia would show a strong canonical hypoxia signature, which is well-characterized. However, more subtle transcriptomic changes that facilitate adaptation under varying durations of hypoxia could be overshadowed in differential expression analysis (DEA) and gene set enrichment analysis (GSEA).

An important consideration in this analysis is the focus on clusters rather than hypoxic timepoints. Initial clustering analysis revealed that hypoxic populations at day 7 (clusters 7 and 4) and day 9 (clusters 9 and 0) exhibit intrinsic heterogeneity. Analysing timepoints as a whole would fail to

characterize this heterogeneity and could compromise the overall analysis if the clusters are substantially different. In such cases, the detected signal would be an average of two heterogeneous distributions.

DEA, pathway enrichment analysis (PEA), and visualizations were performed using ***Seurat::FindMarkers()***, **GSEA**, and **ComplexHeatmap**. These methods have been briefly mentioned above and are covered in extensive detail in Chapter 2.

The results of the DEA analysis in the heatmap below **[Figure 4-9]** reveal a wealth of interesting biological insights. The DEA markers shown in this heatmap are the top 10 markers for each contrast. These DEA markers exhibit very low levels of expression in normoxic cells as well as in cluster 5 (which was expected since cluster 5 was used as the baseline).

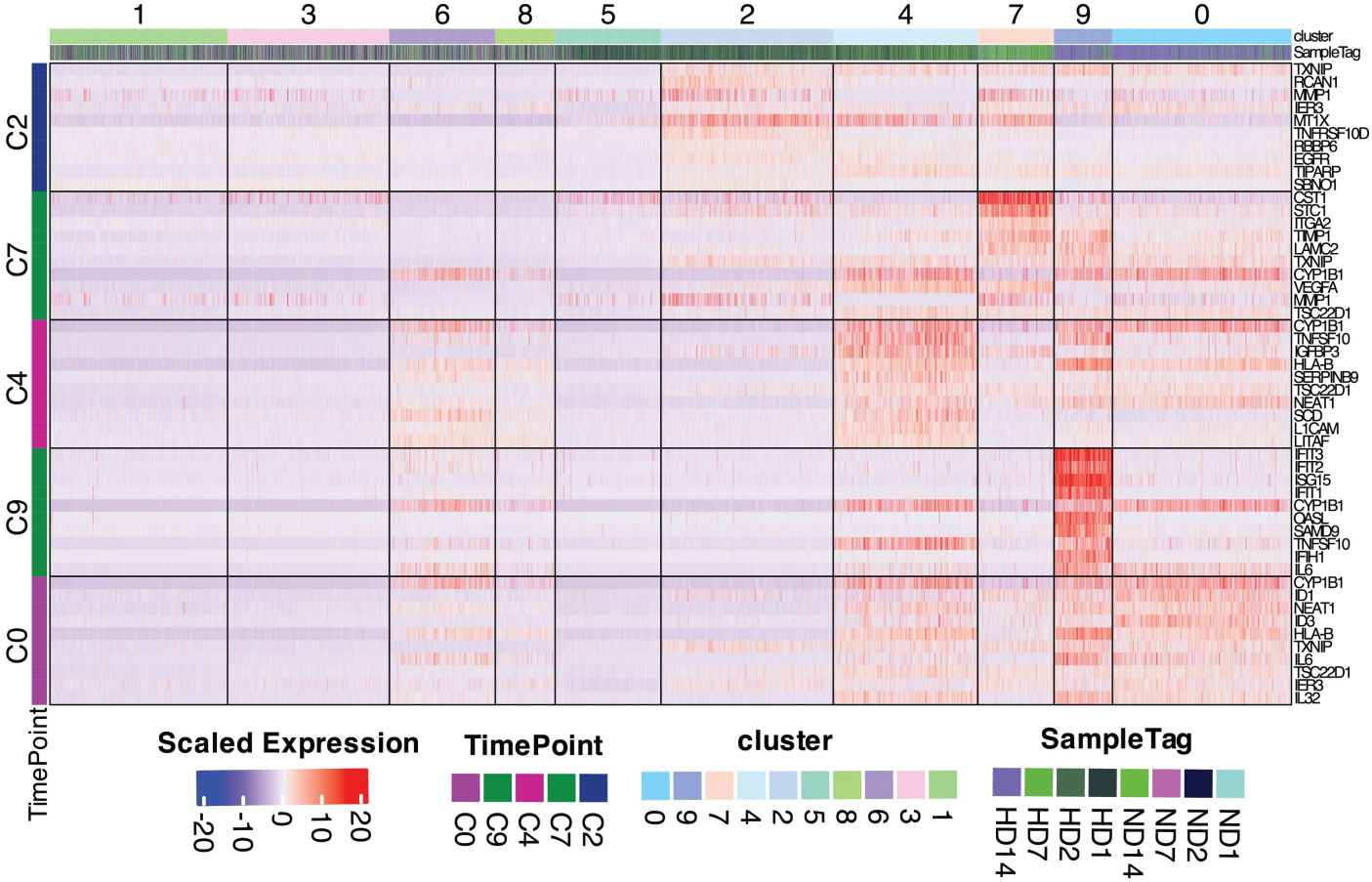


Figure 4-9: Top Markers of Hypoxic Clusters *– This heatmap shows the expression of marker genes for the different clusters. Rows represent the marker genes, with row annotations indicating the specific contrasts to which these genes belong (e.g., C0 and C7 represent the contrasts cluster 0 vs. cluster 5 and cluster 7 vs. cluster 5, respectively). Columns represent individual cells in the dataset, with column annotations displaying the clusters and sample tags to which these cells belong. Hypoxic clusters 2, 7, 4, 9, and 0 were contrasted against cluster 5 to identify key markers that enable adaptation to prolonged durations of hypoxia. The markers identified are distinct from the canonical HIF1A targets, which are highly dominant in cluster 5 (corresponding to hypoxia day 1). This analysis highlights the unique transcriptomic adaptations of these clusters to extended hypoxic conditions, providing insights into the mechanisms underlying cellular responses to chronic hypoxia.*

In cluster 2, primarily composed of cells from day 2 of hypoxia, key markers include Matrix Metallopeptidase 1 (MMP1), Regulator of Calcineurin 1 (RCAN1), and Epidermal Growth Factor Receptor (EGFR). MMP1 plays a crucial role in degrading the extracellular matrix (ECM), promoting invasion and metastasis by enlarging the physical space for cancer cell proliferation and facilitating angiogenic signaling essential for tumor revascularization and extravasation, critical steps in the invasion-metastasis cascade (**Eslami-S et al., 2020; Reunanen & Kähäri, 2013**) (Eslami-S et al., 2020; Reunanen & Kähäri, 2013). Numerous studies have highlighted MMP1's significant role in tumorigenesis, invasion, and metastasis in TNBC (**H. Liu et al., 2012; Q.-M. Wang et al., 2019**) {Citation}.

RCAN1, shown in **[Figure 7-8]**, acts as a direct regulator of calcineurin (CN) signaling by binding competitively to CN's catalytic domain. CN, a serine/threonine protein phosphatase, is heavily influenced by intracellular calcium levels and serves as a key effector in calcium signaling pathways (Lao et al., 2022). Activating CN leads to the dephosphorylation of the nuclear factor of activated T cells (NFAT), facilitating its nuclear translocation and the subsequent activation of genes involved in tumor proliferation and angiogenesis, such as VEGFA (Lao et al., 2021) and IGF1 (H. Jin et al., 2017), which are crucial for tumor growth and angiogenic development (Masaki & Shimada, 2022). Additionally, NFAT upregulates RCAN1, which acts as an indirect suppressor in a feedback loop.

The overexpression of RCAN1 is likely a consequence of hyperactive CN-NFAT signaling, reducing pathway activity and acting in a tumor-suppressing capacity. However, it also aids tumor survival by protecting against hypoxia-induced apoptosis through the inhibition of cytochrome C release and the activation of caspases 9 and 3 in the intrinsic apoptotic pathway (Duan et al., 2015).

EGFR, a receptor tyrosine kinase, is activated by EGF binding, resulting in dimerization, auto-phosphorylation, and subsequent activation of kinase domains. EGFR signaling, as depicted in [**Figure 7-9**], drives the AKT-PI3K and NF-kB pathways, which are significantly pro-tumorigenic (Uribe et al., 2021). AKT-PI3K promotes cell survival and proliferation through MDM2-mediated inhibition of p53 (Abraham & O'Neill, 2014), Rac-mediated reorganization of the actin cytoskeleton to facilitate migration and invasion (Lien et al., 2017), and inhibition of apoptotic signaling by phosphorylating and inhibiting pro-apoptotic BCL-2 family proteins (Bondar et al., 2002).

NF-kB signaling enhances the inflammatory environment by promoting pro-inflammatory cytokines like IL-6 and TNF-α, which drive ECM degradation and cancer invasion (H. Zhao et al., 2021), and upregulates anti-apoptotic BCL-2 family proteins like Bcl-2, A1/Bfl-1, and Bcl-xl, ensuring cell survival under hypoxic conditions (Dolcet et al., 2005, p. 200; C.-Y. Wang et al., 1999).

In cluster 7, primarily composed of cells from hypoxia day 7, key markers include Vascular Endothelial Growth Factor A (VEGFA), TIMP Metallopeptidase Inhibitor 1 (TIMP1), and Integrin Alpha 2 (ITGA2). VEGFA, a growth factor, acts as a primary driver of angiogenic signaling, crucial for tumor re-vascularization to support long-term survival, proliferation, and extravasation—a vital phase in the invasion-metastasis cascade. It is notably one of the targets of HIF1A and a significant component of the hallmark (Liberzon et al., 2015) and buffa (Buffa et al., 2010a) hypoxia signatures. **[Figure 7-9]** highlights VEGFA signaling, which initiates with VEGFA binding to VEGFR, leading to receptor dimerization, activation of its intrinsic tyrosine kinase activity, and autophosphorylation. These phosphorylated residues attract effector proteins that recognize these motifs, ultimately activating PI3K-AKT, Raf-MEK-MAPK, and ERK1/2-FAK signaling axes, fostering endothelial cell proliferation and migration essential for neovascularization (S. Qi et al., 2022).

TIMP1, typically an inhibitor of matrix metalloproteinases (MMPs), also drives cell proliferation by enhancing the PI3K/AKT and ERK/FAK signaling pathways (Forte et al., 2016). A pivotal study on colon cancer, using patient-derived organoids to compare prognoses of right-sided versus left-sided colorectal carcinomas, identified TIMP1 as a crucial factor. Higher TIMP1 levels correlated with increased phosphorylation of pFAK, pAKT, and pBad, and reducing TIMP1 led to lower phosphorylation levels and diminished cell invasion in right-sided PDOs (B. Ma et al., 2022). While TIMP1 does not bind to a known receptor tyrosine kinase, like EGFR or VEGFR, its action could involve c-Src, a non-receptor tyrosine kinase. c-Src activation, suggested by similarities with TIMP2, triggers a signaling cascade similar to that activated by RTKs, involving PI3K/AKT and ERK/FAK pathways (H. I. Kim et al., 2015, p. 201). This proposed action, though speculative for TIMP1, offers a plausible mechanism given their structural and functional similarities. Further studies are required to clarify TIMP1's exact role in these signaling pathways.

In cluster 4, primarily composed of cells experiencing hypoxia on day 7, the dominant features are L1 Cell Adhesion Molecule (L1CAM) and Tumor Necrosis Factor Superfamily Member 10 (TNFSF10). TNFSF10 is a member of the TNF cytokine family and exhibits both tumor-suppressive

and pro-tumorigenic roles. As a tumor suppressor, TNFSF10, also known as TRAIL, activates apoptosis by binding to its death receptors, TRAIL-R1 and TRAIL-R2, on cancer cells. This interaction triggers the extrinsic apoptotic pathway, leading to cell death, a process detailed in the supplementary materials and visualized in **[Figure 7-12]** (Thorburn, 2007). Conversely, TNFSF10 can promote tumor cell survival through autophagy, a survival mechanism that degrades internal organelles to sustain metabolic demands under stress. This less-characterized role involves potential interaction with TRAF2 and activation of MAPK8/JNK, which regulates autophagy (W. He et al., 2012; Yamamoto & Iwakuma, 2017).

L1CAM, a membrane glycoprotein, significantly influences cancer progression and metastasis. It is linked with the upregulation of epithelial-mesenchymal transition (EMT) markers like vimentin, suggesting its role in enhancing metastatic capabilities (Schrevel et al., 2017). Furthermore, L1CAM supports the cancer stem cell (CSC) phenotype, with studies showing that its inhibition can disrupt these cells' properties, underscoring its importance in maintaining CSC characteristics (Giordano & Cavallaro, 2020). The dual functionality of TNFSF10, alongside the pivotal role of L1CAM in cancer biology, illustrates the complex interplay of molecules that can drive or inhibit tumor progression based on cellular context.

In cluster 9, primarily consisting of cells subjected to hypoxia for 14 days, the key markers identified include Interferon Induced Protein with Tetratricopeptide Repeats 2 (IFIT2), Interferon Induced Protein with Tetratricopeptide Repeats 3 (IFIT3), Interferon-Stimulated Gene 15 (ISG15), and Interleukin 6 (IL6). IFIT2 and IFIT3 are part of the IFIT family, notable for their role in the interferon alpha and gamma response pathways (Liberzon et al., 2015). Unlike typical enzymes, IFIT proteins function through protein-protein interactions utilizing their tetratricopeptide repeat (TPR) motifs, chiefly activating innate immune responses against viral threats (Pidugu et al., 2019). Their involvement in cancer, particularly in promoting epithelial-mesenchymal transition (EMT) and therapeutic resistance, as observed in oral squamous cell carcinoma, suggests a complex role in tumorigenesis, though the exact mechanisms remain to be fully understood (Pidugu et al., 2019).

IL6, a cytokine crucial for both innate and adaptive immunity, becomes problematic when its signaling persists in the tumor microenvironment (TME), transforming from a short-term defender into a promoter of chronic inflammation and tumorigenesis (D. E. Johnson et al., 2018). IL6 signaling is mediated by the JAK-STAT3 pathway, initiated when IL6 binds to its receptor IL6R, leading to the activation of JAK and phosphorylation of STAT3 through gp130 interaction **[Figure 7-11]**. Activated STAT3 then enhances the expression of cyclins and anti-apoptotic BCL-2 family proteins, bolstering cell proliferation and survival. It also creates a feedback loop by upregulating IL6 itself, ensuring sustained support for tumor growth and survival (Rose-John, 2020).

IL6, produced by cancer cells and immune cells within the tumor, acts autocrinely and paracrinely to perpetuate its tumorigenic effects, enhancing growth and thwarting apoptosis (Grivennikov & Karin, 2008). Furthermore, chronic activation of the IL6-STAT3 pathway imparts immunosuppressive properties, particularly through the activation of STAT3 in tumor-infiltrating leukocytes, which dampens immune responses against the tumor. This pathway also fosters regulatory T cells that contribute to immune tolerance, enabling tumors to evade immune detection and destruction.

In cluster 0, which primarily consists of cells exposed to 14 days of hypoxia, key markers identified include Nuclear Paraspeckle Assembly Transcript 1 (NEAT1), Inhibitor of DNA-binding/differentiation-1 (ID1), Inhibitor of DNA-binding/differentiation-3 (ID3), and Interleukin 32 (IL32). NEAT1, a long non-coding RNA (lncRNA), is integral to cancer progression, particularly through its involvement in regulatory mechanisms that manipulate the expression of protein-coding genes by interacting directly with mRNAs. This interaction can enhance or repress gene expression, providing a pathway for cancer cells to exploit for metastatic development (Gu et al., 2022). In pancreatic cancer, NEAT1 acts as a competing endogenous RNA (ceRNA) for miR-302a-3p, inhibiting its tumor-suppressive actions on RELA mRNA translation, thus promoting cancer

progression through the NF-kB signaling pathway (Luo et al., 2019). In breast cancer, NEAT1 upregulation correlates with higher levels of cyclin D1 and ZEB1, promoting cell cycle progression and epithelial-mesenchymal transition (EMT) by sequestering miR-448, which typically suppresses ZEB1 expression (Jiang et al., 2018).

ID1 and ID3, belonging to class V of the helix-loop-helix (HLH) transcription factors, impact tumorigenesis by interacting with other proteins rather than binding directly to DNA. These interactions prevent DNA binding by other transcription factors, suppressing differentiation and promoting tumor development (Sachdeva et al., 2019). ID1 is notably associated with poor cancer prognosis due to its regulation of several key signaling pathways, although these mechanisms remain to be fully elucidated (Ling et al., 2002; L. Wang et al., 2015). ID1 also facilitates the development of EMT and a cancer stem cell phenotype, contributing to its tumorigenic effects (A et al., 2001).

IL32, a pro-inflammatory cytokine like IL6, is unique in its action as it does not utilize a traditional extracellular receptor for signaling and shares no homology in primary structure with other cytokines (S.-H. Kim et al., 2005). IL32 is highly expressed in cancer cells and cancer-associated fibroblasts (CAFs), where it promotes tumorigenesis by binding to integrin β3 via the RGD motif, activating focal adhesion kinase (FAK), which then triggers p38 MAPK signaling. This cascade enhances mitotic activity and the expression of EMT markers such as N-cadherin and fibronectin, supporting cancer cell invasion and metastasis (Wen et al., 2019).

The findings from the Pathway Enrichment Analysis (PEA) using Gene Set Enrichment Analysis (GSEA) with hallmark gene sets build upon those from the Differential Expression Analysis (DEA). In cluster 2 **[Figure 4-10A]**, there is significant enrichment in hallmark hypoxia and glycolytic pathways, typical of acute hypoxia experienced for up to 72 hours. Clusters 4 and 7 **[Figure 4-10B&C]** display upregulation in pathways such as EMT, TNFA via NF-kB, and IL-2 STAT5, which are suggestive of a complex immune-modulatory and cancer stem cell (CSC) phenotype emerging after 7 days of hypoxia. Concurrently, these clusters show downregulation of E2F and G2M checkpoint targets, indicating suppression of the cell cycle under extended hypoxic conditions. Clusters 9 and 0 **[Figure 4-10D&E]** continue to show elevated levels of EMT and TNFA via NF-kB pathways, with an addition of interferon alpha and gamma pathways. Although cell cycle suppression is ongoing, it is less marked compared to earlier stages, hinting at a gradual recovery of the cell cycle. These observations are consistent with prior findings. The upregulated pathways for clusters 7, 4, 9, and 0 will be further detailed in subsequent sections to explore the development of emerging phenotypes in depth.

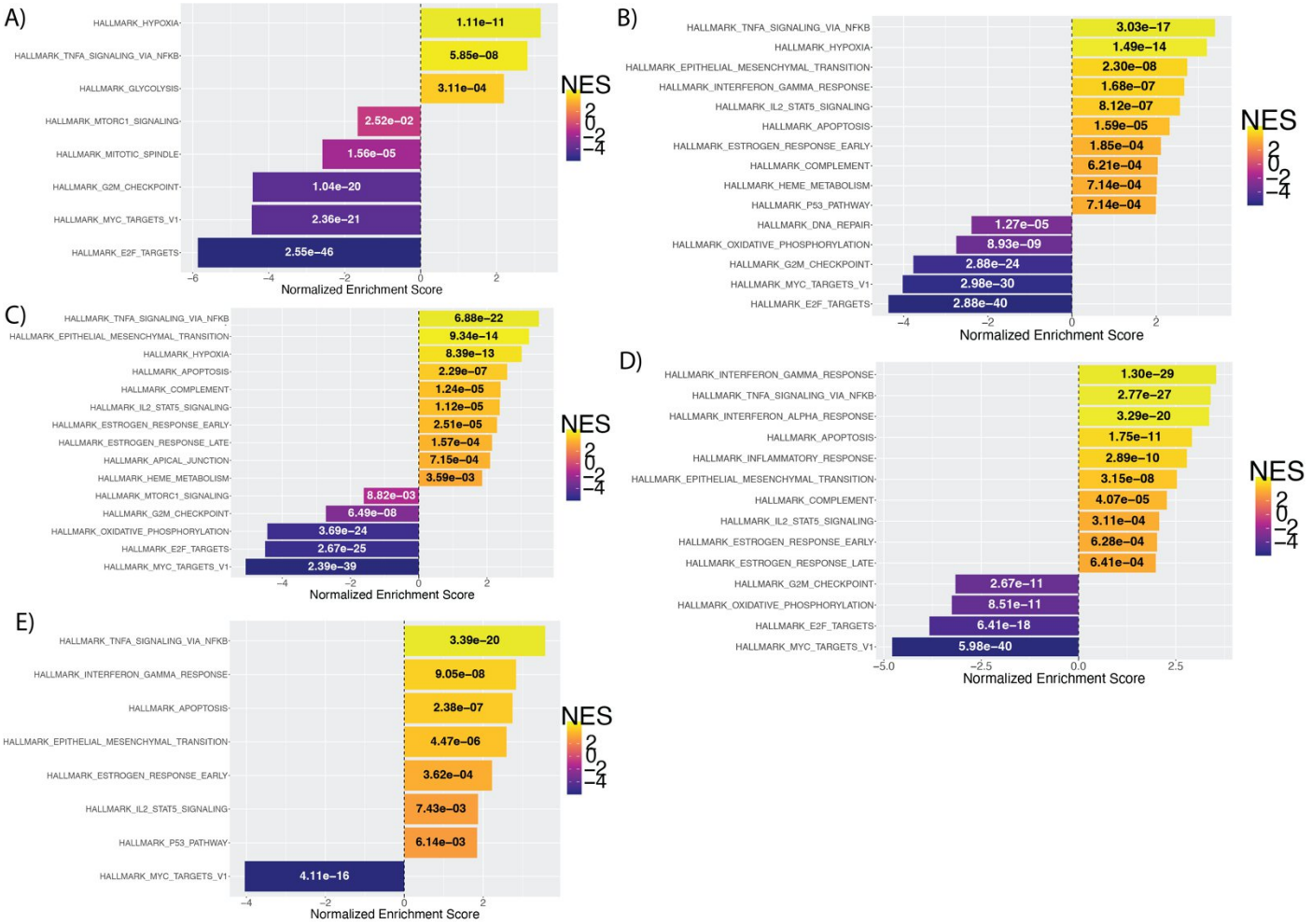


Figure 4-10: Pathway Enrichment Analysis (PEA) Results Using GSEA *– This figure illustrates the results of Pathway Enrichment Analysis conducted using Gene Set Enrichment Analysis (GSEA) with hallmark gene signatures as gene sets. The x-axis represents the normalized enrichment score, while the y-axis lists the pathways. Numbers inside the bars show the adjusted p-values, and only statistically significant pathways are included in this analysis.*

A) Cluster 2 vs Cluster 5: *This panel shows upregulation of the hallmark hypoxia and glycolysis pathways and downregulation of the hallmark G2M checkpoint and E2F targets, reflecting the initial cellular response to hypoxic stress and associated cell cycle arrest, typical of acute hypoxia (up to 72 hours).*

B) Cluster 4 vs Cluster 5: *Notable upregulated pathways include hallmark EMT and hallmark TNFA-signalling via NF-kB. Continued downregulation of hallmark E2F targets and G2M checkpoint is evident, indicating cell cycle arrest and possibly the development of a CSC phenotype in this cell population, hypothesized to start developing in day 7 hypoxia, represented by this cluster along with cluster 7.*

C) Cluster 7 vs Cluster 5: *The upregulation of hallmark EMT and hallmark TNFA-signalling via NF-kB persists, maintaining the pattern observed in cluster 4 (both representing day 7 hypoxia). Downregulation continues for hallmark E2F targets and G2M checkpoint, supporting previous interpretations of prolonged hypoxic response and its effects on cell cycle progression.*

D) Cluster 9 vs Cluster 5: *This panel shows upregulation in hallmark TNFA signalling via NF-kB and hallmark interferon signalling, indicating the development of an immune-inflammatory phenotype. Hallmark Myc targets remain downregulated, consistent with the ongoing suppression of proliferative pathways.*

E) Cluster 0 vs Cluster 5: *Similar to cluster 9, this panel shows upregulation in hallmark TNFA signalling and hallmark interferon signalling, suggesting the development of an immune-inflammatory phenotype. Hallmark Myc targets continue to be downregulated, indicating the suppression of proliferative pathways.*

4.3.5. Tracing Cellular Trajectories: Pseudotime Analysis

After concluding the DEA and PEA analyses, it was particularly intriguing to delve into the upregulated pathways in clusters 7 and 4, which are representative of day 7 hypoxia. Day 7 is considered a key transition point between acute and chronic hypoxia, not only because it serves as an intermediate timepoint but also based on findings from pseudotime analysis. This analysis, conducted using the Monocle3 pipeline, leverages the top 3000 variable features in the data to identify developmental trajectories within the cell population.

Pseudotime analysis **[Figure 4-11]** revealed two distinct trajectories: one corresponding to the normoxic population and the other to the hypoxic population. The identification of these trajectories

by the Monocle algorithm substantiates the biological distinction between the hypoxic and normoxic cell populations.

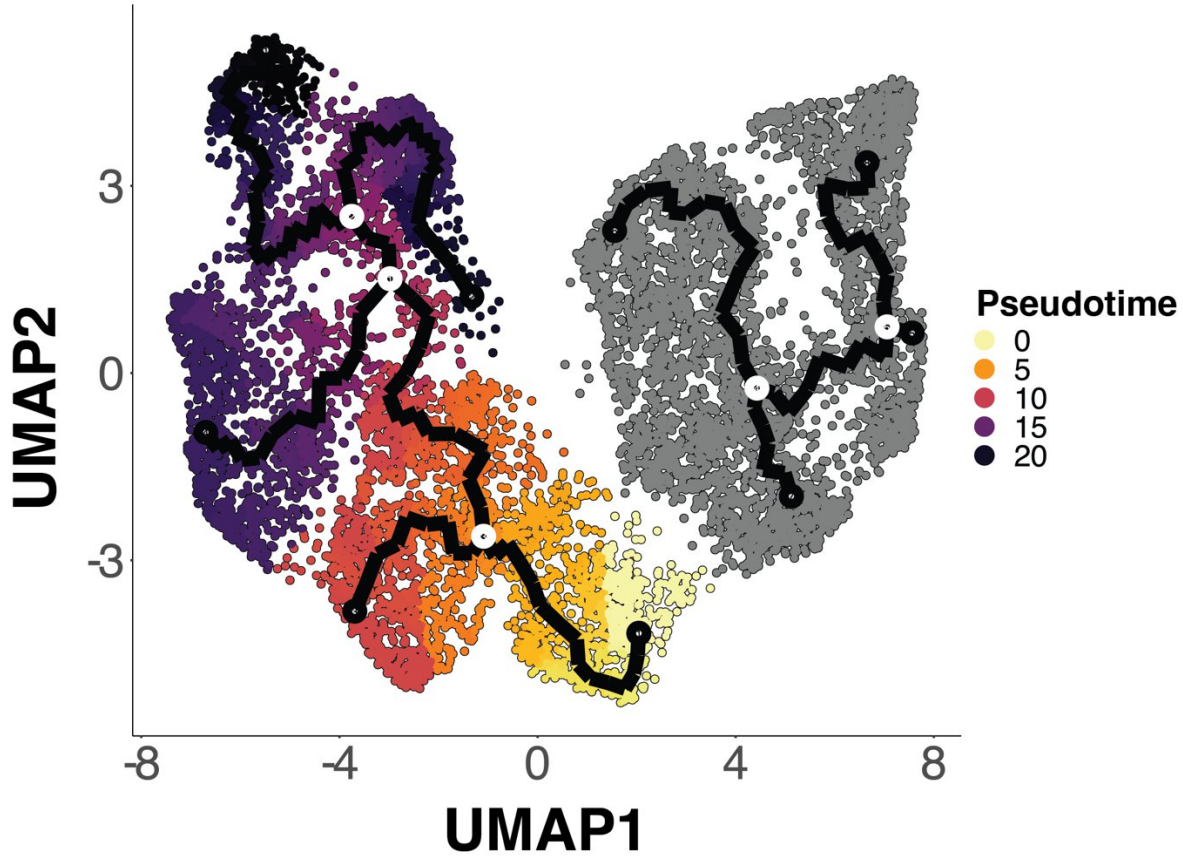


Figure 4-11: Monocle3 Pseudotime Analysis *– This figure presents the pseudotime analysis results for the scsHypoxiaTime dataset, revealing two distinct trajectories within the data: one representing hypoxic conditions and the other normoxic. The trajectories are overlaid on UMAP plots, with cells coloured according to their progression through pseudotime (pseudotime scale shown in the legend, using arbitrary units). In the trajectory graph, black nodes represent branching points where different cell fates diverge, white nodes indicate starting points of the trajectories, and grey nodes mark the terminal points, representing the end stages of cell fate progression.*

Progression within the normoxic trajectory is marked by cell cycle variation, moving from the lower portion of the UMAP (G2M dominated), through the middle portion (S phase dominated), to the top portion (G1 phase dominated). This progression aligns with previous analyses that highlighted the role of the cell cycle in contributing to the heterogeneity of the normoxic data across timepoints. Interestingly, the trajectory identified two terminal branches within the G1 population, suggesting further variation that merits exploration, whether of biological or technical origin. However, this aspect was not the focus of the current analysis and thus was not further discussed.

For the hypoxic trajectory, there is a clear progression from acute hypoxia in clusters 5 and 2, with a core branching point within cluster 5, moving towards clusters 7 and 4. Another significant branching point occurs at cluster 7, with a terminal branch emerging at cluster 4, and the trajectory concludes with an additional branching point and two terminal nodes at clusters 9 and 0, indicative of chronic hypoxia.

Based on the comprehensive findings from Differential Expression Analysis (DEA), Pathway Enrichment Analysis (PEA), and pseudotime analysis, several key hypotheses have been formulated, which are poised for both computational and experimental validation. It is hypothesized that Cluster 7 acts as a critical juncture in cellular decision-making, where cells either progress towards chronic hypoxia, represented by Clusters 9 and 0, or adopt a more quiescent state as stem cells in Cluster 4. This node signifies a strategic point in hypoxic adaptation, directing either towards sustained dormancy or further hypoxic progression. Additionally, Cluster 4 is hypothesized to represent a cancer stem cell (CSC) phenotype, which plays a pivotal role in cancer survival and progression. The characteristics associated with CSCs, such as self-renewal and differentiation

potential, may contribute to the resilience and adaptability of tumors in hypoxic environments. Finally, Clusters 9 and 0 are believed to represent the culmination of chronic hypoxic adaptation, where cells develop an aggressive immune-inflammatory phenotype. This adaptation is thought to be a preparatory phase for aggressive tumorigenesis and subsequent tumor revascularization, marking a transition from survival under extreme conditions to a proactive oncogenic phase.

4.3.7. EMT & Angiogeneic Phenotypes - Day 7

Day 7 hypoxia represents one of the two chronic hypoxic timepoints explored in this dataset, revealing significant biological insights through Differential Expression Analysis (DEA), Pathway Enrichment Analysis (PEA), and pseudotime analysis. To further elucidate these biological phenotypes, we employed signature scoring analysis and transcription factor (TF) pathway activity estimation techniques.

Signature scoring was conducted using **Seurat::AddModuleScores()**, a method previously detailed in Chapter 2 and earlier sections of this document. In contrast, **decoupleR**, a Bioconductor package not previously discussed, was utilized for estimating TF pathway activity. **decoupleR** facilitates the inference of TF activity from transcriptomic data, applicable to both bulk and single-cell analyses. It primarily interprets TF signatures to estimate TF activity, representing a sophisticated extension of signature scoring methods.

Signature scoring with Seurat, using the hallmark gene sets, highlighted hallmark epithelial-to-mesenchymal transition (EMT) **[Figure 4-12A]** and hallmark angiogenesis **[Figure 4-12B]** as pathways with elevated activity in clusters 7 and 4. Hallmark EMT is associated with an aggressive cancer phenotype, characterized by a marked propensity for invasion and expansion. Similarly, hallmark angiogenesis is crucial for metastasis; it supports tumor revascularization after prolonged hypoxia, which is essential for tumor progression and further metastasis via the invasion-metastasis cascade.

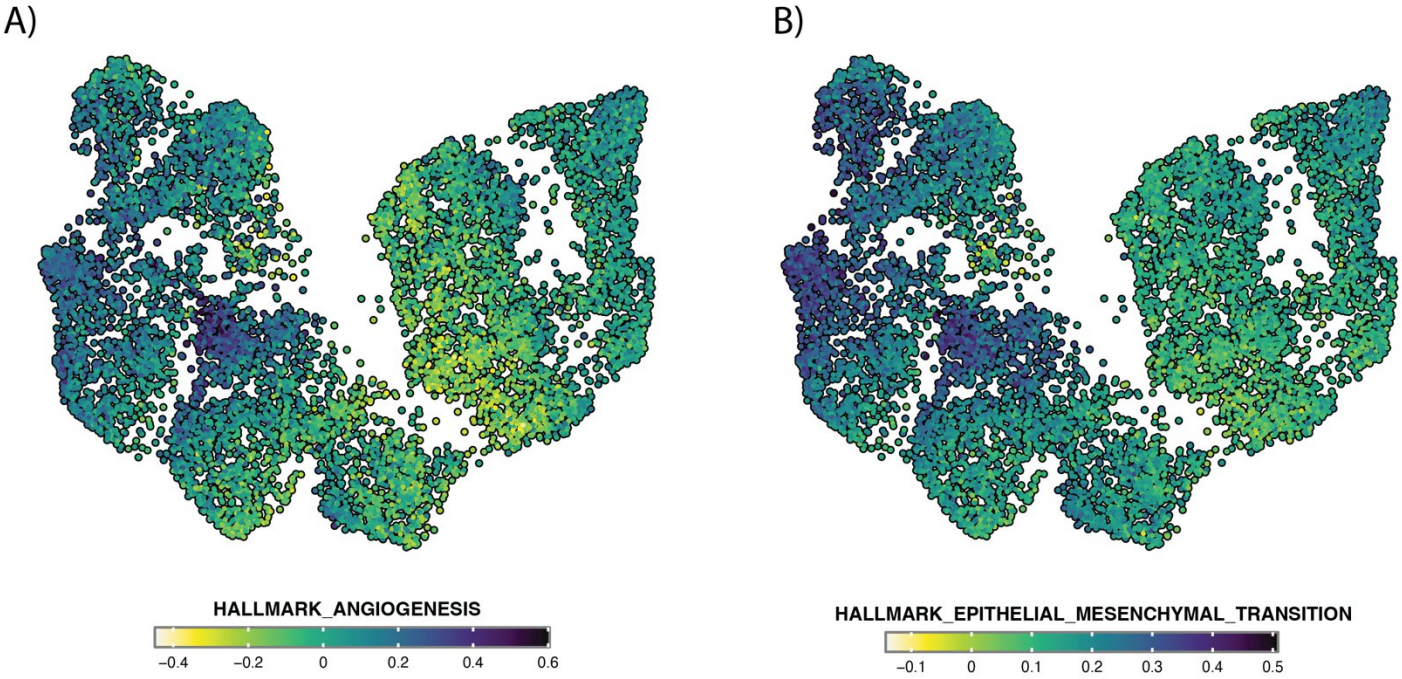


Figure 4-12: UMAP Distribution of EMT and Angiogenesis Signatures *– These UMAP plots represent the distribution of cells based on epithelial mesenchymal transition (EMT) and angiogenesis gene expression signatures. The colour gradient signifies the degree of gene set enrichment within each cell, with deeper colors indicating higher expression levels. Specifically, cluster 7, captured during hypoxic day 7 (HD7), shows pronounced expression of both EMT and angiogenesis-associated genes. Differential expression analysis corroborates these findings, revealing a significant increase in matrix metalloproteinases (such as MMP1 and MMP14) and angiogeneic factors (such as VEGFA), which underscores the emergence of these cellular phenotypes in response to hypoxic stress.*

To fully grasp the biological implications of the signature scoring results, it is crucial to examine the constituent genes within each pathway. This analysis is valuable as it reveals the uniformity or

variability of gene expression patterns across different clusters. Additionally, it identifies the predominant genes driving these pathways. Importantly, since the signature scoring methods applied do not adjust the weighting for individual genes, it is likely that specific genes disproportionately influence the signature score, while others may be less detectably expressed.

In line with this analytical approach, for the visualization within this thesis, the leading-edge genes for these pathways from GSEA PEA in these expression heatmaps were used. This decision is justified because some hallmark signatures, like the Hallmark EMT which includes 187 genes, are too extensive to plot entirely in a practical manner within the limited space of this thesis. Focusing on the leading-edge genes is also advantageous as they are the most significant contributors to the signature score, aligning with the earlier observation regarding the influential roles of specific genes in signature scoring.

For EMT, 55 leading edge genes across both clusters 4 and 7 were identified and plotted, most notably VEGFA and CD44 **[Figure 7-2]**. VEGFA, as the master regulator of angiogenesis, is intricately linked to the EMT process, reflecting the critical interplay between angiogenesis and the metastasis-invasion cascade. CD44, a membrane glycoprotein, is profoundly implicated in oncogenesis and the evolution of the cancer stem cell (CSC) phenotype, mainly through its interaction with extracellular matrix components like hyaluronic acid (HA) (C. Chen et al., 2018). The binding of HA to CD44 induces conformational changes that enhance its interaction with key signaling pathways such as Ras-Raf-MEK-ERK, PI3K-AKT, and MAPK, which are vital for cell survival and proliferation (Zöller, 2011). In breast cancer, CD44 is known to promote glycolytic metabolism, essential under the hypoxic conditions of the TME. Silencing CD44 shifts metabolism from glycolysis to oxidative phosphorylation, decreases the expression of HIF-1A and LDHA, and reduces the activity of c-Src and AKT, demonstrating its central role in cellular metabolism under hypoxic stress (Nam et al., 2016). Additionally, the overexpression of CD44v6 is linked with enhanced EGFR signaling, cancer metastasis, and resistance to endocrine therapy (Bellerby et al., 2016). CD44 also serves as a canonical CSC marker, reflecting its extended role in carcinogenesis (Trapasso & Allegra, 2012).

For angiogenesis, the complete signature was utilized, identifying VEGFA, TIMP1, and STC1 as top features **[Figure 7-3]**. VEGFA's critical role in carcinogenesis is well-established, with its function in promoting angiogenesis thoroughly characterized. TIMP1's involvement in carcinogenesis has been detailed, especially in discussions about top DEA markers for cluster 7. STC1 (Stanniocalcin-1), a homodimeric secreted glycoprotein, varies widely in expression across tissues and significantly contributes to angiogenesis. It regulates cell cycle progression through cyclin and pRb, modulates cell adhesion via vimentin, and enhances VEGFA and VEGFR expression, core components of angiogenic signaling (Law & Wong, 2013). Additionally, STC1 has been shown to increase VEGF expression through ERK1/2 and PKCβII signaling pathways in gastric cancer, further emphasizing its role in angiogenesis (L. He et al., 2011).

4.3.8. Immuno-inflammatory Signalling – Day 14

A comprehensive examination of the biological responses at Day 14 of hypoxia was undertaken to gain deeper insights into the key pathways active at this advanced hypoxic timepoint. This period is particularly notable for the substantial recovery of the cell cycle, suggesting a significant degree of cellular adaptation to the sustained hypoxic stress.

In-depth analysis of hallmark pathways highlighted the interferon alpha and gamma pathways as primary drivers of the observed biological phenomena during Day 14 hypoxia, specifically within cluster 9 **[Figure 4-13A&B]**. Interferons (IFNs) play a pivotal role in the immune system's response to both viral infections and neoplastic growths. These proteins function as cytokines, initiating critical signaling cascades through both paracrine and autocrine communication upon binding to specific IFN receptors on cell surfaces. The engagement of these receptors activates the JAK-STAT signaling pathway, a critical mechanism in immune response modulation **[Figure 7-10]**.

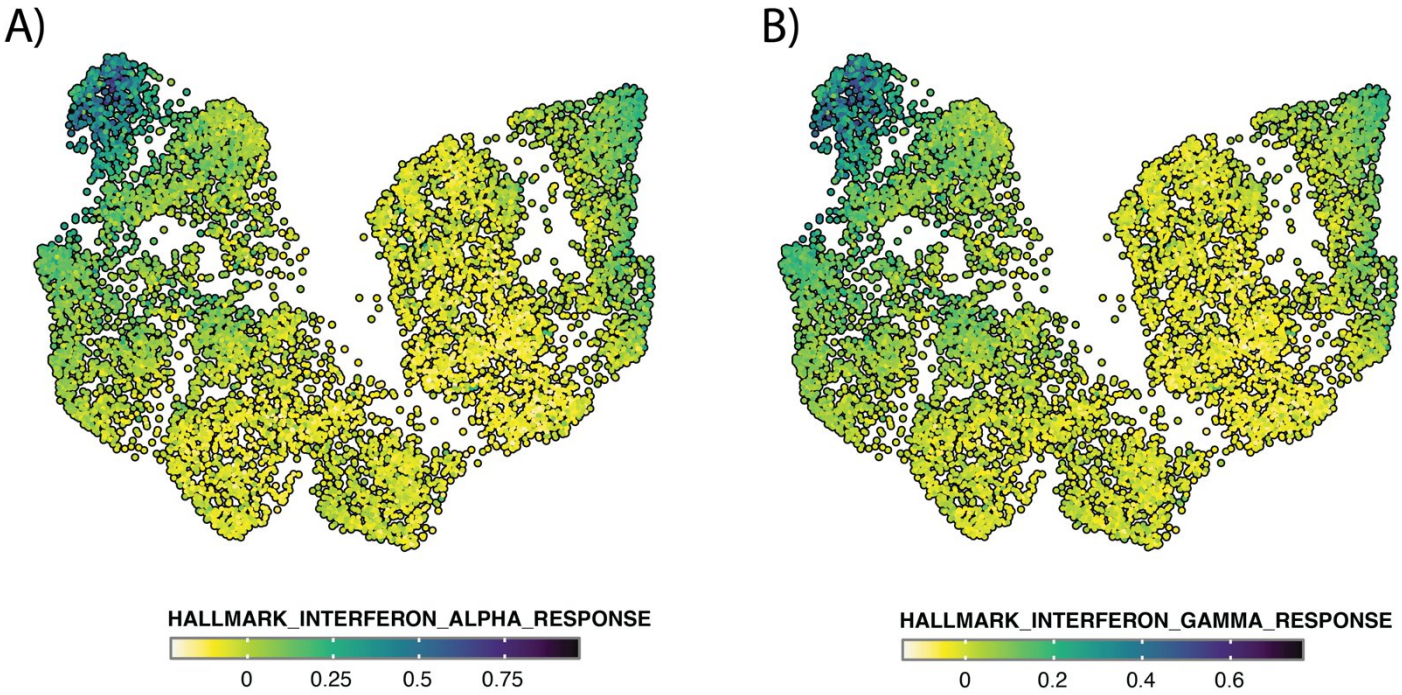


Figure 4-13: UMAP Distribution of Interferon Pathway Signatures - *These UMAP plots represent the distribution of cells based on Hallmark interferon alpha and gamma gene expression signatures. The colour gradient signifies the degree of gene set enrichment within each cell, with deeper colors indicating higher expression levels. Specifically, cluster 9, captured during hypoxic day 14 (HD14), shows pronounced expression of both interferon pathways. Differential expression analysis corroborates these findings, revealing a significant increase in interferon-related genes (IFIT3, IFIT1, IFIT2, and SOD2), which underscores the emergence of these cellular phenotypes in response to chronic hypoxic stress.*

Further analytical efforts were directed towards understanding the specific genes driving these interferon pathways. To this end, UMAP plots of the leading-edge genes for these pathways in cluster 9 were constructed. The IFN alpha pathway was prominently marked by the expression of IFIT2 and IFIT3, genes that have been extensively explored in previous sections dealing with cluster DEA results. These genes are noted for their significant roles in mediating the cellular response to interferon signaling.

The IFN gamma pathway similarly exhibited a notable presence of IFIT family members, underscoring the substantial genetic overlap between the interferon alpha and gamma pathways. These pathways are intricately linked to the cellular defense mechanism against pathogens and abnormal cell growth, yet they also play complex roles in carcinogenesis. They possess dual functionality, acting both as promoters of tumor growth under certain conditions and as tumor suppressors in others. This dual nature adds layers of complexity to the interpretation of their activity, making the assessment of their overall impact on cancer progression both ambiguous and challenging.

These findings not only emphasize the importance of the interferon pathways in cellular adaptation and immune response within the tumor microenvironment but also highlight the nuanced interplay between genetic regulation and cellular behaviour in the context of prolonged hypoxic stress.

4.3.9. ScBrAtlas: Clinical Validation Of Cell Line Findings

Following this analysis, a significant concern with scRNA-Seq cell line datasets is the validation of their clinical relevance. Due to the simplicity of cell line models, the lack of hypoxic gradations, and the absence of the Tumor Microenvironment (TME), it is reasonable to assume that the biological phenotypes and characteristics observed in cell line data may not be highly transferable to clinical settings.

To address this concern, the single-cell breast cancer atlas (ScBrAtlas) dataset was utilized (Y. Chen et al., 2022). This scRNA-Seq dataset encompasses various subtypes of breast cancer, including both tumor and stromal cells, making it an exceptional reference for investigating transcriptomic changes at the single-cell level within breast cancer. For this analysis, only the triple-

negative breast cancer (TNBC) subset was used, which includes both BRCA1 mutated and wild-type (WT) samples **[Figure 4-14]**.

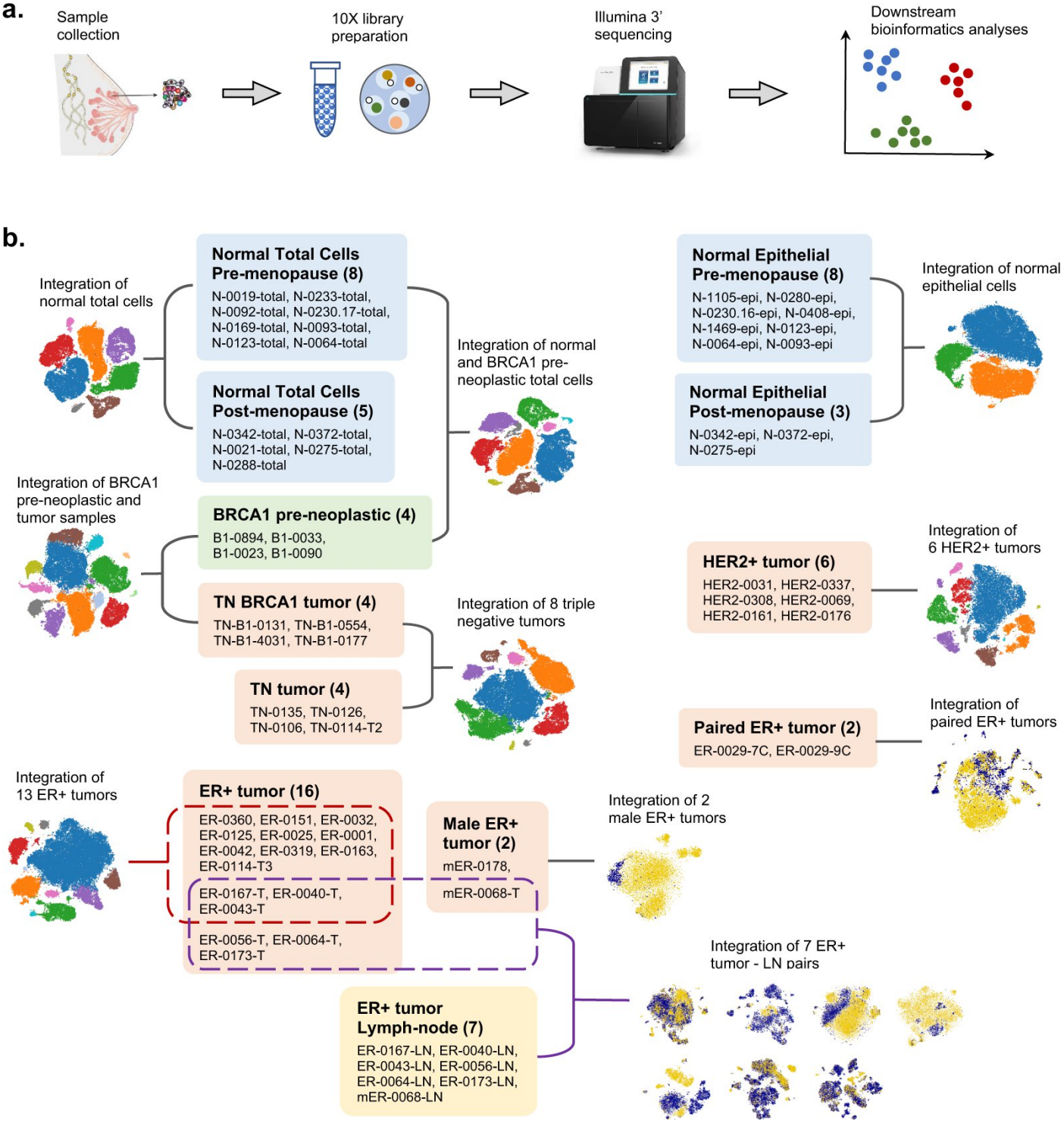


Figure 4-14: ScBrAtlas Dataset – A) Experimental Workflow: *This panel illustrates the experimental procedure, starting with sample collection from the tumor, followed by the preparation of the 10X library, Illumina sequencing, and subsequent bioinformatics analyses.* ***B) Data Structure:*** *This panel details all available scRNA-Seq datasets across various tumor types, including their stromal microenvironment and immune landscapes. Normal breast tissue is also included for comparison. (Figure adapted from (Y. Chen et al., 2022)).*

The TNBC subset was provided as an RDS object. Importing it into R, creating the Seurat object, normalization, variable feature selection, and dimensionality reduction using PCA and UMAP were all performed locally using previously developed pipelines. UMAP plots annotated by primary metadata columns were generated to aid characterization **[Figure 4-15]**.

UMAP annotation by patient sample shows that batch effects are not the main source of variation in the data, providing reassurance that the biological findings reflect true biological variation rather

than technical variance **[Figure 4-15A]**. UMAP annotation by BRCA1 mutational status reveals that this is also not the primary source of stratification in the data **[Figure 4-15B]**. UMAP annotation by Seurat clusters (identified using the lowest clustering resolution of 0.1) distinguishes two distinct populations within the data **[Figure 4-15C]**. UMAP annotation by cell cycle clearly indicates that cell cycle effects are the major driver of variation, which is expected and biologically relevant **[Figure 4-15D]**. Cell cycle variation is closely associated with hypoxic signaling and gradation; tumor regions experiencing more hypoxia (closer to the tumor core and further from the blood supply) exhibit higher levels of cell cycle arrest, while regions with lower levels of hypoxia show proliferative patterns that support tumor growth and progression.

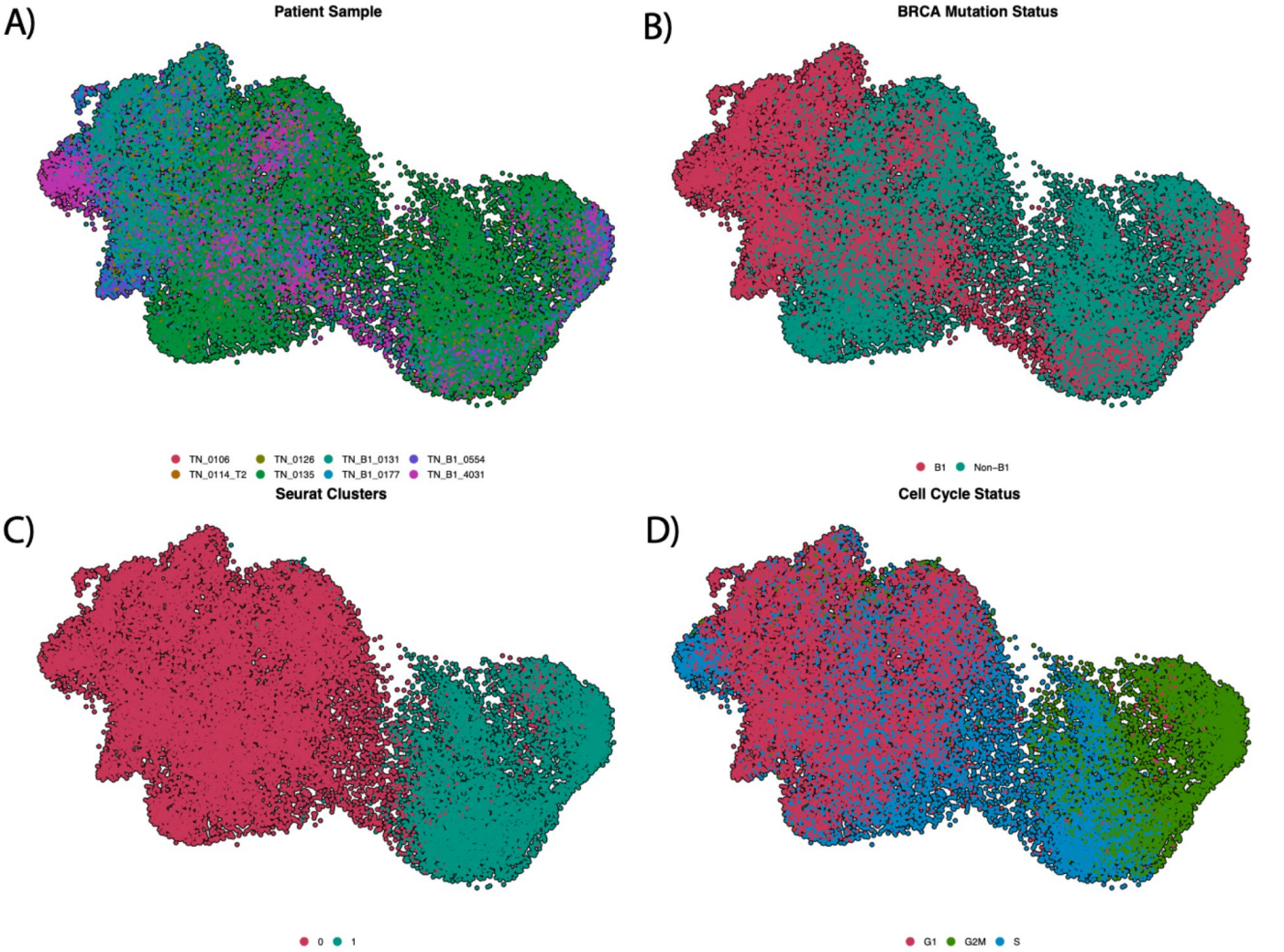


Figure 4-15: UMAP Projections of the Triple Negative Breast Cancer (TNBC) Subset from the Single Cell Breast Cancer Atlas *– **A)** Visualization by Patient ID: Each cell is color-coded to represent individual patient samples, facilitating comparison between patients within the TNBC subset. **B)** Stratification by BRCA1 Mutation Status: Cells are distinguished based on the presence or absence of BRCA1 mutations, highlighting genetic variability within the cohort. **C)** Identification of Seurat Clusters: Cells are color-coded to denote clusters determined by the Louvain algorithm at a fine resolution of 0.1, illustrating the heterogeneity within the TNBC population. **D)** Cell Cycle Phase Distribution: The cells are shaded to indicate their respective phases in the cell cycle, providing insights into the proliferative dynamics of the TNBC subset.*

Subsequent signature scoring analysis using ***Seurat::AddModuleScore()***, the Hallmark gene sets (Liberzon et al., 2015), and the Buffa hypoxia signature (Buffa et al., 2010a) revealed intriguing biological insights **[Figure 4-16]**.

The Hallmark G2M checkpoint signature indicated **[Figure 4-16A]** strong cell cycle activity on the right edge of the UMAP, suggesting substantial proliferation. This finding aligns with the results from the cell cycle scoring analysis.

The Hallmark E2F targets signature **[Figure 4-16C]** showed a similar pattern to the G2M checkpoint signature, which is expected given E2F's critical role in cell cycle progression and activation. E2F

drives proliferative signaling through its interactions with the Retinoblastoma tumor suppressor protein (Rb), p107, and p130, and it is significantly deregulated during malignant transformation and the early stages of tumorigenesis (D. G. Johnson & Schneider-Broussard, 1998).

Both the Hallmark hypoxia **[Figure 4-16B]** and Buffa hypoxia **[Figure 4-16D]** signatures displayed a similar distribution of scores, with strong enrichment in the left branch of the UMAP towards the upper medial edge. It is noteworthy that the Buffa hypoxia signature consistently performed well compared to the hallmark hypoxia signature, underscoring the robustness of the Buffa signature.

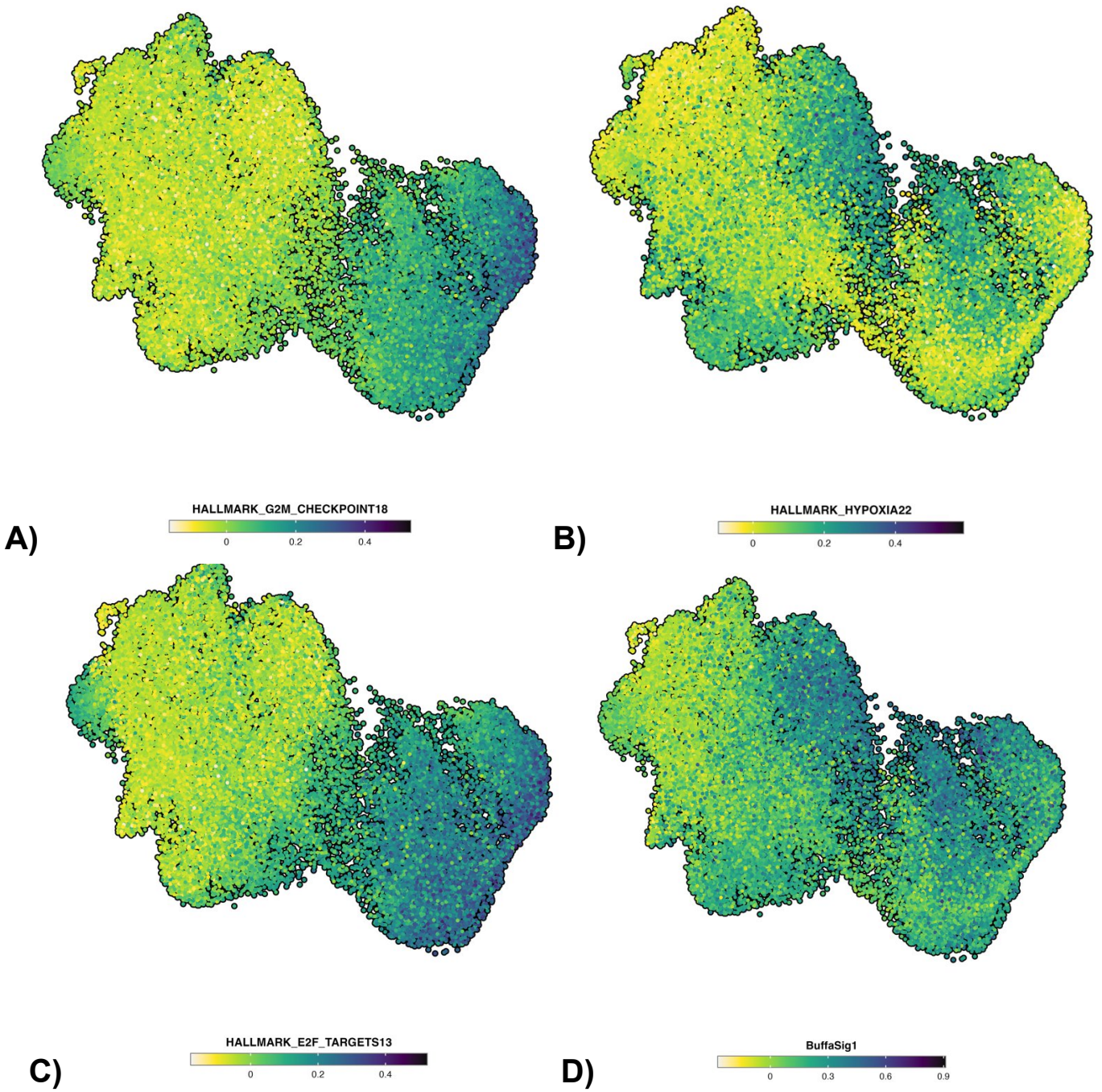


Figure 4-16: ScBrAtlas Signature Scoring Analysis – *Hallmark gene sets and the Buffa hypoxia signature were utilized.* ***A) Hallmark G2M Checkpoint:*** *This figure aligns with the cell cycle scoring results, showing substantial G2M enrichment on the right edge of the UMAP.* ***B) Hallmark Hypoxia:*** *Displays the distribution of hypoxia-related scores.* ***C) Hallmark E2F Targets:*** *The E2F transcription factor, a master regulator of cell cycle activation and progression, mirrors the score distribution pattern seen in cell cycle scoring and the Hallmark G2M Checkpoint signature.* ***D) Buffa Hypoxia Signature:*** *Both hypoxia signatures show similar scoring patterns, highlighting the robustness of the hypoxia signature across the Hallmark standard.*

Other Hallmark gene sets were also analysed, but they are not included here for the sake of clarity and space.

Of particular interest in this analysis were clusters 7 and 9, making a major goal the mapping of these clusters onto the TNBC subset of ScBrAtlas. Cluster mapping is a broad topic in bioinformatics with no current consensus on the most representative method. For this analysis, signature scoring analysis was used as a proxy for cluster mapping by assuming these clusters represent distinct biological states and then using the top 50 marker genes for each cluster (based on average log2 fold change) as independent gene signatures for application on the clinical data.

The assumptions of this approach align with those of signature scoring analysis: the presence of all signature genes in the data and homogeneity of expression patterns for the different genes within the dataset. While these assumptions were not always met, the results still offer valuable insights into the relevance of these clusters. However, these results should be interpreted cautiously due to potential inconsistencies in method assumptions.

For cluster 7 (indicative of chronic hypoxia at day 7) **[Figure 4-17A]**, the signature maps to the left medial branch of the data, the same region where the Hallmark hypoxia and Buffa hypoxia signatures mapped. This indicates that this region of the clinical data is highly hypoxic, yet it still exhibits canonical hypoxic features along with the emergence of EMT phenotypes characteristic of cluster 7.

For cluster 9 (indicative of chronic hypoxia at day 14) **[Figure 4-17B]**, the signature maps to the same region on the far right, lateral edge of the data. This region also shows a high prevalence of cells in the G2M state (from cell cycle scoring analysis) and high scores for the Hallmark G2M checkpoint and E2F targets. This alignment perfectly matches the observed phenotype of cell cycle recovery under chronic hypoxia in the cell line data, validating the capacity of cells experiencing prolonged hypoxia and cellular stress to recover their proliferative phenotype.

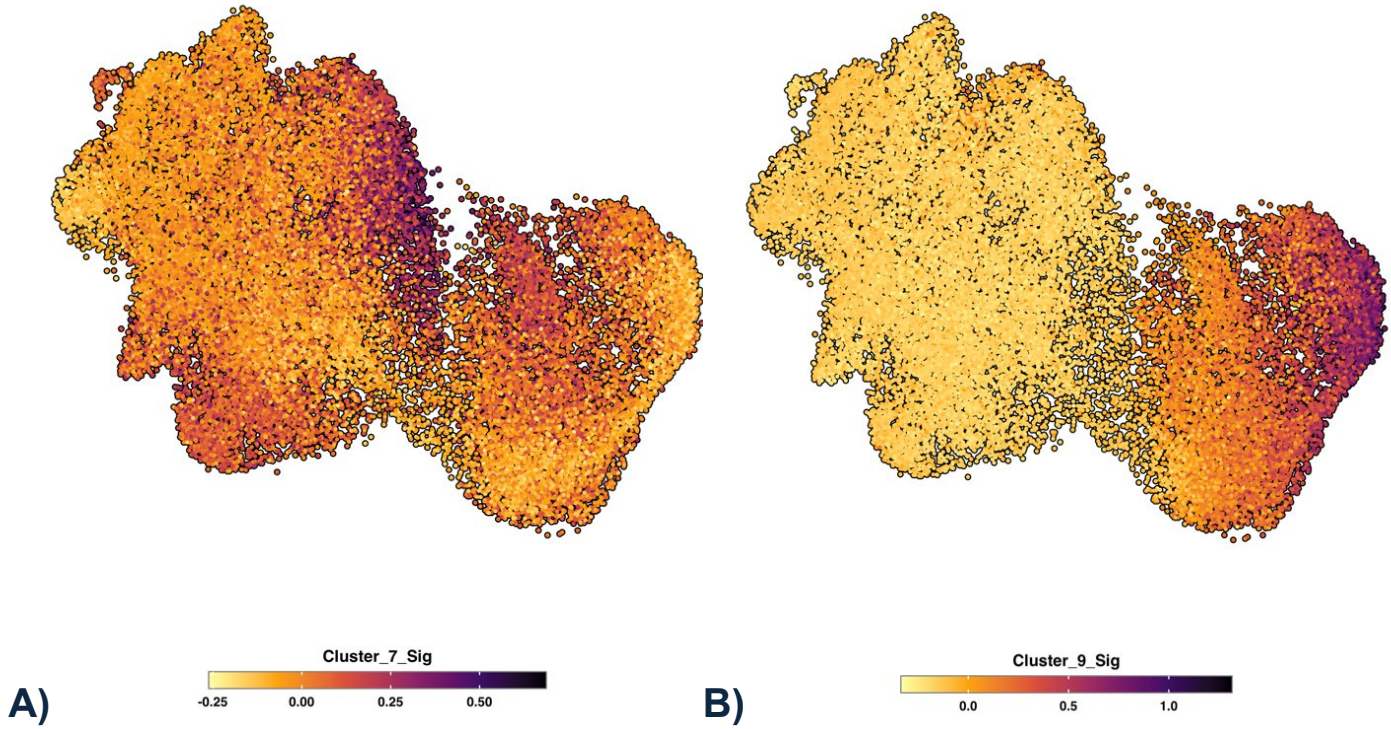


Figure 4-17: Signature Mapping of Hypoxia-Related Clusters onto TNBC Subsets from the Single Cell Breast Cancer Atlas *– **A)** Cluster 7 Signature Mapping: Signature scoring analysis, utilizing the top 50 marker genes identified by average log2 fold change, was applied to map Cluster 7 from the scsHypoxiaTime dataset onto the TNBC subset of the ScBrAtlas dataset. The cluster's signature predominantly localizes to regions corresponding to the G1 phase, aligning with previous observations that cells under day 7 hypoxia experience notable hypoxic stress and are largely in a state of cell cycle arrest. **B)** Cluster 9 Signature Mapping: Similarly, Cluster 9's signature was mapped using the same signature scoring methodology. It shows pronounced enrichment in data regions with active G2M and S cell cycle phases. This supports earlier findings that prolonged hypoxia, characteristic of Cluster 9, is associated with cell cycle progression and recovery.*

4.4. Discussion

This analysis represents a pioneering endeavor to investigate chronic hypoxia at the single-cell level within a TNBC cell line. Given the clinical significance of chronic hypoxia, this dataset emerges as a crucial resource for delineating the biological phenotypes that cancer cells adopt during adaptation to chronic hypoxic stress. These adaptations lay the groundwork for the development of metastatic and aggressive disease. Understanding these states and potentially disrupting their progression could unveil new pharmacological strategies to mitigate metastatic development.

4.4.1. Cell Cycle Arrest & Recovery

A key distinction between acute and chronic hypoxia pertains to cell cycle dynamics. Under acute hypoxia, there's a marked downregulation of the cell cycle with cells predominantly entering the G1 phase. However, extended hypoxia reveals a cell cycle recovery, with a significant number of cells progressing into the S and G2M phases. Initially identified using Seurat's cell cycle scoring functionality, this recovery was further substantiated through the visualization of individual markers and signature scoring using the hallmark G2M checkpoint signature.

Although the precise mechanisms behind cell cycle recovery remain unidentified, they likely involve a synergy of EMT phenotype development and the upregulated expression of cyclins and CDKs to drive the cell cycle, alongside anti-apoptotic BCL2 family members countering apoptotic cell death due to cell cycle stress.

4.4.2. Temporal Dynamics of The Hypoxic Response: Matched Contrasts

DEA and PEA analyses of matched hypoxia vs normoxia contrasts (day 1 hypoxia vs day 1 normoxia, day 2 hypoxia vs day 2 normoxia) underscored the consistency of the hypoxic response. Prominent hypoxia markers such as NDRG1, ADM, and BNIP3 were among the most highly expressed genes across all timepoints. PEA also revealed positive enrichment of hallmark hypoxia and glycolysis pathways throughout these contrasts, suggesting that the gene expression programs during chronic hypoxia are not exclusive to those during acute hypoxia. These results confirm that even with prolonged hypoxia, where HIF1A levels typically diminish, the canonical hypoxic response remains robustly active alongside other pathways facilitating adaptation to varied hypoxia durations.

4.4.3. Phenotypical Evolution Across Different Durations of Hypoxia

With the initial validation of the acute hypoxic response across timepoints, the focus shifted towards elucidating the specific pathways altered by cancer cells in response to varying hypoxia durations. This was conducted through DEA and PEA, comparing hypoxic timepoints beyond the first, with attention turning from timepoints to clusters due to the emergence of distinct clusters at later stages. This analysis proved invaluable, illustrating upregulation of CSC and SASP phenotypes in cluster 4, EMT and angiogenesis in cluster 7, and interferon and EMT pathways in clusters 9 and 0. Pseudotime analysis further revealed a developmental progression from day 1 hypoxia, culminating in a decision point by day 7 in cluster 7, where cells either progress towards a CSC state in cluster 4 or advance to the immune-inflammatory state in clusters 9 and 0.

4.4.4. Clinical Relevance

While cell line data always poses challenges for clinical correlation, many studies attempt to bridge this gap using TCGA data, which is suboptimal due to the substantial differences between bulk and single-cell transcriptomics. Instead, we utilized the TNBC subset of the ScBrAtlas dataset, applying a signature scoring approach to align our clusters with clinical data. This alignment not only affirmed the relevance of the biological phenotypes observed but also corroborated the cell cycle recovery noted in day 14 clusters, which aligned with areas in the ScBrAtlas dataset exhibiting peak G2M signaling.

5. Chapter 5: Stromal Heterogeneity in Triple Negative Breast Cancer Lung Metastases

5.1. Introduction

The tumor micro-environment (TME) plays a dynamic role in tumor progression, characterized by a complex interplay of both physical and cellular components. Critical physical components such as oxygen levels and pH, along with the extensively discussed hypoxic heterogeneity, define the challenging conditions within which tumors thrive. On the cellular front, the TME encompasses a diverse array of cell subtypes—including fibroblasts, endothelial cells, and mesothelial cells—that interact with the tumor core, influencing either the suppression or enhancement of tumor progression and invasion.

Hypoxia, a pivotal physical factor in the tumor microenvironment (TME) of metastases and particularly in triple-negative breast cancer (TNBC) metastases, plays a crucial role in shaping the development and progression of these aggressive tumors. It acts as an inevitable byproduct of tumour proliferation that influences cellular subtype changes, signaling cascade activation, and overall TME progression. Numerous studies have underscored the importance of hypoxia in TNBC metastases, highlighting its role as a driver of compositional shifts and a master regulator of active signaling cascades.

This review (I. Kim et al., 2022) emphasizes hypoxia's regulatory influence on cancer-associated fibroblast (CAF) activity, resulting in significant enrichment of signaling cascades associated with extracellular matrix (ECM) remodeling—such as MMP upregulation in response to hypoxic signaling—cytokine signaling and immune-inflammation, and metabolic reprogramming. Specifically, it facilitates the secretion of CXCL12 from CAFs and stabilizes CXCR4 in cancer cells, promoting proliferation. Hypoxia also supports the metabolic reprogramming of CAFs to produce lactate, thereby sustaining the Warburg effect in cancer cells.

Additionally, another review (Nagl et al., 2020) focuses on the heterogeneity of tumor endothelial cells (TECs), particularly how hypoxia enhances the VEGF signaling cascade, alters endothelial cell morphology, and destabilizes chromosomal stability, all of which favor endothelial proliferation and neo-vascularization as part of the invasion-metastasis cascade.

Thus, characterizing the TME composition and its development under hypoxic conditions is a direct extension of the preceding chapters on hypoxic heterogeneity. This focus shift examines how hypoxia serves as a selective pressure that engenders TME heterogeneity, pivotal for understanding the complex dynamics within TNBC metastases.

Central to aggressive tumorigenesis and the poor outcomes often seen in patients is the invasion-metastasis cascade, where stromal cells contribute crucially to each phase. Epithelial-mesenchymal transition (EMT), a pivotal step in this cascade, is significantly driven by stromal signaling. For instance, IL-32 released from cancer-associated fibroblasts promotes EMT and tumor survival primarily through its interaction with Beta3 integrin via an RGD motif, subsequently activating the p38 MAPK signaling cascade (Wen et al., 2019). Additionally, mesenchymal stem cells are known to enhance cancer cell EMT, while CXCL14, an under-characterized orphan chemokine released from cancer-associated fibroblasts, promotes EMT by activating ACKR2, a similarly under-characterized G-protein-coupled receptor (GPCR) in cancer cells (Sjöberg et al., 2019).

The signaling pathways between cancer cells and the surrounding stroma are intricate, featuring multiple autocrine and paracrine signaling cascades that drive tumorigenesis. Traditional characterization of these pathways through bulk transcriptomics has been challenging; however, the advent of single-cell RNA sequencing (scRNA-Seq) has revolutionized our ability to explore the transcriptomic profiles of these cellular subtypes (Cobos et al., 2020; S. Jin et al., 2021). This

technology enables detailed analysis of gene expression changes over time, with pseudotime analysis pinpointing the key drivers of stromal adaptation under tumorigenic conditions.

Furthermore, sophisticated cell-cell communication analysis now allows for the use of transcriptomic data from different cell subtypes to infer active ligand-receptor communication networks, including both autocrine and paracrine signaling cascades (S. Jin et al., 2021). This detailed characterization of stromal signaling within the TME is essential for identifying the primary signaling pathways that drive advanced tumorigenesis.

This chapter delves into the stromal micro-environment of TNBC metastases in mouse lungs. We distinguish between the immediate surrounding TME (niche) of the tumor core and the distal TME, conducting scRNA-Seq analyses at two pivotal timepoints: 7 days and 14 days post-transfection. The objectives of this analysis are to:

- Characterize the stromal TME in both the niche and distal compartments.
- Explore changes in the composition and transcriptomic profiles of the TME across the specified timepoints.
- Identify primary drivers within the stromal signaling network, focusing on the utilized pathways and the key sender and receiver cells.

5.2. Methods

Experimental and computational methods were as described in chapter 2.

5.3. Results

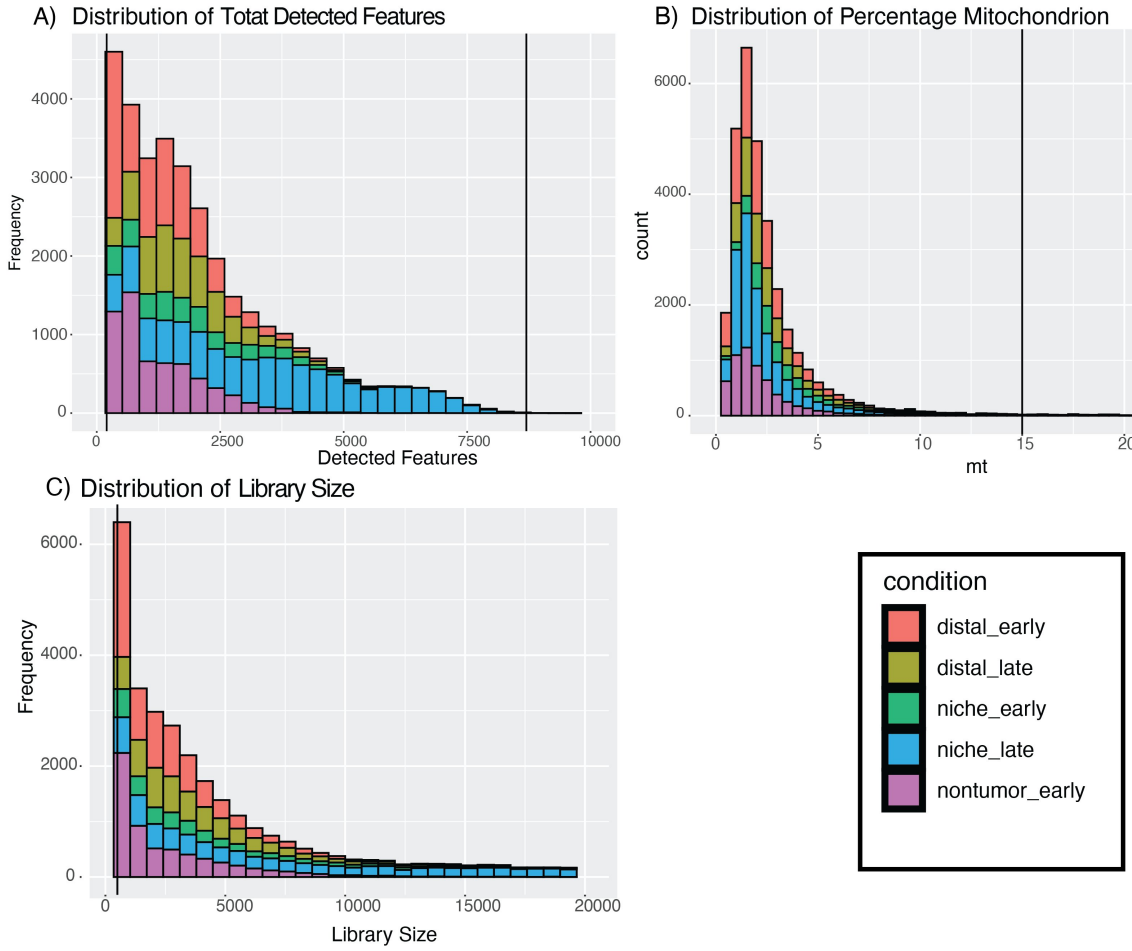
5.3.1. Clustering Analysis Identified 21 Distinct Clusters

The dataset primarily differentiates between four conditions related to stromal cells in a tumor environment: normal, niche_early, niche_late, distal_early, and distal_late. The term “normal” refers to stromal cells that were extracted from mouse pulmonary tissue in the absence of any tumours; this condition serves as the baseline that all other conditions are contrasted against. The term "niche_early" refers to stromal cells located directly adjacent to the tumor, playing a potential role in regulating the tumor microenvironment; these cells were collected seven days post-transfection. "Niche_late" describes a similar group of stromal cells but collected at a later timepoint, fourteen days post-transfection. Conversely, "distal_early" denotes stromal cells situated farther from the tumor, also collected seven days post-transfection, while "distal_late" refers to the same distal stromal cells collected fourteen days post-transfection. The critical distinction between 'niche' and 'distal' stromal cells was made based on mCherry staining, which facilitated further FACS sorting and cell isolation. For a more detailed explanation of these procedures, refer to chapter 2.

The histograms and combined scatterplots **[Figure 5-1]** provide a comprehensive visualization of the quality control (QC) metrics for our samples, detailing library size, number of detected genes, and mitochondrial gene content. The uniformity of histogram shapes across all conditions indicates consistent quality among the various sample groups.

For QC purposes, histograms were utilized to establish upper and lower thresholds for each metric **[Figure 5-1A, B & C]**. Upper thresholds help in identifying and removing doublets and multiplets, while lower thresholds are used to exclude cells likely experiencing stress or undergoing apoptotic or necrotic processes. An exception is made for mitochondrial gene content, where only a lower threshold is applied. This is because elevated mitochondrial content typically indicates mitochondrial permeation, which is only existent under severe cellular stress.

Accompanying these histograms, a final combined scatterplot integrates all QC metrics **[Figure 5-1D]**, displaying the established thresholds. This visualization ensures that the thresholds are appropriately set across all conditions to prevent any bias that could potentially affect subsequent analyses. This approach guarantees balanced filtration, aiding in the accurate interpretation of our experimental data.



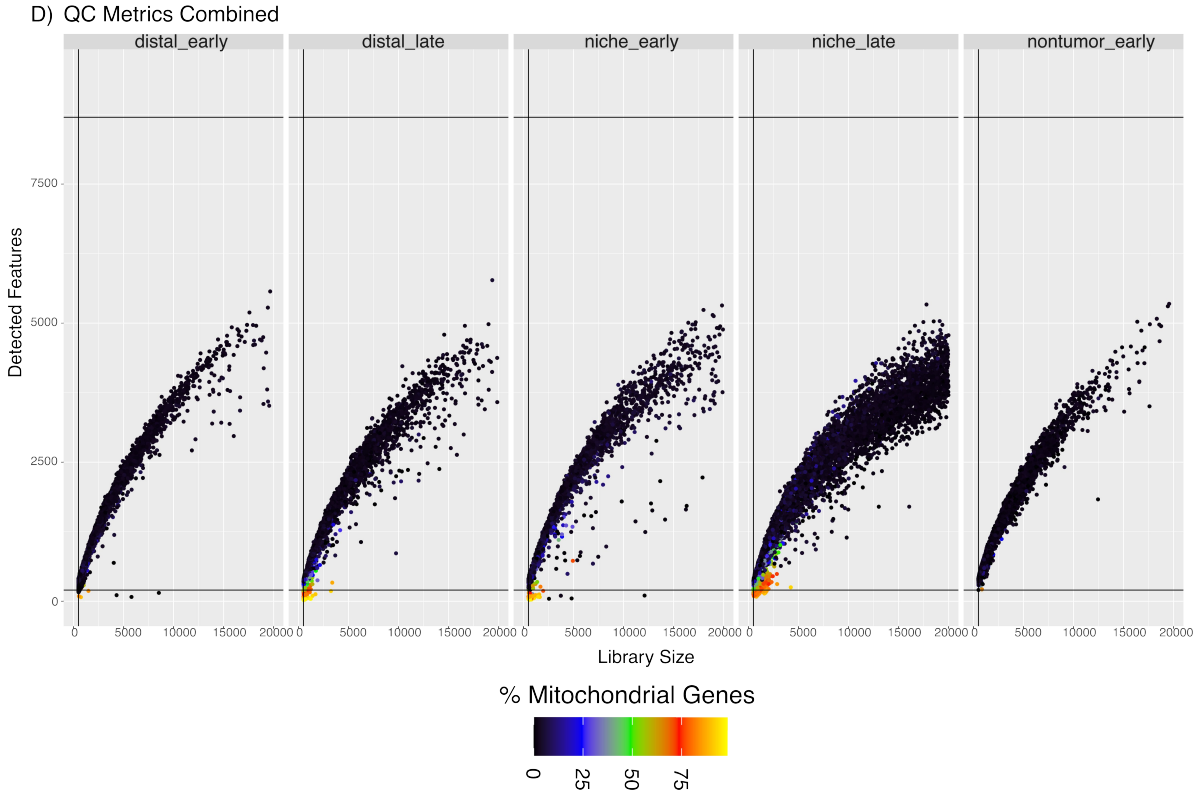


Figure 5-1: QC Plots for the Mlung Dataset *– **A)** The histogram details library sizes for each condition. Thresholds defining the acceptable range for library sizes are indicated, ensuring the exclusion of potential doublets (excessively high counts) and dying cells (extremely low counts). **B)** This histogram displays the number of detected features per cell, with thresholds demarcating the upper and lower limits to identify cells with unusually high or low gene expression diversity. **C)** Shown is a histogram of the percentage of mitochondrial genes by condition, with an upper threshold delineating the acceptable level of mitochondrial content, beyond which may suggest cellular damage or stress. **D)** Scatterplots combine these QC metrics: library size on the x-axis, number of detected features on the y-axis, and colour representing the percentage of mitochondrial genes. Condition-specific plots ensure that the applied thresholds are appropriately calibrated to avoid overly stringent exclusions.*

After the quality control and filtration phases, the datasets were normalized, variable features were identified, and the expression data was scaled. These procedures were executed using the SCTransform pipeline, detailed in chapter 2.

Subsequently, cell clustering was conducted using the Louvain algorithm. The choice of an optimal clustering resolution, set at 0.5, was guided by the clustering tree analysis **[Figure 5-2]**. This specific resolution was selected to effectively capture unique biological variations within the data while preventing over-clustering and minimizing the detection of false biological variation. The final number of clusters identified in the data was 21 clusters.Top of FormBottom of Form

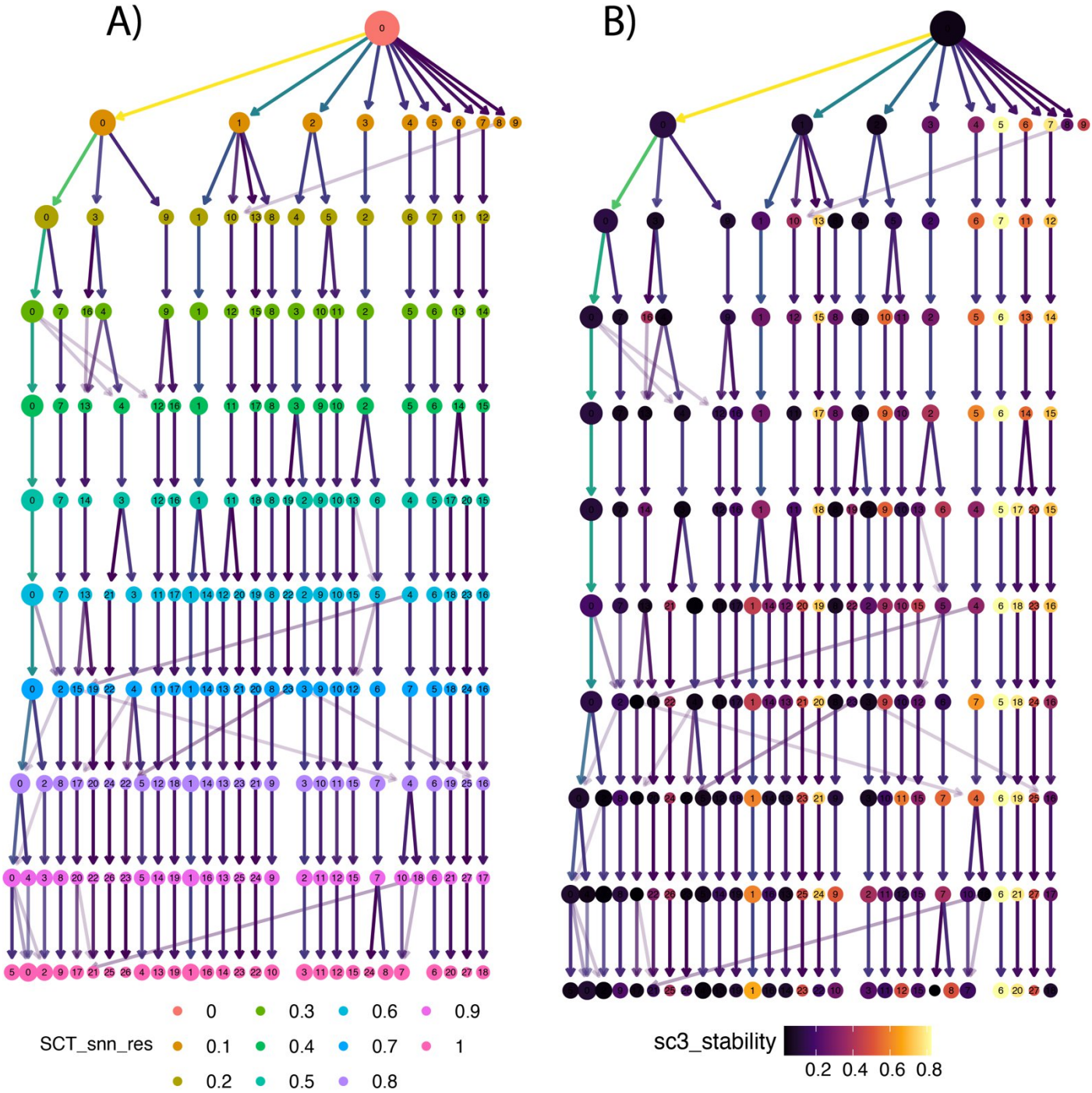


Figure 5-2: Clustering Trees from Louvain Algorithm Analysis - A) *Clustering Tree Coloured by Cluster Assignment: This dendrogram represents the hierarchical clustering structure determined by the Louvain algorithm, with the colour scheme reflecting the cluster identity assigned at various resolutions of shared nearest neighbor (SNN) modularity. The node size correlates with the number of cells in each cluster, and branch colors indicate different resolution parameters, highlighting how the granularity of clustering is influenced by the chosen resolution.* ***B)*** *Clustering Tree coloured by Cluster Stability: Here, the same dendrogram is coloured based on cluster stability scores derived from the SC3 (Single-cell Consensus Clustering) index, providing a visual assessment of cluster robustness. Warmer colors indicate higher stability, suggesting that clusters are consistent and reliable across different resolution parameters. The node size remains representative of cluster population, and the varying hues illustrate the differential stability of clusters, with th e most stable clusters being potential targets for further characterization. The resolution parameter plays a crucial role in the Louvain algorithm's success, as it determines the depth of community detection within the data, affecting the sensitivity and specificity of the identified clusters. Clustree, a tool for visualizing clustering dendrograms across resolutions, provides insights into the optimal resolution setting, balancing between over-clustering and under-clustering, which is vital for accurate downstream analyses. Based on these figures, a final clustering resolution of 0.5 was chosen.*

5.3.2. Cluster Annotation And The Identification of Mesothelial, Endothelial, and Fibroblasts As Major TME Components Across Metastatic Development

After conducting quality control and clustering, UMAP visualizations were generated and color-coded according to condition **[Figure 5-3A]** and by Seurat clusters **[Figure 5-3B]**. The colouring

based on condition reveals that subsets 'distal_early' and 'distal_late' are situated in the lower part of the UMAP, 'normal' in the centre, and 'niche_early' and 'niche_late' in the upper regions. The depiction by Seurat clusters indicates the distribution of the 21 clusters.

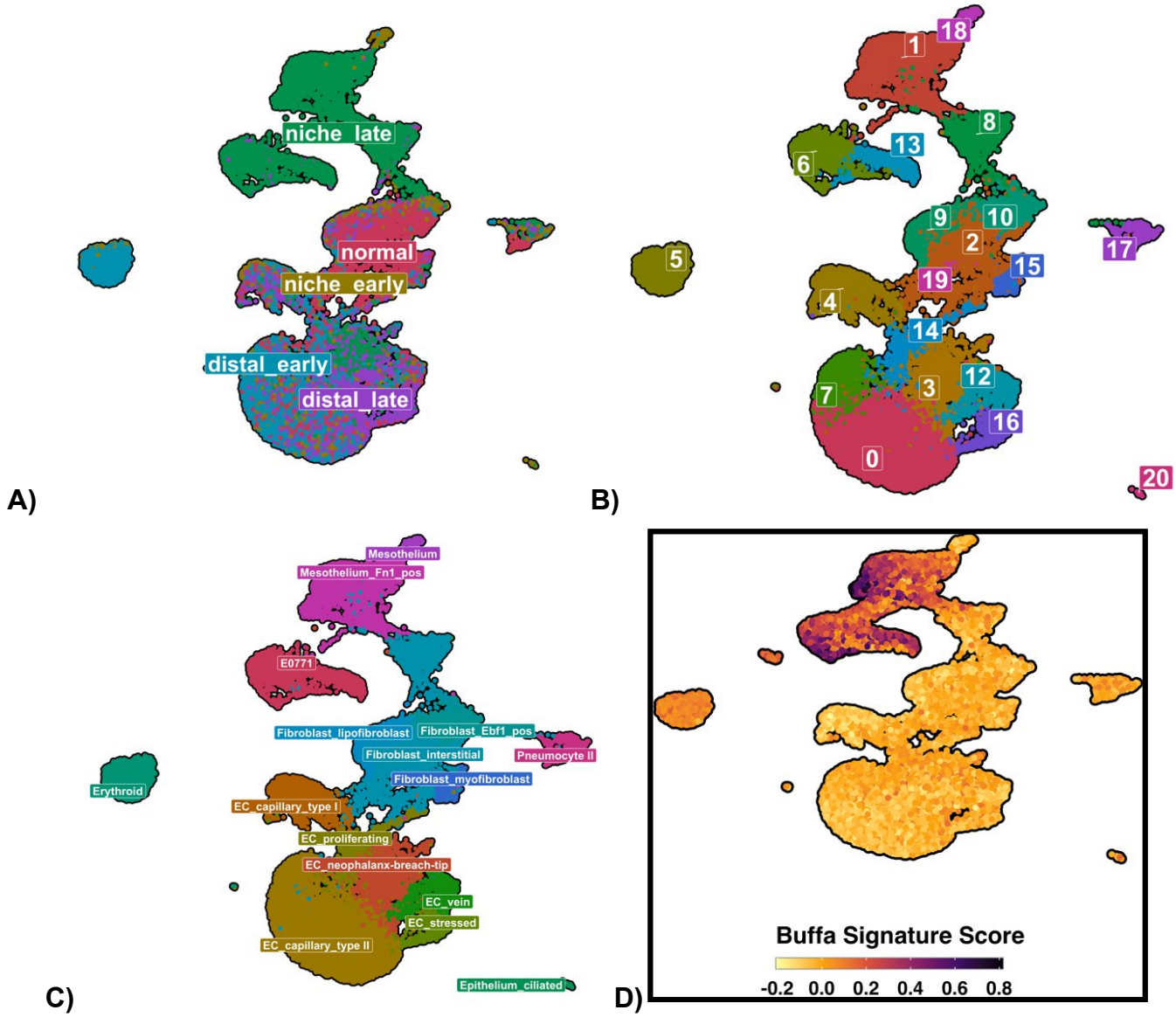


Figure 5-3: Multidimensional Analysis of Cellular Composition in the Mlung Dataset via UMAP – A) *Stromal Cell Distribution by Condition: The UMAP plot showcases stromal cells differentiated by extraction time points within specific tissue compartments. 'Niche early' denotes cells from the niche compartment at day 7, while 'Niche late' refers to day 14. Similarly, 'Distal early' and 'Distal late' represent stromal cells from the distal compartment at days 7 and 14, respectively.* ***B)*** *Seurat Cluster Delineation: Colored by distinct Seurat clusters, this plot reveals the cellular heterogeneity as determined by the Louvain clustering algorithm, with a resolution of 0.5 chosen to optimize cluster distinction based on dendrogram analysis.* ***C)*** *Subtype Identification by DEA Markers: Annotations of cell subtypes are overlaid on the UMAP visualization, guided by differential expression analysis (DEA) markers for each cluster, referencing a comprehensive cell subtype signature from the literature.* ***D)*** *Buffa Hypoxia Signature Scoring: The UMAP highlights cells scored for the Buffa hypoxia signature—a well-validated hypoxia gene signature known for its robustness across benchmarks and internal validation (discussed in Chapter 6). The scoring, performed using the* ***AddModuleScore()*** *method, illustrates a notable overlap between high hypoxia levels and the 'mesothelium_FN1_positive' population. Predominantly observed in the 'Niche late' compartment, this overlap suggests the pivotal role of hypoxia in promoting aggressive, mesothelium-dependent metastatic development within breast cancer metastases.*

Cluster annotation was determined using Differential Expression Analysis (DEA) and a curated list of cell type markers from the literature, detailed in **[Table 2]** and **[Table 3]**.

Table 2: *Marker Profiles for Fibroblasts, Mesothelial Cells, and Other Subtypes. This table lists the specific markers used to identify fibroblasts and mesothelial cells, with additional subtypes included for comprehensive coverage. Markers for endothelial populations are detailed in a separate table.*

Annotated cell type	DEGs	Reference

Fibroblast population		
Interstitial fibroblasts	*Col1a1, Col3a1, Serping1, Bgn*	(Angelidis et al., 2019)
	Igfpb7, Fst1l	(X. Liu et al., 2021)
Lipofibroblasts	*Col13a1, Macf1, G0s2, Tcf21, Gyg*	(X. Liu et al., 2021)
EBF1-positive fibroblasts	*Pi16, Serpinf1, Scara5, Ebf1*	(X. Liu et al., 2021)
Myofibroblasts	*Hhip, Aspn, Mustn1*	(Xie et al., 2018)
	Hhip, Tagln, Enpp2, Wnt5a	(X. Liu et al., 2021)
Mesothelial population		
Mature mesothelial cells	*Igfbp5, Igfbp6, Dcn, Upk3b, Rarres2, Msln, Wt1*	(Angelidis et al., 2019; Xie et al., 2018)
	Upk3b, Msln, Wt1	(X. Liu et al., 2021)
Cancer-associated mesenchymal cells (CAMCs)	*Rarres2, Wt1, Msln, Upk3b*	(Angelidis et al., 2019)
	Fn1	(Kenny et al., 2014; Sakai et al., 2022)
Other cell types		
Ciliated epithelial cells	*Sec14l3, Ccdc153, Tppp3*	(Angelidis et al., 2019)
Pneumocytes type II	*Sftpc, Sftpa1, Sftpb*	
Erythroid cells	*Hba-x, Hbb-bh1, Hbb-y*	(Otsuka et al., 2016)
E0771 tumor cells	Human *CD2* Note: E0771 cells express human *CD2* because they were transduced with mCherry-labelling lentiviral plasmid	(Ombrato et al., 2019)

DEA was performed using ***Seurat::FindAllMarkers()*** function to identify the predominant marker genes for each cluster. These markers were then employed to determine the corresponding cell types for each cluster, using a detailed set of cell type and subtype markers compiled by Salwa Lin from various studies; the list of these marker genes is provided in the supplementary materials for reference. In addition to cluster markers, expression UMAP plots of the canonical markers of the key stromal cell types were used to identify regions of the UMAP that corresponded to these subtypes.

Top of FormThe dataset identified several main cell types **[Figure 5-3C]**, each playing a distinct role in the tumor microenvironment (TME). Endothelial cells (ECs) form a monolayer that interacts directly with the bloodstream, regulating substance transfer between the tissues and the blood, displaying significant heterogeneity and involvement in tumor progression, including metastasis, angiogenesis, therapeutic resistance, and extracellular matrix (ECM) remodelling (Krüger-Genge et al., 2019; X. Yao & Zeng, 2023). Key EC markers such as Pecam1, Cdh5, and Cldn5 were used to delineate EC subtypes, including EC type I, EC type II, and proliferating ECs **[Figure 7-6A]**.

Epithelial cells, or pneumocytes, line the lung cavities and, while included in this study, are not the primary focus. Markers such as Epcam1 and Cdh1 helped identify this cell type **[Figure 7-6B]**. Fibroblasts, critical in forming connective tissue and restructuring the ECM, contribute significantly to cancer progression via mechanisms such as MMP release, secretion of growth factors like HGF and VEGF, and promotion of angiogenesis through SDF-1 (Mao et al., 2013; Orimo et al., 2005; Saito et al., 2010). Key markers like Pdgfra, Fn1, Tcf21, Colla1, and Col1a2 were instrumental in identifying fibroblasts **[Figure 7-6C]**.

Mesothelium cells, lining serous cavities and organs, provide a protective, non-adhesive surface and are a focus due to their role in metastatic development, particularly through a paracrine axis

involving AMH and its receptor AMHR2 in ovarian cancers, highlighting their importance in the study. Markers such as Wt1, Msln, Upk3b, and Krt19 were utilized **[Figure 7-6D]**.

E0771 cells, derived from lung metastases, were included to explore cell-cell communication within the TME, particularly paracrine signaling axes promoting tumorigenesis. Markers such as human CD2 (hCD2) and mCherry differentiated these human cancer cells from mouse stromal cells, with mCherry's expression aiding in marking adjacent stromal cells for fluorescence-activated cell sorting (FACS); thus, transcriptional expression of mCherry also indicated the presence of cancer cells **[Figure 7-6E]**.

5.3.3. From Early to Late: Composition Switch with Advanced Metastatic Development

One of the primary objectives of this study was to identify key alterations in cell type composition within the stromal Tumor Microenvironment (TME) between the niche and distal compartments, as well as changes over time. This characterization is crucial for pinpointing cell types essential for metastatic development and progression. To achieve this, UMAP visualizations **[Figure 5-4]** and composition barplots **[Figure 5-5]** were created, annotated by cell subtype and segmented by conditions, offering a quantitative and qualitative overview of the cell types exhibiting the most significant shifts in prevalence across different conditions.

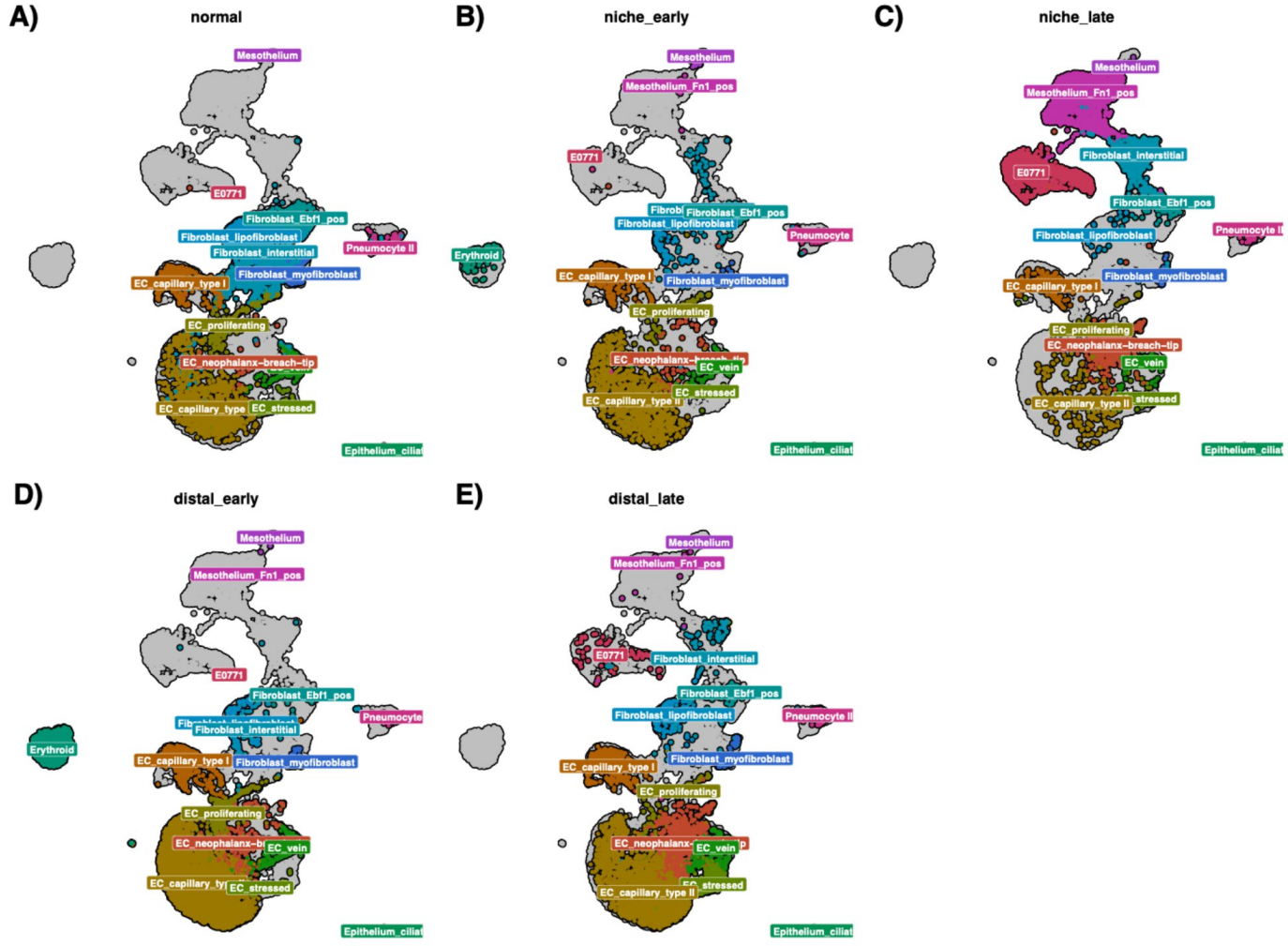


Figure 5-4: UMAP Distribution of Cell Subtypes Across Different Temporal and Spatial Conditions – *This UMAP visualization details cell subtypes annotated and segregated based on their respective conditions, elucidating the temporal dynamics of cell subtype distributions within niche and distal compartments. The plots facilitate a comparative analysis of cellular composition shifts at early and late timepoints of metastatic progression. Notably, within the niche compartment, there's a marked increase in the mesothelium and mesothelium_FN1_positive subtypes over time, signifying their potential roles in advancing metastasis and contributing to angiogenic processes. This observation aligns with the earlier noted hypoxia-related changes (see Figure 12), where hypoxia was seen to influence mesothelium-associated metastatic behaviour. Across both niche and distal compartments, there is a notable elevation in the EC_breach subtype. This upregulation correlates with the subtype's involvement in angiogenesis,*

reinforcing the implications of cellular adaptation in the tumor microenvironment as observed in the broader analysis of the Mlung dataset.

In the normal condition, there is a high prevalence of EC type I and II cells and interstitial fibroblasts, but a low proportion of mesothelial cells (both subtypes) and breach endothelial cells. In the early niche condition, there is a notable presence of EC type II cells and Ebf1 fibroblasts, yet a very low proportion of mesothelial cells and EC breach cells. The late niche condition shows a significant transition, marked by increased proliferation of EC breach cells and mesothelial cells, coupled with a decrease in the prevalence of EC type I and II cells.

For both early and late distal conditions, there is a consistent high proportion of EC type II cells and a low proportion of mesothelial cells. This pattern remains consistent over time, with limited changes in composition, which is expected since this stromal population is spatially distant from the metastatic core and should be less influenced by alterations within the niche stromal compartment. However, it is noteworthy that changes in the distal parts of the lung could occur over an extended period not covered in this study.

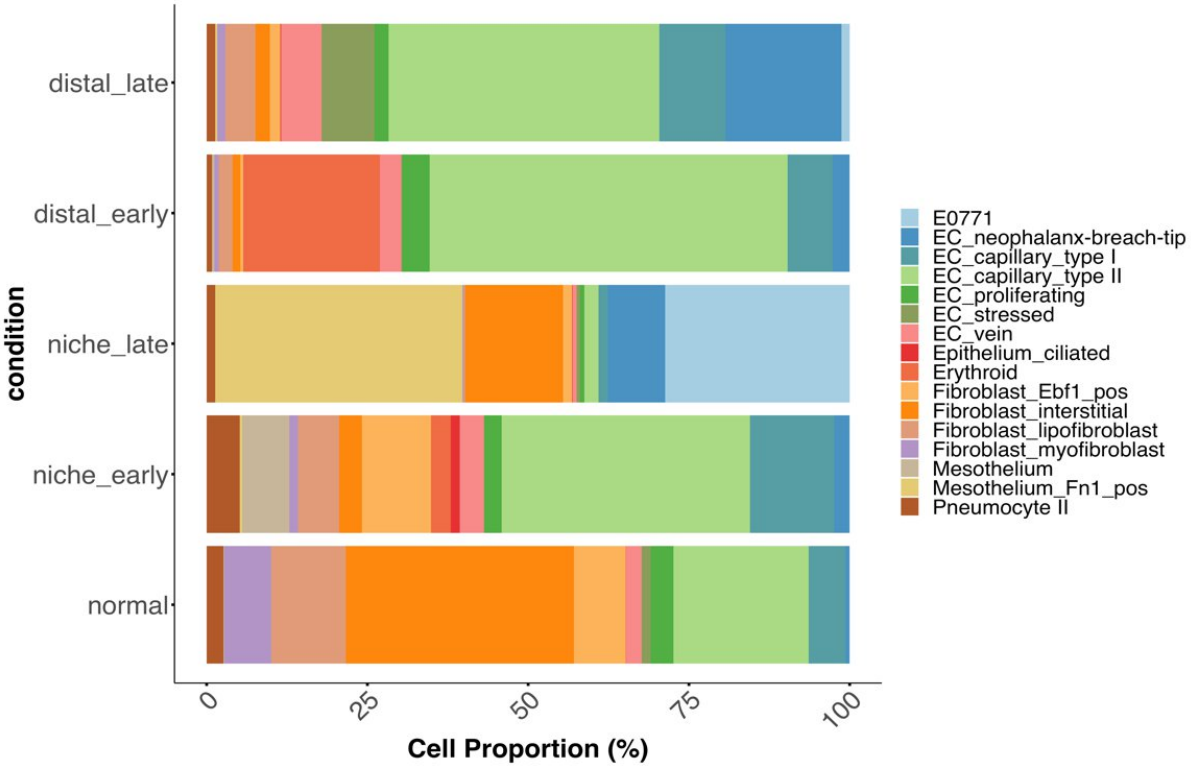


Figure 5-5: Comparative Barplot of Cell Subtype Proportions Across Temporal and Spatial Conditions*– This barplot visualizes the distribution of cell subtypes within the tumor microenvironment across various conditions, offering a quantitative complement to the UMAP analyses presented in Figures 12 and 13. The chart depicts the proportional changes in cell subtype populations from normal tissue to the niche and distal compartments at different stages of metastatic development. Consistent with the spatial analysis in Figure 13, there is a distinct increase in mesothelium and mesothelium_FN1_positive subtypes from the 'niche_early' to 'niche_late' conditions, underscoring their role in the temporal progression of metastasis. This pattern echoes the mesothelium-associated hypoxic response and metastatic behavior identified in Figure 12. The EC_breach subtype shows a significant rise in proportion from 'early' to 'late' conditions in both niche and distal compartments, aligning with the previous observation of its association with angiogenic processes. The proportional representation in this barplot reinforces the cellular dynamics of angiogenesis across the timepoints, further elucidating the angiogenic development patterns revealed by the UMAP visualizations.*

5.3.4. Cell-Cell Communication Analysis

5.3.4.1. General Network Structure

This analysis was conducted to identify crucial ligand-receptor pairs and the most active signaling pathways within a cellular communication network. The primary objectives were to pinpoint the main cell groups contributing to network signaling and to illuminate the critical pathways that facilitate these interactions.

Using CellChat, gene expression data was integrated with established signaling pathways to effectively map intercellular communications. This method helped determine which cell groups are central to the network's signaling and which pathways are essential in mediating these interactions. Although this analysis provides a comprehensive overview, it does not exhaust every detail due to the data's complexity. Nonetheless, it lays a solid foundation for further research into these vital signaling pathways and their roles in health and disease.

A significant limitation of this study is the absence of analysis subsetting into different conditions such as niche_early or niche_late. While crucial, the global comparison of different networks at this level is not feasible with the current data, making further analysis and interpretation challenging.

The circle plots included below offer a visual overview of the interaction frequency **[Figure 5-6A]** and strength **[Figure 5-6B]** across the full communication network. A key observation is that all cell populations, regardless of size, appear to contribute equally to the network. Notably, fibroblasts emerge as strong senders, while endothelial cells (ECs) are prominent receivers.

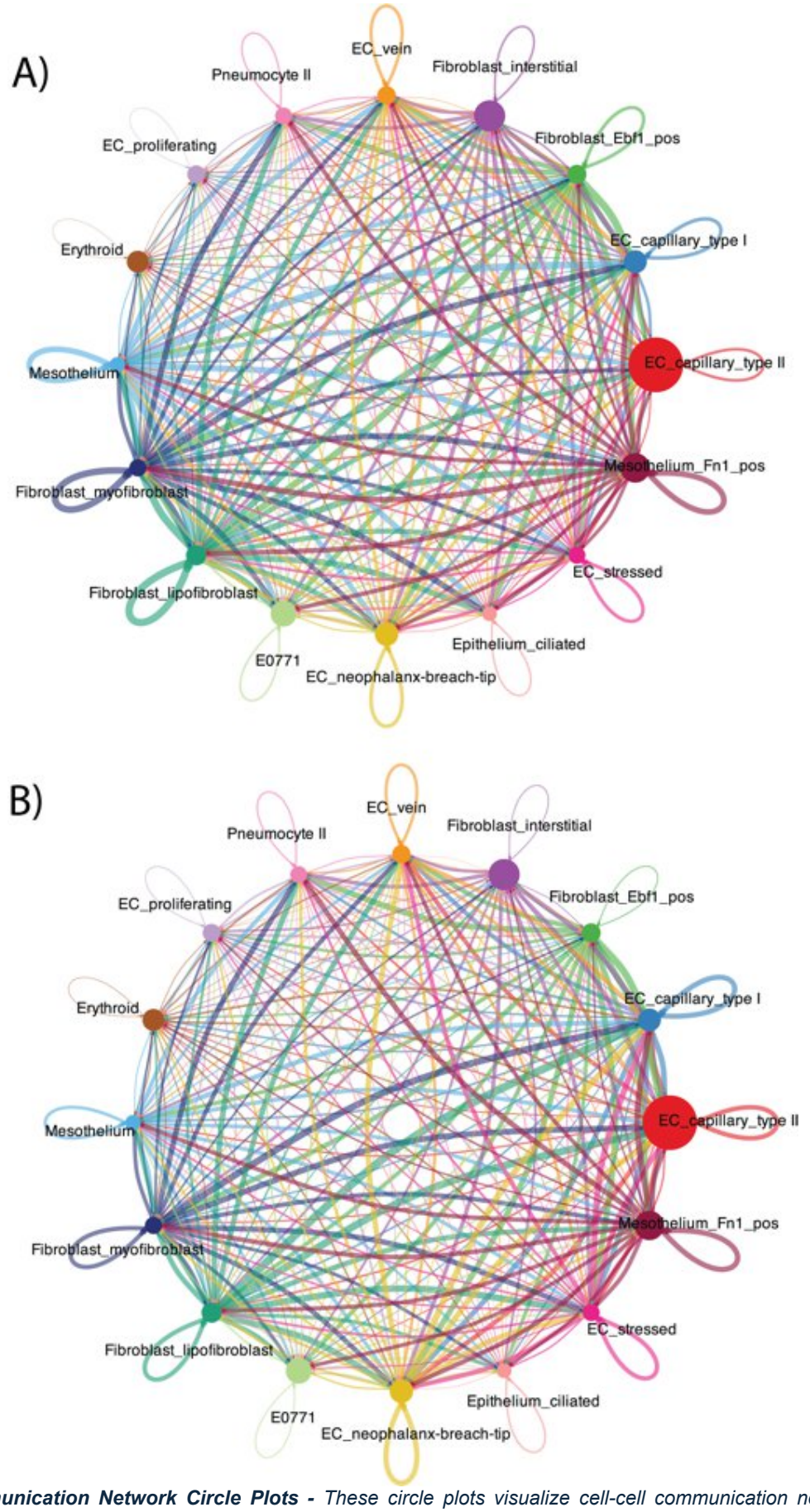


Figure 5-6: Communication Network Circle Plots - *These circle plots visualize cell-cell communication networks, illustrating interactions between different cell groups within the dataset.* ***Node Representation:*** *The size of each node represents the group size, with larger nodes indicating larger groups.* ***Link Representation:*** *Links between nodes represent ligand-receptor (LR) pairs and signaling pathways. The colour of each link corresponds to the originating node, enhancing the visual mapping of communication sources.* ***Weight Metrics: A)*** *The width of the links in panel A represents the number of interactions, with thicker lines indicating a higher number of connections between nodes.* ***B)*** *The width reflects the strength of the interactions, with thicker lines denoting stronger communications.*

To delve deeper, individual circle plots **[Figure 5-7]** were created to examine the connections initiated by different subtypes separately, providing more detailed insights into the dynamics of the network.

The individual circle plots distinctly illustrate the dynamics of cell-cell communication within the network, identifying key sender groups and their primary targets. Notably, fibroblasts, particularly myofibroblasts and lipofibroblasts, along with Mesothelium_FN_1 cells, emerge as the strongest signaling senders in the network. These fibroblasts most frequently communicate with endothelial cells, especially EC_capillary_type_II cells, indicating a significant interaction that holds profound biological implications.

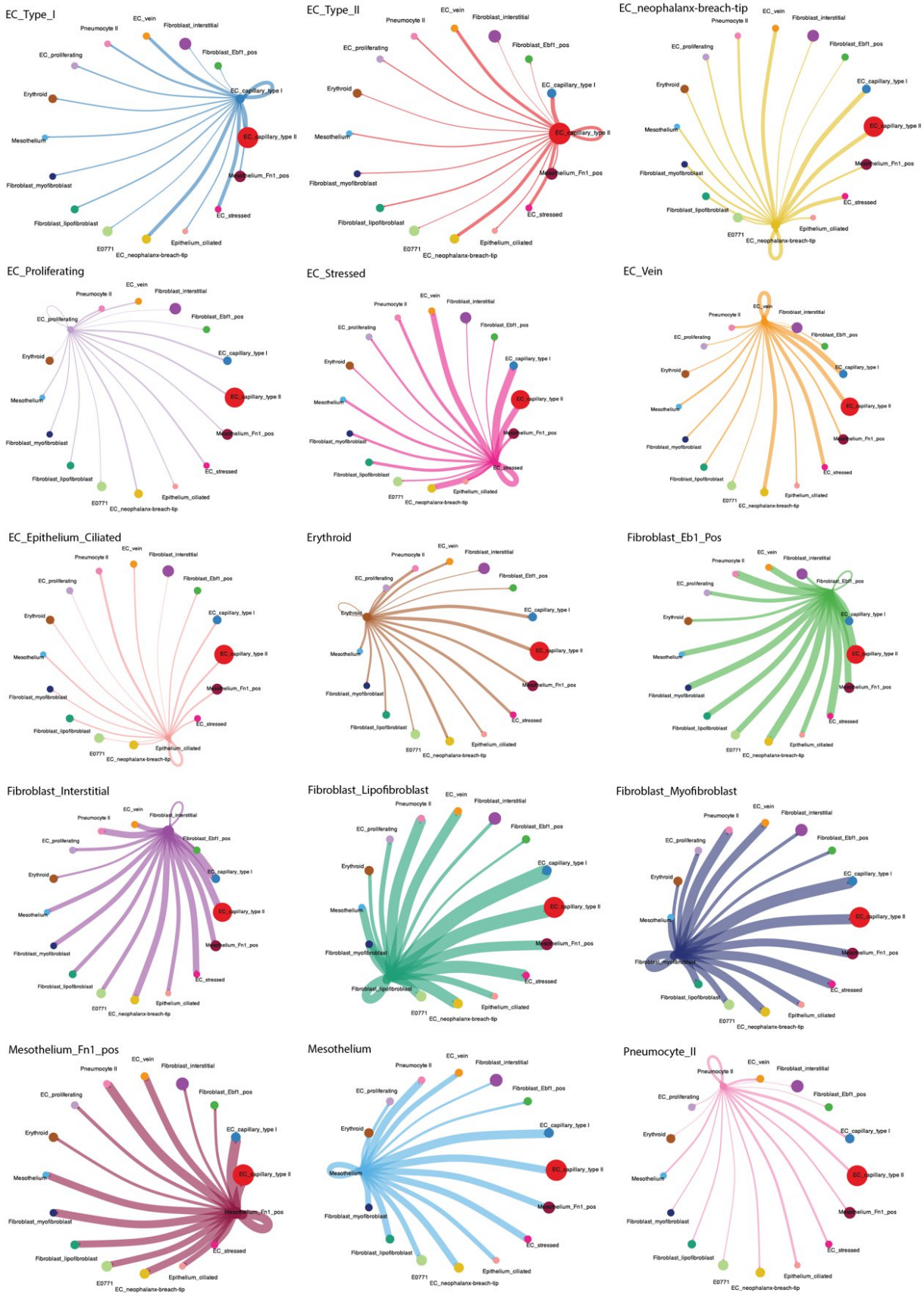


Figure 5-7: Individual Circle Plots of Cell Communication Networks - *These individual circle plots display the communication dynamics within the cellular network, emphasizing the primary sender groups and their main targets. Each plot corresponds to a different cell subtype, illustrating how each interacts within the network.* ***Node Size****: Indicates the relative size of each group, with larger nodes representing larger groups.* ***Link Colour:*** *Matches the colour of the group from which the signal originates, helping to trace the source of each communication pathway.* ***Link Width:*** *Represents the strength of the signaling, with wider links indicating stronger communication. These faceted plots provide a detailed view of the signaling crosstalk between different cell subtypes. They highlight not only the interactions but also identify the most influential signal senders within the network, offering insights into the dominant communication roles played by specific groups.*

Literature suggests that fibroblast signaling through VEGFA, which is mediated by the upregulation of the cytoplasmic hypoxia-induced angiogenesis regulator (HIAR), crucially enhances angiogenesis and blood vessel formation in co-cultured endothelial cells under hypoxic conditions (Kugeratski et al., 2019). This interaction underscores the vital role of fibroblasts in promoting vascular networks necessary for tissue viability and repair under hypoxic conditions (an inevitable byproduct of aggressive tumor growth and progression).

Additionally, mesothelial cells, which are also potent signaling senders within this network, have fibroblasts as one of their direct targets. This reciprocal signaling relationship between mesothelial cells and fibroblasts is intriguing and warrants further investigation. Exploring individual ligand-receptor pairs and signaling pathways in more detail will likely provide deeper insights into these complex intercellular interactions and the biological processes they regulate.

Before delving deeper into individual signaling cascades, identifying the primary senders and receivers within the network is crucial. Utilizing the ***CellChat::netAnalysis_signalingRole_heatmap()*** function, we generated heatmaps that help pinpoint these key participants. For incoming signals **[Figure 5-8A]**, the heatmaps highlight collagen, laminin, cyclophilin A (cypA), and VEGF as the top pathways involved. The prominent receivers are primarily endothelial cells, with EC_capillary_type_I, EC_capillary_type_II, and EC_stressed cells identified as significant targets of these signals.

In terms of outgoing signals **[Figure 5-8B]**, the same pathways are involved, with the addition of PECAM1. Here, mesothelial cells and various subtypes of fibroblasts, including fibroblast_ebf1_positive, fibroblasts_myofibroblasts, and fibroblasts_lipofibroblasts, are the main senders, playing vital roles in transmitting signals across the network.

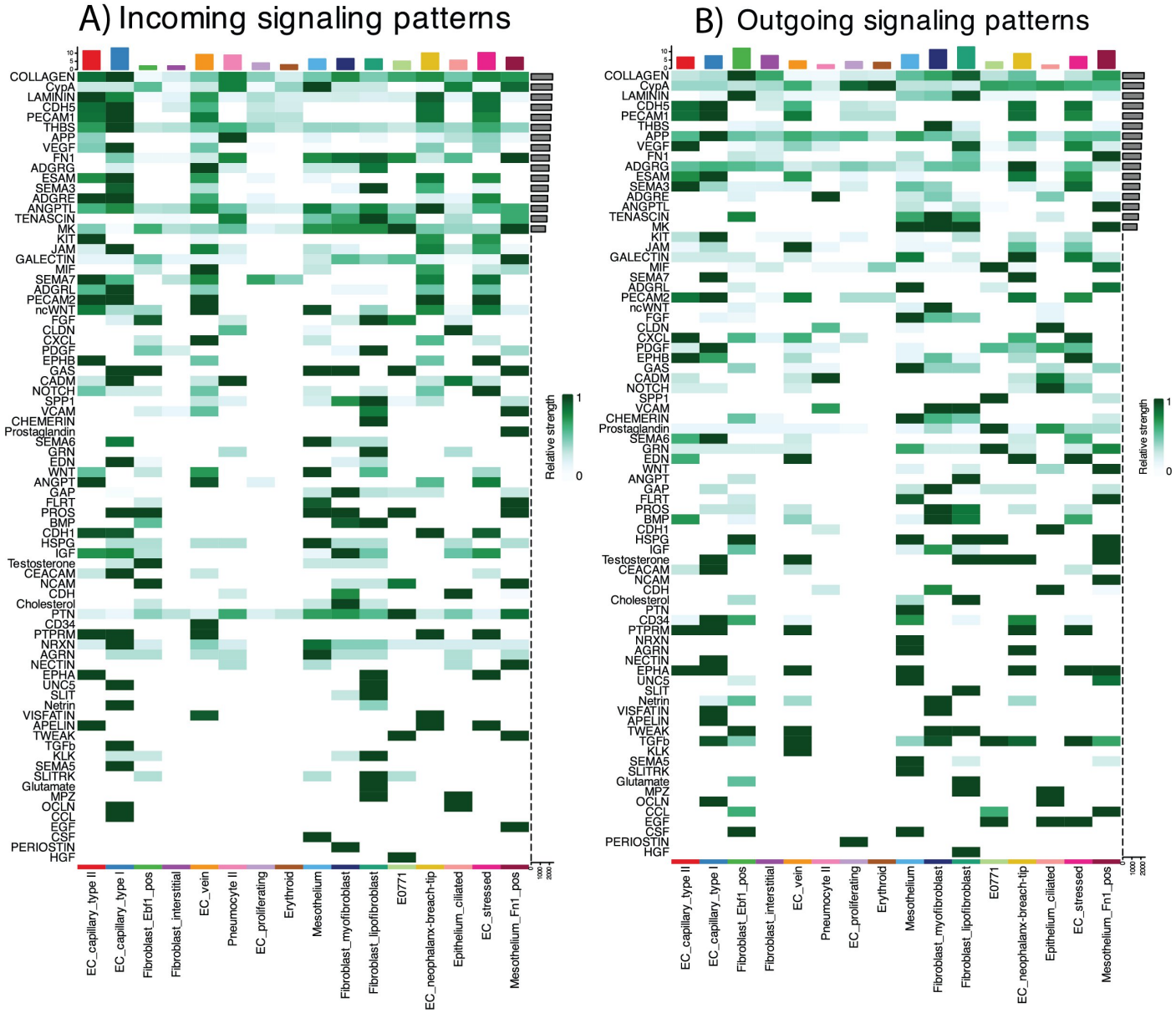


Figure 5-8: Global Heatmaps of Major Senders & Receivers in the Network - *This figure presents a comprehensive view of the communication dynamics within the network through two detailed heatmaps -* ***A) Incoming Signal Patterns:*** *This heatmap on the left shows the signaling strength received by various cell types from different pathways. The x-axis categorizes cell types, and the y-axis lists the pathways involved. The colour of the tiles represents the relative strength of the signals received. The top bar plot aggregates the total signaling strength received by each cell type across all pathways, providing a summary measure of incoming signals. The side bar plot compiles the overall signaling strength of each pathway, indicating how broadly each pathway impacts different cell types.* ***B) Outgoing Signal Patterns:*** *The heatmap on the right displays the signaling strength emitted by different cell types across various pathways. Similar to the incoming signals heatmap, the x-axis specifies cell types, and the y-axis identifies pathways. The colour intensity of the tiles indicates the strength of the signals sent. The top bar plot summarizes the total signaling strength sent by each cell type across all pathways, while the side bar plot shows the aggregate signaling strength of each pathway across all involved cell types.*

This analysis not only sheds light on the significant pathways and cell types involved in signaling but also sets a solid foundation for further exploration into how these interactions are orchestrated and their implications in the broader context of cellular communication.

In the subsequent sections, the analysis will shift to a detailed examination of individual signaling pathways of interest. This deeper dive aims to unravel the dynamics of communication within the cellular network by pinpointing the major senders and receivers. The objective is to identify and understand the key ligand-receptor (LR) pairs these cells use for communication.

5.3.4.2. VEGF Signalling Pathway: Primary Contributors & Key LR Pairs

The VEGF signaling pathway is critically examined due to its fundamental role in promoting angiogenesis and its potential involvement in aggressive metastatic invasion. This pathway is well-

known for mediating communication from cancer cells to endothelial cells, a key process in tumor progression. What makes this analysis particularly compelling is the focus on identifying the primary stromal subtypes within the TME that utilize the VEGF signaling pathway to communicate with endothelial cells, thereby facilitating their proliferation and contributing to neo-angiogenesis.

The provided chord plot **[Figure 5-9A]** and heatmap **[Figure 5-9B]** illustrate the dominant senders within the VEGF signaling pathway. Notably, cancer cells (labelled as E0771) emerge as strong senders, reflecting their central role in driving tumor growth and invasion. Additionally, endothelial cells display both paracrine and autocrine signaling patterns, indicating complex intra- and inter-cellular communication dynamics. Fibroblasts and mesothelial cells expressing fibronectin 1 (mesothelial_FN1_positive) are also highlighted as significant senders within this network.

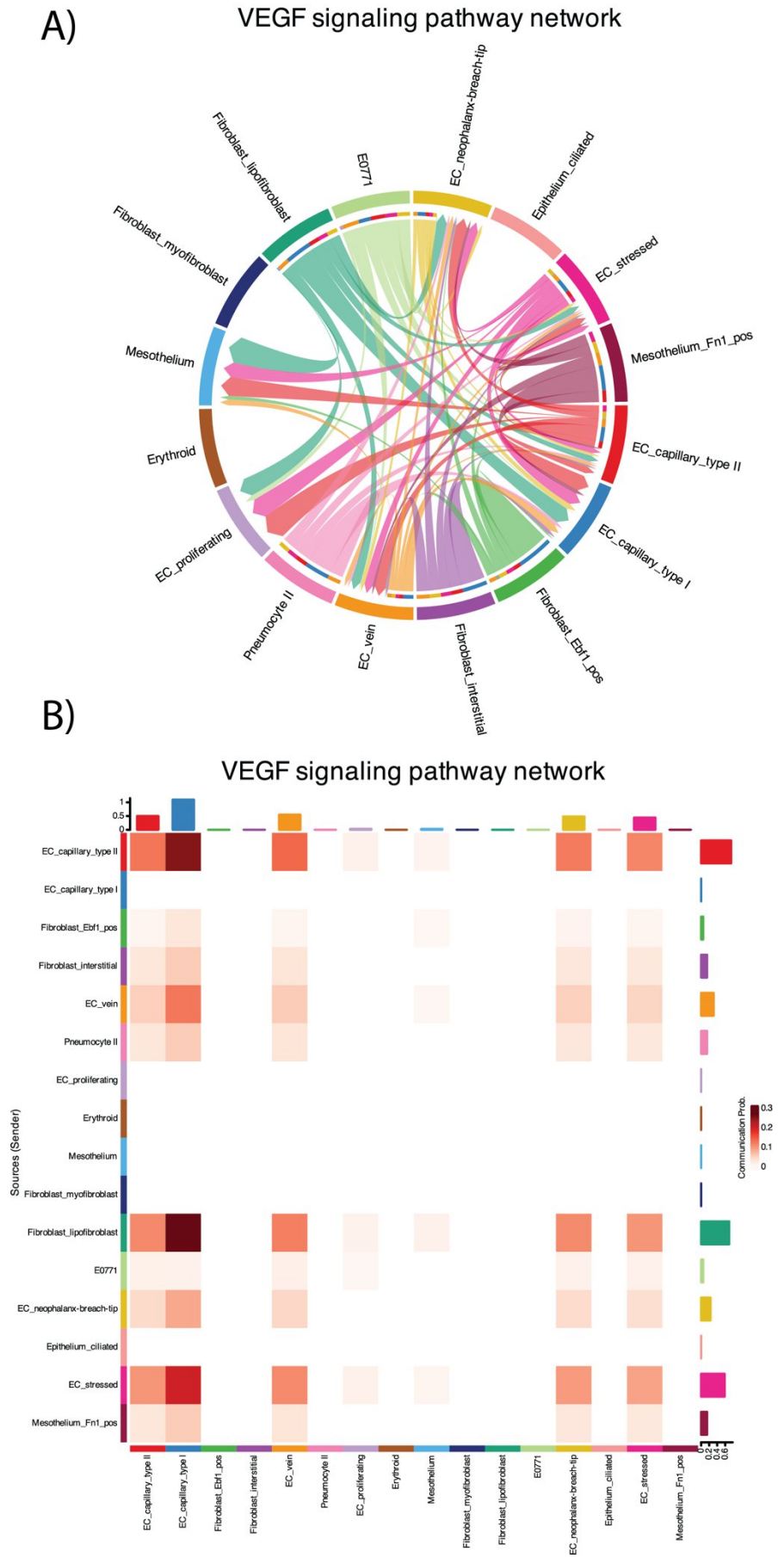


Figure 5-9: Visualization of VEGF Signaling Pathway Communication Dynamics - *This figure illustrates the communication dynamics within the VEGF signaling pathway, highlighting the primary senders and their respective targets.* ***A) Chord Plot:*** *Displays the top senders in the VEGF signaling pathway and the primary cell types receiving these signals. Each chord represents a communication link, with colors indicating the specific sender cell type, making it easy to visualize the flow of signals across the network.* ***B) Heatmap:*** *Complements the chord plot by presenting the same information in a matrix format. Here, the y-axis represents the sender cell types, and the x-axis should be corrected to show the receivers. The colors of the tiles within the heatmap indicate the communication probability, providing a quantitative view of the interaction strengths between different cell types.*

Interestingly, mesothelial cells are not only senders but also prominent receivers of VEGF signals. This dual role suggests a potentially unique and influential part in facilitating invasion and neo-angiogenesis within the TME. Further investigations into the VEGF signaling pathway have pinpointed the VEGFA-VEGFR1 pair as the most influential contributor to this cascade **[Figure 5-10A]**. This finding was substantiated by detailed visualizations using chord plots **[Figure 5-10B]** and circle plots **[Figure 5-10C]**, which mapped the primary senders and receivers utilizing this specific ligand-receptor pair. These visual tools corroborated previous observations across the entire pathway, reaffirming the central role of endothelial cells (ECs) in sustaining both paracrine and autocrine signaling mechanisms within the network. Moreover, these visualizations highlighted the significant involvement of fibroblasts in augmenting the VEGF signaling cascade.

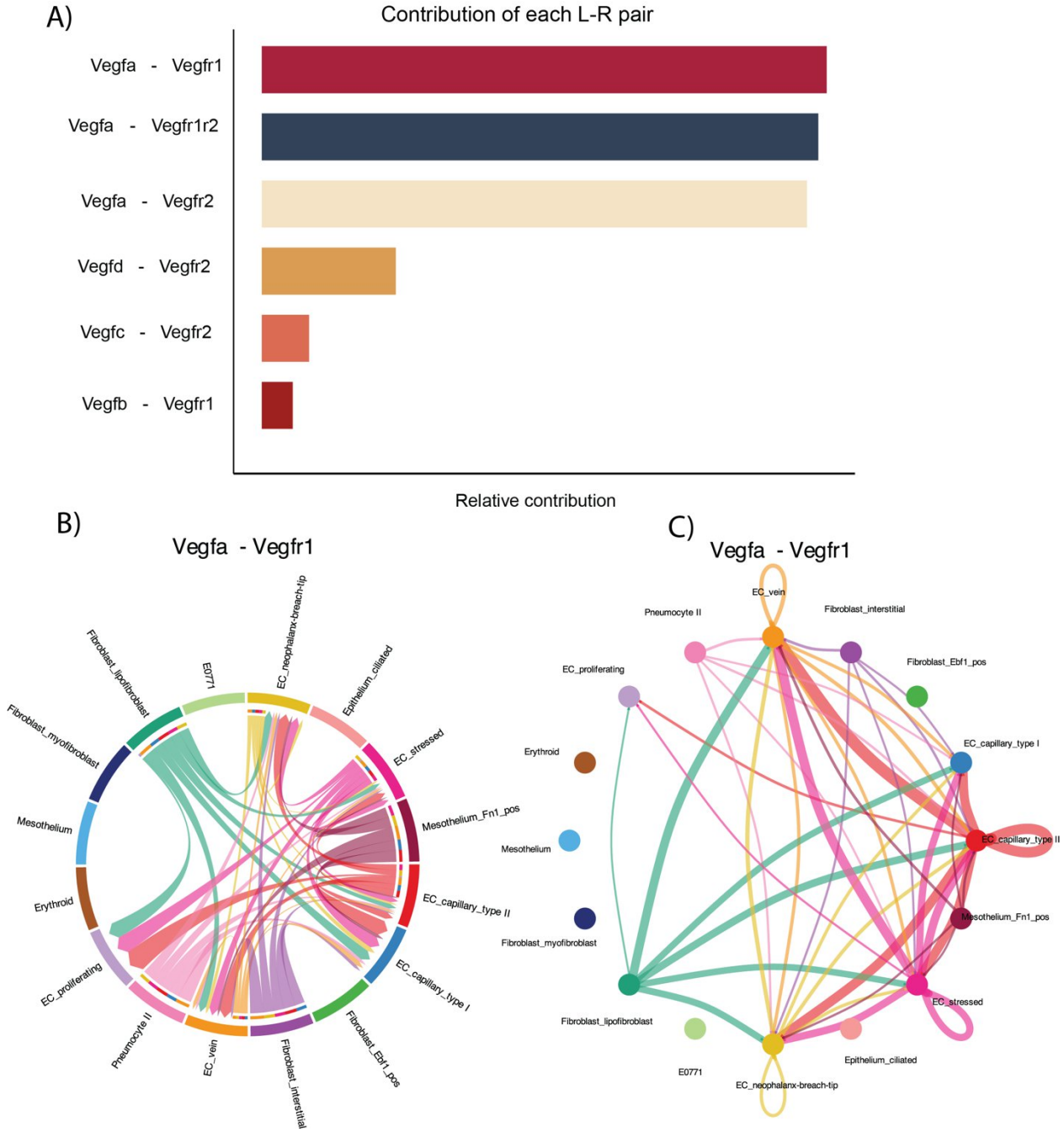


Figure 5-10: Analysis of Top Ligand-Receptor Pairs in the VEGF Signaling Pathway - *This figure highlights the dominant ligand-receptor (LR) pairs within the VEGF signaling pathway and visualizes the primary communication routes among cell types.* ***A) Bar Plots:*** *These plots illustrate the relative contribution of individual LR pairs to the overall VEGF signaling within the dataset. Notably, the pair VEGFA-VEGFR1 emerges as the most significant contributor, indicating its central role in mediating VEGF signaling.* ***B) and C) Chord and Circle Plots:*** *Both plots visualize the primary senders utilizing the VEGFA-VEGFR1 pair for signaling. The chord plot (B) maps the direction and strength of signals between sender and receiver cell types, while the circle plot (C) further elaborates these interactions, showing how signals circulate within the network. Each plot uses colors to represent different cell types, aiding in the identification of specific pathways and interactions facilitated by the VEGFA-VEGFR1 pair.*

5.3.4.3. Collagen Signaling Pathway: Primary Contributors

The collagen signaling pathway is examined next because of its significance in the network and because of the vital role collagen signaling plays in metastasis and angiogenesis.

The provided chord plot **[Figure 5-11A]** and heatmap **[Figure 5-11B]** illustrate the dominant senders within the Collagen signaling pathway. Fibroblasts seem to be the top senders in the network along with mesothelial cells with endothelial cells being the primary receivers. Interestingly, fibroblasts_myobfibroblasts receive a significant amount of collagen signaling, thereby contributing to both autocrine and paracrine signaling cascades.

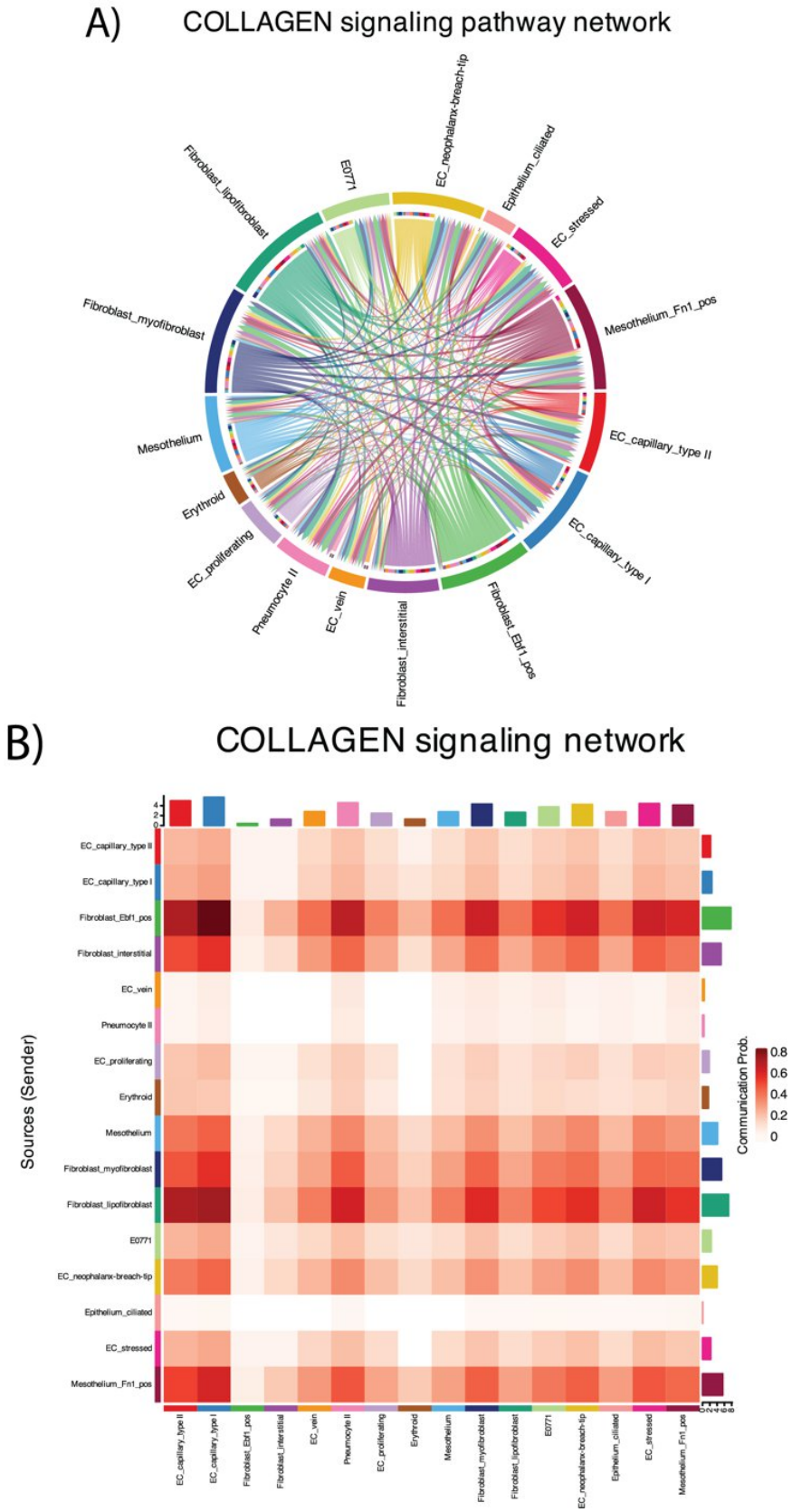


Figure 5-11: Visualization of Collagen Signaling Pathway Communication Dynamics *- This figure illustrates the communication dynamics within the Collagen signaling pathway, highlighting the primary senders and their respective targets -* ***A) Chord Plot:*** *Displays the top senders in the Collagen signaling pathway and the primary cell types receiving these signals. Each chord represents a communication link, with colors indicating the specific sender cell type, facilitating a clear visualization of the flow of signals across the network.* ***B) Heatmap:*** *Complements the chord plot by presenting the same information in a matrix format. The y-axis categorizes the sender cell types, while the x-axis lists the receivers, accurately reflecting the direction of signaling. The color intensity of the tiles within the heatmap indicates the communication probability, providing a quantitative view of the interaction strengths between different cell types.*

Further investigations into the Collagen signaling pathway have pinpointed the Col1a1 - Itga1+Itgb1 pair as the most influential contributor to this cascade **[Figure 5-12A]**. This finding was substantiated

by detailed visualizations using chord plots **[Figure 5-12B]** and circle plots **[Figure 5-12C]**, which mapped the primary senders and receivers utilizing this specific ligand-receptor pair.

In gastric cancer, Col1a1overexpression has been identified through TCGA data analysis, where it is associated with poor prognosis. Further investigation revealed that Col1a1 signaling activates the BCL9L/β-catenin pathway via ITGB1. Col1a1, primarily produced by fibroblasts, interacts with ITGB1, which is abundantly expressed on the surfaces of cancer cells. This interaction promotes enhanced colony formation, indicative of aggressive cancer progression (Lv et al., 2021).

Supporting these findings, our analysis demonstrates that E0771 cells, a model for gastric cancer cells, are potent receivers of general collagen signaling **[Figure 5-11B]**. However, it's important to note that while these cells are significantly impacted by collagen pathways, they do not directly contribute to the Col1a1 - Itga1+Itgb1 ligand-receptor interaction **[Figure 5-12B&C]**. This distinction highlights that while E0771 cells benefit from the pro-tumorigenic environment fostered by Col1a1 signaling, they are not the originators of this critical ligand-receptor interaction.

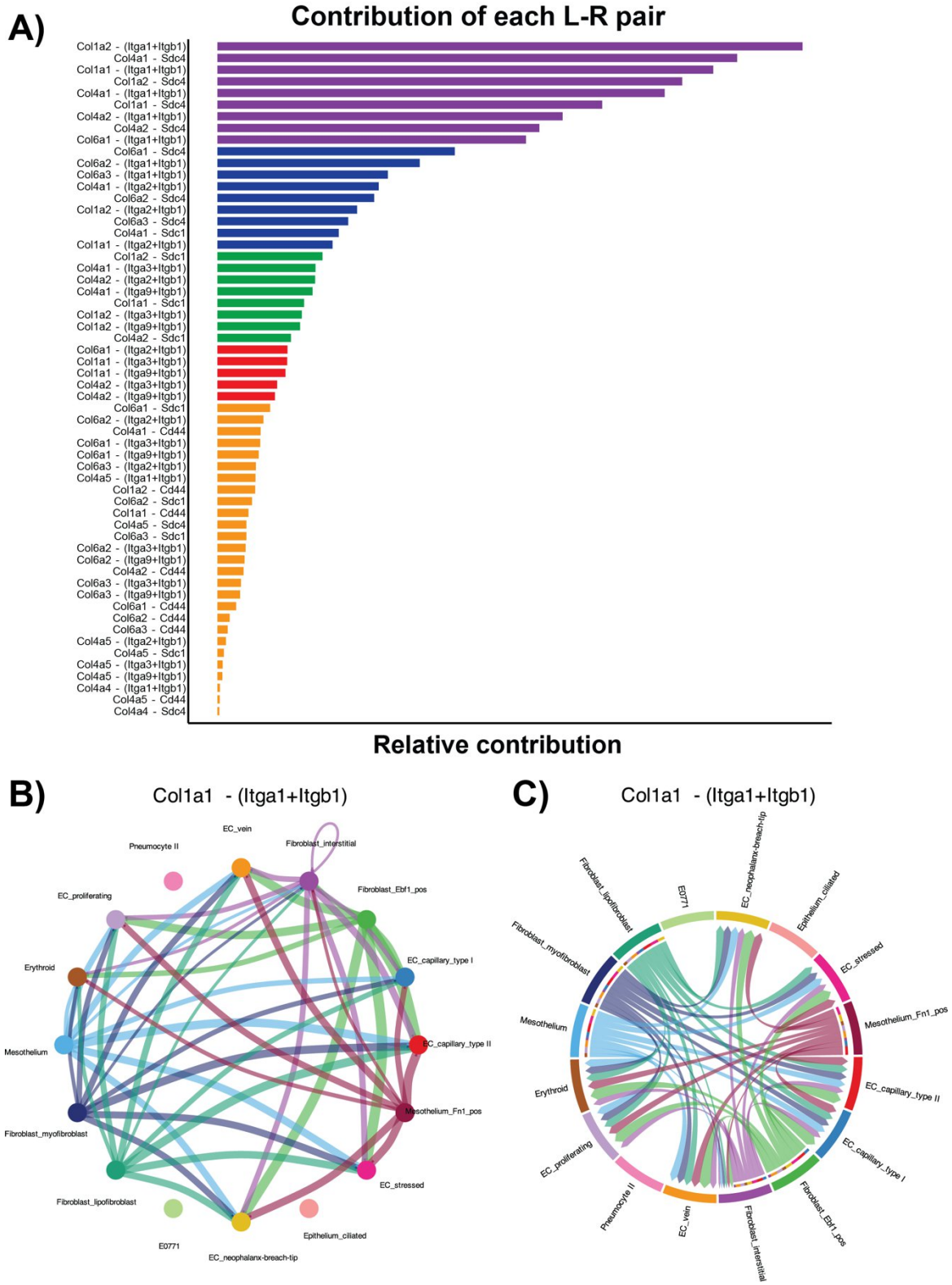


Figure 5-12: Analysis of Top Ligand-Receptor Pairs in the Collagen Signaling Pathway - *This figure highlights the dominant ligand-receptor (LR) pairs within the Collagen signaling pathway and visualizes the primary communication routes among cell types:* ***A) Bar Plots:*** *These plots illustrate the relative contribution of individual LR pairs to the overall Collagen signaling within the dataset. Notably, the pair Col1a1 - Itga1+Itgb1 emerges as the most significant contributor, indicating its central role in mediating Collagen signaling.* ***B) and C) Chord and Circle Plots:*** *Both plots visualize the primary senders utilizing the Col1a1 - Itga1+Itgb1 pair for signaling. The chord plot (C) maps the direction and strength of signals between sender and receiver cell types, while the circle plot (B) further elaborates these interactions, showing how signals circulate within the network. Each plot uses colors to represent different cell types, aiding in the identification of specific pathways and interactions facilitated by the Col1a1 - Itga1+Itgb1 pair.*

5.3.5. Endothelial Cell Subtype Heterogeneity Sub-Clustering Analysis

5.3.5.1. Data Structure

Given the marked increase in endothelial cell breaches (EC_breach cells) from early to late niche conditions, a more detailed examination of the endothelial compartment was deemed necessary. This examination aimed to dissect the dynamics within this compartment, free from the interference of other cell types. Consequently, a focused strategy involving isolation and sub-clustering of endothelial cells was adopted. This method involved segregating the endothelial cells for preprocessing, followed by clustering analysis, to gain deeper insights into the variations among these cells across different conditions.

To this end, endothelial cells were isolated, and standard procedures such as normalization, log-transformation, and dimensionality reduction were applied as previously described. The analysis of clusters was performed using the Louvain clustering algorithm. Additionally, the clustering resolution was fine-tuned using clustering trees **[Figure 5-13A&B]**, also previously detailed. A clustering resolution of 0.5 was selected to maintain consistency with the main dataset and to strike a balance between stability and heterogeneity of the data clusters. It is crucial to highlight that the clustering exhibited some instability at lower resolutions within this subset, although this phenomenon was not further explored in the study.

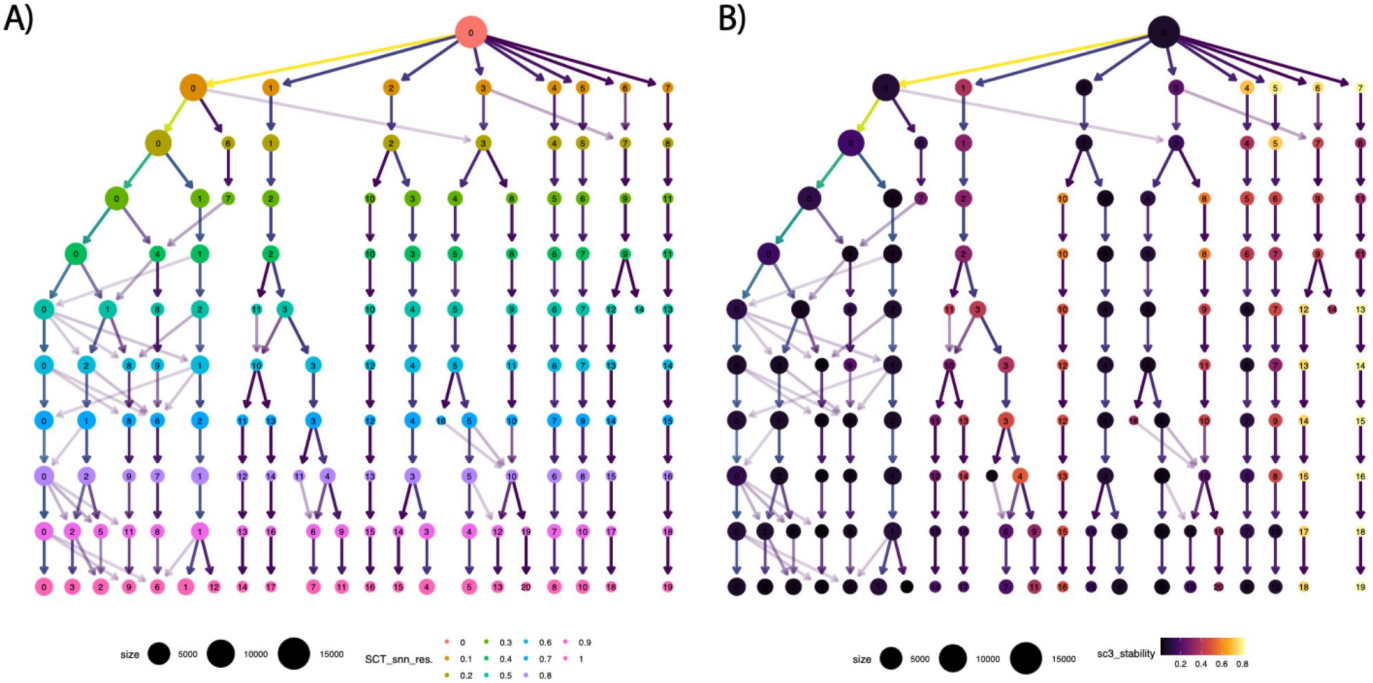
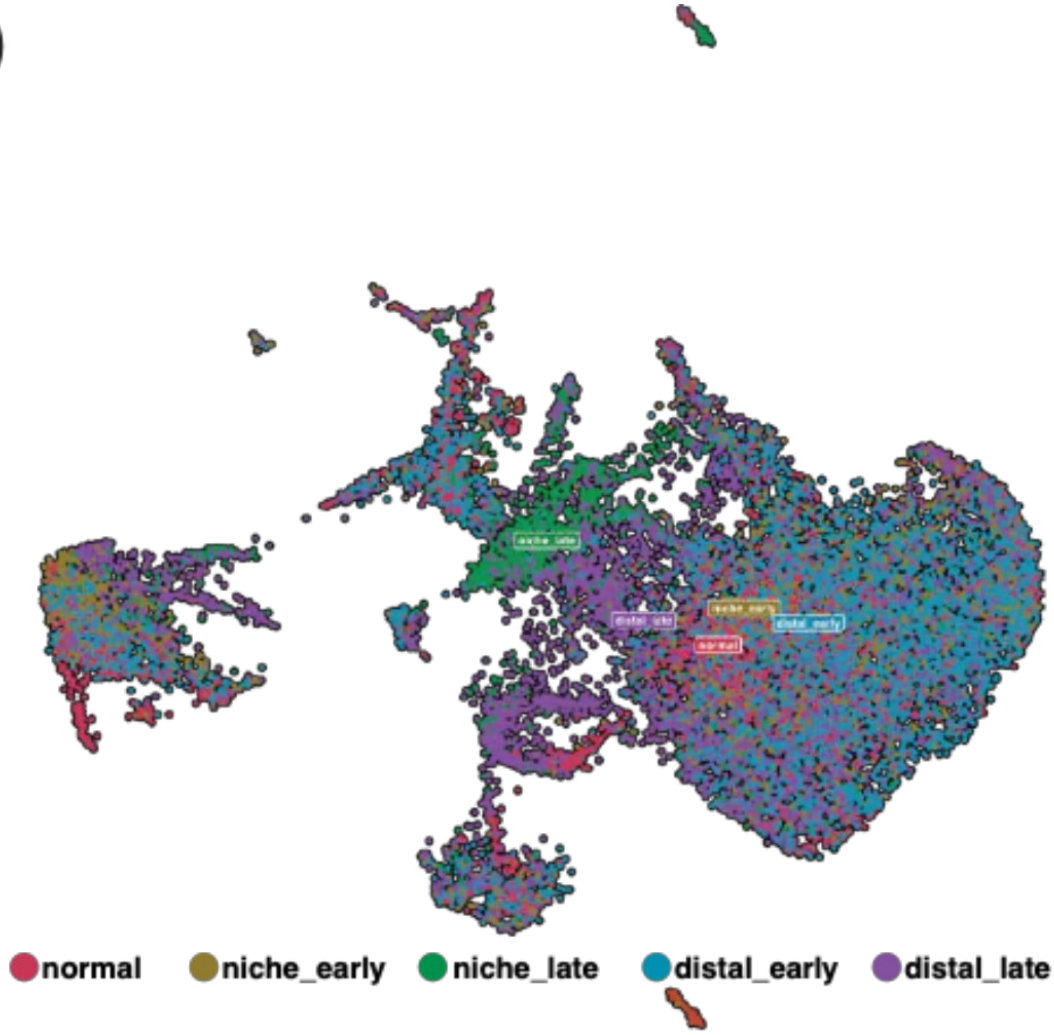


Figure 5-13: Resolution Trees for Endothelial Cell Subclustering Analysis - *A) Clustering tree illustrating subgroups within the endothelial cell population, coloured by clustering resolution. Each circle's colour represents a distinct clustering resolution level, while the size of the circle indicates the relative size of each cluster within the resolution. B) Similar clustering tree as in panel A, but here, each circle is coloured according to the SC3 stability index, illustrating the stability of clusters across different resolutions. This color-coding aids in identifying which clusters remain consistent and which are more variable when the resolution settings are adjusted.*

The UMAP plots annotated by both condition **[Figure 5-14A]** and cell subtype **[Figure 5-14B]** (with subtype labels derived from the original analysis of the main data) clearly demonstrate that the primary source of variation in the UMAP is cell subtype identity rather than condition, which aligns with expectations.

A)


normal
niche_early
niche_late
distal_early
distal_late

B)

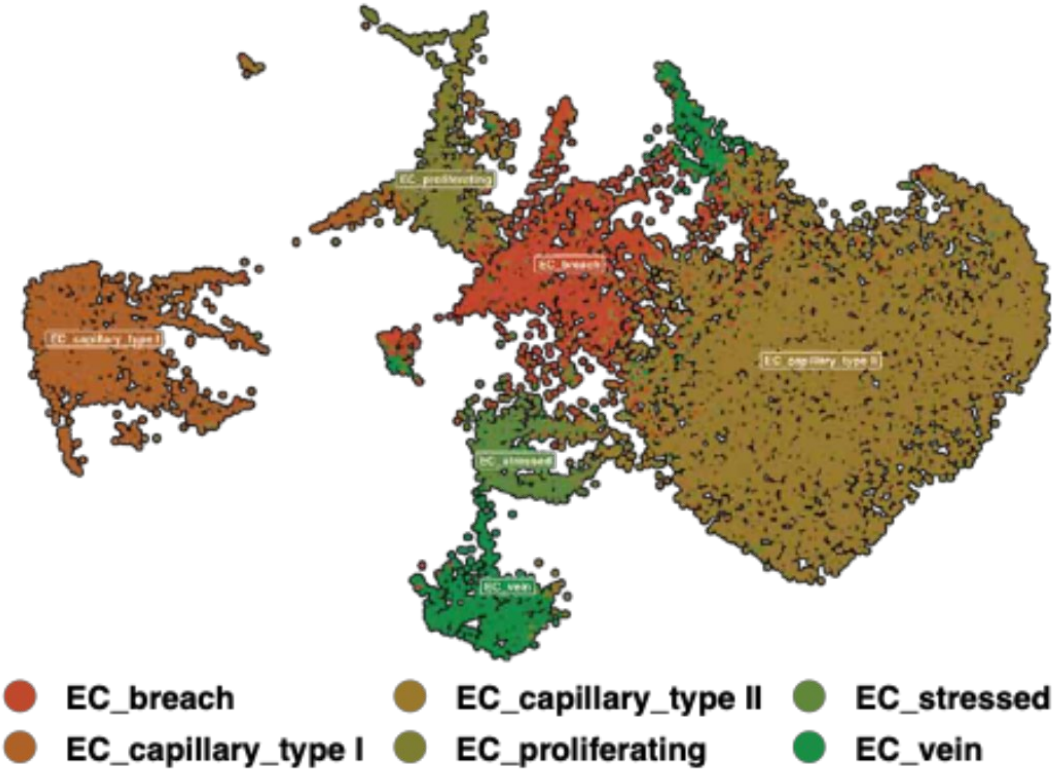

EC_breach
EC_capillary_type I
EC_capillary_type II
EC_proliferating
EC_stressed
EC_vein

Figure 5-14: UMAP Plots of Endothelial Subclustering Analysis - *A) UMAP plot of the endothelial subset of the data, annotated by condition. Each point represents a cell, coloured according to its specific condition: normal, niche_early, niche_late, distal_early, and distal_late. This visualisation aids in understanding how endothelial cells cluster based on their condition. B) UMAP plot of the same endothelial subset, coloured by the original endothelial subtype labels from the main dataset analysis. This panel provides insight into the distribution of endothelial subtypes across different conditions, highlighting the consistency and variability in subtype classification.*

However, a composition barplot **[Figure 5-15]** offers a more detailed perspective, revealing significant enrichment of the EC_breach compartment in the niche_late condition. Additionally, there is a notable increase in the EC_breach population during the transition from distal_early to distal_late, although this increase is less pronounced compared to the changes observed in the niche compartment. These alterations are also visible in split UMAP plots **[Figure 5-16]**, which illustrate the distribution of different endothelial subtypes across various conditions within the UMAP space.

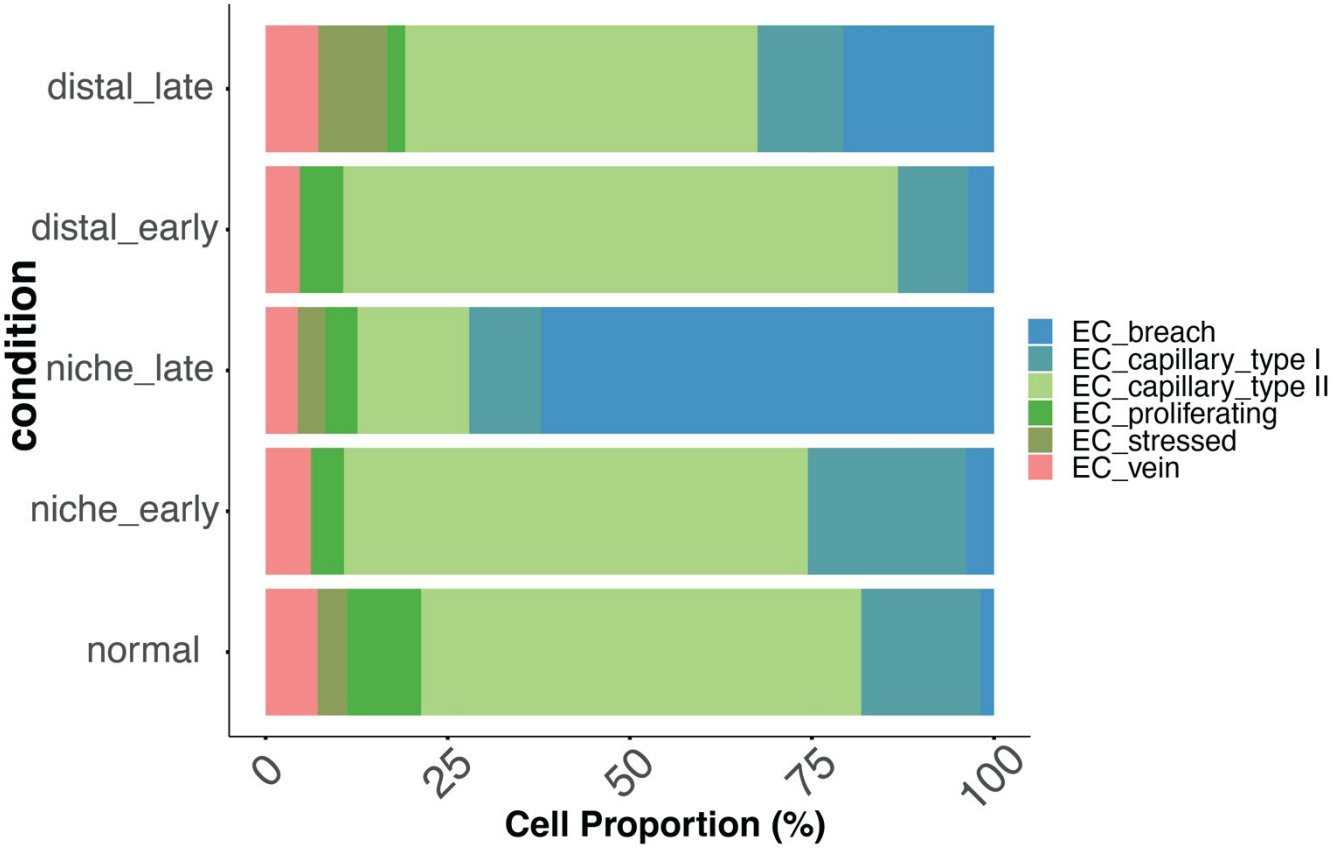


Figure 5-15: Composition Plot for Endothelial Subset *This bar plot illustrates the composition of endothelial subtypes across different conditions: normal, niche_early, niche_late, distal_early, and distal_late. Each bar represents a condition, and the segments within each bar correspond to the proportion of various endothelial subtypes: EC_breach, EC_capillary_type I, EC_capillary_type II, EC_proliferating, EC_stressed, and EC_vein, as labelled in the original dataset. This visualisation highlights the observed expansion of the EC_breach compartment from niche_early to niche_late, which was the primary observation driving this analysis. A similar, though less pronounced, increase in the EC_breach compartment can also be seen in the transition from distal_early to distal_late. This plot underscores the dynamic changes in endothelial subtype composition across different conditions, providing a foundation for further analysis and refinement of cell subtype identities in subsequent studies.*

Another significant observation between the niche_early and niche_late compartments is the substantial reduction in EC_capillary_type II cells. This downregulation was not evident in the main dataset and represents a new finding that merits further investigation.

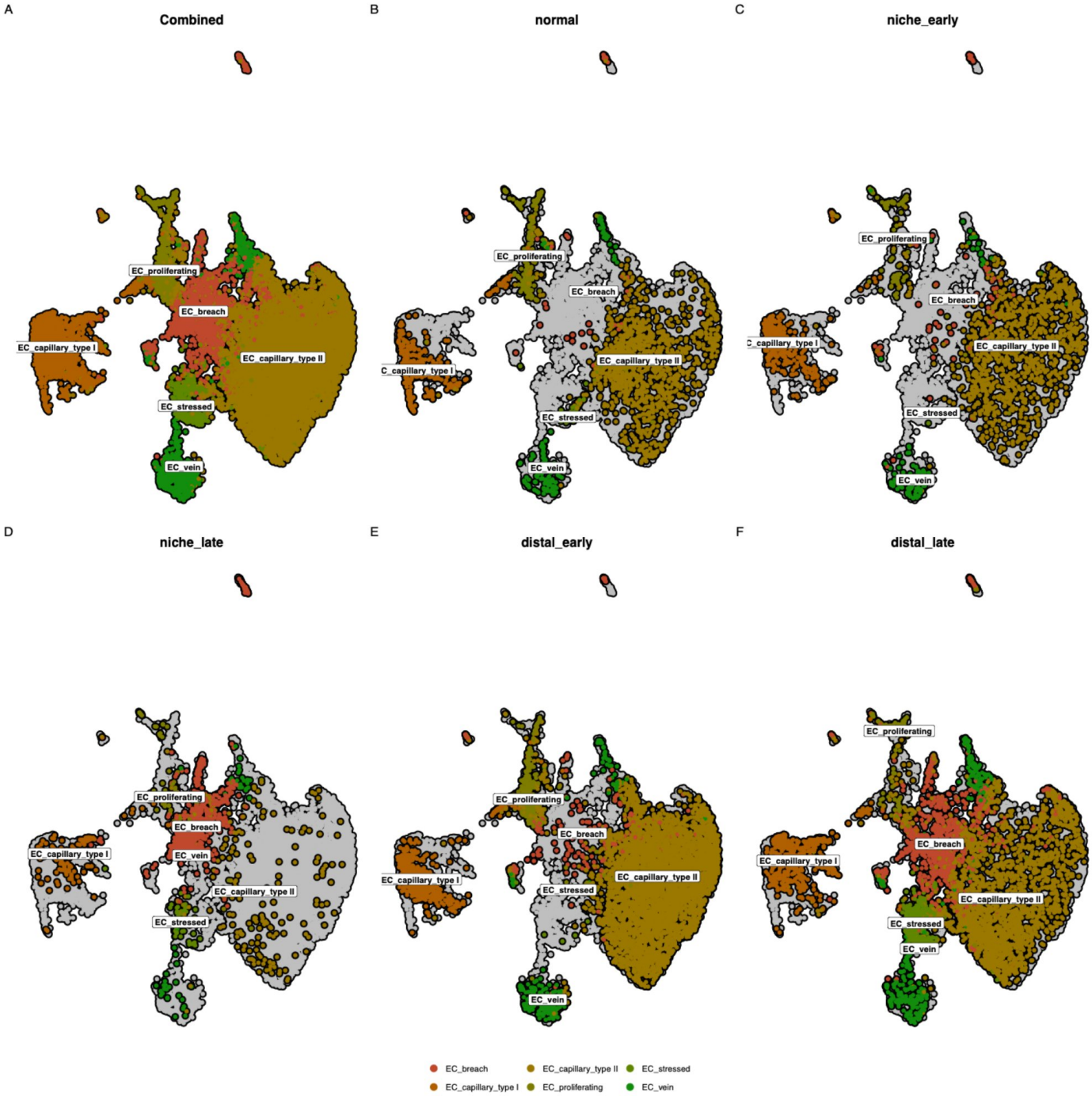


Figure 5-16: Split UMAP Plot – Distribution of Endothelial Subtypes by Condition *This figure presents UMAP plots split by condition—normal, niche_early, niche_late, distal_early, and distal_late—to illustrate the distribution of various endothelial subtypes across different conditions. Each plot highlights the distribution of subtypes, including EC_breach, EC_capillary_type I, EC_capillary_type II, EC_proliferating, and EC_stressed cells. Notably, the plots emphasize the significant rise in the EC_breach compartment from niche_early to niche_late, and a similar, though less pronounced, increase from distal_early to distal_late. This expansion of the EC_breach compartment is a focal observation in this analysis. Additionally, a substantial reduction in EC_capillary_type II cells is observed between the niche_early and niche_late conditions, indicating a compositional shift that warrants further investigation.*

5.3.5.2. Endothelial Subtype Annotation Refinement

The next step in this analysis was the annotation of the resulting Seurat clusters to further refine the identification of endothelial subtypes. The rationale behind this approach is that these clusters represent more specific subpopulations within the endothelial subtypes, whose identification and characterization should enhance our understanding of how the metastatic microenvironment evolves over time.

Differential expression analysis (DEA) was performed using ***Seurat::FindAllMarkers()*** for all the different clusters. Cluster annotation was conducted using a list of known cell subtype markers

curated from the literature, and these are shown in the marker table below. This curation of cell subtype markers and subsequent cluster annotation were carried out by Salwa Lin.

The Expression UMAP plots of marker genes [**Figure 7-7**] for various identified subtypes are designed to aid in verifying the accuracy of the annotations performed. The final UMAP, coloured by the refined subtypes, reveals significant endothelial heterogeneity that was not characterized in the main dataset **[Figure 5-17]**. It is essential to define each subtype individually. **EC_Capillary_Type I** (aerocyte capillary) are endothelial cells of capillaries specializing in facilitating gas exchange and work alongside alveolar type I epithelial cells to create an efficient interface for oxygen and carbon dioxide diffusion. They also contribute to pulmonary immune defence by facilitating leukocyte migration through the expression of adhesion markers, as highlighted in (Gillich et al., 2020). Conversely, **EC_Capillary_Type II** (general capillary) cells, less specialized for gas exchange, play various roles in the lung, such as acting as stem cells for damaged capillary regeneration (angiogenesis) and controlling vasomotor activities, as noted in (Gillich et al., 2020).

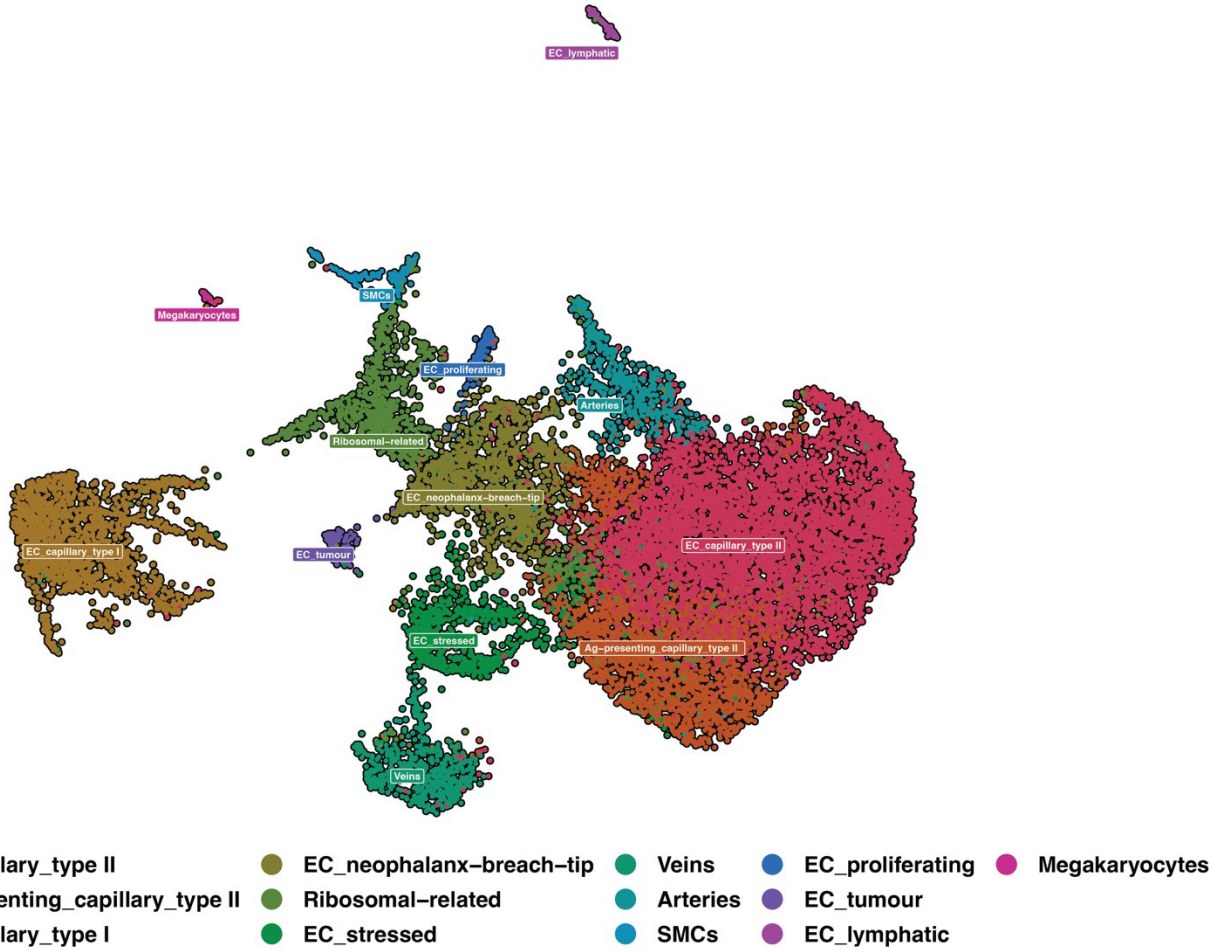


Figure 5-17: UMAP Plot Annotated by Refined Cell Subtypes - *This UMAP plot illustrates the refined annotation of endothelial subtypes, revealing multiple new subtypes that significantly contribute to endothelial heterogeneity across different conditions. The refined subtypes include capillary type I, capillary type II, arterial cells, venous cells, EC_breach cells, EC_stressed cells, proliferating ECs, ribosomal-related genes, SMA or myofibroblasts, tumour ECs, and megakaryocytes.*

Arterial ECs are endothelial cells lining the walls of pulmonary arteries, crucial for regulating arterial blood flow and pressure within the lungs. **Venous ECs**, which line the walls of pulmonary veins, are involved in returning oxygen-depleted blood to the heart, integral to the circulatory pathway. **Breach ECs** show significant invasive potential, capable of breaching physical barriers, which suggests a role in metastasis or tissue invasion. **Stressed ECs** are under significant environmental pressure from tumor growth, exhibiting upregulated stress markers that reflect their struggle to adapt to a hostile environment dominated by cancer cell proliferation and invasion.

Proliferating ECs display numerous proliferative markers and exhibit marked proliferation, often driven by angiogenic signaling activated in response to hypoxia by cancer cells. **Myofibroblasts** represent a sub-class of fibroblasts with contractile functionality; traditionally crucial in wound healing and tissue repair, their role can be subverted to drive tumor microenvironment (TME) reshaping and cancer invasion through their ability to express and release MMPs and remodel the TME with mechanical pressure. **Megakaryocytes**, platelet progenitors that typically reside within the bone marrow but also mimic endothelial cells in terms of surface antigens, cytoplasmic signaling, and regulatory pathways, are present in the TME of lung metastases, raising intriguing questions about their roles and interactions in such settings.

Table 3: Endothelial Subtypes Marker Features - *This table summarizes the marker features for various endothelial subtypes, curated from extensive literature reviews and annotated based on differential expression analysis (DEA) results. The curation and annotation process was meticulously carried out by Salwa Lin.*

Annotated cell type	DEGs		Reference
Vascular population			
Capillary type I ECs	*Emp2, Car4, Fibin, Cyp4b1*		(Gillich et al., 2020; Goveia et al., 2020)
Capillary type II ECs	*Plvap*		(Gillich et al., 2020; Vila Ellis et al., 2020)
	Sema3c, Hmcn1, Clec14a		(Gillich et al., 2020; Goveia et al., 2020)
IFN-Ag-presenting Capillary type II ECs	MHC-II antigen presentation marker genes:	*CD74, H2-Ab1, H2-Eb1,*	(Wolf et al., 2018)
		H2-DMa, H2-DMb1	(Goveia et al., 2020; Roewe et al., 2020)
	Capillary type II ECs marker genes: *Sema3c, Pltp, Hmcn1*		(Gillich et al., 2020; Goveia et al., 2020)
Veins	*Vwf, Slc6a2, Amigo2*		
Arteries	*Atp13a3, Ltbp4, Gja5, Bmx*		
Lymphatic ECs	*Maf, Prox1*		
	Ccl21a		(Goveia et al., 2020)
Stressed ECs	Stressed-induced marker genes: *Ucp2, Dusp1, Ifitm3*		(L. Zhao et al., 2020)
	EC marker genes: *Klf4, Klf2, Foxf1*		(Goveia et al., 2020)
Proliferating ECs	*Top2a, Mki67, Ube2c*		(Goveia et al., 2020)
Tumour ECs	*Fabp4, Cd300lg, Fabp5*		
Neophalanx/breach/tip ECs	Neophalanx ECs marker genes: *Sparcl1, S100a6*		
	Breach ECs marker genes: *Col4a1, Col4a2*		
	Tip ECs marker genes: *Tmsb10, Sparc, Trp53i11*		
Ribosomal-related cells	*Rpl3, Rpl7a*		
SMCs or pericytes	SMCs marker genes: *Cox4i2, Postn, Pdgfrb, Tagln, Acta2, Myh9*		(X. Liu et al., 2021; Teuwen et al., 2021)
	Pericytes marker genes: *Cox4i2, Postn, Pdgfrb, Acta2, Myh9*		(X. Liu et al., 2021; Teuwen et al., 2021)
Megakaryocytes	*Ppbp, Nrgn, Plek, Alox12*		(Angelidis et al., 2019)

5.3.5.3. Endothelial Subtype Composition Distribution

Following the annotation of these endothelial sub-clusters, the next logical step of the analysis was a characterisation of the distribution these refined subtypes across conditions. This was done using composition barplots as well as split UMAP plots.

The main findings from UMAP plots **[Figure 5-18]** and the composition barplots **[Figure 5-19]** provide a comprehensive overview of the dynamic changes within the endothelial cell subtypes across various conditions, highlighting specific patterns of interest. Particularly notable is the

significant enrichment of EC_breach cells observed moving from niche_early to niche_late conditions, indicating an increase in invasive potential that may be linked to pathological changes within the endothelial landscape. This is contrasted with the marked decrease in capillary types I and II cells in these same conditions, suggesting a shift away from typical vascular functions towards more aggressive phenotypes under metastatic conditions. Additionally, the minimal presence of EC_breach cells in the normal condition serves as a stark baseline, emphasizing the drastic alterations that occur in response to cancer cell invasion.

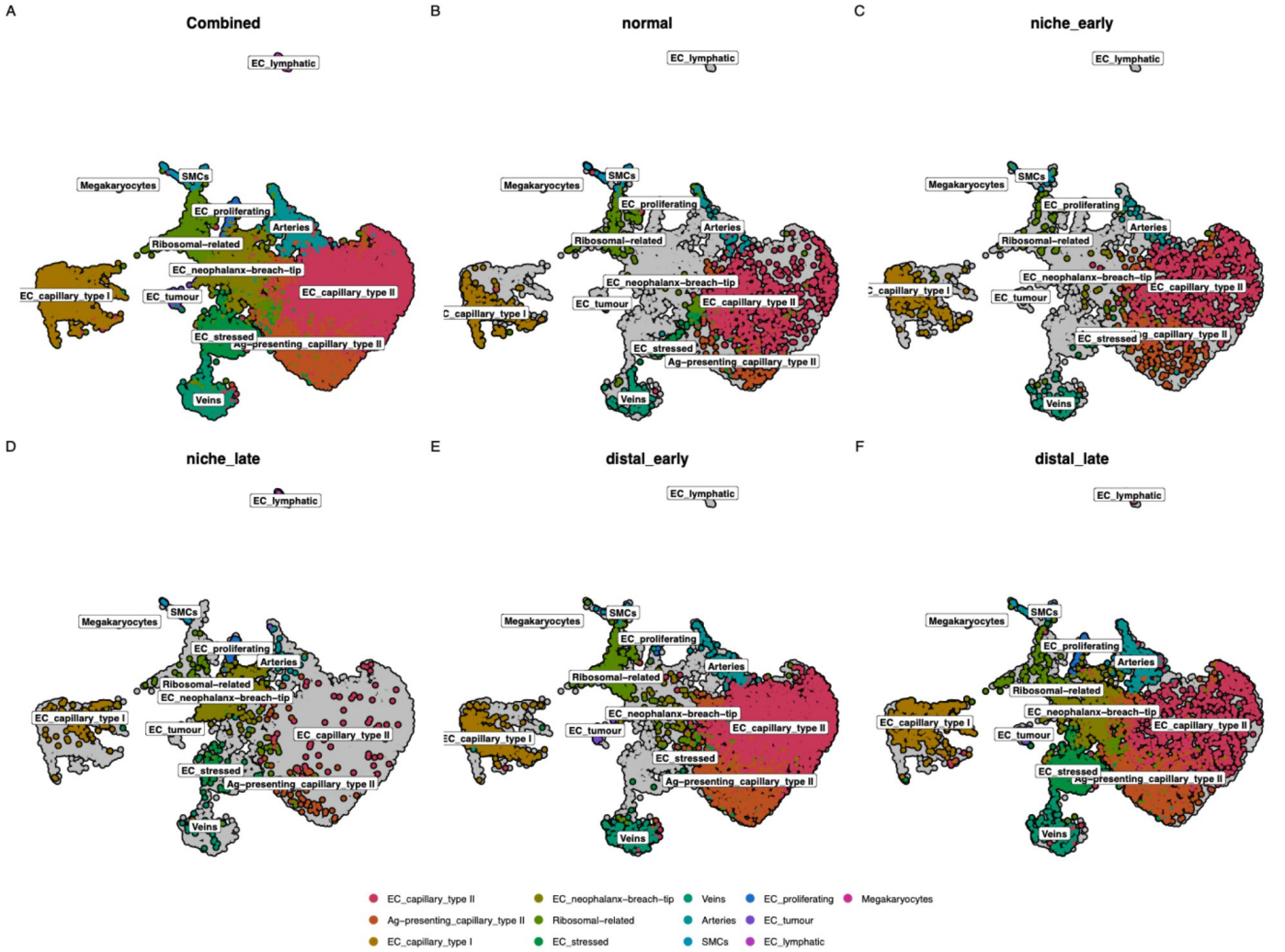


Figure 5-18: Split UMAP Plots by Condition - *These UMAP plots demonstrate the distribution of newly refined endothelial subtypes across various conditions: normal, niche_early, niche_late, distal_early, and distal_late. The plots highlight significant changes in subtype distributions through different stages, most notably the marked increase in EC_breach cells from niche_early to niche_late and the corresponding decrease in capillary types I and II cells in these conditions. It is also important to highlight that the normal condition, which serves as the baseline or background, shows a very minimal presence of EC_breach cells.*

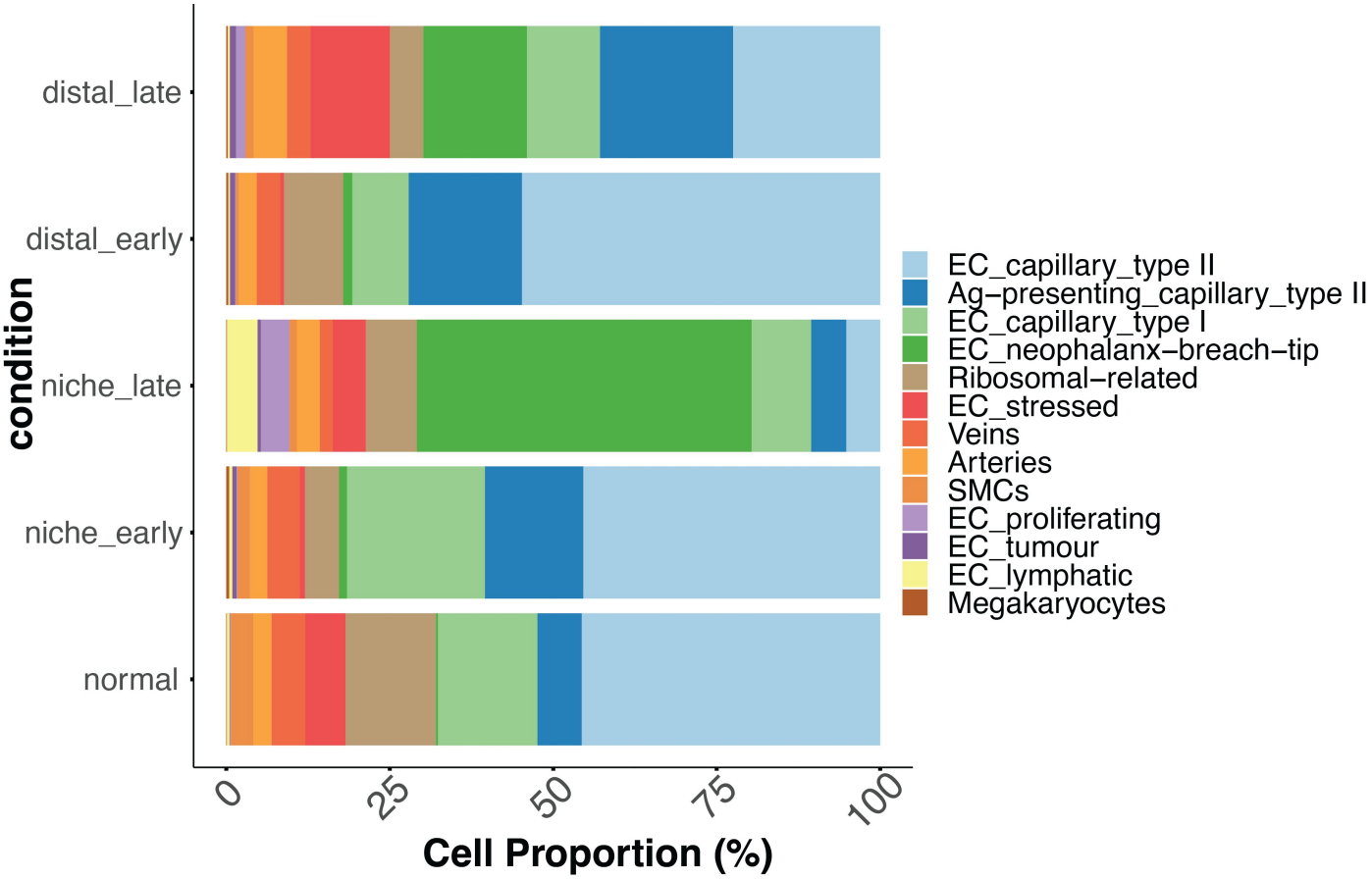


Figure 5-19: Composition Barplot for Refined Endothelial Subtypes Across Conditions *- This barplot visualizes the cell proportion (%) of refined endothelial subtypes across different conditions: normal, niche_early, niche_late, distal_early, and distal_late. Notably, the plot underscores the dynamic shifts observed in the endothelial cell landscape, particularly emphasizing the significant enrichment of EC_breach cells from niche_early to niche_late conditions, mirroring observations from the UMAP analyses. Similarly, it shows a reduction in capillary types I and II cells in these conditions. The normal condition, depicted here as the baseline, contains only a minimal amount of EC_breach cells, contrasting sharply with the changes seen in the TME of this breast cancer-derived lung metastasis.*

5.3.5.4. DEA & PEA

Building on the observations from the endothelial subtype analyses, the next critical phase involves differential expression analysis (DEA) and pathway enrichment analysis to further elucidate the molecular underpinnings of the observed phenotypic changes. Particularly, the focus will be on contrasting the EC_breach cells between the niche_early and niche_late compartments. This emphasis stems from the hypothesis that the proliferation of this cell subtype over time is accompanied by significant alterations in gene expression patterns, which are likely pivotal in enhancing its invasive capabilities and its role in angiogenesis associated with metastatic progression.

However, this analysis faces a notable limitation due to the inherent scarcity of cells within certain conditions, particularly for the EC_breach subtype and, conversely, for the downregulated EC_capillary_type II cells. The limited sample size for these specific contrasts inherently reduces the statistical power of the Wilcoxon signed rank test used in the DEA. This reduction in statistical power could impact the reliability of identifying true gene expression differences, potentially masking subtle yet biologically significant changes. Therefore, while the results may highlight critical trends and potential pathways involved in endothelial subtype transitions, they must be interpreted with caution, considering the constraints imposed by the available sample sizes.

The differential expression analysis (DEA) results were effectively visualized using volcano plots **[Figure 5-20]**, which highlighted significant findings concerning gene expression changes. The plots notably identified **S100a6** and **Rpl23a** as the top upregulated targets, while **Clic4** and **Tm4sf1**

emerged as the top downregulated targets. These results offer critical insights into the molecular adjustments occurring under the studied conditions.

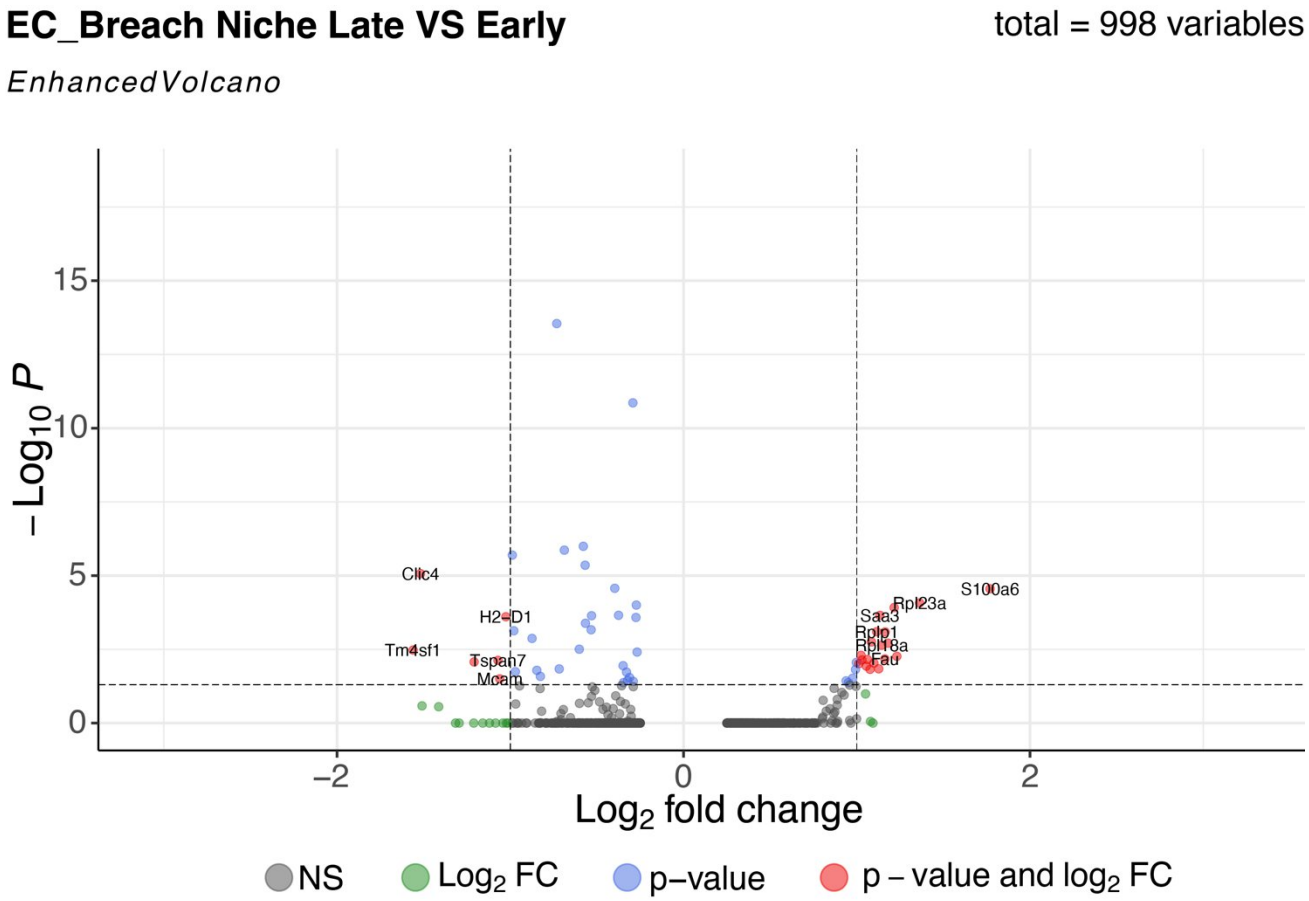


Figure 5-20: Volcano Plot for EC_Breach in Niche Late vs. Early Comparison *- This volcano plot depicts differential gene expression between late and early niche environments for EC_breach. The x-axis represents the log2 fold change, indicating the magnitude of gene upregulation or downregulation. The y-axis shows the -log10 of the adjusted p-value, which transforms the p-values into an ascending scale where higher values indicate greater statistical significance. This plot effectively illustrates which genes are significantly regulated between these conditions, highlighting key candidates such as Rpl23a and S100a6 for further investigation.*

S100a6 is a calcium-sensing protein that utilizes the EF-hand calcium-binding motif to detect elevated intracellular signaling levels. As a crucial effector of calcium signaling, it responds to changes typically mediated by ligand-gated calcium channels, which transiently increase intracellular calcium in response to external stimuli (Bao et al., 2012). In non-small cell lung cancer (NSCLC) cell lines, particularly the A5489 line, S100A6 is significantly upregulated. The overexpression of S100A6 is linked to enhanced proliferation, migration, and invasion capabilities, along with the degradation of acetylated p53. This upregulation is controlled at the transcriptomic level by miR193a, which is found to be downregulated (P. Li et al., 2019). In endothelial cells, S100A6 promotes cell proliferation, and its depletion leads to extended cell cycle arrest in the G1/G0 state, reducing progression to the S phase and mitosis. The impact on endothelial cells is primarily through the downregulation of CDK1, likely mediated by interactions with HSP90 (Bao et al., 2012).

Rpl23a is an integral component of the 60S subunit of ribosomes, which are essential for protein translation within cells. As a ribosomal protein, Rpl23a plays a critical role in the assembly and functional integrity of the ribosome, facilitating the translation process central to cellular metabolism and growth. In the context of hepatocellular carcinoma, the overexpression of RPL23 (the human equivalent of Rpl23a) has been found to significantly influence tumor progression. RPL23 enhances tumor invasion and metastasis by binding to and stabilizing the 3' UTR of MMP9, a matrix metalloproteinase that plays a crucial role in the degradation of the extracellular matrix (ECM), thereby facilitating cancer cell invasion and metastasis (M. Yang et al., 2021). Additionally, Rpl23 is upregulated in myelodysplastic syndrome, where it contributes to apoptotic resistance. This effect is mediated by the downregulation of Miz-1, a transcription factor that normally activates CDK

inhibitors such as p15Ink4b and p21Cip1 and suppresses c-myc. Given that c-myc is also a driver of Rpl23 expression, a negative feedback loop is established where high levels of Rpl23 induced by c-myc lead to the suppression of Miz-1. This suppression in turn reduces the levels of p15Ink4b and p21Cip1, attenuating Miz-1-dependent downregulation of c-myc (Y. Qi et al., 2017).

Clic4 is a chloride intracellular channel that exhibits poor selectivity for ions, with a particular affinity for chloride ions. Its activity is regulated by pH levels, enabling adaptation to varying intracellular environments (Suh et al., 2005). In terms of angiogenesis, Clic4 plays a critical role in endothelial tubulogenesis, a vital precursor in the angiogenic cascade, which is the process by which new blood vessels form from pre-existing vessels. This function highlights its importance in angiogenesis, where its proper functioning is crucial for the formation and stabilization of new vascular structures. Disruption of Clic4 expression has been observed to impair angiogenic capabilities, as demonstrated in Matrigel plug angiogenesis assays (Ulmasov et al., 2009).

In the context of breast cancer, Clic4 exhibits dual behaviour. Within breast cancer epithelial cells, it functions as a potent tumor suppressor, potentially contributing to apoptosis and cell cycle arrest (Suh et al., 2005). This suppressive action makes its inhibition and downregulation a target for cancer cells seeking to overcome constraints on cellular proliferation and stress adaptation. Conversely, cancer cells can manipulate their surrounding stromal environment to upregulate Clic4 expression, thereby promoting tumor progression (Suh et al., 2007). This complex interaction between Clic4 expression in cancer cells and their microenvironment underscores the protein's significant role in cancer biology, influencing both tumor suppression and promotion depending on the cellular context.

Tm4sf1 (Transmembrane 4 L Six Family Member 1) is a member of the transmembrane 4 superfamily, playing a key role in signal transduction and cell-cell adhesion, which are essential for cellular development. In the cancer stromal microenvironment, Tm4sf1 is highly overexpressed; knockdown of this protein leads to reduced angiogenesis, decreased cell motility, inhibition of cytokinesis, and impaired endothelial angiogenesis (Shih et al., 2009). In pancreatic cancer, upregulation of Tm4sf1, along with concurrent downregulation of miR-141 (which inhibits Tm4sf1 expression), correlates with poor prognosis, contributing to increased angiogenesis and invasion (Xu et al., 2020). Additionally, in colorectal cancer, upregulation of miR-9 results in Tm4sf1 inhibition, leading to reduced angiogenesis and invasion, thereby demonstrating its pivotal role in the dynamics of tumor progression (Park et al., 2016).

Subsequent Pathway Enrichment Analysis (PEA) was conducted using the ***clusterProfiler*** package, leveraging the mouse equivalent of the hallmark gene set from MSigDB, which was imported via the ***msigdbr*** package. Results from the Differential Expression Analysis (DEA), organized by average log2 fold change (log2FC), were analysed. Visualization of the data was achieved through ridge plots **[Figure 5-21]**, which distinctly illustrated that the predominant and singularly upregulated pathway in this targeted DEA list is the "Hallmark Myc Targets" pathway. This pathway's upregulation is strongly indicative of substantial proliferative capacity within the cell population.

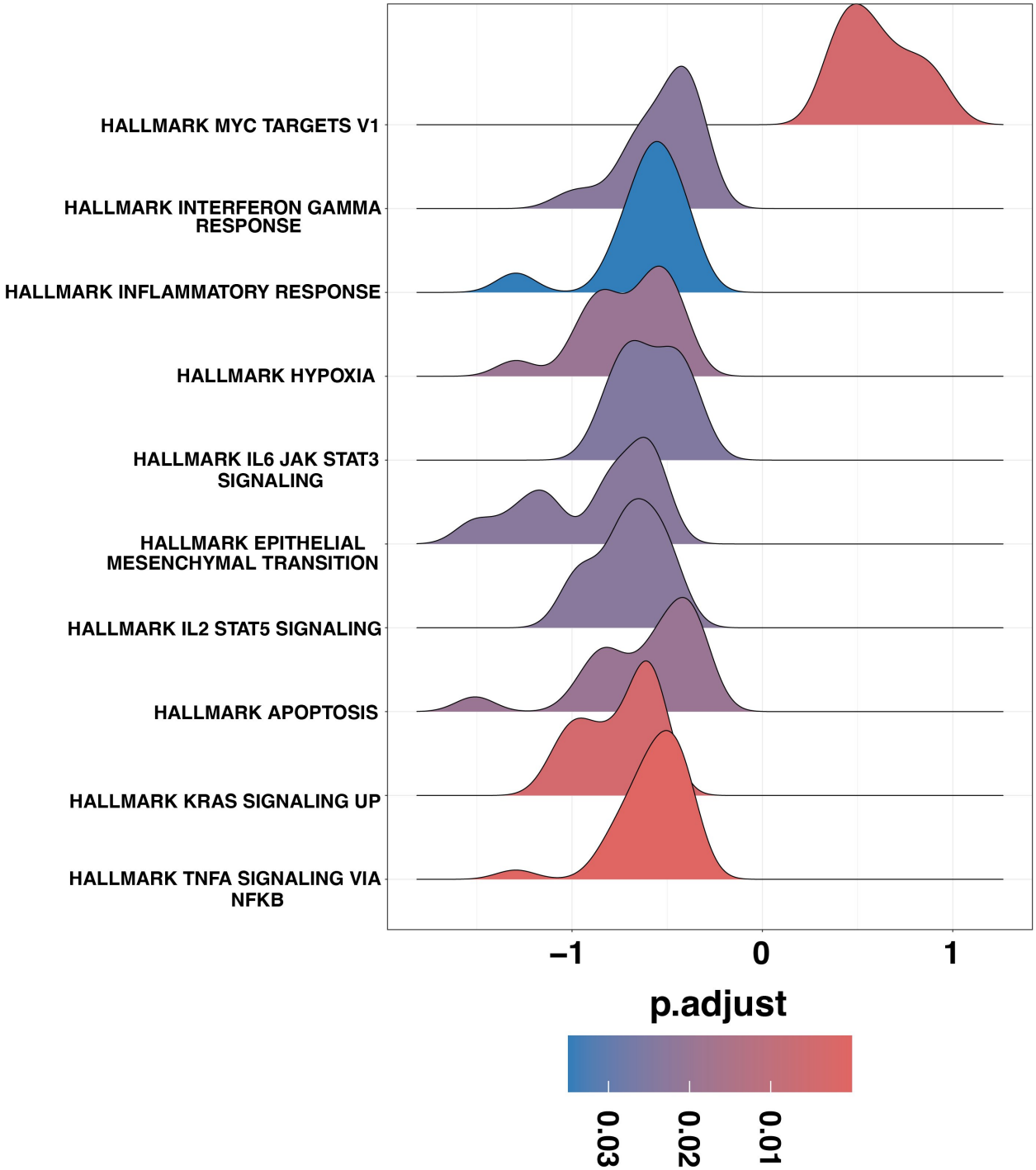


Figure 5-21: Ridge Plot of Pathway Enrichment Analysis (PEA) Results – *This ridge plot displays the results of Gene Set Enrichment Analysis (GSEA) performed using Differential Expression Analysis (DEA) results, which were organized by the average log2 fold change (log2FC) across different pathways. The x-axis represents the normalized enrichment score, with negative values indicating pathway downregulation and positive values indicating upregulation. The colour gradient represents the adjusted p-value, highlighting the statistical significance of the observations. Notably, the only pathway significantly upregulated is the "Hallmark Myc Targets," indicative of substantial proliferation for this contrast.*

Interestingly, the analysis also revealed that both the Hallmark hypoxia and apoptosis pathways were downregulated. This pattern likely reflects the active proliferative state adopted by the cellular population during the early stages of tumor metastasis integration, invasion, and initiation of neo-angiogenesis. Such insights highlight the dual nature of cellular adaptation in response to tumorigenic stress, emphasizing an increase in proliferation alongside a reduction in hypoxia response and apoptosis—a combination that potentially facilitates aggressive tumor expansion and adaptation to new microenvironments.

Further contrasts, such as comparing EC_breach cells between distal_early and distal_late conditions, presented intriguing possibilities for investigation. However, these contrasts were excluded from the current analysis due to substantial differences in the number of cells available for

each condition. The significant disparity in cell counts between these groups renders the statistical power insufficient to conduct a robust analysis. This limitation affects the reliability of detecting true biological differences, necessitating the exclusion of these contrasts to maintain the integrity of the analytical outcomes.

5.3.5.5. Cell-Cell Communication Analysis

Cell-cell communication analysis was also carried out here on the endothelial subset of the data to further understand the signaling cascades driving the development and activity of individual endothelial subsets.

High levels of attention are, of course, warranted to the **EC_neophalanx/tip/breach** subpopulation because of its interesting proliferation between niche_early and niche_late conditions.

5.3.5.5.1. General Network Structure

The general network structure is analysed using a combination of visualizations. Circle visualizations provide an aggregate overview of the network, highlighting the most prominent senders and receivers in terms of signaling frequency **[Figure 5-22A]** and intensity **[Figure 5-22B]**.

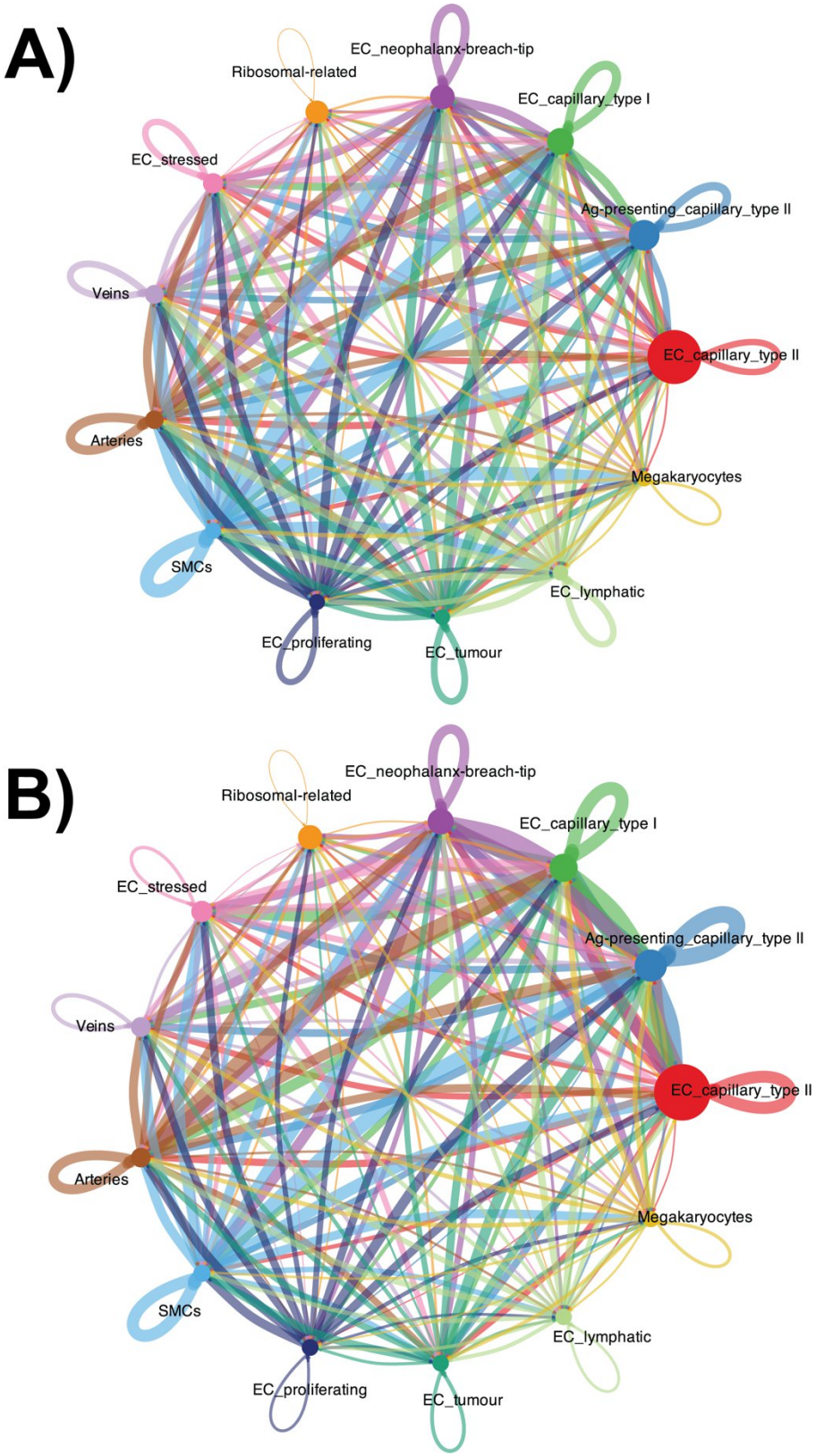


Figure 5-22: *These circle plots visualize cell-cell communication within the endothelial subset of the dataset, illustrating the interactions between different endothelial cell groups.* ***Node Representation:*** *The size of each node indicates the group size within this subset, with larger nodes representing larger subpopulations of endothelial cells.* ***Link Representation:*** *Links between nodes depict ligand-receptor (LR) pairs and signaling pathways relevant to endothelial interactions. The colour of each link corresponds to the originating node, facilitating the identification of signal sources within this specific subset.* ***Weight Metrics:*** *A) The width of the links in panel A indicates the number of interactions, with thicker lines representing a higher number of connections between nodes, highlighting the frequency of communication. B) In panel B, the width of the links denotes the strength of these interactions, with thicker lines indicating stronger communications among endothelial cells.*

In this network, EC_proliferating and the two EC_capillary_type_II subtypes emerge as significant senders with high frequency and vivid intensity. Interestingly, these subtypes demonstrate

pronounced autocrine signaling patterns, engaging multiple ligand-receptor (LR) pairs at high intensity.

Circle plots of individual subtypes [**Figure 5-23**] reveal their detailed signaling patterns and are instrumental in identifying the network's strongest senders; these plots are scaled relative to the highest intensity link in the full global network. The strongest senders include EC_capillary_type_II, EC_neophalanx-breach-tip, EC_proliferating, and SMCs.

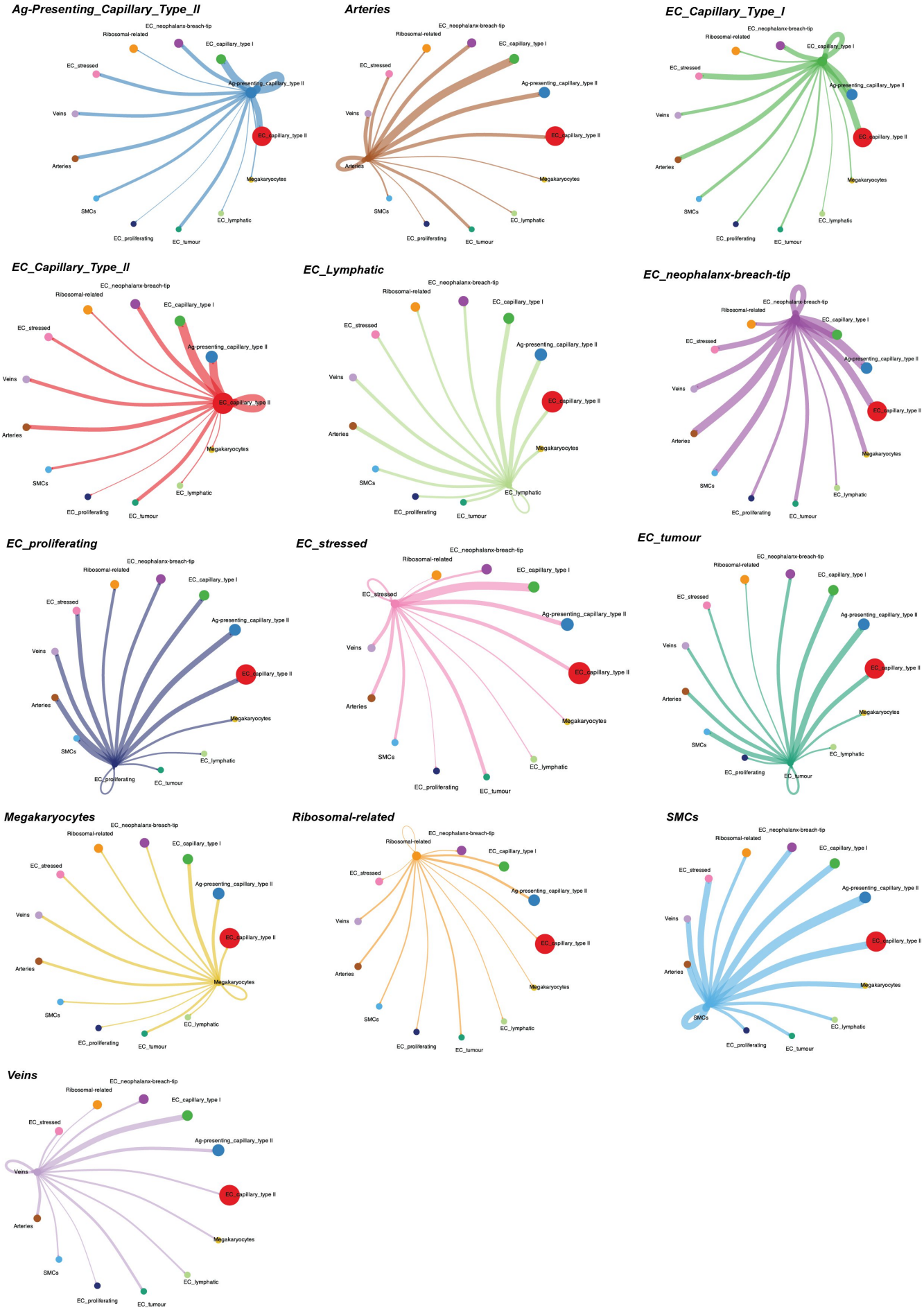


Figure 5-23: Individual Circle Plots of Endothelial Cell Communication Networks – *These individual circle plots illustrate the communication dynamics specifically within the endothelial subset of the cellular network, highlighting how various endothelial cell subtypes interact. Each plot is dedicated to a different endothelial cell subtype, showcasing the unique communication patterns of each.* ***Node Size:*** *Indicates the relative size of each endothelial group, with larger nodes representing more substantial groups within*

this subset. ***Link Colour:*** *Corresponds to the cell subtype from which the signal originates, aiding in tracing the source of each communication pathway within the endothelial context.* ***Link Width:*** *Represents the strength of the signaling, with wider links denoting more robust communication between endothelial cells.*

Heatmap plots are used to visualize the top incoming **[Figure 5-24A]** and outgoing **[Figure 5-24B]** signals for individual pathways. The dominant pathways in this network are Collagen, Laminin, VEGF, and CypA. Top senders include EC_capillary_type_I, EC_capillary_type_II, and EC_neophalanx-breach-tip. Notably, EC_capillary_type_I is a particularly strong sender and receiver, engaging in significant paracrine and autocrine signaling cascades.

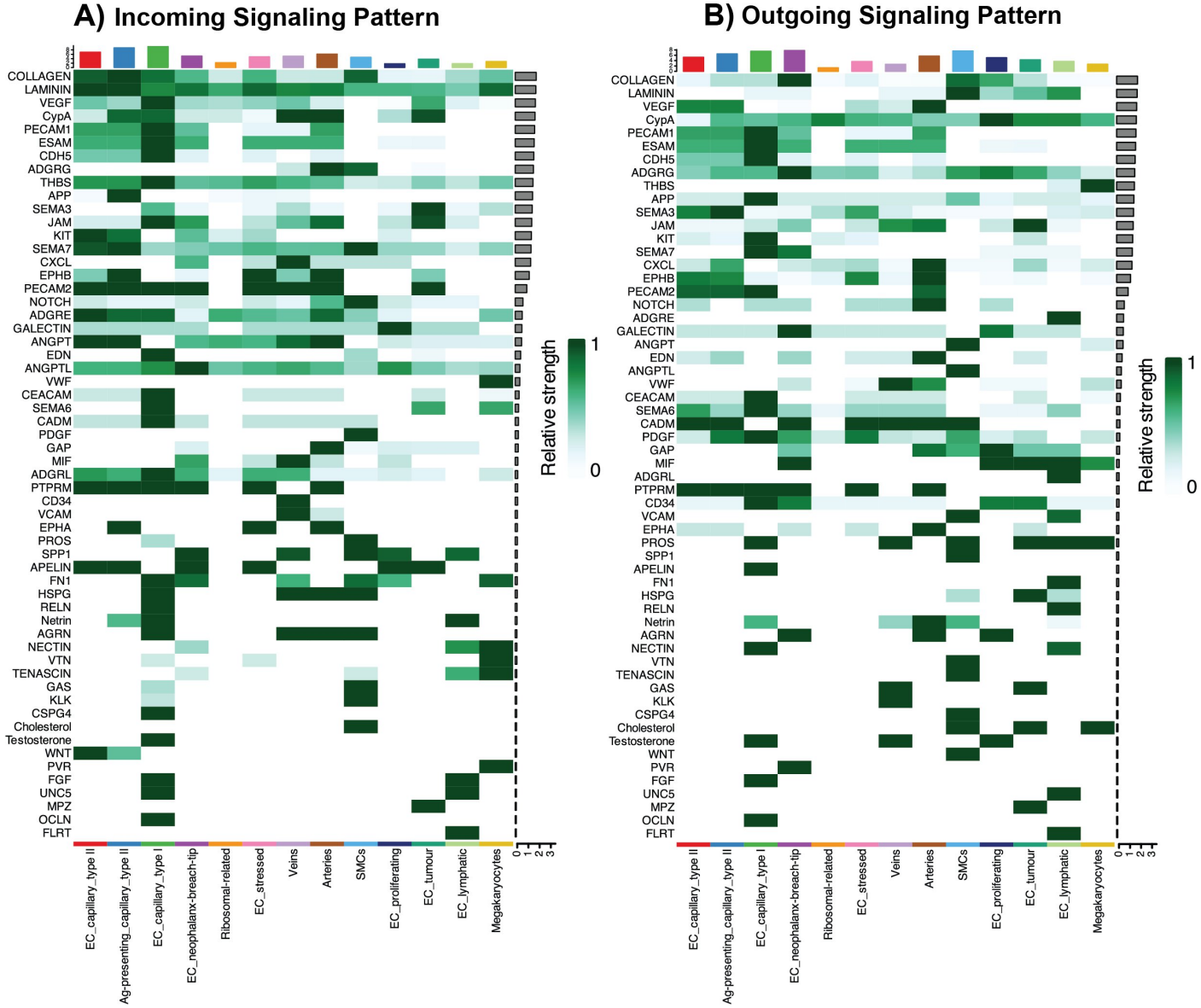


Figure 5-24: Global Heatmaps of Major Senders & Receivers in the Endothelial Network *– This figure provides a comprehensive view of the communication dynamics specifically within the endothelial subset of the network, illustrated through two detailed heatmaps:* ***A) Incoming Signal Patterns:*** *This heatmap on the left shows the signaling strength received by various endothelial cell types from different pathways. The x-axis categorizes the endothelial cell types, and the y-axis lists the pathways involved. The colour of the tiles represents the relative strength of the signals received by these endothelial cells. The top bar plot aggregates the total signaling strength received by each endothelial cell type across all pathways, providing a summary measure of incoming signals. The side bar plot compiles the overall signaling strength of each pathway, indicating how broadly each pathway impacts different endothelial cell types.* ***B) Outgoing Signal Patterns****: The heatmap on the right displays the signaling strength emitted by different endothelial cell types across various pathways. The x-axis specifies endothelial cell types, and the y-axis identifies pathways. The colour intensity of the tiles indicates the strength of the signals sent. The top bar plot summarizes the total signaling strength sent by each endothelial cell type across all pathways, while the side bar plot shows the aggregate signaling strength of each pathway across all involved endothelial cell types.*

The next step involves a detailed examination of the Collagen and VEGFA signaling cascades in this subset of the data.

5.3.5.5.2. VEGF Signalling Pathway

The analysis of the VEGF signaling cascade highlights it as a prominently utilized pathway within the tumor microenvironment (TME), especially relevant in aggressive breast cancer metastasis. This is biologically consistent and expected given the TME's inherent proliferative nature, which is crucial for tumor growth and vascular development.

The chord **[Figure 5-25A]** and heatmap **[Figure 5-25B]** plots provide a detailed overview of the VEGF signaling network, identifying key senders and receivers within this pathway. Arteries and EC_capillary_type_II cells emerge as significant senders, actively driving the signaling processes. Conversely, EC_capillary_type_I and EC_tumor cells stand out as the most responsive receivers, underlining their vital roles in the signaling dynamics of the TME.

Further exploration through bar plots **[Figure 5-26A]** visualizes the contributions of specific ligand-receptor (LR) pairs to the network. Among these, Vegfa – Vegfr1, Vegfa – Vegfr1r2, and Vegfa – Vegfr2 are notably prominent, with Vegfa – Vegfr1 identified as a critical mediator. Subsequent chord **[Figure 5-26B]** and circle plots **[Figure 5-26C]** reveal that capillaries and EC_capillary_type_II cells are primary senders of Vegfa – Vegfr1 signals. Interestingly, EC_proliferating cells are not significant receivers of this LR interaction, suggesting that their proliferation may be driven by alternative signaling pathways within the network.

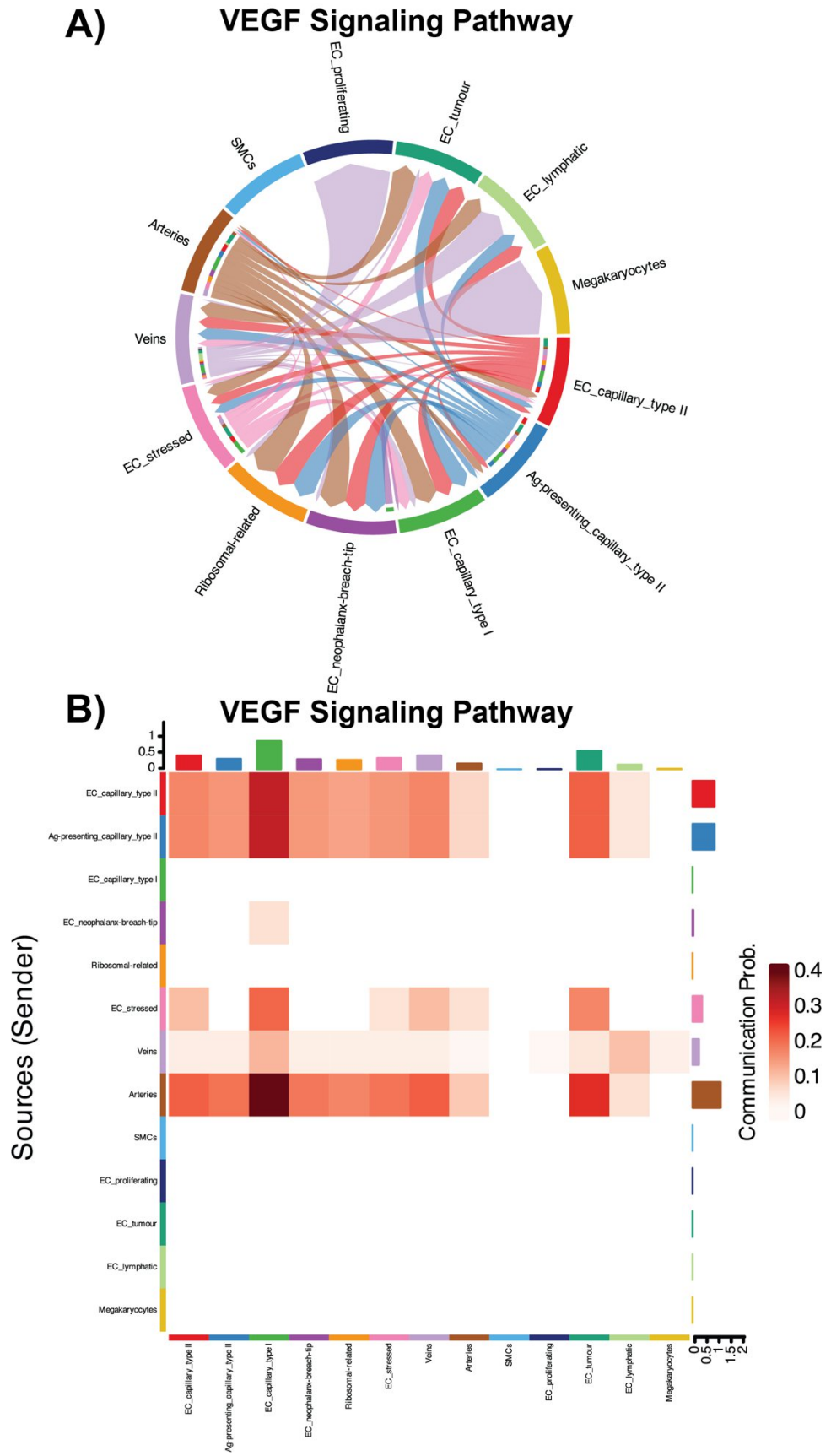


Figure 5-25: Visualization of VEGF Signaling Pathway Communication Dynamics Within the Endothelial Subset – *This figure highlights the communication dynamics within the endothelial subset of the VEGF signaling pathway, emphasizing the primary senders and their targets:* ***A) Chord Plot:*** *This plot displays the top senders within the endothelial subset of the VEGF signaling pathway and the primary endothelial cell types receiving these signals. Each chord represents a communication link, with colors indicating the specific sender endothelial cell type. This visual arrangement facilitates an easy understanding of how signals are distributed across the endothelial network.* ***B) Heatmap:*** *Complementing the chord plot, this heatmap presents the same interaction data in a matrix format. The y-axis lists the sender endothelial cell types, and the x-axis, corrected here, shows the receiver endothelial cell types. The colors of the tiles reflect the communication probability, providing a quantitative measure of the interaction strengths between different endothelial cells.*

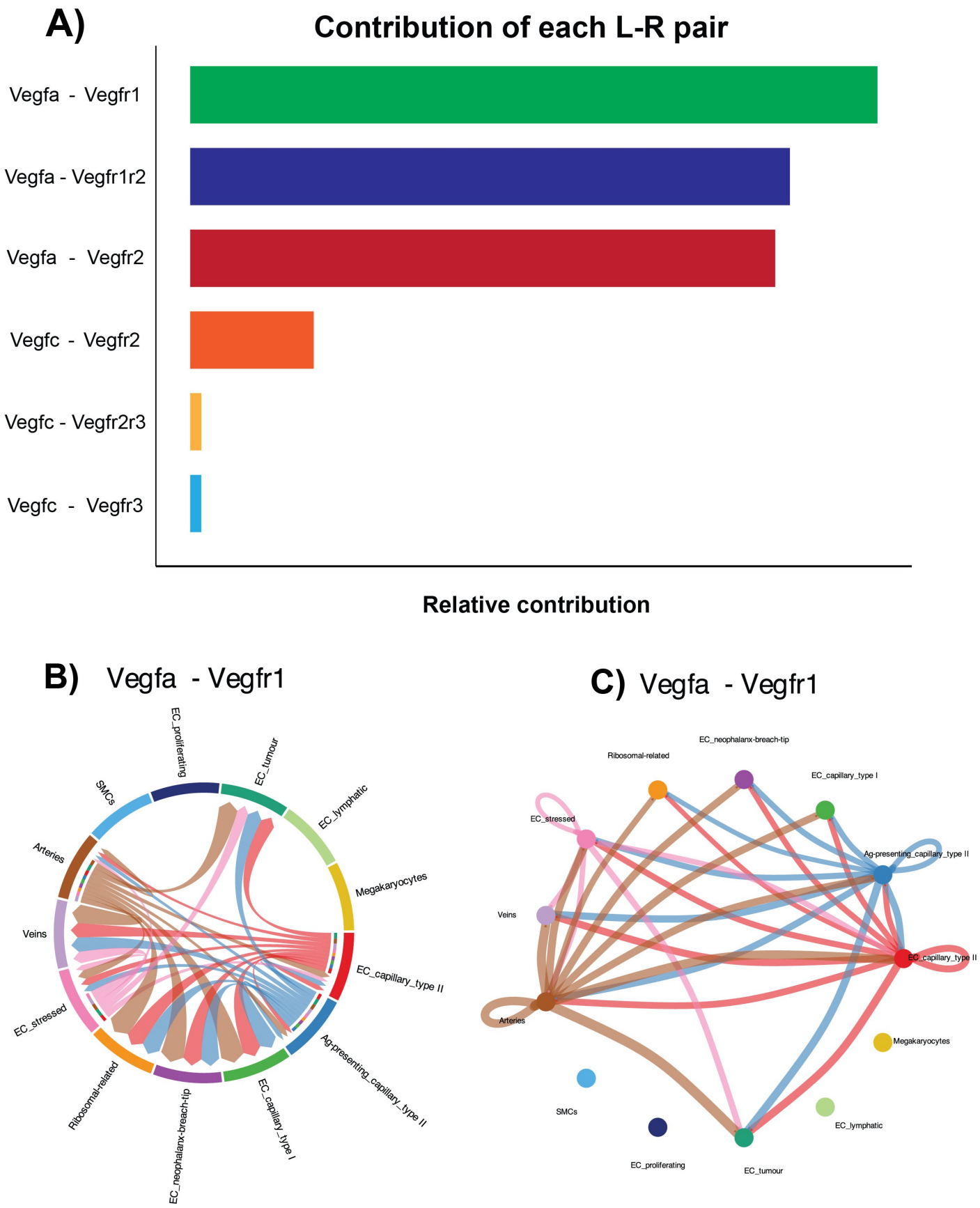


Figure 5-26: Analysis of Top Ligand-Receptor Pairs in the VEGF Signaling Pathway within the Endothelial Subset – *This figure underscores the key ligand-receptor (LR) pairs driving VEGF signaling specifically within the endothelial subset:* ***A) Bar Plots****: Illustrate the relative contribution of individual LR pairs to VEGF signaling in the endothelial subset. Notably, the VEGFA – VEGFR1 pair is highlighted as the most significant contributor, underscoring its pivotal role in mediating VEGF signaling among endothelial cells.* ***B) Chord Plot:*** *Displays the primary endothelial senders utilizing the VEGFA – VEGFR1 pair, mapping the direction and strength of signals between sender and receiver endothelial cell types. This visualization facilitates an understanding of how this dominant LR pair influences communication within the network.* ***C) Circle Plot:*** *Complements the chord plot by showing the flow of*

signals among endothelial cells mediated by the VEGFA – VEGFR1 pair, further detailing these interactions. The colors in both plots correspond to different endothelial cell types involved, aiding in the visualization of specific pathways and interactions.

5.3.5.5.3. Collagen Signalling Pathway

The second pathway of interest is the collagen signaling cascade, recognized for its crucial role in angiogenesis and its significant performance in both incoming and outgoing signals. The chord **[Figure 5-27A]** and heatmap **[Figure 5-27B]** plots below clearly demonstrate the extensive nature of the collagen pathway signaling. Top senders in this pathway include EC_neophalanx-tip-breach, SMCs, and EC_proliferating, while the top receivers are EC_capillary_type_I, EC_capillary_type_II, and SMCs. Notably, the autocrine signaling pattern for SMCs is both prominent and strong.

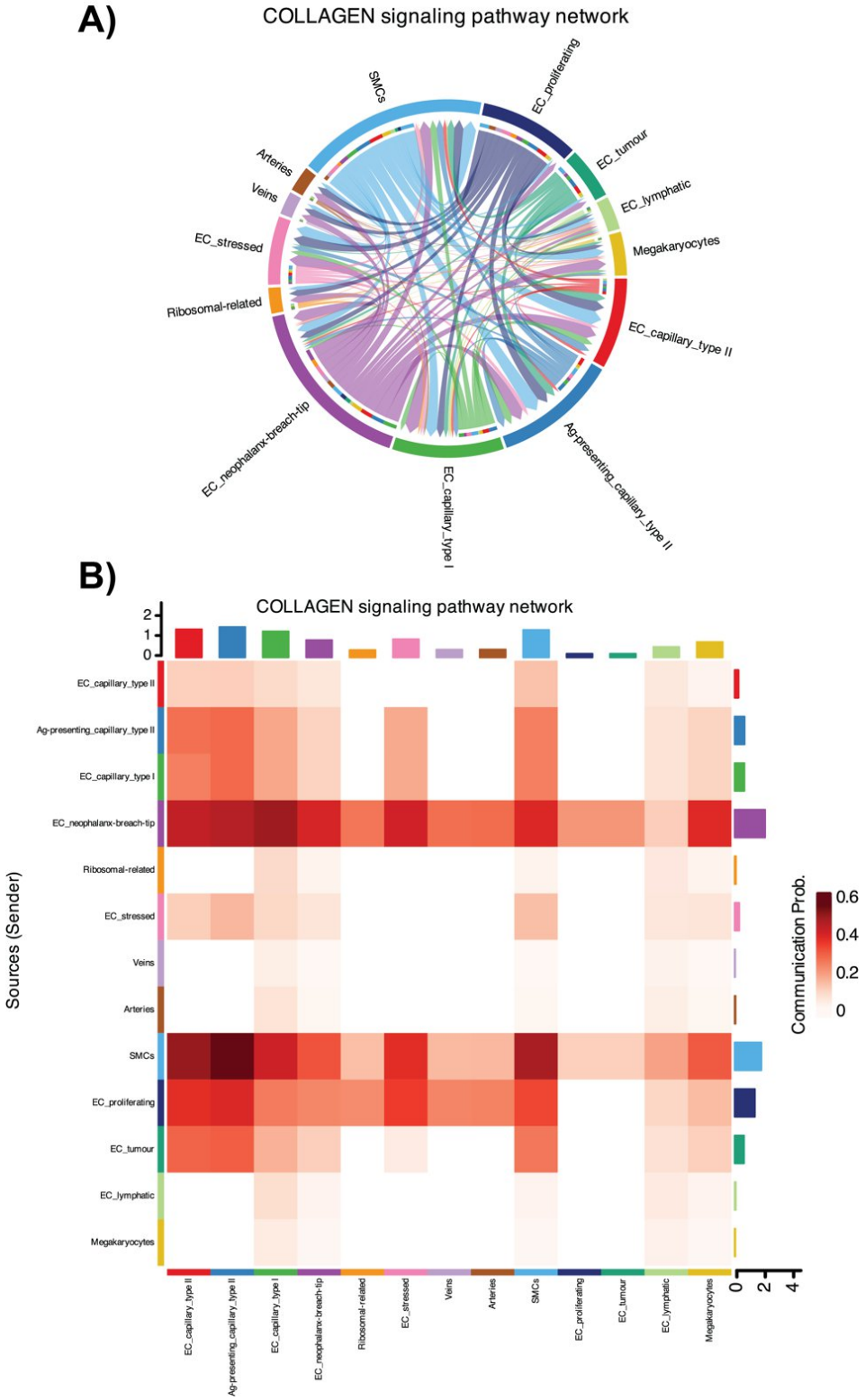


Figure 5-27: Visualization of Collagen Signaling Pathway Communication Dynamics Within the Endothelial Subset – *This figure illustrates the communication dynamics specifically within the endothelial subset of the Collagen signaling pathway, highlighting the primary endothelial senders and their targets:* ***A) Chord Plot:*** *This plot displays the top senders within the endothelial cells of the*

Collagen signaling pathway and the primary endothelial cell types receiving these signals. Each chord represents a communication link, with colors indicating the specific endothelial sender cell type, facilitating a clear visualization of the flow of signals across the endothelial network. ***B) Heatmap:*** *Complements the chord plot by presenting the same interaction data in a matrix format. The y-axis lists the sender endothelial cell types, while the x-axis shows the receiving endothelial cell types. The color intensity of the tiles reflects the communication probability, providing a quantitative view of the interaction strengths between different endothelial cells involved in the pathway.*

Further analysis of individual ligand-receptor (LR) pairs and their contribution to the network underscores the prolific nature of this signaling cascade, with multiple LR pairs showing high levels of contribution. The relative contribution bar plot **[Figure 5-28A]** highlights col4a1 – (itga1 + itgb1) as the top LR pair.

Additional investigation of this particular pair using chord **[Figure 5-28B]** and circle **[Figure 5-28C]** plots reveals that SMCs are the top and sole sender, with EC_capillary_type_II cells (both sub-populations) being the strongest receivers. This finding emphasizes the primary role of SMCs as key drivers of the collagen network, with capillary type II cells as the most prominent receivers. This LR pair's signaling is recognized as a key mediator of angiogenesis and capillary formation, primarily because its inhibition results in stunted angiogenesis (Sweeney et al., 2003) and its expression tends to be upregulated by VEGF signaling (Senger et al., 1997).

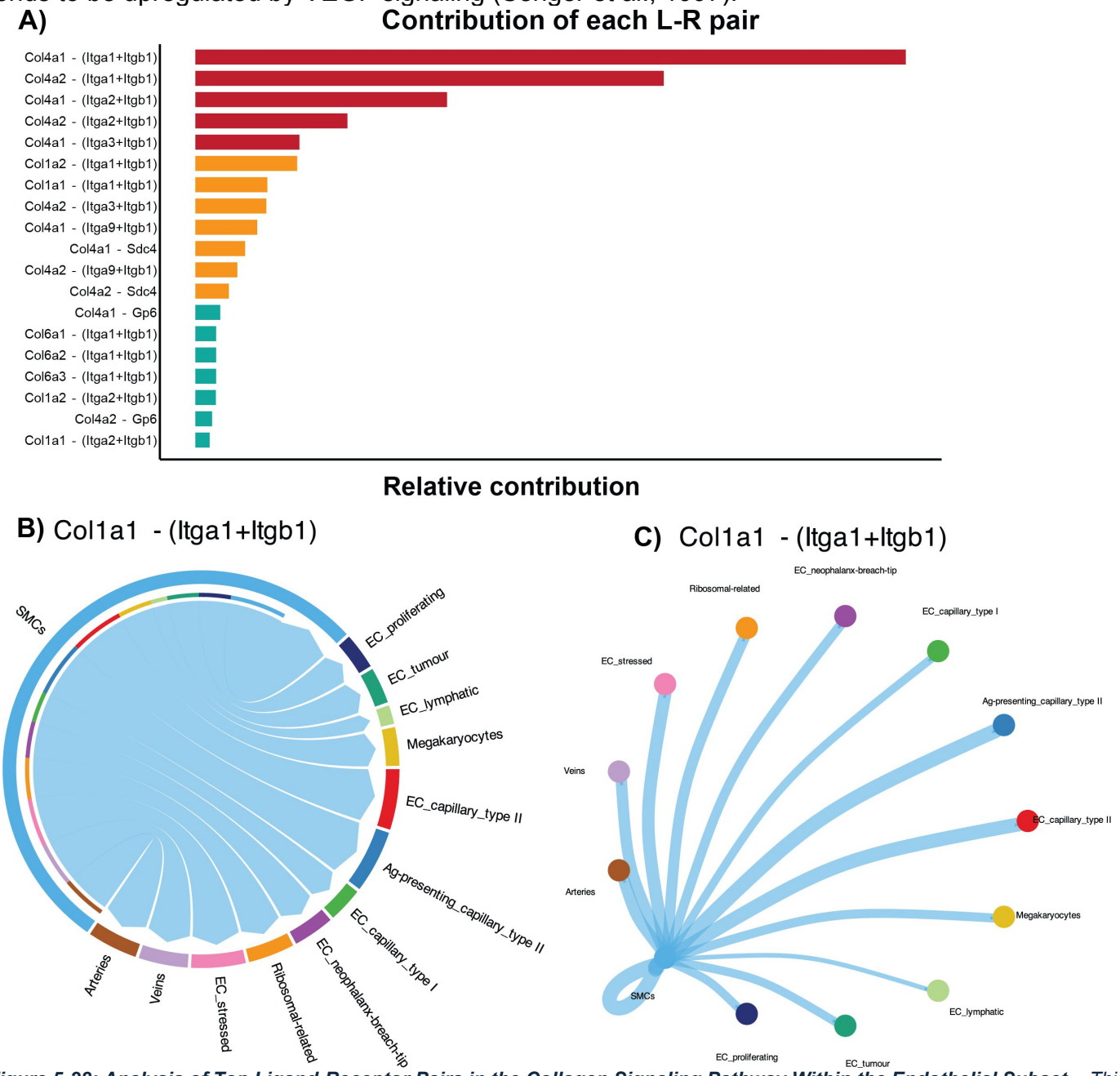


Figure 5-28: Analysis of Top Ligand-Receptor Pairs in the Collagen Signaling Pathway Within the Endothelial Subset – *This figure delves into the dominant ligand-receptor pairs within the Collagen signaling pathway, focusing on the endothelial subset:* ***A)***

Bar Plots: *These plots illustrate the relative contribution of individual ligand-receptor (LR) pairs to Collagen signaling within the endothelial cells. Notably, the pair Col1a1 – Itga1+Itgb1 emerges as the most significant contributor, underscoring its central role in mediating Collagen signaling within this subset.* ***B) Chord Plot:*** *This visualization maps the primary endothelial senders utilizing the Col1a1 – Itga1+Itgb1 pair for signaling. It details the direction and strength of signals between specific endothelial sender and receiver types, using colors to represent different endothelial cell types. This plot highlights the intricate communication links facilitated by this key ligand-receptor pair, illustrating how signals are directed within the endothelial network.* ***C) Circle Plot:*** *Complementing the chord plot, this circle plot further elaborates on the interactions shown in B. It displays how signals circulate within the endothelial network, emphasizing the flow and connectivity between the various endothelial cells involved in the signaling. Each colour corresponds to a different endothelial cell type, aiding in the visualization of specific pathways and interactions driven by the Col1a1 – Itga1+Itgb1 pair.*

5.4. Discussion

Understanding the structural composition of the TME and its evolution across timeframes provides a unique perspective not previously explored at the single-cell level for TNBC subtypes. This analysis opens avenues to identify primary stromal cell types involved in the development of aggressive metastases and the signaling cascades they employ to initiate and maintain this aggression.

5.4.1. Dynamic Changes in EC Composition Across Timepoints

Our analysis, following clustering and cluster annotation, revealed significant changes in the composition of the stromal compartment between early and late timepoints for both niche and distal compartments. Notably, there was a significant increase in EC_breach cells and a proliferation of mesothelial cells from early to late timepoints, coupled with a noticeable reduction in type II ECs. The proliferation of the mesothelial compartment and the emergence of the EC_breach subtype as critical drivers of angiogenic signaling in late metastatic stages prompted further investigation into their signaling roles.

Given the dataset's complexity, conventional DEA and PEA were deemed insufficient for elucidating the key genes driving specialized metastatic and angiogenic signaling. This realization led to the adoption of cell-cell communication analysis to better understand signaling dynamics.

5.4.2. Global Cell-Cell Communication Patterns

Cell-cell communication analysis aimed to identify primary signaling senders and receivers within the network, emphasizing the dominant signaling pathways mediating aggressive metastasis. Although the analysis was not condition-specific (e.g., niche_early, distal_early), it highlighted significant interactions. Fibroblasts and mesothelial cells emerged as principal signaling senders, with ECs being the main receivers. Key signaling pathways involved included VEGF, Collagen, Laminin, and PECAM1, with VEGF and Collagen pathways further dissected to pinpoint crucial ligand-receptor (LR) pairs.

5.4.3. Detailed Analysis of EC Subtypes

To further understand the role of ECs in angiogenic signaling, a subset analysis focused on refining the annotation of EC subtypes and validating the role of the EC_breach compartment at the single-cell level. This analysis distinguished multiple EC subtypes such as proliferating, stressed, tumor-associated, and lymphatic ECs, highlighting the significant rise of EC_breach cells across the observed timepoints. Subsequent DEA and PEA for the EC_breach compartment between niche_early and niche_late conditions identified significant upregulation in targets like S100a6 and Rpl23—both known drivers of cellular proliferation—and showed positive enrichment in the hallmark Myc targets pathway.

5.4.4. Intra-EC Communication

The final aspect of our investigation focused on the signaling network within the EC subset to identify key signaling senders, receivers, and the primary pathways utilized. The dominant pathways in this network included VEGF and Collagen, consistent with findings from the broader global network analysis. Among these, type I and type II ECs, along with EC_breach, were identified as primary senders. Notably, Vegfa-Vegfr1 emerged as the top LR pair in the VEGF pathway critical for

angiogenesis, while Col4a1—paired with Itga1 and Itgb1—was a key mediator in the Collagen pathway, further emphasizing the interconnected role of these pathways in angiogenesis driven by VEGF signaling.

6. Chapter 6: Shiny Apps

Throughout my DPhil and my development as a bioinformatician, I have realized that the efficacy of bioinformatic analysis significantly hinges on how it is presented. One of the major challenges in interpreting sophisticated analyses is the lack of accessible presentation formats. Since bioinformaticians often collaborate closely with experimental scientists, having an accessible medium that allows for the visualization and exploration of analysis results is crucial for advancing biological research, particularly in the field of transcriptomics.

Given this context, I have started using Shiny apps (refer to chapter 2 for further details) as a medium to share scRNA-Seq analysis results. This approach specifically aims to facilitate the interpretation and exploration of these results. Consequently, for every data chapter in my thesis, Shiny apps were created to aid in the analysis interpretation and exploration. I have standardized the usage of Shiny apps within our lab as a primary collaborative tool. Initially shared locally, these apps have been streamlined to be hosted on **ShinyApp.io** to allow any collaborators to access the data without needing to have R or any of its packages set up on their system. The links for these Shiny apps are the following:

- **MCF7 Dataset:** **https://wolf5996.shinyapps.io/mcf7/**
- **scsHypoxiaTime Dataset:** **https://wolf5996.shinyapps.io/scshypoxiatime/**
- **Mlung Dataset:** **https://wolf5996.shinyapps.io/mlung/**
- **EndoMlung Dataset:** **https://wolf5996.shinyapps.io/endomlung/**

7. Appendices

7.1. Cell Cycle Regression Analysis

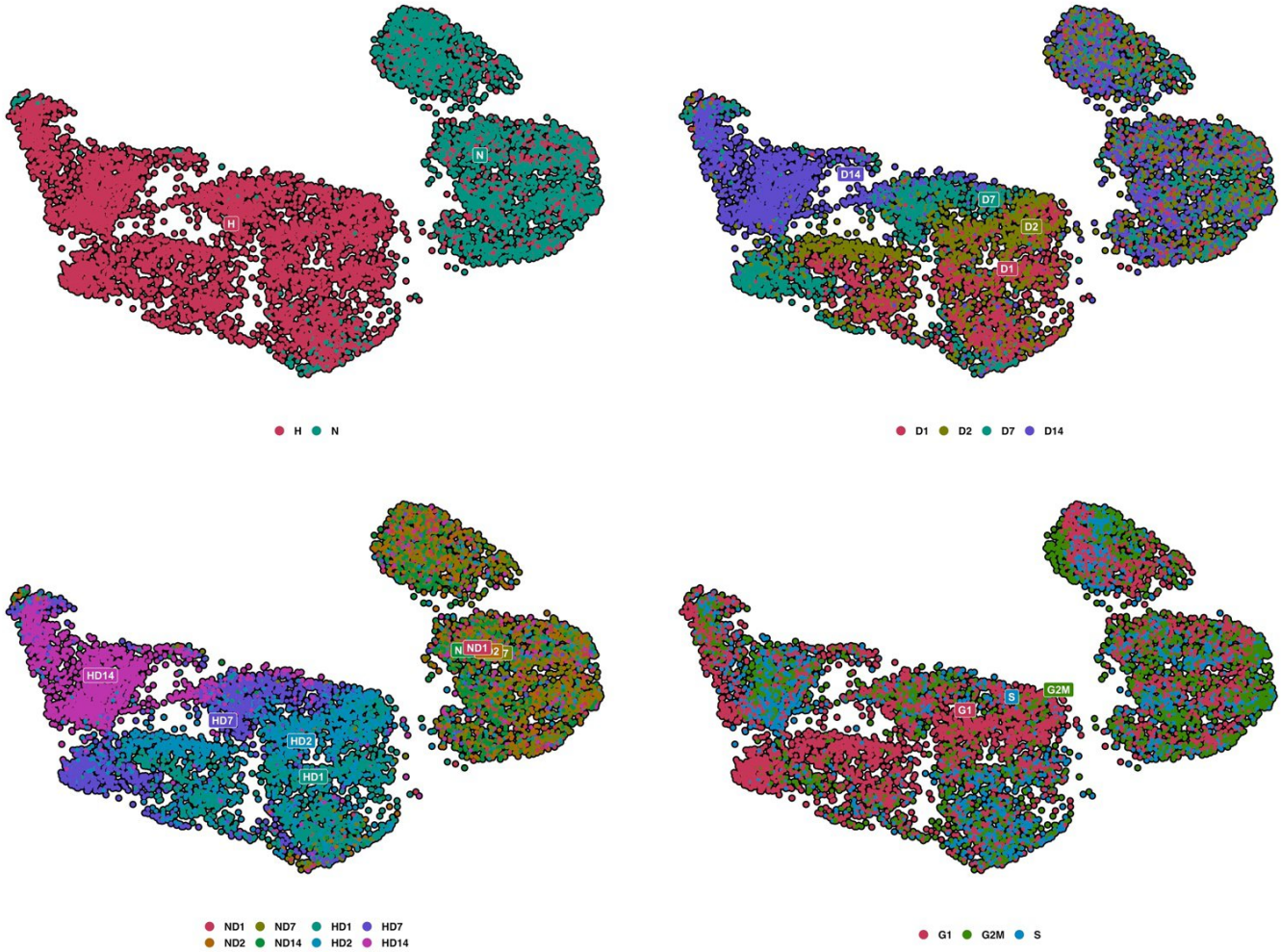
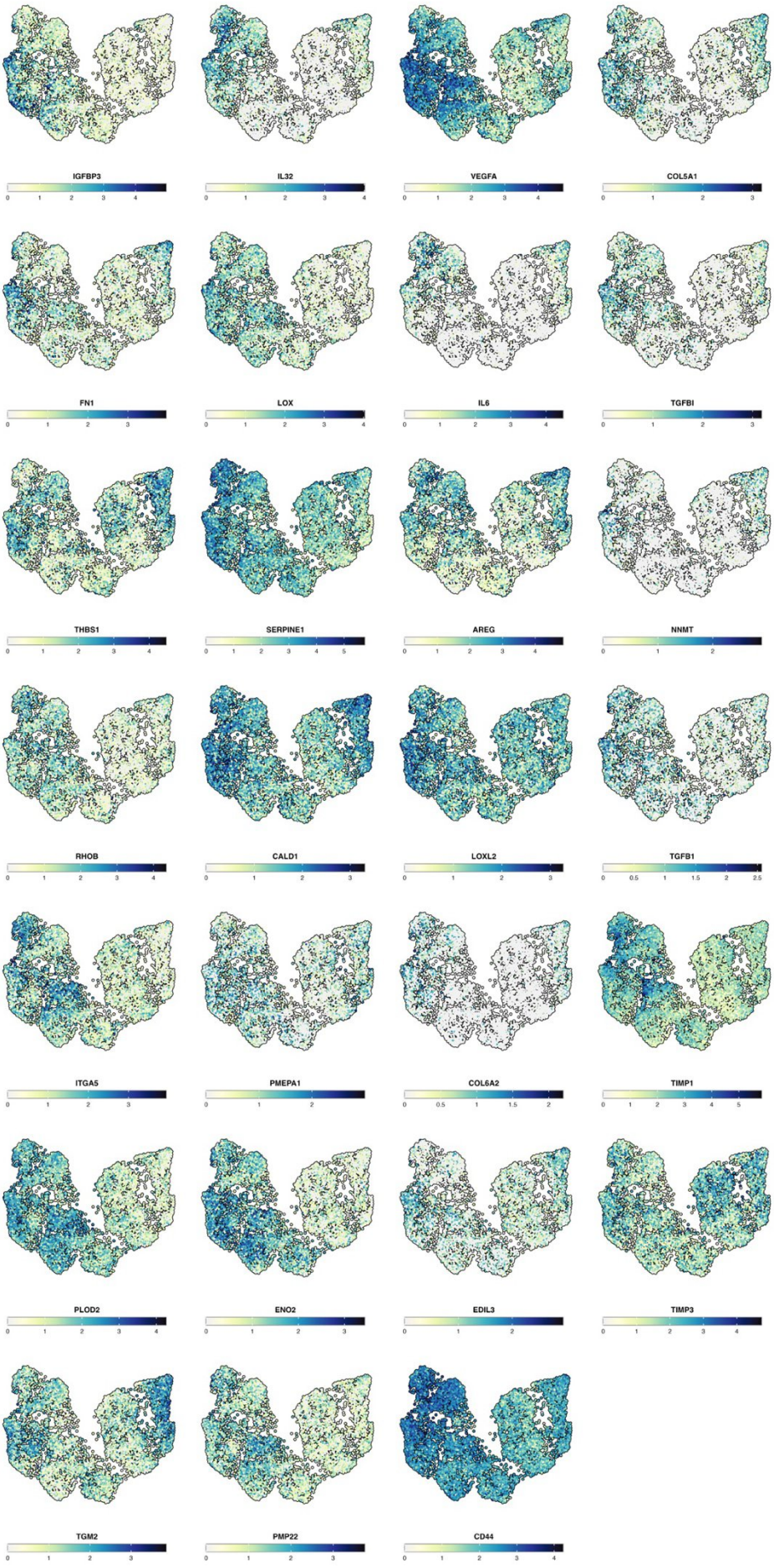


Figure 7-1: scsHypoxiaTime Cell Cycle Regression Analysis: *This figure presents UMAP plots demonstrating the outcomes of cell cycle regression analysis, conducted using the SCTransform pipeline. The analysis specifically involved regressing out the effects of the cell cycle state using non-regularized linear regression to isolate other sources of biological variation.* ***A) Hypoxic Populations:*** *This plot highlights the hypoxic populations, showing that despite the removal of cell cycle effects, these populations still exhibit considerable biological heterogeneity and form completely isolated branches of the UMAP plot.* ***B) Timepoint:*** *Displays the distribution of cells across different timepoints.* ***C) Combinations of Condition and Timepoint:*** *This plot identifies different hypoxic and normoxic populations belonging to the 4 different timepoints used in the experiment.* ***D) Cell Cycle State:*** *Classified using Seurat cell cycle scoring analysis, this plot confirms that the regression effectively removes cell cycle-related variation, particularly evident in the normoxic populations which appear more homogeneous post-analysis. This finding was very important for future experimental designs which will use only a single normoxic timepoint to maximize the number of hypoxic cells and minimize experimental costs.*

7.2. scsHypoxiaTime Leading Edge Expression UMAP Plots



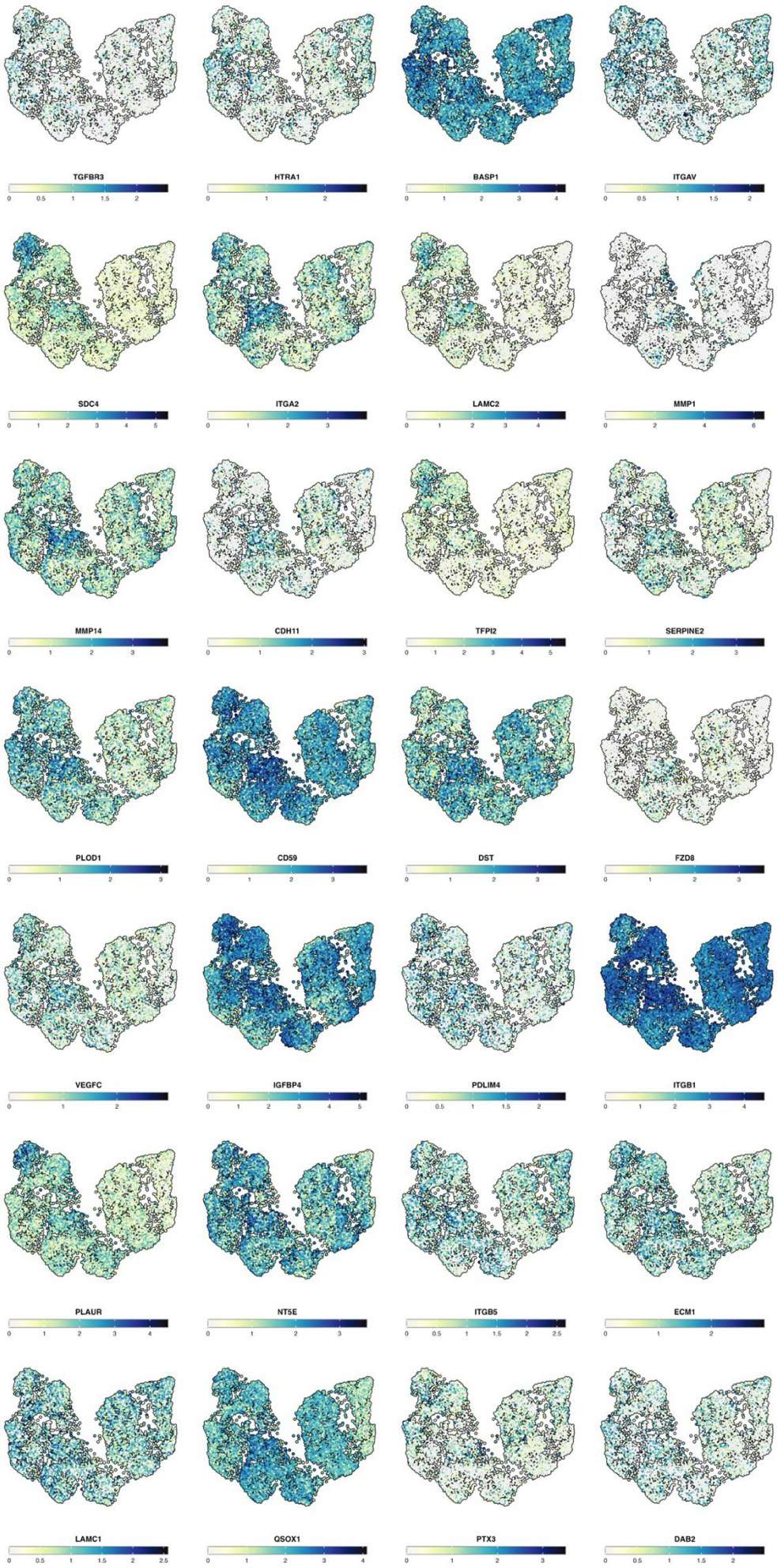

TGFBR3
HTRA1
BASP1
ITGAV
SDC4
ITGA2
LAMC2
MMP1
MMP14
CDH11
TFPI2
SERPINE2
PLOD1
CD59
DST
FZD8
VEGFC
IGFBP4
PDLIM4
ITGB1
PLAUR
NT5E
ITGB5
ECM1
LAMC1
QSOX1
PTX3
DAB2

Figure 7-2: Hallmark EMT Signature Genes Expression UMAP Plots - *This series of UMAP plots illustrates the expression levels of EMT signature genes across clusters 4 and 7. These signature genes, identified as leading-edge genes in GSEA, highlight the differential expression patterns that are critical for understanding the EMT pathway's role in these specific clusters.*

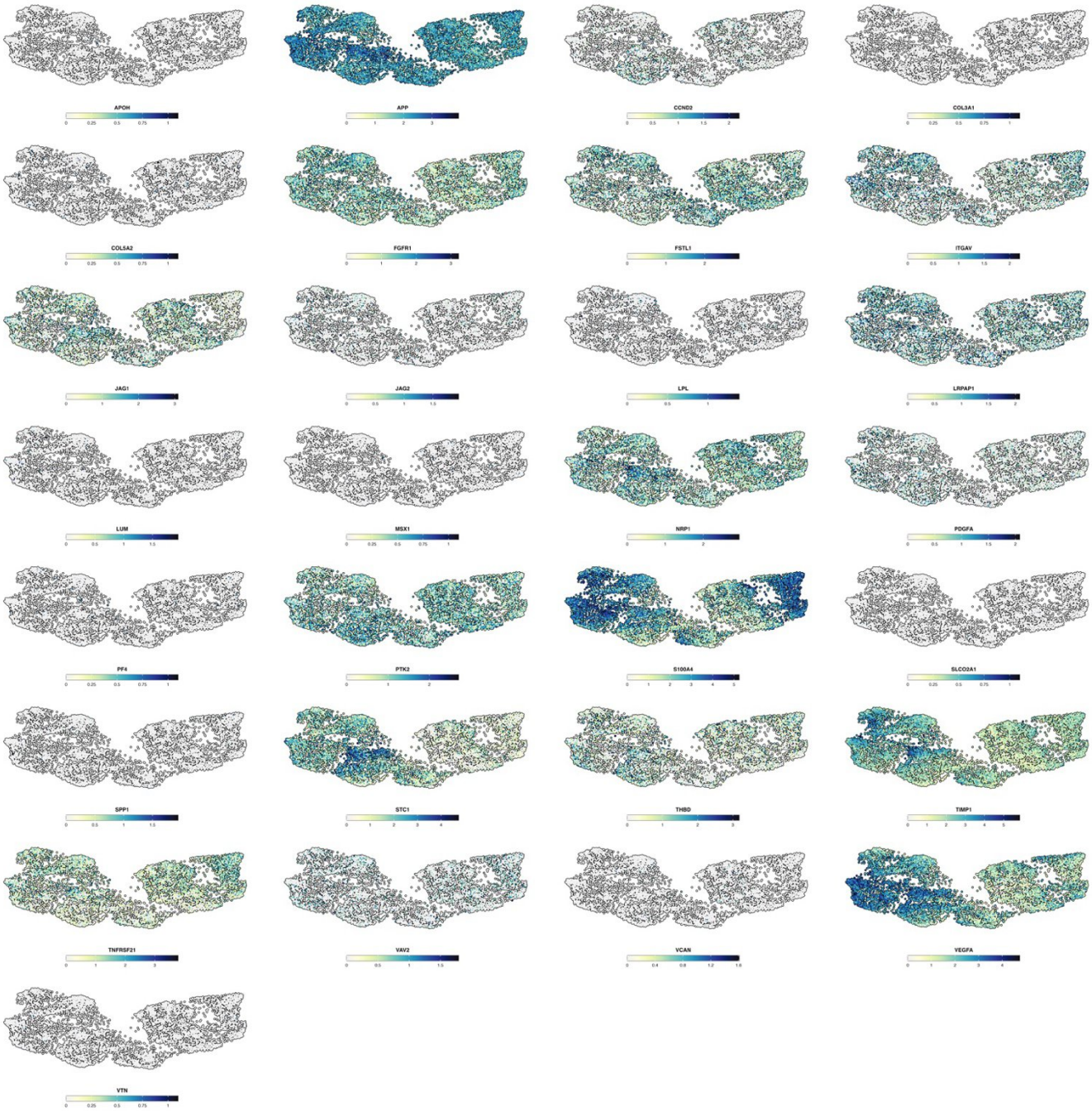


Figure 7-3: Hallmark Angiogenesis Genes Expression UMAP Plots - *This series of UMAP plots illustrates the expression levels of angiogenesis signature genes across clusters 4 and 7. Given the relatively small size of the angiogenesis signature, the full signature was shown here instead of just the leading edge.*

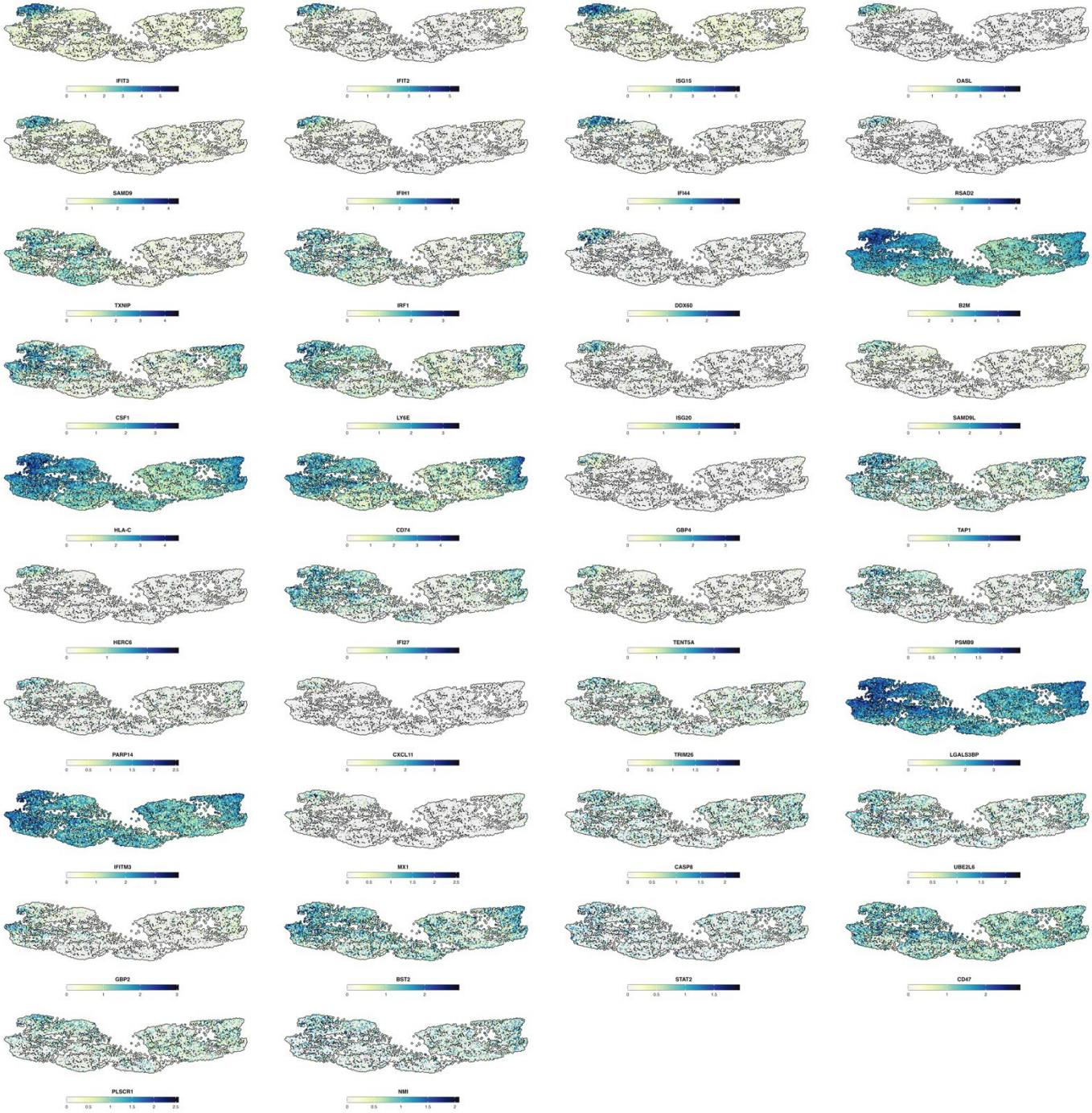


Figure 7-4: Hallmark Interferon Alpha Genes Expression UMAPs - *Signature genes were selected based on those identified as leading-edge genes in the Gene Set Enrichment Analysis (GSEA) for clusters 9 and 0. These individual expression UMAP plots effectively illustrate the variation in feature expression across different subsets of the data, highlighting the distinct biological insights that can be derived from each cluster.*

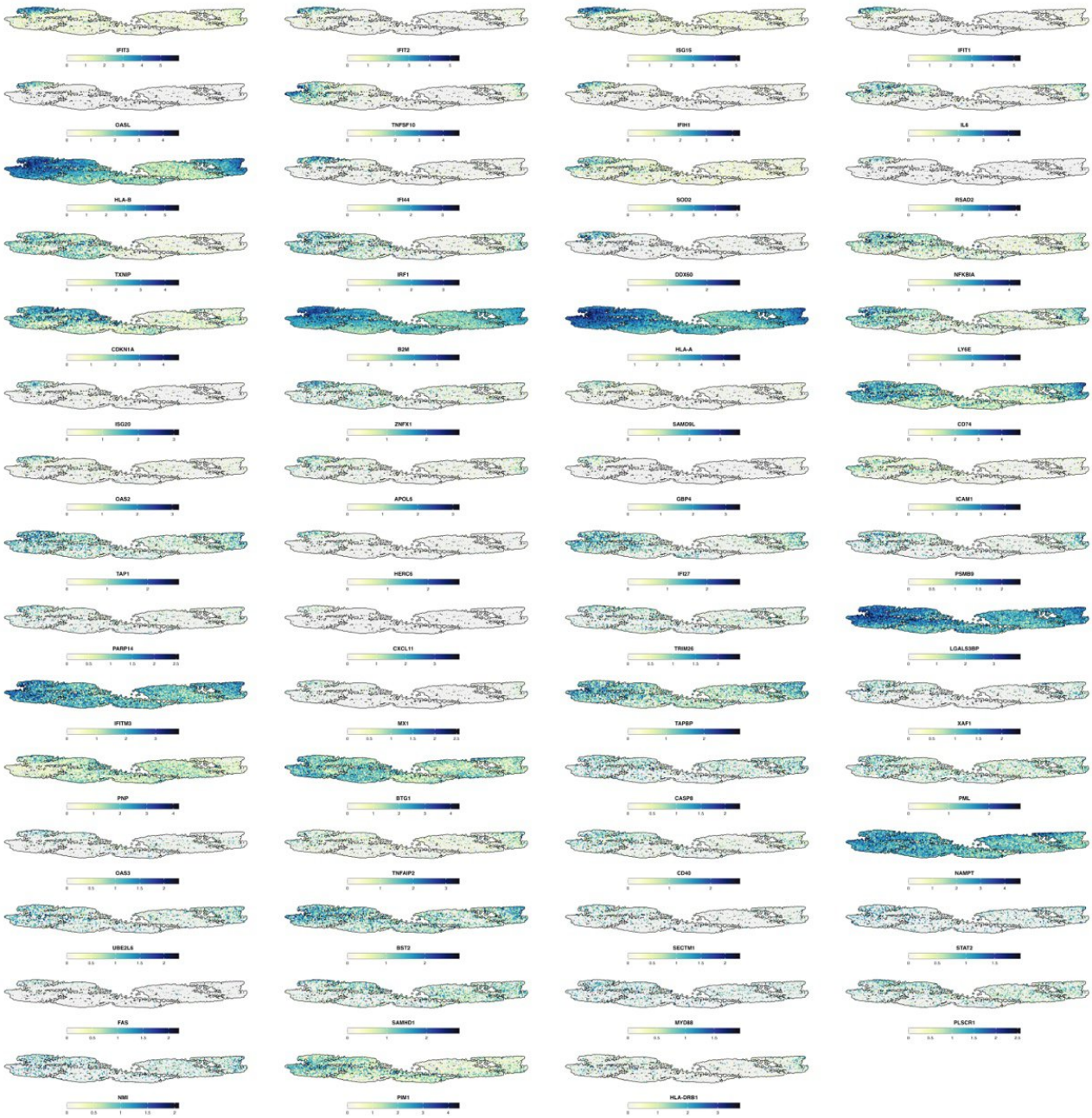
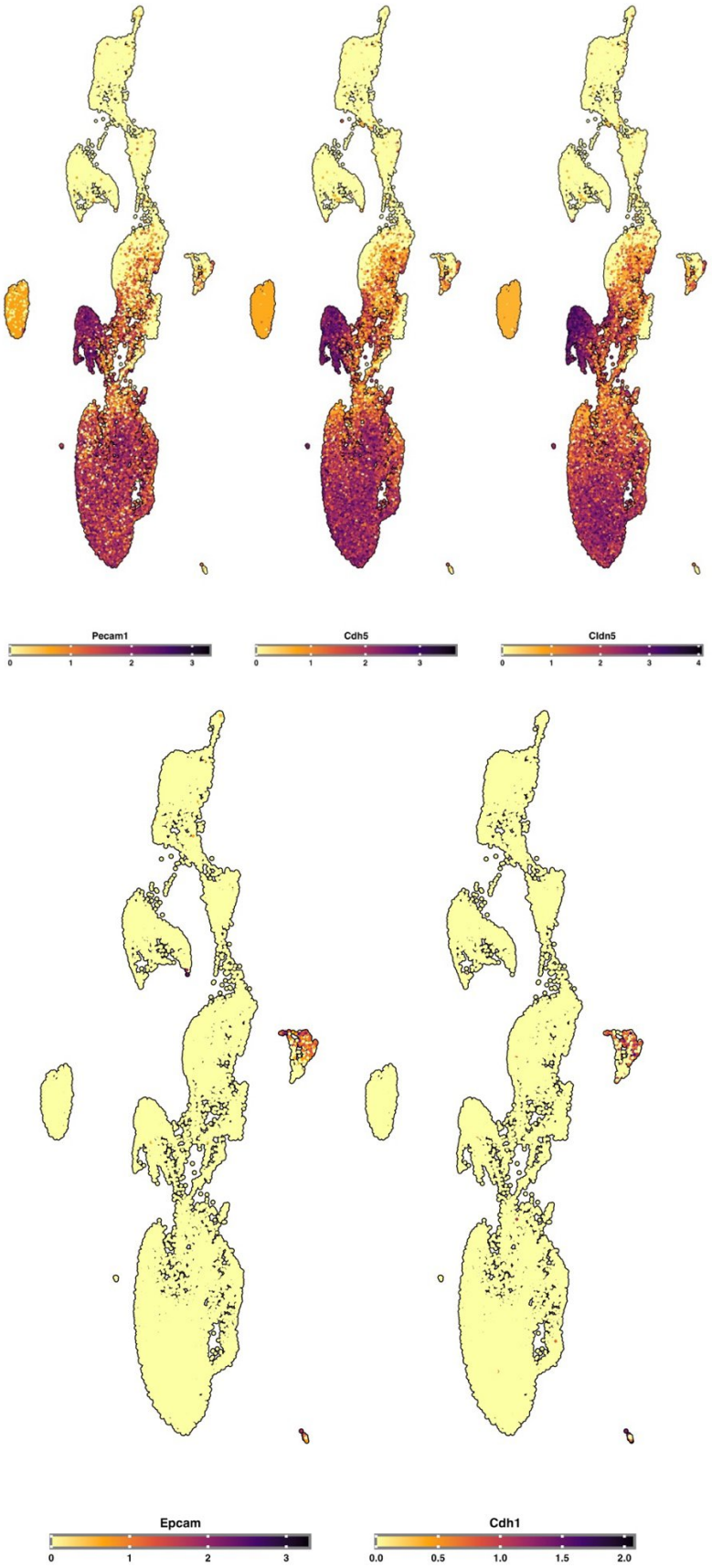
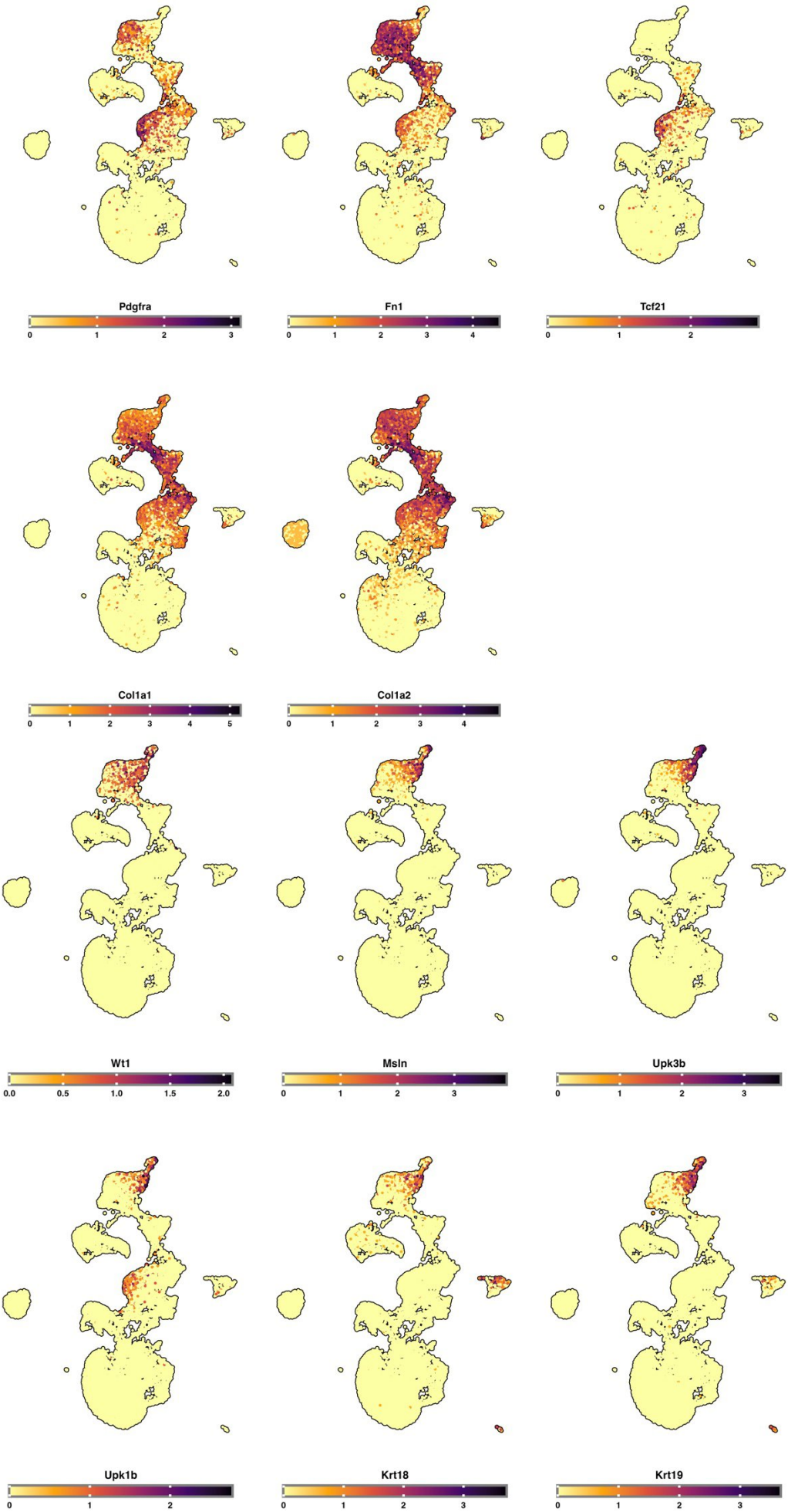


Figure 7-5: Hallmark Interferon Gamma Genes Expression UMAPs - *Signature genes were selected based on those identified as leading-edge genes in the Gene Set Enrichment Analysis (GSEA) for clusters 9 and 0. These individual expression UMAP plots effectively illustrate the variation in feature expression across different subsets of the data, highlighting the distinct biological insights that can be derived from each cluster.*

7.3. Mlung Stromal Cell Subtype Markers




Pdgfra
0 1 2 3
Fn1
0 1 2 3 4
Tcf21
0 1 2
Col1a1
0 1 2 3 4 5
Col1a2
0 1 2 3 4
Wt1
0.0 0.5 1.0 1.5 2.0
Msln
0 1 2 3
Upk3b
0 1 2 3
Upk1b
0 1 2
Krt18
0 1 2 3
Krt19
0 1 2 3

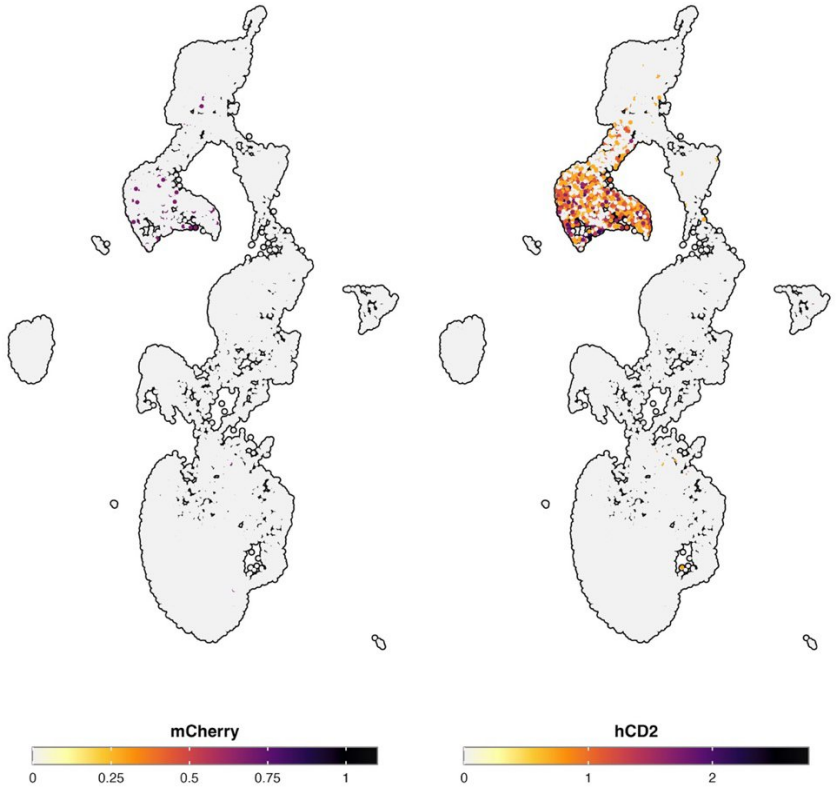
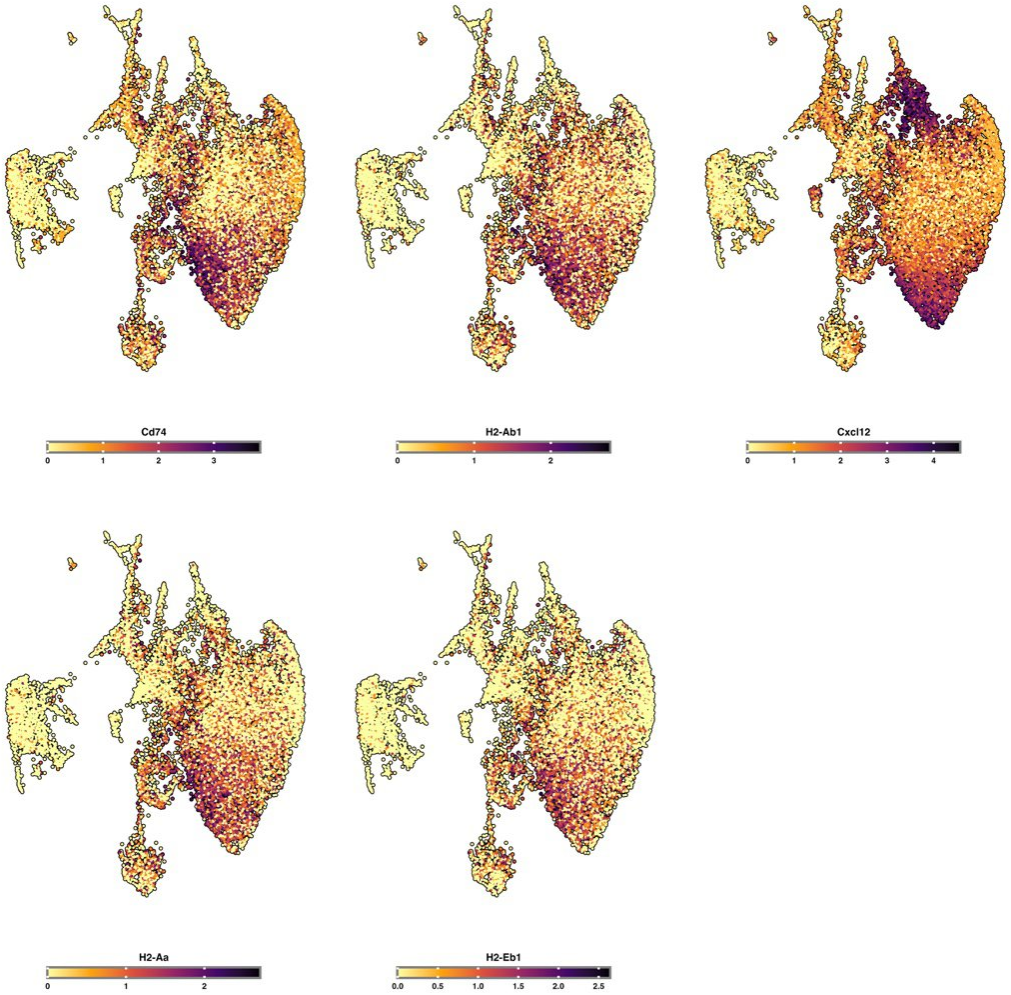
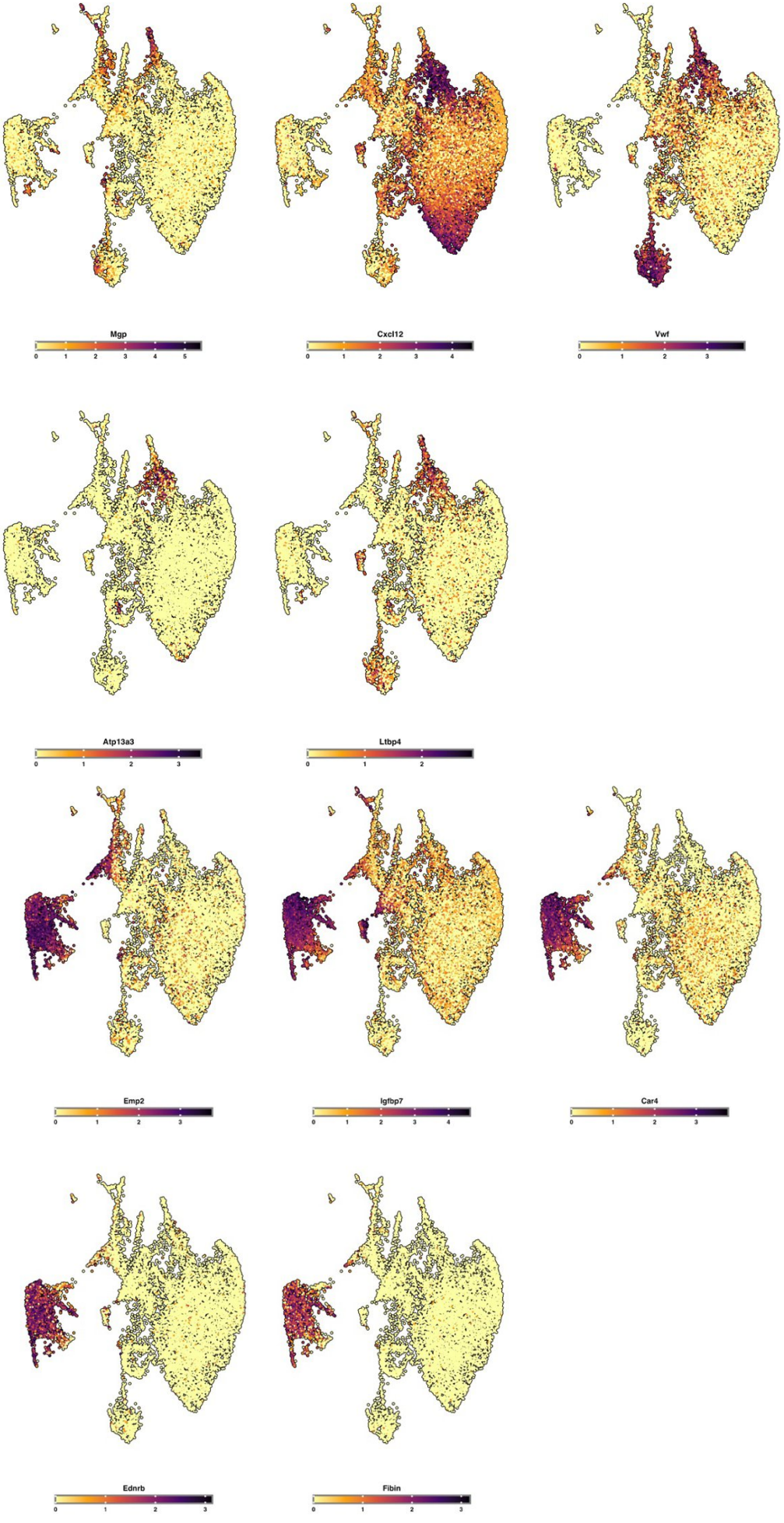
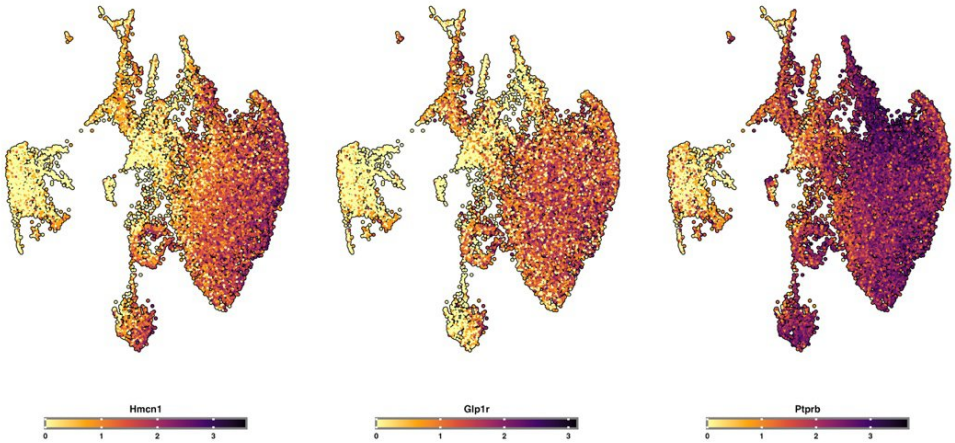
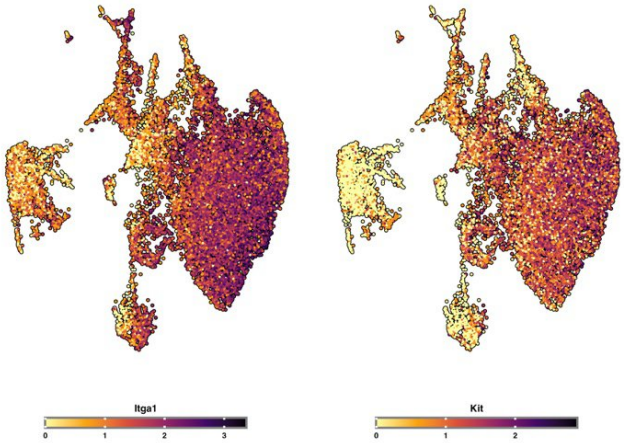


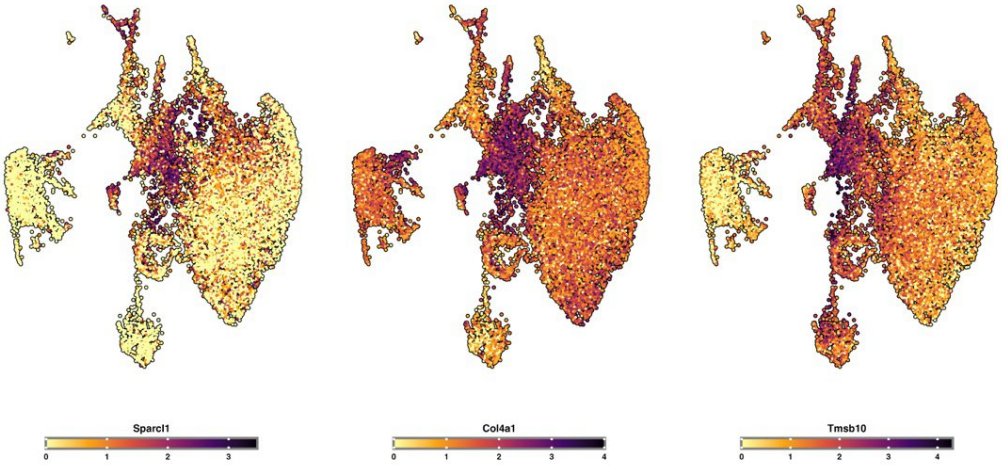
Figure 7-6: UMAP Visualization of Cell Type-Specific Marker Expression *- These UMAP plots highlight regions within the dataset that exhibit high expression levels of canonical markers associated with stromal cell types. A) Endothelial markers, B) Epithelial markers, C) Fibroblasts, D) Mesothelial markers, E) E0771 cancer cell markers, displaying the diverse cellular landscapes and their distribution patterns across different cell types.*

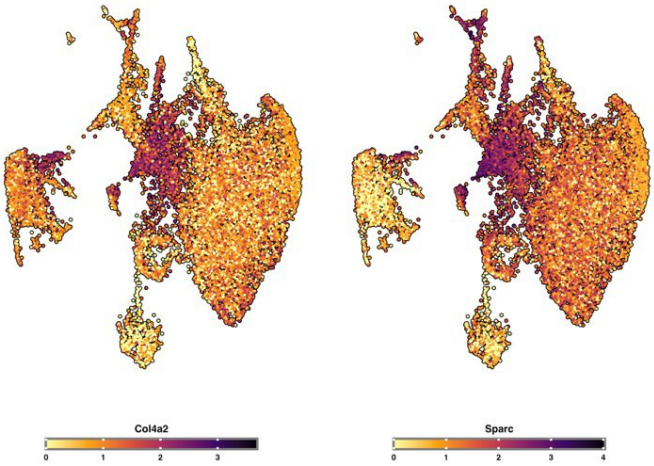


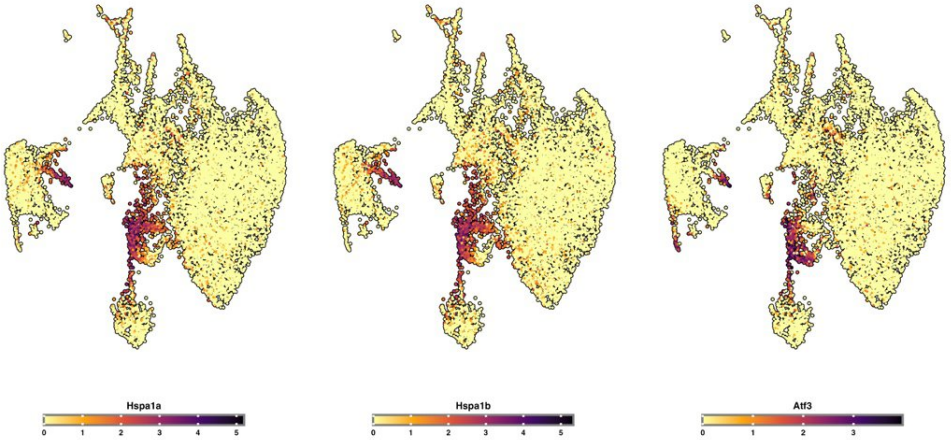

Mgp
0 1 2 3 4 5
Cxcl12
0 1 2 3 4
Vwf
0 1 2 3
Atp13a3
0 1 2 3
Ltbp4
0 1 2
Emp2
0 1 2 3
Igfbp7
0 1 2 3 4
Car4
0 1 2 3
Ednrb
0 1 2 3
Fibin
0 1 2 3

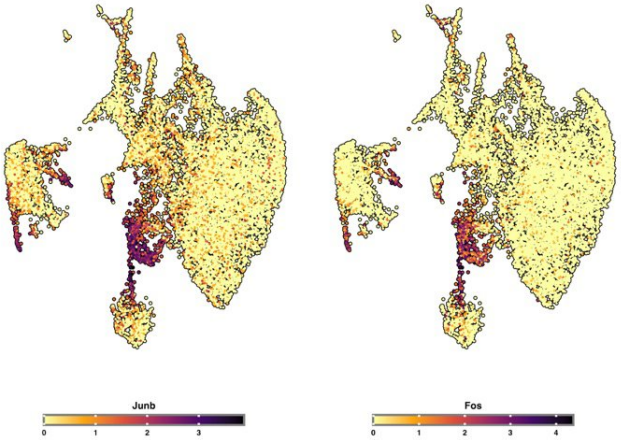

Hmcn1
0 1 2 3
Glp1r
0 1 2 3
Ptprb
0 1 2 3

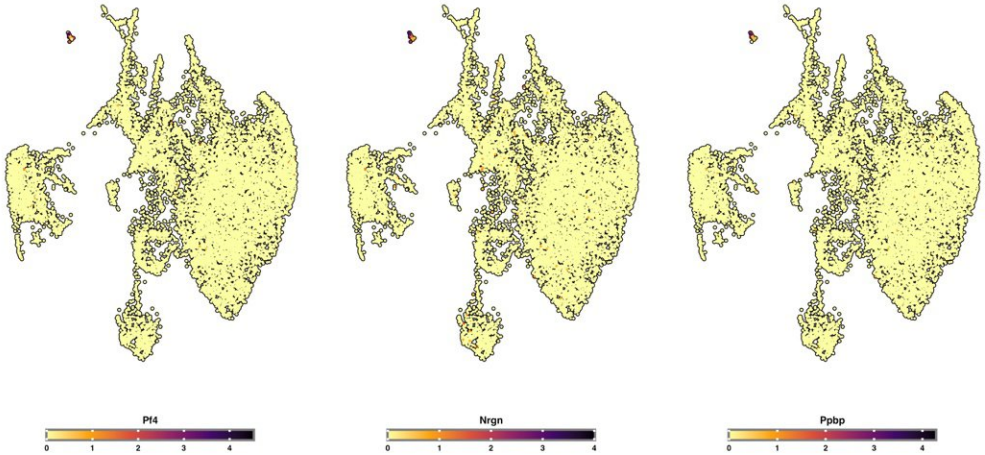

Itga1
0 1 2 3
Kit
0 1 2

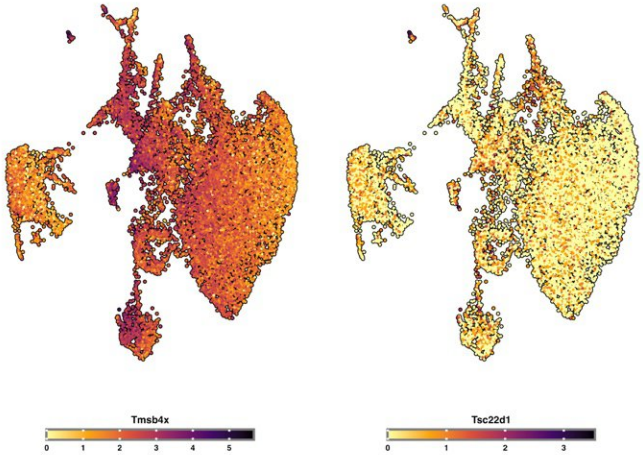

Sparcl1
0 1 2 3
Col4a1
0 1 2 3 4
Tmsb10
0 1 2 3 4

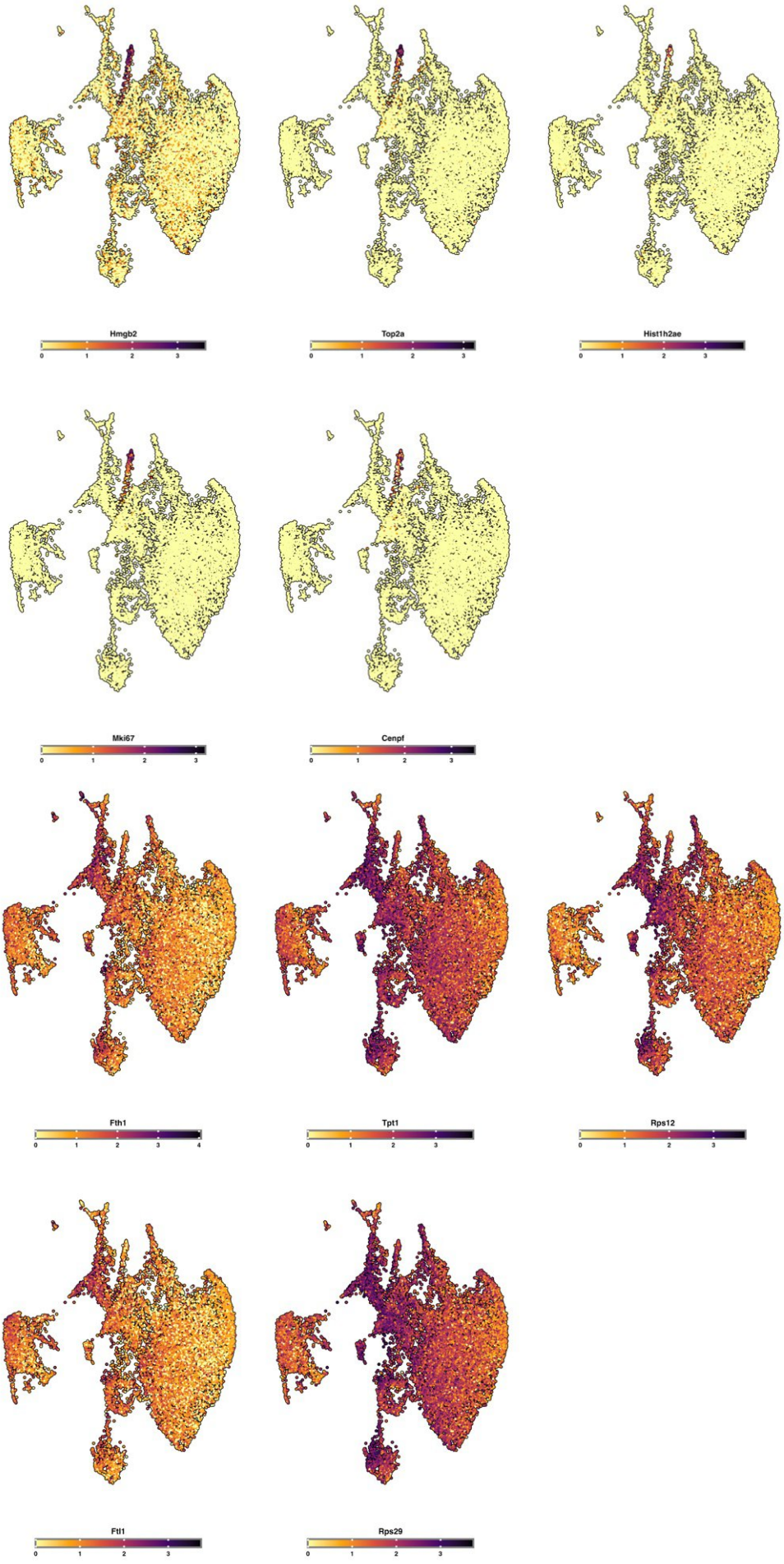

Col4a2
0 1 2 3
Sparc
0 1 2 3 4

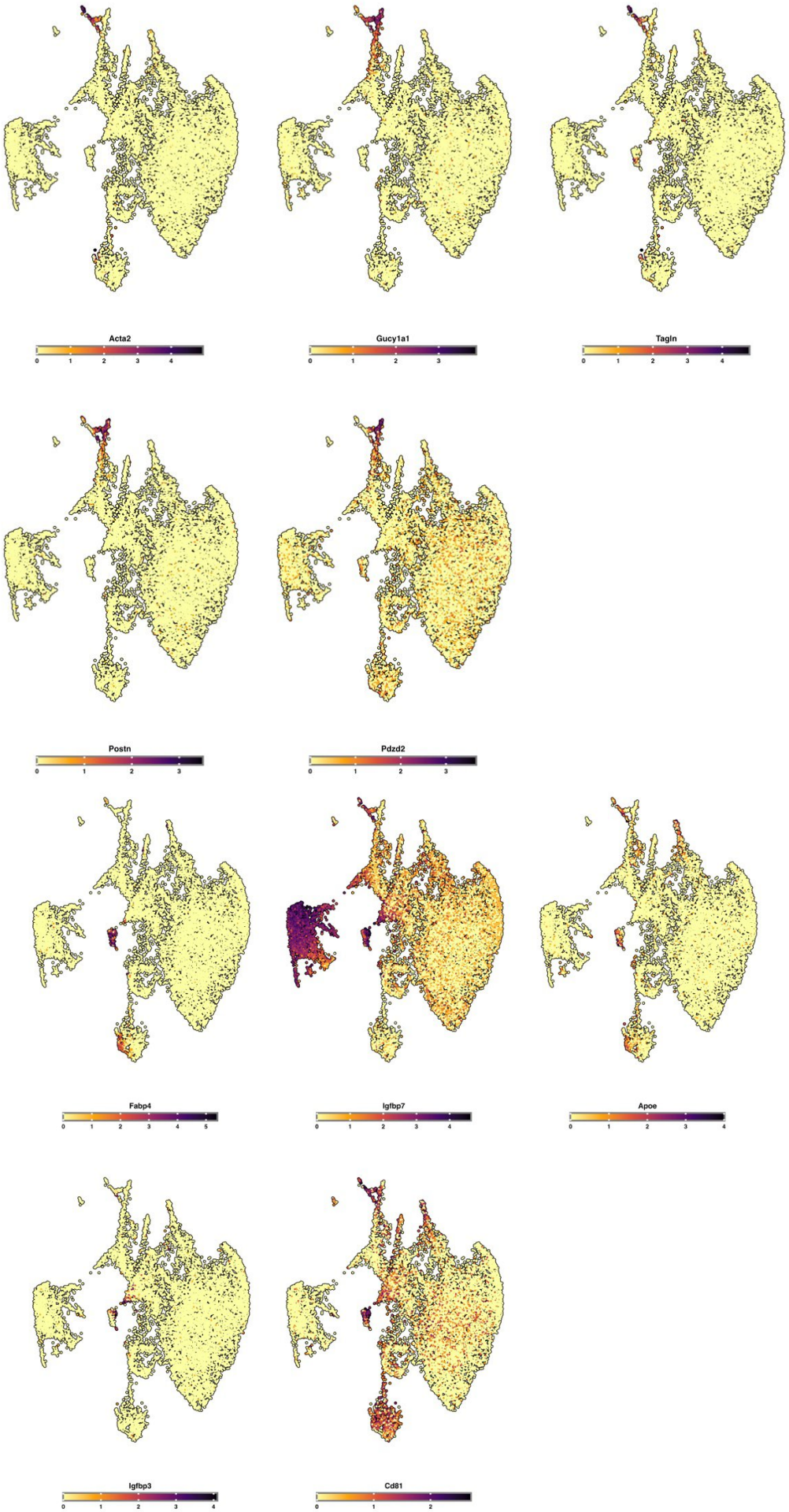

Hspa1a
0 1 2 3 4 5
Hspa1b
0 1 2 3 4 5
Atf3
0 1 2 3


Junb
0 1 2 3
Fos
0 1 2 3 4


Pf4
0 1 2 3 4
Nrgn
0 1 2 3 4
Ppbp
0 1 2 3 4


Tmsb4x
0 1 2 3 4 5
Tsc22d1
0 1 2 3


Hmgb2
0 1 2 3
Top2a
0 1 2 3
Hist1h2ae
0 1 2 3
Mki67
0 1 2 3
Cenpf
0 1 2 3
Fth1
0 1 2 3 4
Tpt1
0 1 2 3
Rps12
0 1 2 3
Ftl1
0 1 2 3
Rps29
0 1 2 3


Acta2
Gucy1a1
Tagln
Postn
Pdzd2
Fabp4
Igfbp7
Apoe
Igfbp3
Cd81

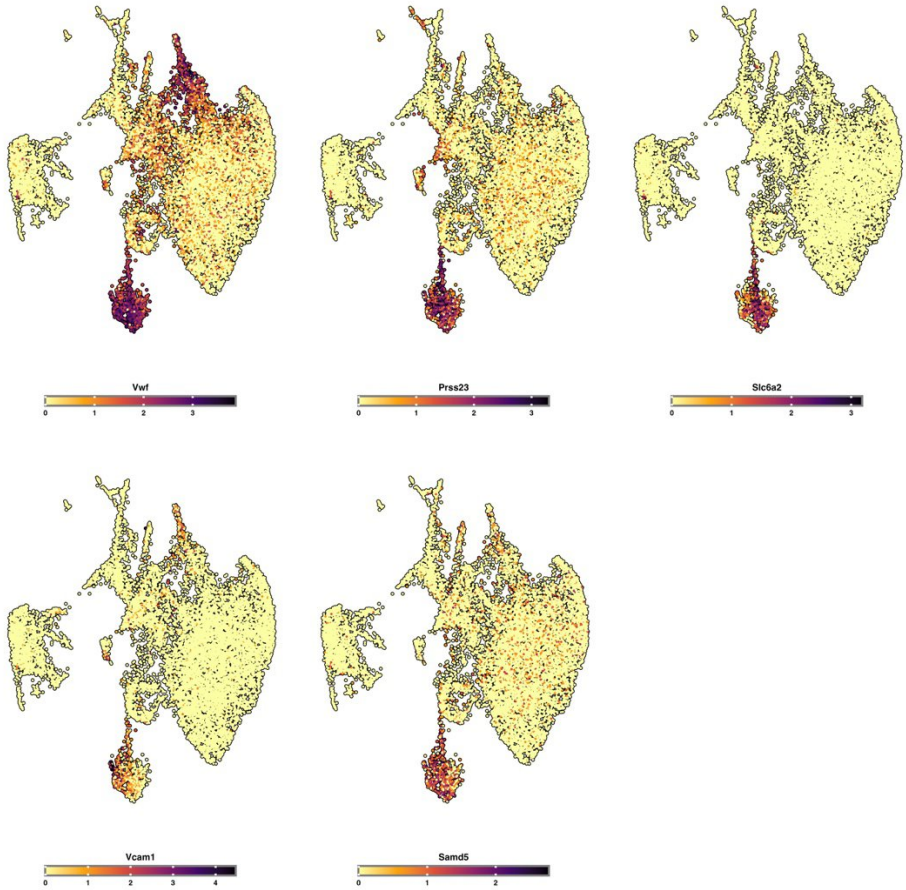


Figure 7-7: Expression UMAP Plots for Cell Subtype Markers - *This figure presents UMAP plots displaying the expression patterns of various cell subtype markers across different endothelial cell populations. Each panel highlights the expression levels of specific markers, which are crucial for identifying and differentiating between endothelial subtypes: A) Markers for antigen-presenting capillary type II cells. B) Markers for arterial cells. C) Markers for capillary type I cells. D) Markers for capillary type II cells. E) Markers for EC_breach cells. F) Markers for EC_stressed cells. G) Markers for megakaryocytes. H) Markers for proliferating endothelial cells (ECs). I) Markers for ribosomal-related genes. J) Markers for smooth muscle actin (SMA) or myofibroblasts. K) Markers for tumour endothelial cells (ECs). L) Markers for venous cells.*

7.4. Stromal Cell Type and Subtype Markers: Manual Curation by Salwa Lin

Cell type	Markers
Pan-endothelial	CD31 (Pecam1)/Ve-cad(Cdh5) (CD144)/Cldn5
Capillary type I ECs	Car4/Fibin/Scn7a/Cav2/Acvrl1/Klf2/Foxf1
Capillary type II ECs	Plvap/Tmem100/Sema3/Clec14a/Hmcn1/Epas1/Bmpr2 /Tubb2a/Pltp
Lymphatic ECs	Ccl21a/Prox1/Reln/Flt4/Lyve1/Fgl2/Gng11
Veins ECs	Vwf/Adgrg6/Slc6a2/Hspa1a/Car8/Ackr3/CD200/Vegfc /Plac8/Klf4/Car8/Foxp1/Junb
Arteries ECs	Gja4/Cytl1/Ltbp4/Bmx/Pde3a/Tinagl1/Bmp4/Fos
Activated arteries	Mecom/Atp13a3/Gja5/Efna1/Sox17/Ptprr

Interferone ECs	Iigp1/Parp14/Gbp7/Isg15/Igtp/Ifit3
Tip ECs	Trp53i11/Prss23/Tmsb4x/Tmsb10/Actb/Mest
Breach ECs	Kit/Dll4/Esm1/Col4a1/Col4a2
postcapillary venules	Ifitm3/Ctla2a/Socs3/Zfp36/Aqp1/Jun
Neophalanx	Cxcl12/Btg2
Immature ECs	Pdlim3/Ccnd1
Stressed ECs	Atf3
Proliferating ECs	Npm1/H2afz/Mif/Hmgb2
Pericytes	Cox4i2/Higd1b/Pag1/Cspg4/Fam162b/Ndufa4l2
SMC	Acta2/Myh11/Actg2/Actc1/Aspn/Rgs2
Myofibroblasts	Tagln/Pdgfra/Tgfbi/Hhip/Enpp2/Wnt5a/Acta2/Myh11
Fibroblasts	Pdgfra/Fn1
Lipofibroblasts	Tcf21/Plin2/Fgf10/G0s2/Gyg/Macf1/Wnt2/Col13a1/Mfap4
Interstitial fibroblasts	Agtr2/Prss35/Igfbp7/Fbln5/Ptn/Heyl/Fstl1Tm4sf1/Timp1/Inmt/Npnt/Gsn/Mm p3
Ebf1 positive fibroblasts	Scara5/Serpinf1/Pdgfrb/Higd1b/Cox4i2/Notch3/Ebf1/Gucy1a3/Pdzd2/Postn/ Pi16
Cancer associated Fibroblasts	S100s4
Mesothelial cells	Upk3b/Wt1/Msln/Upk1b/Aldh1a2/Cpe/Gm12840/Aqp1/Igfbp6/Cfb/Ccdc80/C lec3b
Proliferating cells	Mki67/Hmmr/Top2a/Spc25/Cdca3/Ccnb2/Hist1h2apRps29/Rpl37/Actg1/Gim ap4/Ube2a/Nusap1/Prc1/Cenpf/Cenpe/Lgals1/Ay036118/smc4
Proliferative Mesenchymal Progenitors	Hlst1h2ap
Epithelial cells	Epcam/Krt8

Type I pneumocytes	Anxa1/Cystm1/Ahnak/Hopx/Gipc2/Osbpl6/Akap5
Type II pneumocytes	Etv5/Cbr2/Sftpc, Sftpa1, Sftpb
Ciliated cells	Sec14l3/Anxa1/Cystm1/Gipc2/Osbpl6/Ccdc153/Dynlrb2/Tmem212
Clara cells	Krt15/Scgb3a2/
Goblet cells	Aldh3a1/
Red blood cells (RBCs)	Hba-a2/Epor/Hba-a1
Erythroid progenitor or Erythroblasts	Hba-x/Hbb-bh1/Hbb-y
Megakaryocytes	Itga2b/Gng11/Ppbp/Clec1b/Mest/Dut/Nrgn/Alox12

7.5. Signalling Cascades

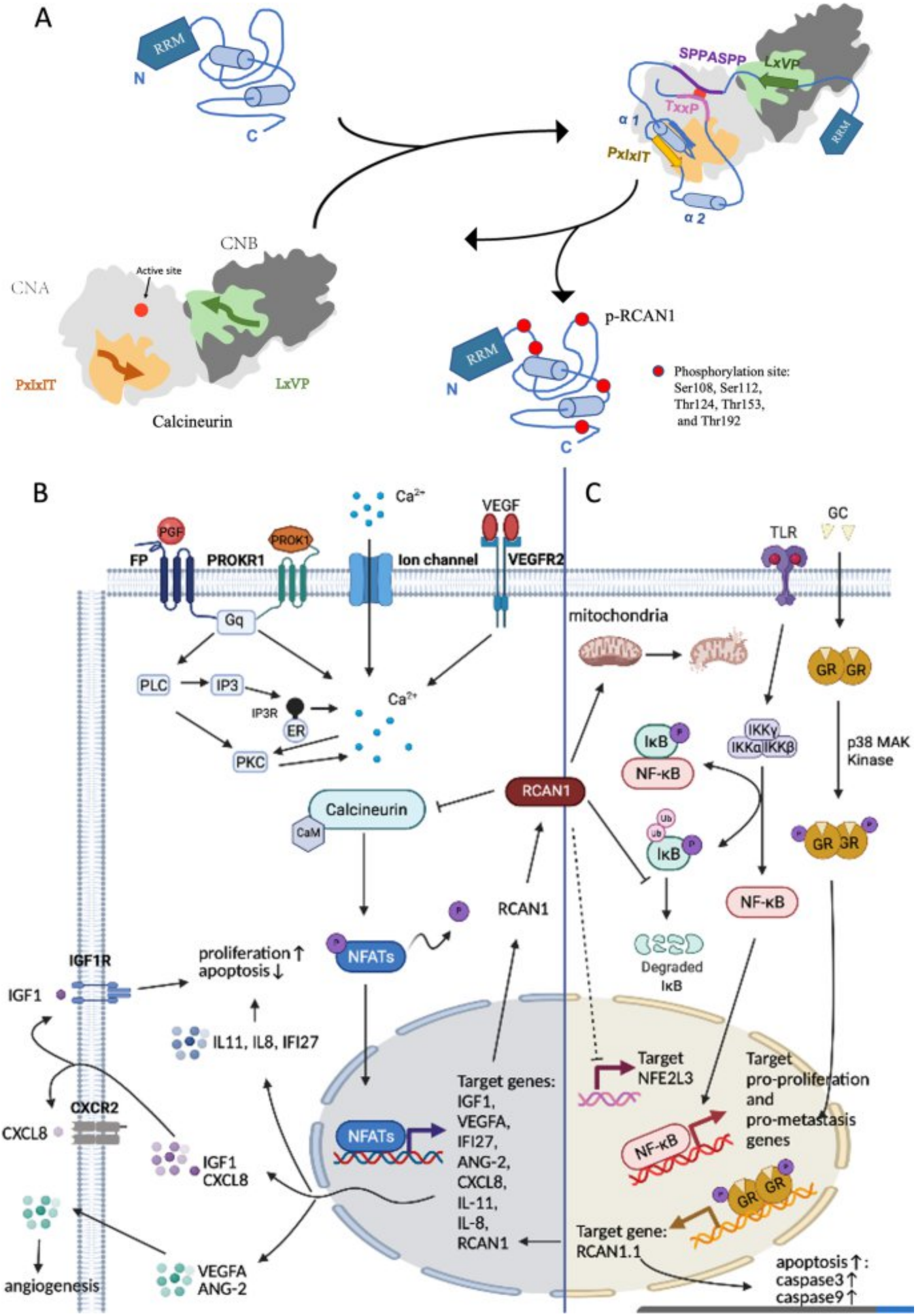

A
RRM
N
C
SPPASPP
LxVP
TxxP
α 1
PxIxIT
α 2
RRM
CNB
CNA
Active site
PxIxIT
LxVP
Calcineurin
p-RCAN1
RRM
N
C
Phosphorylation site:
Ser108, Ser112,
Thr124, Thr153,
and Thr192
B
C
Ca2+
VEGF
PGF
PROK1
FP
PROKR1
Ion channel
VEGFR2
Gq
PLC
IP3
IP3R
ER
Ca2+
PKC
RCAN1
Calcineurin
CaM
NFATs
RCAN1
IGF1R
IGF1
proliferation ↑
apoptosis ↓
IL11, IL8, IFI27
CXCR2
CXCL8
IGF1
CXCL8
VEGFA
ANG-2
angiogenesis
Target genes:
IGF1,
VEGFA,
IFI27,
ANG-2,
CXCL8,
IL-11,
IL-8,
RCAN1
NFATs
TLR
GC
mitochondria
GR GR
IκB
NF-κB
IKKγ
IKKα IKKβ
p38 MAK
Kinase
Ub
IκB
GR GR
NF-κB
Degraded
IκB
Target
NFE2L3
Target
pro-proliferation
and
pro-metastasis
genes
NF-κB
GR GR
Target gene:
RCAN1.1
apoptosis ↑:
caspase3 ↑
caspase9 ↑

Figure 7-8: RCAN1 Signalling Cascade – *RCAN1 acts as a potent tumour suppressor gene through its direct inhibition of the CN-NFAT pathway, which is responsible for the upregulation of IGF1, VEGFA, and CXCL8, among other targets. These genes play key roles in proliferative, angiogenic, and anti-apoptotic signalling, making RCAN1-mediated inhibition of CN-NFAT signalling strongly tumour-suppressive. RCAN1 also functions through alternative mechanisms that do not involve CN binding and inhibition. RCAN1 inhibits cytochrome C release, a crucial step in mitochondrial permeabilization and the execution of the intrinsic apoptotic pathway; this is one of the few functions that RCAN1 performs that might aid tumour survival. RCAN1 overexpression is also implicated in the upregulation of caspases 9 and 3, the primary effectors of both the intrinsic and extrinsic apoptotic cascades. Figure adapted from (Lao et al., 2022).*

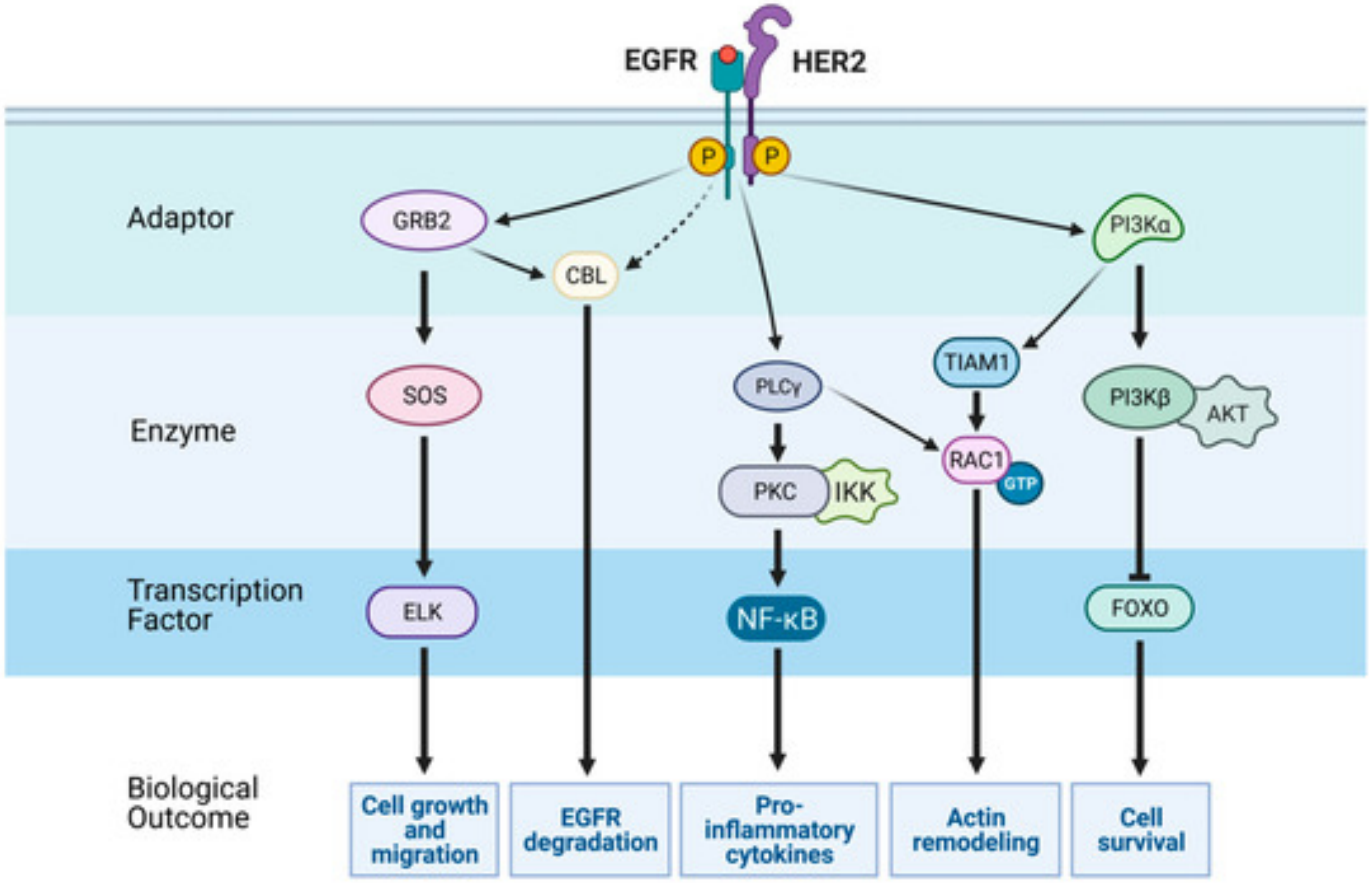


Figure 7-9: EGFR Signalling Cascade *– EGF binding to EGFR results in receptor dimerization, auto-phosphorylation, and activation of the tyrosine kinase domains. These phosphorylate various targets that contribute to cell survival, actin remodelling, cell growth and proliferation, and the production of pro-inflammatory cytokines. All these signalling cascades are highly tumorigenic, hence the potentiation of EGFR signalling during cancer development. Figure adapted from (Uribe et al., 2021).*

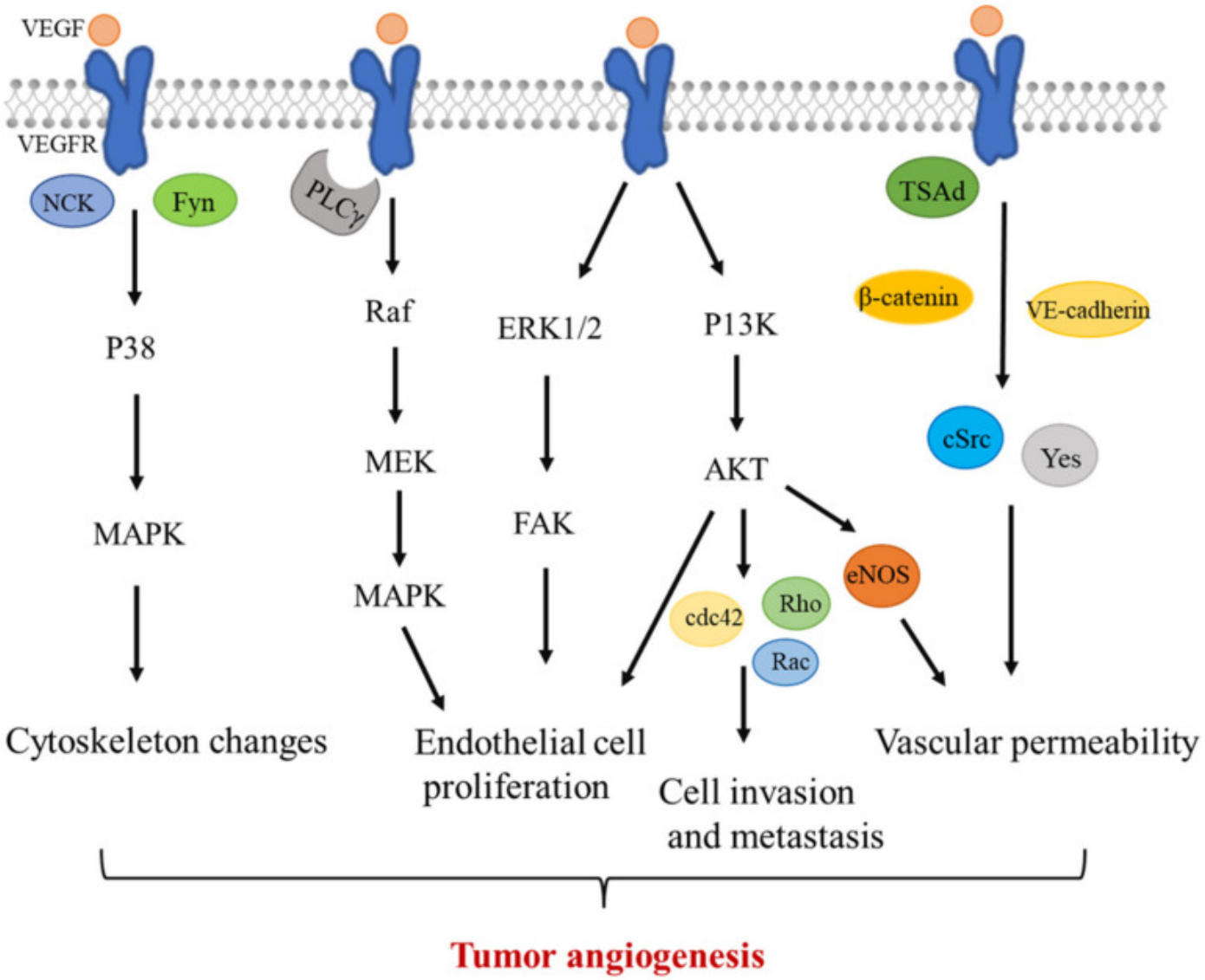
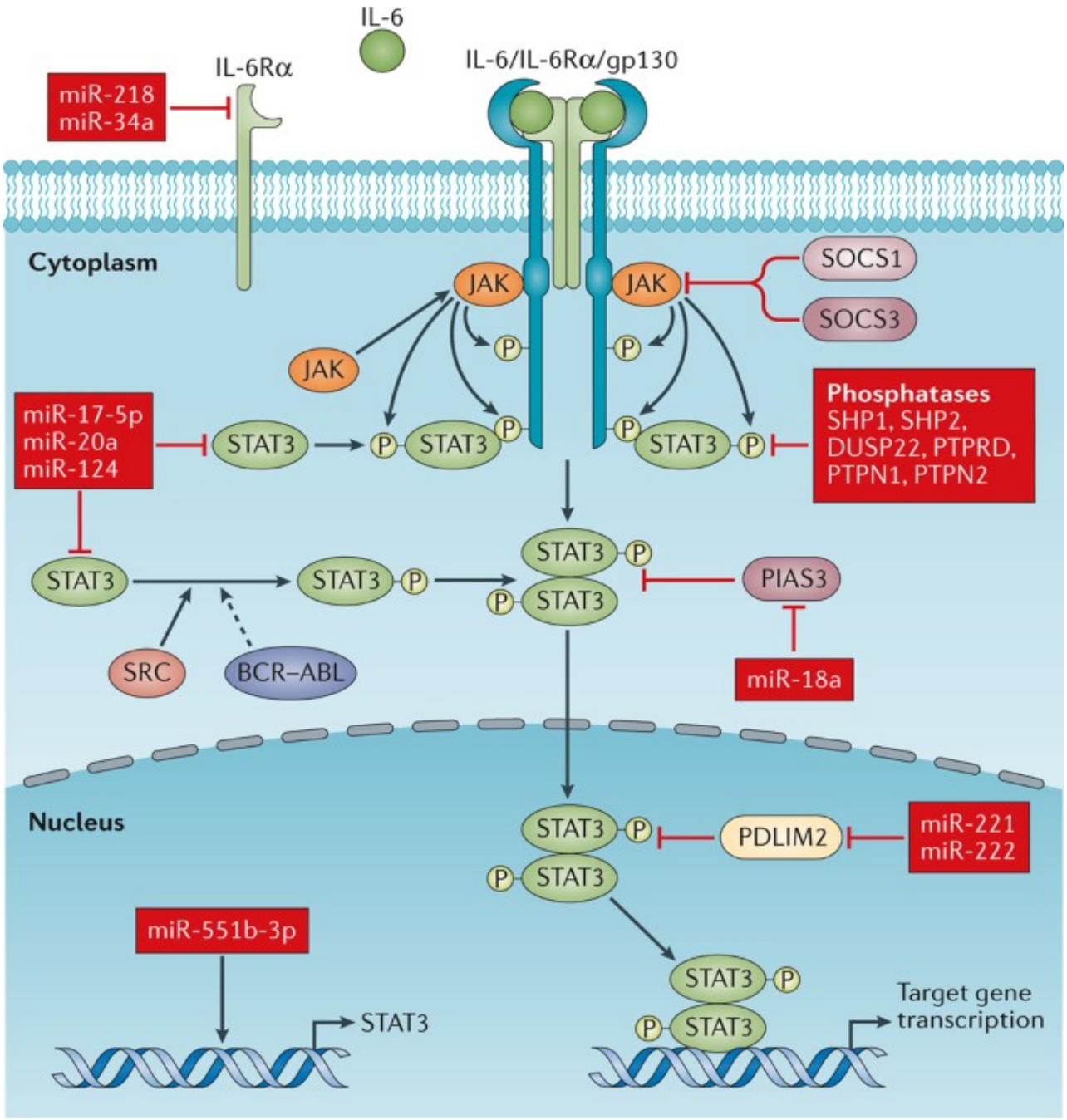


Figure 7-10: VEGFA Signalling Cascade and Tumour Angiogenesis – *VEGFA binds to VEGFR on the surface of endothelial cells, promoting endothelial cell motility, proliferation, and angiogenic capacity. VEGFR is a receptor tyrosine kinase (RTK); thus, its activation by VEGFA binding results in dimerization, auto-phosphorylation, and receptor activation. The binding of interacting partner proteins with binding domains (e.g., SH2 and PTB domains) that have a high affinity for phosphorylated tyrosine residues leads to the activation of these interacting elements through physical proximity. Key activated signalling pathways in this process include the PI3K-AKT signalling cascade, a potent promoter of cell survival and proliferation, and the ERK-FAK signalling axis, a strong promoter of cellular motility through actin cytoskeleton restructuring. These changes facilitate tumour neovascularisation through endothelial cell proliferation, supporting tumour growth and progression. Figure adapted from (S. Qi et al., 2022).*



Nature Reviews | Clinical Oncology

Figure 7-11: IL-6 Signalling Cascade – *IL-6 binds to its receptor IL-6R on the plasma membrane. Once IL-6 binds to IL-6R, the ligand-receptor complex associates with the membrane glycoprotein gp130 and induces its dimerisation. On the C-domain of gp130, Janus kinase (JAK) is constitutively bound, and IL-6R-dependent dimerisation of gp130 results in the transphosphorylation of the JAK kinases. Signal transducer and activator of transcription (STAT) proteins are transcription factors containing the SH2 binding domain, which has a high affinity for phosphotyrosine residues. This drives the translocation of STAT3 proteins to the plasma membrane, where the phosphorylated gp130 dimers are located. This translocation results in the binding and subsequent phosphorylation of STATs by JAK, leading to the dimerisation of STAT proteins through SH2-phosphotyrosine interactions. STAT3 dimerisation activates these proteins, which is rapidly followed by their nuclear translocation to regulate gene expression. Figure adapted from (D. E. Johnson et al., 2018).*

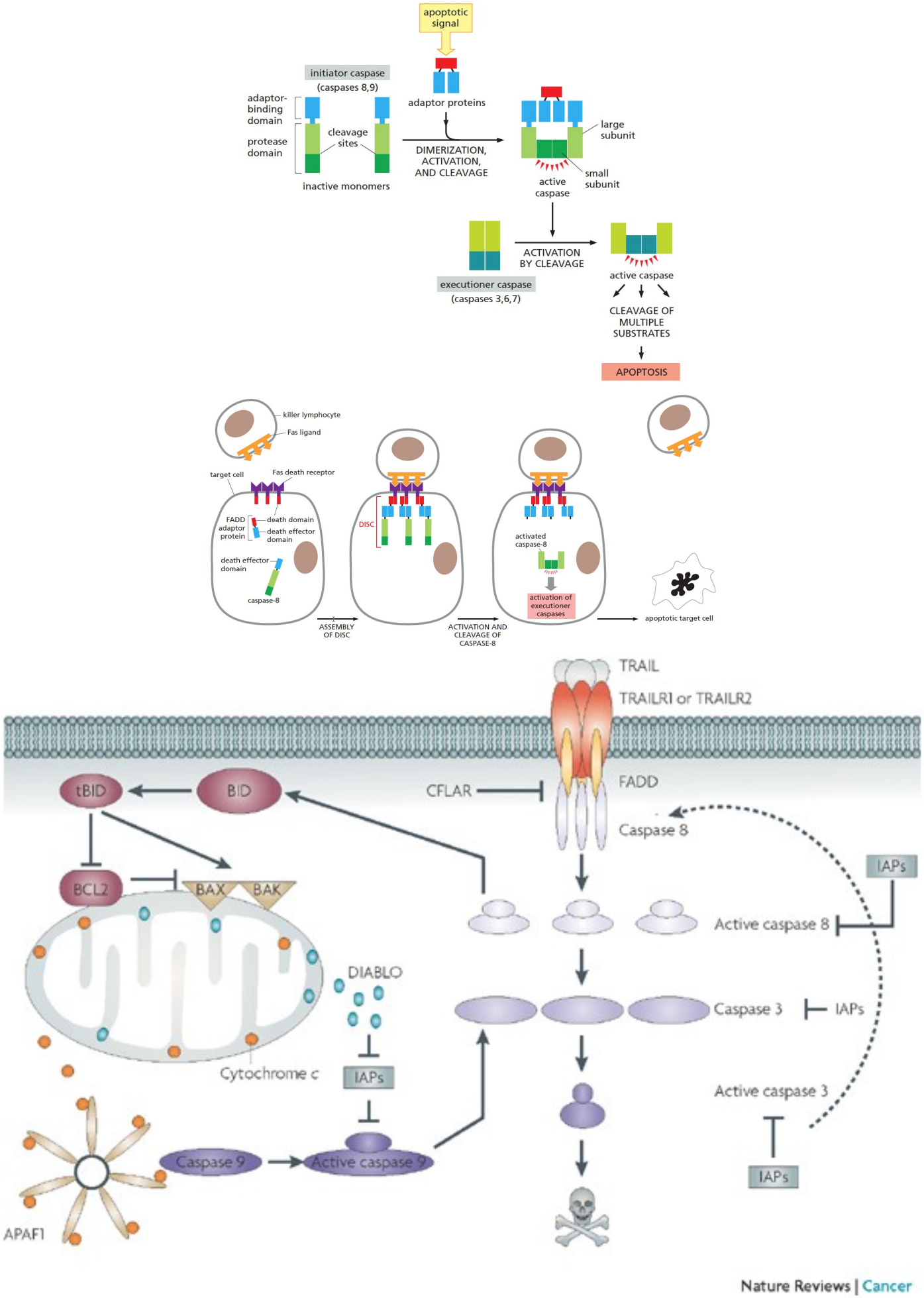

apoptotic signal
initiator caspase
(caspases 8,9)
adaptor-binding domain
protease domain
cleavage sites
inactive monomers
adaptor proteins
DIMERIZATION, ACTIVATION, AND CLEAVAGE
large subunit
small subunit
active caspase
executioner caspase
(caspases 3,6,7)
ACTIVATION BY CLEAVAGE
active caspase
CLEAVAGE OF MULTIPLE SUBSTRATES
APOPTOSIS
killer lymphocyte
Fas ligand
target cell
Fas death receptor
FADD adaptor protein
death domain
death effector domain
death effector domain
caspase-8
DISC
activated caspase-8
activation of executioner caspases
apoptotic target cell
ASSEMBLY OF DISC
ACTIVATION AND CLEAVAGE OF CASPASE-8
TRAIL
TRAILR1 or TRAILR2
FADD
CFLAR
tBID
BID
Caspase 8
BCL2
BAX
BAK
IAPs
Active caspase 8
DIABLO
Caspase 3
IAPs
Cytochrome c
IAPs
Active caspase 3
IAPs
Caspase 9
Active caspase 9
APAF1
Nature Reviews | Cancer

Figure 7-12: Extrinsic Pathway of Cell Death – *Apoptotic cell death is mediated by both extrinsic and intrinsic pathways. This figure focuses on the extrinsic pathway.*
A) Caspase Activation Cascade: *Caspases are the effector proteins in apoptotic signalling. They are proteases, specializing in cleaving other proteins. Regardless of the apoptotic signal, it results in the dimerisation, activation, and trans-cleavage of initiator caspases 8 and 9. This is followed by the cleavage and activation of effector caspases 3, 6, and 7, which then target multiple cellular proteins to mediate apoptosis.*
B) Extrinsic Apoptotic Pathway: *This panel illustrates the signal mediated by the Fas ligand on the surface of killer lymphocytes binding to the Fas death receptor on the surface of infected or cancer cells. Ligand binding and receptor activation lead to the recruitment of the FADD adaptor protein, which has a death domain and a death effector domain. The death domain is essential for adaptor protein translocation to the cell surface to bind to the Fas death receptor, while the death effector domain binds to caspase 8. The binding of caspase 8 to the FADD adaptor protein results in caspase 8 cleavage, dimerisation, and activation, which subsequently activates the effector caspases, thereby mediating the apoptotic pathway as explained in panel A.*
C) TNFSF10/TRAIL Extrinsic Apoptotic Pathway: *TNFSF10/TRAIL is a soluble ligand that binds to TRAIL-R1 and TRAIL-R2 receptors on cancer cells, activating the extrinsic apoptotic pathway through the same sequence of molecular events described previously. Panels A and B adapted from (Alberts et al., 2014). Panel C adapted from (Johnstone et al., 2008).*

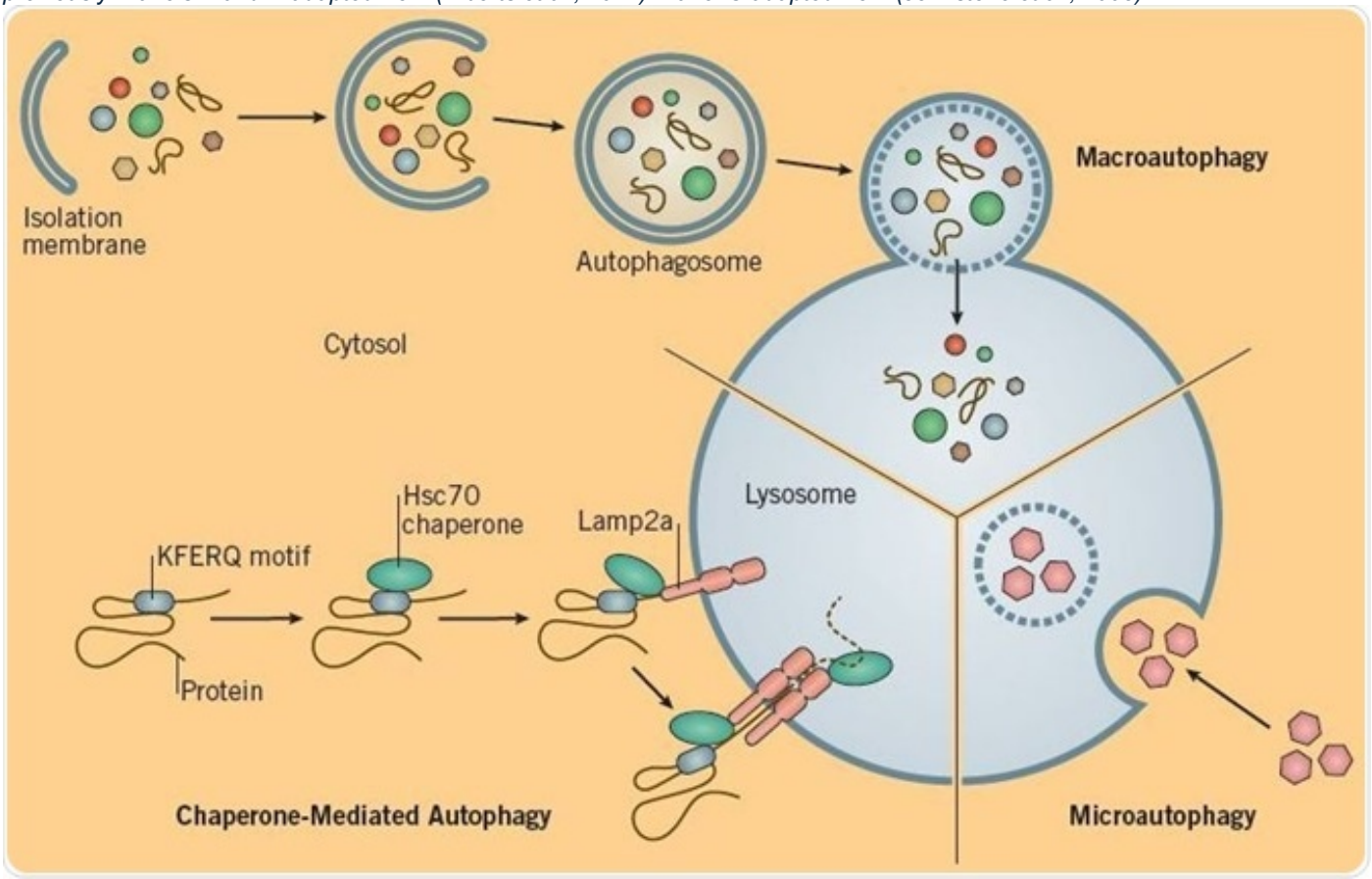


Figure 7-13: Overview of Autophagy *– Autophagy is a survival pathway used by cells, both normal and cancerous, to adapt to cellular stress. Pathway activation results in the creation of an isolation membrane within the cytoplasm, which engulfs all cytosolic components bound for degradation. Once the isolation membrane closes, it forms the autophagosome, which then fuses with a lysosome to form an autolysosome. Lysosomes provide acid hydrolases and caspases, responsible for degrading the engulfed components of the autophagosome. Broadly, there are three types of autophagy: macro-autophagy, micro-autophagy, and chaperone-mediated autophagy. Macro-autophagy is responsible for the degradation of organelles. Micro-autophagy degrades smaller cytoplasmic components, and chaperone-mediated autophagy targets specific cellular components for degradation by chaperone proteins. The outcome of autophagy is the degradation of the engulfed components and the release of macromolecules to support metabolism under conditions of cellular stress. Figure adapted from (Function and Mechanisms of Autophagy, 2019).*

8. Bibliography

A, L., T, U., & A, I. (2001). Id proteins at the cross-road of development and cancer. *Oncogene*, *20*(58). https://doi.org/10.1038/sj.onc.1205093

Abraham, A. G., & O'Neill, E. (2014). PI3K/Akt-mediated regulation of p53 in cancer. *Biochemical Society Transactions*, *42*(4), 798–803. https://doi.org/10.1042/BST20140070

Abu-Jamous, B., Buffa, F. M., Harris, A. L., & Nandi, A. K. (2017). In vitro downregulated hypoxia transcriptome is associated with poor prognosis in breast cancer. *Molecular Cancer*, *16*(1), 105. https://doi.org/10.1186/s12943-017-0673-0

Alberts, B., Johnson, A., Lewis, J., Morgan, D., Raff, M., Roberts, K., & Walter, P. (2014). *Molecular Biology of the Cell* (6th ed.). Garland Science.

Ali, Z. (2018). *Investigating mechanisms of angiogenesis in health and disease using zebrafish models*. https://doi.org/10.3384/diss.diva-153266

Andrews, S., Krueger, F., Segonds-Pichon, A., Biggins, L., Krueger, C., & Wingett, S. (2012, January). *FastQC*.

Angelidis, I., Simon, L. M., Fernandez, I. E., Strunz, M., Mayr, C. H., Greiffo, F. R., Tsitsiridis, G., Ansari, M., Graf, E., Strom, T.-M., Nagendran, M., Desai, T., Eickelberg, O., Mann, M., Theis, F. J., & Schiller, H. B. (2019). An atlas of the aging lung mapped by single cell transcriptomics and deep tissue proteomics. *Nature Communications*, *10*(1), 963. https://doi.org/10.1038/s41467-019-08831-9

Bagegni, N., Thomas, S., Liu, N., Luo, J., Hoog, J., Northfelt, D. W., Goetz, M. P., Forero, A., Bergqvist, M., Karen, J., Neumüller, M., Suh, E. M., Guo, Z., Vij, K., Sanati, S., Ellis, M., & Ma, C. X. (2017). Serum thymidine kinase 1 activity as a pharmacodynamic marker of cyclin-dependent kinase 4/6 inhibition in patients with early-stage breast cancer receiving neoadjuvant palbociclib. *Breast Cancer Research*, *19*(1), 123. https://doi.org/10.1186/s13058-017-0913-7

Bao, L., Odell, A. F., Stephen, S. L., Wheatcroft, S. B., Walker, J. H., & Ponnambalam, S. (2012). The S100A6 calcium-binding protein regulates endothelial cell-cycle progression and senescence. *The FEBS Journal*, *279*(24), 4576–4588. https://doi.org/10.1111/febs.12044

Barberis, A., Aerts, H. J. W. L., & Buffa, F. M. (2024). Robustness and reproducibility for AI learning in biomedical sciences: RENOIR. *Scientific Reports*, *14*(1), 1–13. https://doi.org/10.1038/s41598-024-51381-4

Bardia, A., Hurvitz, S. A., Tolaney, S. M., Loirat, D., Punie, K., Oliveira, M., Brufsky, A., Sardesai, S. D., Kalinsky, K., Zelnak, A. B., Weaver, R., Traina, T., Dalenc, F., Aftimos, P., Lynce, F., Diab, S., Cortés, J., O'Shaughnessy, J., Diéras, V., … Rugo, H. S. (2021). Sacituzumab Govitecan in Metastatic Triple-Negative Breast Cancer. *New England Journal of Medicine*, *384*(16), 1529–1541. https://doi.org/10.1056/NEJMoa2028485

Bellerby, R., Smith, C., Kyme, S., Gee, J., Günthert, U., Green, A., Rakha, E., Barrett-Lee, P., & Hiscox, S. (2016). Overexpression of Specific CD44 Isoforms Is Associated with Aggressive Cell Features in Acquired Endocrine Resistance. *Frontiers in Oncology*, *6*, 145. https://doi.org/10.3389/fonc.2016.00145

Bhatia, N., Hazra, S., & Thareja, S. (2023). Selective Estrogen receptor degraders (SERDs) for the treatment of breast cancer: An overview. *European Journal of Medicinal Chemistry*, *256*, 115422. https://doi.org/10.1016/j.ejmech.2023.115422

Bhatia, N., & Thareja, S. (2023). Elacestrant: A new FDA-approved SERD for the treatment of breast cancer. *Medical Oncology (Northwood, London, England)*, *40*(6), 180. https://doi.org/10.1007/s12032-023-02045-2

Bondar, V. M., Sweeney-Gotsch, B., Andreeff, M., Mills, G. B., & McConkey, D. J. (2002). Inhibition of the phosphatidylinositol 3'-kinase-AKT pathway induces apoptosis in pancreatic carcinoma cells in vitro and in vivo. *Molecular Cancer Therapeutics*, *1*(12), 989–997.

Buffa, F. M., Harris, A. L., West, C. M., & Miller, C. J. (2010a). Large meta-analysis of multiple cancers reveals a common, compact and highly prognostic hypoxia metagene. *British Journal of Cancer*, *102*(2), 428–435. https://doi.org/10.1038/sj.bjc.6605450

Buffa, F. M., Harris, A. L., West, C. M., & Miller, C. J. (2010b). Large meta-analysis of multiple cancers reveals a common, compact and highly prognostic hypoxia metagene. *British Journal of Cancer*, *102*(2), 428–435. https://doi.org/10.1038/sj.bjc.6605450

Cao, J., Spielmann, M., Qiu, X., Huang, X., Ibrahim, D. M., Hill, A. J., Zhang, F., Mundlos, S., Christiansen, L., Steemers, F. J., Trapnell, C., & Shendure, J. (2019). The single-cell transcriptional landscape of mammalian organogenesis. *Nature*, *566*(7745), 496–502. https://doi.org/10.1038/s41586-019-0969-x

Chen, C., Zhao, S., Karnad, A., & Freeman, J. W. (2018). The biology and role of CD44 in cancer progression: Therapeutic implications. *Journal of Hematology & Oncology*, *11*, 64. https://doi.org/10.1186/s13045-018-0605-5

Chen, Y., Pal, B., Lindeman, G. J., Visvader, J. E., & Smyth, G. K. (2022). R code and downstream analysis objects for the scRNA-seq atlas of normal and tumorigenic human breast tissue. *Scientific Data*, *9*(1), 96. https://doi.org/10.1038/s41597-022-01236-2

Chen, Z., Han, F., Du, Y., Shi, H., & Zhou, W. (2023). Hypoxic microenvironment in cancer: Molecular mechanisms and therapeutic interventions. *Signal Transduction and Targeted Therapy*, *8*(1), 1–23. https://doi.org/10.1038/s41392-023-01332-8

Cheng, Y., Li, S., Gao, L., Zhi, K., & Ren, W. (2021). The Molecular Basis and Therapeutic Aspects of Cisplatin Resistance in Oral Squamous Cell Carcinoma. *Frontiers in Oncology*, *11*. https://doi.org/10.3389/fonc.2021.761379

Chia, S. K., Wykoff, C. C., Watson, P. H., Han, C., Leek, R. D., Pastorek, J., Gatter, K. C., Ratcliffe, P., & Harris, A. L. (2001). Prognostic significance of a novel hypoxia-regulated marker, carbonic anhydrase IX, in invasive breast carcinoma. *Journal of Clinical Oncology: Official Journal of the American Society of Clinical Oncology*, *19*(16), 3660–3668. https://doi.org/10.1200/JCO.2001.19.16.3660

Cobos, F. A., Alquicira-Hernández, J., Powell, J., Mestdagh, P., & Preter, K. D. (2020). Benchmarking of cell type deconvolution pipelines for transcriptomics data. *Nature Communications*, *11*. https://doi.org/10.1038/s41467-020-19015-1

Comşa, Ş., Cîmpean, A. M., & Raica, M. (2015). The Story of MCF-7 Breast Cancer Cell Line: 40 years of Experience in Research. *Anticancer Research*, *35*(6), 3147–3154.

Cortes, J., Rugo, H. S., Cescon, D. W., Im, S.-A., Yusof, M. M., Gallardo, C., Lipatov, O., Barrios, C. H., Perez-Garcia, J., Iwata, H., Masuda, N., Torregroza Otero, M., Gokmen, E., Loi, S., Guo, Z., Zhou, X., Karantza, V., Pan, W., Schmid, P., & KEYNOTE-355 Investigators. (2022). Pembrolizumab plus Chemotherapy in Advanced Triple-Negative Breast Cancer. *The New England Journal of Medicine*, *387*(3), 217–226. https://doi.org/10.1056/NEJMoa2202809

Coughlin, S. (2013). Epidemiology of Breast Cancer in Women. *Advances in Experimental Medicine and Biology*, *1152*, 9–29. https://doi.org/10.1007/978-1-4614-5647-6_2

Dasari, S., & Tchounwou, P. B. (2014). Cisplatin in cancer therapy: Molecular mechanisms of action. *European Journal of Pharmacology*, *740*, 364–378. https://doi.org/10.1016/j.ejphar.2014.07.025

Devor, E. J., Gonzalez-Bosquet, J., Warrier, A., Reyes, H. D., Ibik, N. V., Schickling, B. M., Newtson, A., Goodheart, M. J., & Leslie, K. K. (2017). P53 mutation status is a primary determinant of placenta-specific protein 1 expression in serous ovarian cancers. *International Journal of Oncology*, *50*(5), 1721–1728. https://doi.org/10.3892/ijo.2017.3931

Dhani, N., Fyles, A., Hedley, D., & Milosevic, M. (2015). The clinical significance of hypoxia in human cancers. *Seminars in Nuclear Medicine*, *45*(2), 110–121. https://doi.org/10.1053/j.semnuclmed.2014.11.002

Ding, X., Park, S. I., McCauley, L. K., & Wang, C.-Y. (2013). Signaling between Transforming Growth Factor β (TGF-β) and Transcription Factor SNAI2 Represses Expression of MicroRNA miR-203 to Promote Epithelial-Mesenchymal Transition and Tumor Metastasis. *The Journal of Biological Chemistry*, *288*(15), 10241–10253. https://doi.org/10.1074/jbc.M112.443655

Dittmer, J. (2022). Biological effects and regulation of IGFBP5 in breast cancer. *Frontiers in Endocrinology*, *13*. https://doi.org/10.3389/fendo.2022.983793

Dobin, A., Davis, C. A., Schlesinger, F., Drenkow, J., Zaleski, C., Jha, S., Batut, P., Chaisson, M., & Gingeras, T. R. (2013a). STAR: Ultrafast universal RNA-seq aligner. *Bioinformatics*, *29*(1), 15–21. https://doi.org/10.1093/bioinformatics/bts635

Dobin, A., Davis, C. A., Schlesinger, F., Drenkow, J., Zaleski, C., Jha, S., Batut, P., Chaisson, M., & Gingeras, T. R. (2013b). STAR: Ultrafast universal RNA-seq aligner. *Bioinformatics*, *29*(1), 15–21. https://doi.org/10.1093/bioinformatics/bts635

Dolcet, X., Llobet, D., Pallares, J., & Matias-Guiu, X. (2005). NF-kB in development and progression of human cancer. *Virchows Archiv: An International Journal of Pathology*, *446*(5), 475–482. https://doi.org/10.1007/s00428-005-1264-9

Druker, J., Wilson, J. W., Child, F., Shakir, D., Fasanya, T., & Rocha, S. (2021). Role of Hypoxia in the Control of the Cell Cycle. *International Journal of Molecular Sciences*, *22*(9), 4874. https://doi.org/10.3390/ijms22094874

Duan, H., Li, Y., Yan, L., Yang, H., Wu, J., Qian, P., Li, B., & Wang, S. (2015). Rcan1-1L overexpression induces mitochondrial autophagy and improves cell survival in angiotensin II-exposed cardiomyocytes. *Experimental Cell Research*, *335*(1), 99–106. https://doi.org/10.1016/j.yexcr.2015.05.003

Eslami-S, Z., Cortés-Hernández, L. E., & Alix-Panabières, C. (2020). The Metastatic Cascade as the Basis for Liquid Biopsy Development. *Frontiers in Oncology*, *10*, 1055. https://doi.org/10.3389/fonc.2020.01055

Ewels, P., Magnusson, M., Lundin, S., & Käller, M. (2016a). MultiQC: summarize analysis results for multiple tools and samples in a single report. *Bioinformatics (Oxford, England)*, *32*(19), 3047–3048. https://doi.org/10.1093/bioinformatics/btw354

Ewels, P., Magnusson, M., Lundin, S., & Käller, M. (2016b). 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

Fan, L.-C., Jeng, Y.-M., Lu, Y.-T., & Lien, H.-C. (2016). SPOCK1 Is a Novel Transforming Growth Factor-β–Induced Myoepithelial Marker That Enhances Invasion and Correlates with Poor

Prognosis in Breast Cancer. *PLoS ONE*, *11*(9), e0162933. https://doi.org/10.1371/journal.pone.0162933

Favaro, E., Lord, S., Harris, A. L., & Buffa, F. M. (2011). Gene expression and hypoxia in breast cancer. *Genome Medicine*, *3*(8), 55. https://doi.org/10.1186/gm271

Forker, L., Gaunt, P., Sioletic, S., Shenjere, P., Potter, R., Roberts, D., Irlam, J., Valentine, H., Hughes, D., Hughes, A., Billingham, L., Grimer, R., Seddon, B., Choudhury, A., Robinson, M., & West, C. M. L. (2018). The hypoxia marker CAIX is prognostic in the UK phase III VorteX-Biobank cohort: An important resource for translational research in soft tissue sarcoma. *British Journal of Cancer*, *118*(5), 698–704. https://doi.org/10.1038/bjc.2017.430

Forte, D., Salvestrini, V., Corradi, G., Rossi, L., Catani, L., Lemoli, R. M., Cavo, M., & Curti, A. (2016). The tissue inhibitor of metalloproteinases-1 (TIMP-1) promotes survival and migration of acute myeloid leukemia cells through CD63/PI3K/Akt/p21 signaling. *Oncotarget*, *8*(2), 2261–2274. https://doi.org/10.18632/oncotarget.13664

Function and Mechanisms of Autophagy. (2019, August 17). News-Medical. https://www.news-medical.net/whitepaper/20190817/Function-and-Mechanisms-of-Autophagy.aspx

Gee, H. E., Camps, C., Buffa, F. M., Patiar, S., Winter, S. C., Betts, G., Homer, J., Corbridge, R., Cox, G., West, C. M. L., Ragoussis, J., & Harris, A. L. (2010). Hsa-mir-210 is a marker of tumor hypoxia and a prognostic factor in head and neck cancer. *Cancer*, *116*(9), 2148–2158. https://doi.org/10.1002/cncr.25009

Gillich, A., Zhang, F., Farmer, C. G., Travaglini, K. J., Tan, S. Y., Gu, M., Zhou, B., Feinstein, J. A., Krasnow, M. A., & Metzger, R. J. (2020). Capillary cell-type specialization in the alveolus. *Nature*, *586*(7831), 785–789. https://doi.org/10.1038/s41586-020-2822-7

Giordano, M., & Cavallaro, U. (2020). Different Shades of L1CAM in the Pathophysiology of Cancer Stem Cells. *Journal of Clinical Medicine*, *9*(5), 1502. https://doi.org/10.3390/jcm9051502

Gordan, J. D., Bertovrt, J. A., Hu, C.-J., Diehl, J. A., & Simon, M. C. (2007). HIF-2α promotes hypoxic cell proliferation by enhancing c-Myc transcriptional activity. *Cancer Cell*, *11*(4), 335–347. https://doi.org/10.1016/j.ccr.2007.02.006

Goveia, J., Rohlenova, K., Taverna, F., Treps, L., Conradi, L.-C., Pircher, A., Geldhof, V., de Rooij, L. P. M. H., Kalucka, J., Sokol, L., García-Caballero, M., Zheng, Y., Qian, J., Teuwen, L.-A., Khan, S., Boeckx, B., Wauters, E., Decaluwé, H., De Leyn, P., … Carmeliet, P. (2020). An Integrated Gene Expression Landscape Profiling Approach to Identify Lung Tumor Endothelial Cell Heterogeneity and Angiogenic Candidates. *Cancer Cell*, *37*(1), 21-36.e13. https://doi.org/10.1016/j.ccell.2019.12.001

Greaves, M., & Maley, C. C. (2012). CLONAL EVOLUTION IN CANCER. *Nature*, *481*(7381), 306–313. https://doi.org/10.1038/nature10762

Grivennikov, S., & Karin, M. (2008). Autocrine IL-6 Signaling: A Key Event in Tumorigenesis? *Cancer Cell*, *13*(1), 7–9. https://doi.org/10.1016/j.ccr.2007.12.020

Gu, J., Zhang, B., An, R., Qian, W., Duan, W., Wang, Z., & Ma, Q. (2022). Molecular Interactions of the Long Noncoding RNA NEAT1 in Cancer. *Cancers*, *14*(16), 4009. https://doi.org/10.3390/cancers14164009

Guaita, S., Puig, I., Franci, C., Garrido, M., Dominguez, D., Batlle, E., Sancho, E., Dedhar, S., De Herreros, A. G., & Baulida, J. (2002). Snail induction of epithelial to mesenchymal transition in tumor cells is accompanied by MUC1 repression and ZEB1 expression. *The Journal of Biological Chemistry*, *277*(42), 39209–39216. https://doi.org/10.1074/jbc.M206400200

Hao, Y., Hao, S., Andersen-Nissen, E., Mauck, W. M., Zheng, S., Butler, A., Lee, M. J., Wilk, A. J., Darby, C., Zager, M., Hoffman, P., Stoeckius, M., Papalexi, E., Mimitou, E. P., Jain, J., Srivastava, A., Stuart, T., Fleming, L. M., Yeung, B., … Satija, R. (2021a). Integrated analysis of multimodal single-cell data. *Cell*, *184*(13), 3573-3587.e29. https://doi.org/10.1016/j.cell.2021.04.048

Hao, Y., Hao, S., Andersen-Nissen, E., Mauck, W. M., Zheng, S., Butler, A., Lee, M. J., Wilk, A. J., Darby, C., Zager, M., Hoffman, P., Stoeckius, M., Papalexi, E., Mimitou, E. P., Jain, J.,

Srivastava, A., Stuart, T., Fleming, L. M., Yeung, B., … Satija, R. (2021b). Integrated analysis of multimodal single-cell data. *Cell*, *184*(13), 3573-3587.e29. https://doi.org/10.1016/j.cell.2021.04.048

Hapke, R., & Haake, S. M. (2020). Hypoxia-induced epithelial to mesenchymal transition in cancer. *Cancer Letters*, *487*, 10–20. https://doi.org/10.1016/j.canlet.2020.05.012

He, L., Wang, T., Gao, Q., Zhao, G., Huang, Y., Yu, L., & Hou, Y. (2011). Stanniocalcin-1 promotes tumor angiogenesis through up-regulation of VEGF in gastric cancer cells. *Journal of Biomedical Science*, *18*(1), 39. https://doi.org/10.1186/1423-0127-18-39

He, W., Wang, Q., Xu, J., Xu, X., Padilla, M. T., Ren, G., Gou, X., & Lin, Y. (2012). Attenuation of TNFSF10/TRAIL-induced apoptosis by an autophagic survival pathway involving TRAF2- and RIPK1/RIP1-mediated MAPK8/JNK activation. *Autophagy*, *8*(12), 1811–1821. https://doi.org/10.4161/auto.22145

Heddleston, J. M., Li, Z., Lathia, J. D., Bao, S., Hjelmeland, A. B., & Rich, J. N. (2010). Hypoxia inducible factors in cancer stem cells. *British Journal of Cancer*, *102*(5), 789–795. https://doi.org/10.1038/sj.bjc.6605551

Hu, Y., Gao, J., Wang, M., & Li, M. (2021). Potential Prospect of CDK4/6 Inhibitors in Triple-Negative Breast Cancer. *Cancer Management and Research*, *13*, 5223–5237. https://doi.org/10.2147/CMAR.S310649

Huang, F., Chen, L., Guo, W., Huang, T., & Cai, Y. (2022). Identification of Human Cell Cycle Phase Markers Based on Single-Cell RNA-Seq Data by Using Machine Learning Methods. *BioMed Research International*, *2022*, 2516653. https://doi.org/10.1155/2022/2516653

Hubbi, M. E., Kshitiz, Gilkes, D. M., Rey, S., Wong, C. C., Luo, W., Kim, D.-H., Dang, C. V., Levchenko, A., & Semenza, G. L. (2013). A Nontranscriptional Role for HIF-1α as a Direct Inhibitor of DNA Replication. *Science Signaling*, *6*(262), ra10. https://doi.org/10.1126/scisignal.2003417

Jiang, X., Zhou, Y., Sun, A.-J., & Xue, J.-L. (2018). NEAT1 contributes to breast cancer progression through modulating miR-448 and ZEB1. *Journal of Cellular Physiology*, *233*(11), 8558–8566. https://doi.org/10.1002/jcp.26470

Jimenez-García, M. P., Lucena-Cacace, A., Otero-Albiol, D., & Carnero, A. (2021). Empty spiracles homeobox genes EMX1 and EMX2 regulate WNT pathway activation in sarcomagenesis. *Journal of Experimental & Clinical Cancer Research: CR*, *40*(1), 247. https://doi.org/10.1186/s13046-021-02048-9

Jin, H., Wang, C., Jin, G., Ruan, H., Gu, D., Wei, L., Wang, H., Wang, N., Arunachalam, E., Zhang, Y., Deng, X., Yang, C., Xiong, Y., Feng, H., Yao, M., Fang, J., Gu, J., Cong, W., & Qin, W. (2017). Regulator of Calcineurin 1 Gene Isoform 4, Down-regulated in Hepatocellular Carcinoma, Prevents Proliferation, Migration, and Invasive Activity of Cancer Cells and Metastasis of Orthotopic Tumors by Inhibiting Nuclear Translocation of NFAT1. *Gastroenterology*, *153*(3), 799-811.e33. https://doi.org/10.1053/j.gastro.2017.05.045

Jin, S., Guerrero-Juarez, C. F., Zhang, L., Chang, I., Ramos, R., Kuan, C.-H., Myung, P., Plikus, M. V., & Nie, Q. (2021). Inference and analysis of cell-cell communication using CellChat. *Nature Communications*, *12*(1), 1088. https://doi.org/10.1038/s41467-021-21246-9

Johnson, D. E., O'Keefe, R. A., & Grandis, J. R. (2018). Targeting the IL-6/JAK/STAT3 signalling axis in cancer. *Nature Reviews Clinical Oncology*, *15*(4), 234–248. https://doi.org/10.1038/nrclinonc.2018.8

Johnson, D. G., & Schneider-Broussard, R. (1998). Role of E2F in cell cycle control and cancer. *Frontiers in Bioscience: A Journal and Virtual Library*, *3*, d447-448. https://doi.org/10.2741/a291

Johnstone, R. W., Frew, A. J., & Smyth, M. J. (2008). The TRAIL apoptotic pathway in cancer onset, progression and therapy. *Nature Reviews Cancer*, *8*(10), 782–798. https://doi.org/10.1038/nrc2465

Kakkad, S., Krishnamachary, B., Jacob, D., Pacheco-Torres, J., Goggins, E., Bharti, S. K., Penet, M.-F., & Bhujwalla, Z. M. (2019). Molecular and Functional Imaging Insights Into the Role of

Hypoxia in Cancer Aggression. *Cancer Metastasis Reviews*, *38*(1–2), 51–64. https://doi.org/10.1007/s10555-019-09788-3

Kaplan, A. R., & Glazer, P. M. (2020). Impact of hypoxia on DNA repair and genome integrity. *Mutagenesis*, *35*(1), 61–68. https://doi.org/10.1093/mutage/gez019

Kenny, H. A., Chiang, C.-Y., White, E. A., Schryver, E. M., Habis, M., Romero, I. L., Ladanyi, A., Penicka, C. V., George, J., Matlin, K., Montag, A., Wroblewski, K., Yamada, S. D., Mazar, A. P., Bowtell, D., & Lengyel, E. (2014). Mesothelial cells promote early ovarian cancer metastasis through fibronectin secretion. *The Journal of Clinical Investigation*, *124*(10), 4614–4628. https://doi.org/10.1172/JCI74778

Kim, H. I., Lee, H.-S., Kim, T. H., Lee, J.-S., Lee, S.-T., & Lee, S.-J. (2015). Growth-stimulatory activity of TIMP-2 is mediated through c-Src activation followed by activation of FAK, PI3-kinase/AKT, and ERK1/2 independent of MMP inhibition in lung adenocarcinoma cells. *Oncotarget*, *6*(40), 42905. https://doi.org/10.18632/oncotarget.5466

Kim, I., Choi, S., Yoo, S.-Y., Lee, M., & Kim, I. (2022). Cancer-Associated Fibroblasts in the Hypoxic Tumor Microenvironment. *Cancers*, *14*. https://doi.org/10.3390/cancers14143321

Kim, M.-C., Hwang, S.-H., Yang, Y., Kim, N.-Y., & Kim, Y. (2021). Reduction in mitochondrial oxidative stress mediates hypoxia-induced resistance to cisplatin in human transitional cell carcinoma cells. *Neoplasia (New York, N.Y.)*, *23*(7), 653–662. https://doi.org/10.1016/j.neo.2021.05.013

Kim, S.-H., Han, S.-Y., Azam, T., Yoon, D.-Y., & Dinarello, C. A. (2005). Interleukin-32: A Cytokine and Inducer of TNFα. *Immunity*, *22*(1), 131–142. https://doi.org/10.1016/j.immuni.2004.12.003

King, R., Hayes, C., Donohoe, C., Dunne, M., Davern, M., & Donlon, N. (2021). Hypoxia and its Impact on the Tumour Microenvironment of Gastroesophageal Cancers. *World Journal of Gastrointestinal Oncology*, *13*, 312–331. https://doi.org/10.4251/wjgo.v13.i5.312

Korotkevich, G., Sukhov, V., Budin, N., Shpak, B., Artyomov, M. N., & Sergushichev, A. (2021). *Fast gene set enrichment analysis* (p. 060012). bioRxiv. https://doi.org/10.1101/060012

Koshiji, M., Kageyama, Y., Pete, E. A., Horikawa, I., Barrett, J. C., & Huang, L. E. (2004). HIF-1α induces cell cycle arrest by functionally counteracting Myc. *The EMBO Journal*, *23*(9), 1949–1956. https://doi.org/10.1038/sj.emboj.7600196

Köster, J., & Rahmann, S. (2012). Snakemake—A scalable bioinformatics workflow engine. *Bioinformatics (Oxford, England)*, *28*(19), 2520–2522. https://doi.org/10.1093/bioinformatics/bts480

Krock, B. L., Skuli, N., & Simon, M. C. (2011). Hypoxia-Induced Angiogenesis. *Genes & Cancer*, *2*(12), 1117–1133. https://doi.org/10.1177/1947601911423654

Kronblad, A., Jirström, K., Rydén, L., Nordenskjöld, B., & Landberg, G. (2006). Hypoxia inducible factor-1alpha is a prognostic marker in premenopausal patients with intermediate to highly differentiated breast cancer but not a predictive marker for tamoxifen response. *International Journal of Cancer*, *118*(10), 2609–2616. https://doi.org/10.1002/ijc.21676

Krüger-Genge, A., Blocki, A., Franke, R.-P., & Jung, F. (2019). Vascular Endothelial Cell Biology: An Update. *International Journal of Molecular Sciences*, *20*(18), 4411. https://doi.org/10.3390/ijms20184411

Kugeratski, F. G., Atkinson, S. J., Neilson, L. J., Lilla, S., Knight, J. R. P., Serneels, J., Juin, A., Ismail, S., Bryant, D. M., Markert, E. K., Machesky, L. M., Mazzone, M., Sansom, O. J., & Zanivan, S. (2019). Hypoxic cancer-associated fibroblasts increase NCBP2-AS2/HIAR to promote endothelial sprouting through enhanced VEGF signaling. *Science Signaling*, *12*(567), eaan8247. https://doi.org/10.1126/scisignal.aan8247

Lal, A., Peters, H., St Croix, B., Haroon, Z. A., Dewhirst, M. W., Strausberg, R. L., Kaanders, J. H., van der Kogel, A. J., & Riggins, G. J. (2001). Transcriptional response to hypoxia in human tumors. *Journal of the National Cancer Institute*, *93*(17), 1337–1343. https://doi.org/10.1093/jnci/93.17.1337

Lao, M., Zhang, X., Ma, T., Xu, J., Yang, H., Duan, Y., Ying, H., Zhang, X., Guo, C., Qiu, J., Bai, X., & Liang, T. (2021). Regulator of calcineurin 1 gene isoform 4 in pancreatic ductal

adenocarcinoma regulates the progression of tumor cells. *Oncogene*, *40*(17), 3136–3151. https://doi.org/10.1038/s41388-021-01763-z

Lao, M., Zhang, X., Yang, H., Bai, X., & Liang, T. (2022). RCAN1-mediated calcineurin inhibition as a target for cancer therapy. *Molecular Medicine*, *28*(1), 69. https://doi.org/10.1186/s10020-022-00492-7

Law, A. Y. S., & Wong, C. K. C. (2013). Stanniocalcin-1 and -2 promote angiogenic sprouting in HUVECs via VEGF/VEGFR2 and angiopoietin signaling pathways. *Molecular and Cellular Endocrinology*, *374*(1–2), 73–81. https://doi.org/10.1016/j.mce.2013.04.024

Ledaki, I., McIntyre, A., Wigfield, S., Buffa, F., McGowan, S., Baban, D., Li, J., & Harris, A. L. (2015). Carbonic anhydrase IX induction defines a heterogeneous cancer cell response to hypoxia and mediates stem cell-like properties and sensitivity to HDAC inhibition. *Oncotarget*, *6*(23), 19413–19427.

Li, C.-J., Tzeng, Y.-D. T., Chiu, Y.-H., Lin, H.-Y., Hou, M.-F., & Chu, P.-Y. (2021). Pathogenesis and Potential Therapeutic Targets for Triple-Negative Breast Cancer. *Cancers*, *13*(12), Article 12. https://doi.org/10.3390/cancers13122978

Li, H.-S., Zhou, Y.-N., Li, L., Li, S.-F., Long, D., Chen, X.-L., Zhang, J.-B., Feng, L., & Li, Y.-P. (2019). HIF-1α protects against oxidative stress by directly targeting mitochondria. *Redox Biology*, *25*, 101109. https://doi.org/10.1016/j.redox.2019.101109

Li, J., Ke, J., Fang, J., & Chen, J.-P. (2020). A potential prognostic marker and therapeutic target: SPOCK1 promotes the proliferation, metastasis, and apoptosis of pancreatic ductal adenocarcinoma cells. *Journal of Cellular Biochemistry*, *121*(1), 743–754. https://doi.org/10.1002/jcb.29320

Li, P., Lv, X., Zhang, Z., & Xie, S. (2019). S100A6/miR193a regulates the proliferation, invasion, migration and angiogenesis of lung cancer cells through the P53 acetylation. *American Journal of Translational Research*, *11*(8), 4634–4649.

Li, S., Wei, Q., Li, Q., Zhang, B., & Xiao, Q. (2015). Down-regulating HIF-1α by lentivirus-mediated shRNA for therapy of triple negative breast cancer. *Cancer Biology & Therapy*, *16*(6), 866–875. https://doi.org/10.1080/15384047.2015.1040958

Li, W., Bazaz, S. R., Mayoh, C., & Salomon, R. (2024). Analytical Workflows for Single-Cell Multiomic Data Using the BD Rhapsody Platform. *Current Protocols*, *4*(2), e963. https://doi.org/10.1002/cpz1.963

Liao, Y., Smyth, G. K., & Shi, W. (2014). featureCounts: An efficient general purpose program for assigning sequence reads to genomic features. *Bioinformatics (Oxford, England)*, *30*(7), 923–930. https://doi.org/10.1093/bioinformatics/btt656

Liberti, M. V., & Locasale, J. W. (2016). The Warburg Effect: How Does it Benefit Cancer Cells? *Trends in Biochemical Sciences*, *41*(3), 211–218. https://doi.org/10.1016/j.tibs.2015.12.001

Liberzon, A., Birger, C., Thorvaldsdóttir, H., Ghandi, M., Mesirov, J. P., & Tamayo, P. (2015). The Molecular Signatures Database (MSigDB) hallmark gene set collection. *Cell Systems*, *1*(6), 417–425. https://doi.org/10.1016/j.cels.2015.12.004

Lidegaard, Ø., & Kroman, N. (2005). The epidemiology of breast cancer. *European Clinics in Obstetrics and Gynaecology*, *1*, 24–28. https://doi.org/10.1007/S11296-004-0007-9

Lien, E. C., Dibble, C. C., & Toker, A. (2017). PI3K signaling in cancer: Beyond AKT. *Current Opinion in Cell Biology*, *45*, 62–71. https://doi.org/10.1016/j.ceb.2017.02.007

Ling, M.-T., Wang, X., Ouyang, X.-S., Lee, T. K., Fan, T.-Y., Xu, K., Tsao, S.-W., & Wong, Y. C. (2002). Activation of MAPK signaling pathway is essential for Id-1 induced serum independent prostate cancer cell growth. *Oncogene*, *21*(55), 8498–8505. https://doi.org/10.1038/sj.onc.1206007

Litton, J. K., Rugo, H. S., Ettl, J., Hurvitz, S. A., Gonçalves, A., Lee, K.-H., Fehrenbacher, L., Yerushalmi, R., Mina, L. A., Martin, M., Roché, H., Im, Y.-H., Quek, R. G. W., Markova, D., Tudor, I. C., Hannah, A. L., Eiermann, W., & Blum, J. L. (2018). Talazoparib in Patients with Advanced Breast Cancer and a Germline BRCA Mutation. *The New England Journal of Medicine*, *379*(8), 753–763. https://doi.org/10.1056/NEJMoa1802905

Liu, X., Rowan, S. C., Liang, J., Yao, C., Huang, G., Deng, N., Xie, T., Wu, D., Wang, Y., Burman, A., Parimon, T., Borok, Z., Chen, P., Parks, W. C., Hogaboam, C. M., Weigt, S. S., Belperio, J., Stripp, B. R., Noble, P. W., & Jiang, D. (2021). Categorization of lung mesenchymal cells in development and fibrosis. *iScience*, *24*(6), 102551. https://doi.org/10.1016/j.isci.2021.102551

Liu, Y., Hu, Y., Xue, J., Li, J., Yi, J., Bu, J., Zhang, Z., Qiu, P., & Gu, X. (2023). Advances in immunotherapy for triple-negative breast cancer. *Molecular Cancer*, *22*(1), 145. https://doi.org/10.1186/s12943-023-01850-7

Loncaster, J. A., Harris, A. L., Davidson, S. E., Logue, J. P., Hunter, R. D., Wycoff, C. C., Pastorek, J., Ratcliffe, P. J., Stratford, I. J., & West, C. M. L. (2001). Carbonic Anhydrase (CA IX) Expression, a Potential New Intrinsic Marker of Hypoxia: Correlations with Tumor Oxygen Measurements and Prognosis in Locally Advanced Carcinoma of the Cervix1. *Cancer Research*, *61*(17), 6394–6399.

Lord, C. J., & Ashworth, A. (2017). PARP inhibitors: Synthetic lethality in the clinic. *Science (New York, N.Y.)*, *355*(6330), 1152–1158. https://doi.org/10.1126/science.aam7344

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

Luo, Z., Yi, Z.-J., Ou, Z.-L., Han, T., Wan, T., Tang, Y.-C., Wang, Z.-C., & Huang, F.-Z. (2019). RELA/NEAT1/miR-302a-3p/RELA feedback loop modulates pancreatic ductal adenocarcinoma cell proliferation and migration. *Journal of Cellular Physiology*, *234*(4), 3583–3597. https://doi.org/10.1002/jcp.27039

Lv, Y., Shan, Y., Song, L., Zhao, Y., Lai, R., Su, J., & Zhang, X. (2021). Type I collagen promotes tumor progression of integrin β1 positive gastric cancer through a BCL9L/β-catenin signaling pathway. *Aging (Albany NY)*, *13*, 19064–19076. https://doi.org/10.18632/aging.203355

Ma, B., Ueda, H., Okamoto, K., Bando, M., Fujimoto, S., Okada, Y., Kawaguchi, T., Wada, H., Miyamoto, H., Shimada, M., Sato, Y., & Takayama, T. (2022). TIMP1 promotes cell proliferation and invasion capability of right-sided colon cancers via the FAK/Akt signaling pathway. *Cancer Science*, *113*(12), 4244–4257. https://doi.org/10.1111/cas.15567

Ma, C. X., Reinert, T., Chmielewska, I., & Ellis, M. J. (2015). Mechanisms of aromatase inhibitor resistance. *Nature Reviews Cancer*, *15*(5), 261–275. https://doi.org/10.1038/nrc3920

Ma, Z., Xiang, X., Li, S., Xie, P., Gong, Q., Goh, B.-C., & Wang, L. (2022). Targeting hypoxia-inducible factor-1, for cancer treatment: Recent advances in developing small-molecule inhibitors from natural compounds. *Seminars in Cancer Biology*, *80*, 379–390. https://doi.org/10.1016/j.semcancer.2020.09.011

Mahmoudian, J., Ghods, R., Nazari, M., Jeddi-Tehrani, M., Ghahremani, M. H., Ghaffari-Tabrizi-Wizsy, N., Ostad, S. N., & Zarnani, A.-H. (2019). PLAC1: Biology and potential application in cancer immunotherapy. *Cancer Immunology, Immunotherapy: CII*, *68*(7), 1039–1058. https://doi.org/10.1007/s00262-019-02350-8

Mandapati, A., & Lukong, K. (2022). Triple negative breast cancer: Approved treatment options and their mechanisms of action. *Journal of Cancer Research and Clinical Oncology*, *149*. https://doi.org/10.1007/s00432-022-04189-6

Mandapati, A., & Lukong, K. E. (2023). Triple negative breast cancer: Approved treatment options and their mechanisms of action. *Journal of Cancer Research and Clinical Oncology*, *149*(7), 3701–3719. https://doi.org/10.1007/s00432-022-04189-6

Mao, Y., Keller, E. T., Garfield, D. H., Shen, K., & Wang, J. (2013). Stroma Cells in Tumor Microenvironment and Breast Cancer. *Cancer Metastasis Reviews*, *32*(0), 303–315. https://doi.org/10.1007/s10555-012-9415-3

Martin, M. (2011). Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet.Journal*, *17*(1), Article 1. https://doi.org/10.14806/ej.17.1.200

Masaki, T., & Shimada, M. (2022). Decoding the Phosphatase Code: Regulation of Cell Proliferation by Calcineurin. *International Journal of Molecular Sciences*, *23*(3), 1122. https://doi.org/10.3390/ijms23031122

Masoud, G. N., & Li, W. (2015). HIF-1α pathway: Role, regulation and intervention for cancer therapy. *Acta Pharmaceutica Sinica. B*, *5*(5), 378–389. https://doi.org/10.1016/j.apsb.2015.05.007

McInnes, L., Healy, J., & Melville, J. (2020). *UMAP: Uniform Manifold Approximation and Projection for Dimension Reduction* (arXiv:1802.03426). arXiv. http://arxiv.org/abs/1802.03426

McIntyre, A., Hulikova, A., Ledaki, I., Snell, C., Singleton, D., Steers, G., Seden, P., Jones, D., Bridges, E., Wigfield, S., Li, J.-L., Russell, A., Swietach, P., & Harris, A. L. (2016). Disrupting Hypoxia-Induced Bicarbonate Transport Acidifies Tumor Cells and Suppresses Tumor Growth. *Cancer Research*, *76*(13), 3744–3755. https://doi.org/10.1158/0008-5472.CAN-15-1862

McKeown, S. R. (2014). Defining normoxia, physoxia and hypoxia in tumours—Implications for treatment response. *The British Journal of Radiology*, *87*(1035), 20130676. https://doi.org/10.1259/bjr.20130676

Menck, K., Heinrichs, S., Baden, C., & Bleckmann, A. (2021). The WNT/ROR Pathway in Cancer: From Signaling to Therapeutic Intervention. *Cells*, *10*(1), 142. https://doi.org/10.3390/cells10010142

Min, J. Y., Lee, G. H., Khanal, T., Jin, S. W., Lee, S.-Y., Kim, H. G., Hyon, J.-Y., Chung, Y.-H., Ha, S. K., Han, E. H., & Jeong, H. G. (2022). Upregulation of CYP1B1 by hypoxia is mediated by ERα activation in breast cancer cells. *American Journal of Cancer Research*, *12*(6), 2798–2816.

Mölder, F., Jablonski, K. P., Letcher, B., Hall, M. B., Tomkins-Tinch, C. H., Sochat, V., Forster, J., Lee, S., Twardziok, S. O., Kanitz, A., Wilm, A., Holtgrewe, M., Rahmann, S., Nahnsen, S.,

& Köster, J. (2021). *Sustainable data analysis with Snakemake* (10:33). F1000Research. https://doi.org/10.12688/f1000research.29032.1

Müller-Dott, S., Tsirvouli, E., Vazquez, M., Ramirez Flores, R. O., Badia-i-Mompel, P., Fallegger, R., Türei, D., Lægreid, A., & Saez-Rodriguez, J. (2023). Expanding the coverage of regulons from high-confidence prior knowledge for accurate estimation of transcription factor activities. *Nucleic Acids Research*, *51*(20), 10934–10949. https://doi.org/10.1093/nar/gkad841

Mustacchi, G., Pronzato, P., Arpino, G., D'Alonzo, A., Piezzo, M., Natoli, C., Milani, A., Bologna, A., Blasi, L., Placido, S., Diodati, L., Beano, A., Romagnoli, E., & Cazzaniga, M. (2019). Triple-negative (TNBC) metastatic breast cancer (MBC) patients (pts): Is chemotherapy (CHT) choice influenced by adjuvant (adj) treatments? Results from the GIM-13 AMBRA study. *Journal of Clinical Oncology*. https://doi.org/10.1200/JCO.2019.37.15_SUPPL.E12549

Nagl, L., Horvath, L., Pircher, A., & Wolf, D. (2020). Tumor Endothelial Cells (TECs) as Potential Immune Directors of the Tumor Microenvironment – New Findings and Future Perspectives. *Frontiers in Cell and Developmental Biology*, *8*. https://doi.org/10.3389/fcell.2020.00766

Nam, K., Oh, S., & Shin, I. (2016). Ablation of CD44 induces glycolysis-to-oxidative phosphorylation transition via modulation of the c-Src-Akt-LKB1-AMPKα pathway. *The Biochemical Journal*, *473*(19), 3013–3030. https://doi.org/10.1042/BCJ20160613

Nami, B., Maadi, H., & Wang, Z. (2018). Mechanisms Underlying the Action and Synergism of Trastuzumab and Pertuzumab in Targeting HER2-Positive Breast Cancer. *Cancers*, *10*(10), Article 10. https://doi.org/10.3390/cancers10100342

Neumeister, V. M., Sullivan, C. A., Lindner, R., Lezon-Geyda, K., Li, J., Zavada, J., Martel, M., Glazer, P. M., Tuck, D. P., Rimm, D. L., & Harris, L. (2012). Hypoxia-induced protein CAIX is associated with somatic loss of BRCA1 protein and pathway activity in triple negative

breast cancer. *Breast Cancer Research and Treatment*, *136*(1), 67–75. https://doi.org/10.1007/s10549-012-2232-0

Nicolau-Neto, P., Palumbo, A., De Martino, M., Esposito, F., de Almeida Simão, T., Fusco, A., Nasciutti, L. E., Meireles Da Costa, N., & Ribeiro Pinto, L. F. (2018). UBE2C Is a Transcriptional Target of the Cell Cycle Regulator FOXM1. *Genes*, *9*(4), 188. https://doi.org/10.3390/genes9040188

Ombrato, L., Nolan, E., Kurelac, I., Mavousian, A., Bridgeman, V., Heinze, I., Chakravarty, P., Horswell, S., Gonzalez-Gualda, E., Matacchione, G., Weston, A., Kirkpatrick, J., Husain, E., Speirs, V., Collinson, L., Ori, A., Lee, J.-H., & Malanchi, I. (2019). Metastatic niche labelling reveals tissue parenchyma stem cell features. *Nature*, *572*(7771), 603–608. https://doi.org/10.1038/s41586-019-1487-6

Opavský, R., Pastoreková, S., Zelník, V., Gibadulinová, A., Stanbridge, E. J., Závada, J., Kettmann, R., & Pastorek, J. (1996). Human MN/CA9 gene, a novel member of the carbonic anhydrase family: Structure and exon to protein domain relationships. *Genomics*, *33*(3), 480–487. https://doi.org/10.1006/geno.1996.0223

Orimo, A., Gupta, P. B., Sgroi, D. C., Arenzana-Seisdedos, F., Delaunay, T., Naeem, R., Carey, V. J., Richardson, A. L., & Weinberg, R. A. (2005). Stromal fibroblasts present in invasive human breast carcinomas promote tumor growth and angiogenesis through elevated SDF-1/CXCL12 secretion. *Cell*, *121*(3), 335–348. https://doi.org/10.1016/j.cell.2005.02.034

Otsuka, H., Takito, J., Endo, Y., Yagi, H., Soeta, S., Yanagisawa, N., Nonaka, N., & Nakamura, M. (2016). The expression of embryonic globin mRNA in a severely anemic mouse model induced by treatment with nitrogen-containing bisphosphonate. *BMC Hematology*, *16*, 4. https://doi.org/10.1186/s12878-016-0041-0

Oun, R., Moussa, Y. E., & Wheate, N. J. (2018). The side effects of platinum-based chemotherapy drugs: A review for chemists. *Dalton Transactions (Cambridge, England: 2003)*, *47*(19), 6645–6653. https://doi.org/10.1039/c8dt00838h

Ouyang, J. F., Kamaraj, U. S., Cao, E. Y., & Rackham, O. J. L. (2021). ShinyCell: Simple and sharable visualization of single-cell gene expression data. *Bioinformatics*, *37*(19), 3374–3376. https://doi.org/10.1093/bioinformatics/btab209

Pan, H., Wang, X., Huang, W., Dai, Y., Yang, M., Liang, H., Wu, X., Zhang, L., Huang, W., Yuan, L., Wu, Y., Wang, Y., Liao, L., Huang, J., & Guan, J. (2020). Interferon-Induced Protein 44 Correlated With Immune Infiltration Serves as a Potential Prognostic Indicator in Head and Neck Squamous Cell Carcinoma. *Frontiers in Oncology*, *10*, 557157. https://doi.org/10.3389/fonc.2020.557157

Park, Y. R., Lee, S.-T., Kim, S. L., Liu, Y. C., Lee, M., Shin, J. H., Seo, S., Kim, S., Kim, I., Lee, S. O., & Kim, S. W. (2016). MicroRNA-9 suppresses cell migration and invasion through downregulation of TM4SF1 in colorectal cancer. *International Journal of Oncology*, *48 5*, 2135–2143. https://doi.org/10.3892/ijo.2016.3430

Patel, H. K., & Bihani, T. (2018). Selective estrogen receptor modulators (SERMs) and selective estrogen receptor degraders (SERDs) in cancer treatment. *Pharmacology & Therapeutics*, *186*, 1–24. https://doi.org/10.1016/j.pharmthera.2017.12.012

Pellecchia, S., Sepe, R., Federico, A., Cuomo, M., Credendino, S. C., Pisapia, P., Bellevicine, C., Nicolau-Neto, P., Severo Ramundo, M., Crescenzi, E., De Vita, G., Terracciano, L. M., Chiariotti, L., Fusco, A., & Pallante, P. (2019). The Metallophosphoesterase-Domain-Containing Protein 2 (MPPED2) Gene Acts as Tumor Suppressor in Breast Cancer. *Cancers*, *11*(6), 797. https://doi.org/10.3390/cancers11060797

Pettitt, S. J., Krastev, D. B., Brandsma, I., Dréan, A., Song, F., Aleksandrov, R., Harrell, M. I., Menon, M., Brough, R., Campbell, J., Frankum, J., Ranes, M., Pemberton, H. N., Rafiq, R., Fenwick, K., Swain, A., Guettler, S., Lee, J.-M., Swisher, E. M., … Lord, C. J. (2018). Genome-wide and high-density CRISPR-Cas9 screens identify point mutations in PARP1 causing PARP inhibitor resistance. *Nature Communications*, *9*(1), 1849. https://doi.org/10.1038/s41467-018-03917-2

Picelli, S., Björklund, Å. K., Faridani, O. R., Sagasser, S., Winberg, G., & Sandberg, R. (2013a). Smart-seq2 for sensitive full-length transcriptome profiling in single cells. *Nature Methods*, *10*(11), 1096–1098. https://doi.org/10.1038/nmeth.2639

Picelli, S., Björklund, Å. K., Faridani, O. R., Sagasser, S., Winberg, G., & Sandberg, R. (2013b). Smart-seq2 for sensitive full-length transcriptome profiling in single cells. *Nature Methods*, *10*(11), 1096–1098. https://doi.org/10.1038/nmeth.2639

Pidugu, V. K., Pidugu, H. B., Wu, M.-M., Liu, C.-J., & Lee, T.-C. (2019). Emerging Functions of Human IFIT Proteins in Cancer. *Frontiers in Molecular Biosciences*, *6*, 148. https://doi.org/10.3389/fmolb.2019.00148

Polyak, K., & Garber, J. (2011). Targeting the missing links for cancer therapy. *Nature Medicine*, *17*(3), 283–284. https://doi.org/10.1038/nm0311-283

Qi, S., Deng, S., Lian, Z., & Yu, K. (2022). Novel Drugs with High Efficacy against Tumor Angiogenesis. *International Journal of Molecular Sciences*, *23*(13), Article 13. https://doi.org/10.3390/ijms23136934

Qi, Y., Li, X., Chang, C., Xu, F., He, Q., Zhao, Y., & Wu, L. (2017). Ribosomal protein L23 negatively regulates cellular apoptosis via the RPL23/Miz-1/c-Myc circuit in higher-risk myelodysplastic syndrome. *Scientific Reports*, *7*(1), 2323. https://doi.org/10.1038/s41598-017-02403-x

Qian, J., Beullens, M., Huang, J., De Munter, S., Lesage, B., & Bollen, M. (2015). Cdk1 orders mitotic events through coordination of a chromosome-associated phosphatase switch. *Nature Communications*, *6*(1), 10215. https://doi.org/10.1038/ncomms10215

Ren, M., Gao, Y., Chen, Q., Zhao, H., Zhao, X., & Yue, W. (2020). The Overexpression of Keratin 23 Promotes Migration of Ovarian Cancer via Epithelial-Mesenchymal Transition. *BioMed Research International*, *2020*, 8218735. https://doi.org/10.1155/2020/8218735

Reunanen, N., & Kähäri, V. (2013). Matrix Metalloproteinases in Cancer Cell Invasion. In *Madame Curie Bioscience Database [Internet]*. Landes Bioscience. https://www.ncbi.nlm.nih.gov/books/NBK6598/

Robson, M., Im, S.-A., Senkus, E., Xu, B., Domchek, S. M., Masuda, N., Delaloge, S., Li, W., Tung, N., Armstrong, A., Wu, W., Goessl, C., Runswick, S., & Conte, P. (2017). Olaparib for Metastatic Breast Cancer in Patients with a Germline BRCA Mutation. *The New England Journal of Medicine*, *377*(6), 523–533. https://doi.org/10.1056/NEJMoa1706450

Roewe, J., Stavrides, G., Strueve, M., Sharma, A., Marini, F., Mann, A., Smith, S. A., Kaya, Z., Strobl, B., Mueller, M., Reinhardt, C., Morrissey, J. H., & Bosmann, M. (2020). Bacterial polyphosphates interfere with the innate host defense to infection. *Nature Communications*, *11*(1), 4035. https://doi.org/10.1038/s41467-020-17639-x

Rofstad, E. K. (2000). Microenvironment-induced cancer metastasis. *International Journal of Radiation Biology*, *76*(5), 589–605. https://doi.org/10.1080/095530000138259

Rose-John, S. (2020). Interleukin-6 signalling in health and disease. *F1000Research*, *9*, F1000 Faculty Rev-1013. https://doi.org/10.12688/f1000research.26058.1

Sachdeva, R., Wu, M., Smiljanic, S., Kaskun, O., Ghannad-Zadeh, K., Celebre, A., Isaev, K., Morrissy, A. S., Guan, J., Tong, J., Chan, J., Wilson, T. M., Al-Omaishi, S., Munoz, D. G., Dirks, P. B., Moran, M. F., Taylor, M. D., Reimand, J., & Das, S. (2019). ID1 Is Critical for Tumorigenesis and Regulates Chemoresistance in Glioblastoma. *Cancer Research*, *79*(16), 4057–4071. https://doi.org/10.1158/0008-5472.CAN-18-1357

Saelens, W., Cannoodt, R., Todorov, H., & Saeys, Y. (2019). A comparison of single-cell trajectory inference methods. *Nature Biotechnology*, *37*(5), 547–554. https://doi.org/10.1038/s41587-019-0071-9

Saito, R.-A., Micke, P., Paulsson, J., Augsten, M., Peña, C., Jönsson, P., Botling, J., Edlund, K., Johansson, L., Carlsson, P., Jirström, K., Miyazono, K., & Ostman, A. (2010). Forkhead box F1 regulates tumor-promoting properties of cancer-associated fibroblasts in lung cancer. *Cancer Research*, *70*(7), 2644–2654. https://doi.org/10.1158/0008-5472.CAN-09-3644

Sakai, H., Yasuda, S., Okuda, C., Yamada, T., Owaki, K., & Miwa, Y. (2022). Examination of central nervous system by functional observation battery after massive intravenous infusion of carbon monoxide-bound and oxygen-bound hemoglobin vesicles in rats. *Current*

Research in Pharmacology and Drug Discovery, *3*, 100135. https://doi.org/10.1016/j.crphar.2022.100135

Saul, D., Kosinsky, R. L., Atkinson, E. J., Doolittle, M. L., Zhang, X., LeBrasseur, N. K., Pignolo, R. J., Robbins, P. D., Niedernhofer, L. J., Ikeno, Y., Jurk, D., Passos, J. F., Hickson, L. J., Xue, A., Monroe, D. G., Tchkonia, T., Kirkland, J. L., Farr, J. N., & Khosla, S. (2022). A new gene set identifies senescent cells and predicts senescence-associated pathways across tissues. *Nature Communications*, *13*, 4827. https://doi.org/10.1038/s41467-022-32552-1

Saxena, K., & Jolly, M. K. (2019). Acute vs. Chronic vs. Cyclic Hypoxia: Their Differential Dynamics, Molecular Mechanisms, and Effects on Tumor Progression. *Biomolecules*, *9*(8), 339. https://doi.org/10.3390/biom9080339

Schmid, P., Rugo, H. S., Adams, S., Schneeweiss, A., Barrios, C. H., Iwata, H., Diéras, V., Henschel, V., Molinero, L., Chui, S. Y., Maiya, V., Husain, A., Winer, E. P., Loi, S., Emens, L. A., & IMpassion130 Investigators. (2020). Atezolizumab plus nab-paclitaxel as first-line treatment for unresectable, locally advanced or metastatic triple-negative breast cancer (IMpassion130): Updated efficacy results from a randomised, double-blind, placebo-controlled, phase 3 trial. *The Lancet. Oncology*, *21*(1), 44–59. https://doi.org/10.1016/S1470-2045(19)30689-8

Schödel, J., Bardella, C., Sciesielski, L. K., Brown, J. M., Pugh, C. W., Buckle, V., Tomlinson, I. P., Ratcliffe, P. J., & Mole, D. R. (2012). Common genetic variants at the 11q13.3 renal cancer susceptibility locus influence binding of HIF to an enhancer of cyclin D1 expression. *Nature Genetics*, *44*(4), 420-S2. https://doi.org/10.1038/ng.2204

Schrevel, M., Corver, W. E., Vegter, M. E., Ter Haar, N. T., Dreef, E. J., Beltman, J. J., Kenter, G., Bosse, T., de Kroon, C. D., & Jordanova, E. S. (2017). L1 cell adhesion molecule (L1CAM) is a strong predictor for locoregional recurrences in cervical cancer. *Oncotarget*, *8*(50), 87568–87581. https://doi.org/10.18632/oncotarget.20976

Senger, D., Claffey, K., Benes, J. E., Perruzzi, C., Sergiou, A. P., & Detmar, M. (1997). Angiogenesis promoted by vascular endothelial growth factor: Regulation through

alpha1beta1 and alpha2beta1 integrins. *Proceedings of the National Academy of Sciences of the United States of America*, *94 25*, 13612–13617. https://doi.org/10.1073/PNAS.94.25.13612

Shih, S.-C., Zukauskas, A., Li, D., Liu, G., Ang, L.-H., Nagy, J. A., Brown, L. F., & Dvorak, H. F. (2009). The L6 Protein TM4SF1 Is Critical for Endothelial Cell Function and Tumor Angiogenesis. *Cancer Research*, *69*(8), 3272–3277. https://doi.org/10.1158/0008-5472.CAN-08-4886

Singh, D. D., & Yadav, D. K. (2021). TNBC: Potential Targeting of Multiple Receptors for a Therapeutic Breakthrough, Nanomedicine, and Immunotherapy. *Biomedicines*, *9*(8), Article 8. https://doi.org/10.3390/biomedicines9080876

Sjöberg, E., Meyrath, M., Milde, L., Herrera, M., Lövrot, J., Hägerstrand, D., Frings, O., Bartish, M., Rolny, C., Sonnhammer, E., Chevigné, A., Augsten, M., & Östman, A. (2019). A Novel ACKR2-Dependent Role of Fibroblast-Derived CXCL14 in Epithelial-to-Mesenchymal Transition and Metastasis of Breast Cancer. *Clinical Cancer Research*, *25*(12), 3702–3717. https://doi.org/10.1158/1078-0432.CCR-18-1294

Škubník, J., Pavlíčková, V., Ruml, T., & Rimpelová, S. (2021). Current Perspectives on Taxanes: Focus on Their Bioactivity, Delivery and Combination Therapy. *Plants*, *10*(3), 569. https://doi.org/10.3390/plants10030569

Slawski, J., Jaśkiewicz, M., Barton, A., Kozioł, S., Collawn, J. F., & Bartoszewski, R. (2024). Regulation of the HIF switch in human endothelial and cancer cells. *European Journal of Cell Biology*, *103*(2), 151386. https://doi.org/10.1016/j.ejcb.2024.151386

Span, P. N., Bussink, J., Manders, P., Beex, L. V. A. M., & Sweep, C. G. J. (2003). Carbonic anhydrase-9 expression levels and prognosis in human breast cancer: Association with treatment outcome. *British Journal of Cancer*, *89*(2), 271–276. https://doi.org/10.1038/sj.bjc.6601122

Srivastava, N., Usmani, S. S., Subbarayan, R., Saini, R., & Pandey, P. K. (2023). Hypoxia: Syndicating triple negative breast cancer against various therapeutic regimens. *Frontiers in Oncology*, *13*, 1199105. https://doi.org/10.3389/fonc.2023.1199105

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 of the United States of America*, *102*(43), 15545–15550. https://doi.org/10.1073/pnas.0506580102

Suh, K. S., Crutchley, J. M., Koochek, A., Ryscavage, A., Bhat, K., Tanaka, T., Oshima, A., Fitzgerald, P., & Yuspa, S. H. (2007). Reciprocal Modifications of CLIC4 in Tumor Epithelium and Stroma Mark Malignant Progression of Multiple Human Cancers. *Clinical Cancer Research*, *13*(1), 121–131. https://doi.org/10.1158/1078-0432.CCR-06-1562

Suh, K. S., Mutoh, M., Gerdes, M., & Yuspa, S. H. (2005). CLIC4, an Intracellular Chloride Channel Protein, Is a Novel Molecular Target for Cancer Therapy. *Journal of Investigative Dermatology Symposium Proceedings*, *10*(2), 105–109. https://doi.org/10.1111/j.1087-0024.2005.200402.x

Sun, S., Xu, Y., Fu, P., Chen, M., Sun, S., Zhao, R., Wang, J., Liang, X., & Wang, S. (2018). Ultrasound-targeted photodynamic and gene dual therapy for effectively inhibiting triple negative breast cancer by cationic porphyrin lipid microbubbles loaded with HIF1α-siRNA. *Nanoscale*, *10 42*, 19945–19956. https://doi.org/10.1039/c8nr03074j

Sun, X., & Kaufman, P. D. (2018). Ki-67: More than a proliferation marker. *Chromosoma*, *127*(2), 175–186. https://doi.org/10.1007/s00412-018-0659-8

Sweeney, S. M., Dilullo, G., Slater, S., Martinez, J., Iozzo, R., Lauer-Fields, J. L., Fields, G., & Antonio, J. S. S. (2003). Angiogenesis in collagen I requires alpha2beta1 ligation of a GFP*GER sequence and possibly p38 MAPK activation and focal adhesion disassembly. *The Journal of Biological Chemistry*. https://www.semanticscholar.org/paper/Angiogenesis-

in-collagen-I-requires-alpha2beta1-of-Sweeney-Dilullo/b28822bee28e6ea1d334d315e56e84a6c01694b9

Swietach, P., Hulikova, A., Vaughan-Jones, R. D., & Harris, A. L. (2010). New insights into the physiological role of carbonic anhydrase IX in tumour pH regulation. *Oncogene*, *29*(50), 6509–6521. https://doi.org/10.1038/onc.2010.455

Swietach, P., Wigfield, S., Cobden, P., Supuran, C. T., Harris, A. L., & Vaughan-Jones, R. D. (2008). Tumor-associated carbonic anhydrase 9 spatially coordinates intracellular pH in three-dimensional multicellular growths. *The Journal of Biological Chemistry*, *283*(29), 20473–20483. https://doi.org/10.1074/jbc.M801330200

Telarovic, I., Wenger, R. H., & Pruschy, M. (2021). Interfering with Tumor Hypoxia for Radiotherapy Optimization. *Journal of Experimental & Clinical Cancer Research*, *40*(1), 197. https://doi.org/10.1186/s13046-021-02000-x

Teuwen, L.-A., De Rooij, L. P. M. H., Cuypers, A., Rohlenova, K., Dumas, S. J., García-Caballero, M., Meta, E., Amersfoort, J., Taverna, F., Becker, L. M., Veiga, N., Cantelmo, A. R., Geldhof, V., Conchinha, N. V., Kalucka, J., Treps, L., Conradi, L.-C., Khan, S., Karakach, T. K., … Carmeliet, P. (2021). Tumor vessel co-option probed by single-cell analysis. *Cell Reports*, *35*(11), 109253. https://doi.org/10.1016/j.celrep.2021.109253

Thorburn, A. (2007). Tumor Necrosis Factor-Related Apoptosis-Inducing Ligand (TRAIL) Pathway Signaling. *Journal of Thoracic Oncology*, *2*(6), 461–465. https://doi.org/10.1097/JTO.0b013e31805fea64

Tirosh, I., Izar, B., Prakadan, S. M., Wadsworth, M. H., Treacy, D., Trombetta, J. J., Rotem, A., Rodman, C., Lian, C., Murphy, G., Fallahi-Sichani, M., Dutton-Regester, K., Lin, J.-R., Cohen, O., Shah, P., Lu, D., Genshaft, A. S., Hughes, T. K., Ziegler, C. G. K., … Garraway, L. A. (2016). Dissecting the multicellular ecosystem of metastatic melanoma by single-cell RNA-seq. *Science (New York, N.Y.)*, *352*(6282), 189–196. https://doi.org/10.1126/science.aad0501

Traag, V. A., Waltman, L., & van Eck, N. J. (2019). From Louvain to Leiden: Guaranteeing well-connected communities. *Scientific Reports*, *9*(1), 5233. https://doi.org/10.1038/s41598-019-41695-z

Tran, H. T. N., Ang, K. S., Chevrier, M., Zhang, X., Lee, N. Y. S., Goh, M., & Chen, J. (2020). A benchmark of batch-effect correction methods for single-cell RNA sequencing data. *Genome Biology*, *21*, 12. https://doi.org/10.1186/s13059-019-1850-9

Trapasso, S., & Allegra, E. (2012). Role of CD44 as a marker of cancer stem cells in head and neck cancer. *Biologics : Targets & Therapy*, *6*, 379–383. https://doi.org/10.2147/BTT.S37906

Ulmasov, B., Bruno, J., Gordon, N., Hartnett, M. E., & Edwards, J. C. (2009). Chloride Intracellular Channel Protein-4 Functions in Angiogenesis by Supporting Acidification of Vacuoles Along the Intracellular Tubulogenic Pathway. *The American Journal of Pathology*, *174*(3), 1084–1096. https://doi.org/10.2353/ajpath.2009.080625

Uribe, M. L., Marrocco, I., & Yarden, Y. (2021). EGFR in Cancer: Signaling Mechanisms, Drugs, and Acquired Resistance. *Cancers*, *13*(11), Article 11. https://doi.org/10.3390/cancers13112748

Van den Eynden, G. G., Colpaert, C. G., Couvelard, A., Pezzella, F., Dirix, L. Y., Vermeulen, P. B., Van Marck, E. A., & Hasebe, T. (2007). A fibrotic focus is a prognostic factor and a surrogate marker for hypoxia and (lymph)angiogenesis in breast cancer: Review of the literature and proposal on the criteria of evaluation. *Histopathology*, *51*(4), 440–451. https://doi.org/10.1111/j.1365-2559.2007.02761.x

van Kuijk, S. J. A., Yaromina, A., Houben, R., Niemans, R., Lambin, P., & Dubois, L. J. (2016). Prognostic Significance of Carbonic Anhydrase IX Expression in Cancer Patients: A Meta-Analysis. *Frontiers in Oncology*, *6*, 69. https://doi.org/10.3389/fonc.2016.00069

Venkatesh, P., & Kasi, A. (2024). Anthracyclines. In *StatPearls*. StatPearls Publishing. http://www.ncbi.nlm.nih.gov/books/NBK538187/

Vila Ellis, L., Cain, M. P., Hutchison, V., Flodby, P., Crandall, E. D., Borok, Z., Zhou, B., Ostrin, E. J., Wythe, J. D., & Chen, J. (2020). Epithelial Vegfa Specifies a Distinct Endothelial Population in the Mouse Lung. *Developmental Cell*, *52*(5), 617-630.e6. https://doi.org/10.1016/j.devcel.2020.01.009

Volkova, M., & Russell, R. (2011). Anthracycline Cardiotoxicity: Prevalence, Pathogenesis and Treatment. *Current Cardiology Reviews*, *7*(4), 214–220. https://doi.org/10.2174/157340311799960645

Wang, C.-Y., Guttridge, D. C., Mayo, M. W., & Baldwin, A. S. (1999). NF-κB Induces Expression of the Bcl-2 Homologue A1/Bfl-1 To Preferentially Suppress Chemotherapy-Induced Apoptosis. *Molecular and Cellular Biology*, *19*(9), 5923–5929.

Wang, J., Ding, N., Li, Y., Cheng, H., Wang, D., Yang, Q., Deng, Y., Yang, Y., Li, Y., Ruan, X., Xie, F., Zhao, H., & Fang, X. (2015). Insulin-like growth factor binding protein 5 (IGFBP5) functions as a tumor suppressor in human melanoma cells. *Oncotarget*, *6*(24), 20636–20649.

Wang, L., Man, N., Sun, X.-J., Tan, Y., Cao, M. G., Liu, F., Hatlen, M., Xu, H., Huang, G., Mattlin, M., Mehta, A., Rampersaud, E., Benezra, R., & Nimer, S. D. (2015). Regulation of AKT signaling by Id1 controls t(8;21) leukemia initiation and progression. *Blood*, *126*(5), 640–650. https://doi.org/10.1182/blood-2015-03-635532

Welford, S. M., & Giaccia, A. J. (2011). Hypoxia and Senescence: The impact of oxygenation on tumor suppression. *Molecular Cancer Research : MCR*, *9*(5), 538–544. https://doi.org/10.1158/1541-7786.MCR-11-0065

Wen, S., Hou, Y., Fu, L., Xi, L., Yang, D., Zhao, M., Qin, Y., Sun, K., Teng, Y., & Liu, M. (2019). Cancer-associated fibroblast (CAF)-derived IL32 promotes breast cancer cell invasion and metastasis via integrin β3-p38 MAPK signalling. *Cancer Letters*, *442*, 320–332. https://doi.org/10.1016/j.canlet.2018.10.015

Wolf, Y., Shemer, A., Levy-Efrati, L., Gross, M., Kim, J.-S., Engel, A., David, E., Chappell-Maor, L., Grozovski, J., Rotkopf, R., Biton, I., Eilam-Altstadter, R., & Jung, S. (2018). Microglial MHC

class II is dispensable for experimental autoimmune encephalomyelitis and cuprizone-induced demyelination. *European Journal of Immunology*, *48*(8), 1308–1318. https://doi.org/10.1002/eji.201847540

Xie, T., Wang, Y., Deng, N., Huang, G., Taghavifar, F., Geng, Y., Liu, N., Kulur, V., Yao, C., Chen, P., Liu, Z., Stripp, B., Tang, J., Liang, J., Noble, P. W., & Jiang, D. (2018). Single-Cell Deconvolution of Fibroblast Heterogeneity in Mouse Pulmonary Fibrosis. *Cell Reports*, *22*(13), 3625–3640. https://doi.org/10.1016/j.celrep.2018.03.010

Yamamoto, S., & Iwakuma, T. (2017). RIPK1-TRAF2 interplay on the TNF/NF-κB signaling, cell death, and cancer development in the liver. *Translational Cancer Research*, *6*(Suppl 3), 94–109. https://doi.org/10.21037/tcr.2017.04.01

Yang, J., AlTahan, A., Jones, D. T., Buffa, F. M., Bridges, E., Interiano, R. B., Qu, C., Vogt, N., Li, J.-L., Baban, D., Ragoussis, J., Nicholson, R., Davidoff, A. M., & Harris, A. L. (2015). Estrogen receptor-α directly regulates the hypoxia-inducible factor 1 pathway associated with antiestrogen response in breast cancer. *Proceedings of the National Academy of Sciences of the United States of America*, *112*(49), 15172–15177. https://doi.org/10.1073/pnas.1422015112

Yang, M., Zhou, Y., Deng, H., Zhou, H., Cheng, S., Zhang, D., He, X., Mai, L., Chen, Y., & Chen, J. (2021). Ribosomal Protein L23 Drives the Metastasis of Hepatocellular Carcinoma via Upregulating MMP9. *Frontiers in Oncology*, *11*, 779748. https://doi.org/10.3389/fonc.2021.779748

Yao, X., & Zeng, Y. (2023). Tumour associated endothelial cells: Origin, characteristics and role in metastasis and anti-angiogenic resistance. *Frontiers in Physiology*, *14*, 1199225. https://doi.org/10.3389/fphys.2023.1199225

Yao, Y., & Dai, W. (2014). Genomic Instability and Cancer. *Journal of Carcinogenesis & Mutagenesis*, *5*, 1000165. https://doi.org/10.4172/2157-2518.1000165

Yi, H., Li, Y., Tan, Y., Fu, S., Tang, F., & Deng, X. (2021). Immune Checkpoint Inhibition for Triple-Negative Breast Cancer: Current Landscape and Future Perspectives. *Frontiers in Oncology*, *11*, 648139. https://doi.org/10.3389/fonc.2021.648139

Yu, G., Wang, L.-G., Han, Y., & He, Q.-Y. (2012). clusterProfiler: An R Package for Comparing Biological Themes Among Gene Clusters. *OMICS : A Journal of Integrative Biology*, *16*(5), 284–287. https://doi.org/10.1089/omi.2011.0118

Yu, X., Abbas-Aghababazadeh, F., Chen, Y. A., & Fridley, B. L. (2021). Statistical and Bioinformatics Analysis of Data from Bulk and Single-Cell RNA Sequencing Experiments. *Methods in Molecular Biology (Clifton, N.J.)*, *2194*, 143–175. https://doi.org/10.1007/978-1-0716-0849-4_9

Zappia, L., & Oshlack, A. (2018). Clustering trees: A visualization for evaluating clusterings at multiple resolutions. *GigaScience*, *7*(7), giy083. https://doi.org/10.1093/gigascience/giy083

Zeng, Y., Chen, H., Zhang, Z., Fan, J., Li, J., Zhou, S., Wang, N., Yan, S., Cao, J., Liu, J., Zhou, Z., & Liu, W. (2023). IFI44L as a novel epigenetic silencing tumor suppressor promotes apoptosis through JAK/STAT1 pathway during lung carcinogenesis. *Environmental Pollution*, *319*, 120943. https://doi.org/10.1016/j.envpol.2022.120943

Zhang, N., Zhang, R., Zou, K., Yu, W., Guo, W., Gao, Y., Li, J., Li, M., Tai, Y., Huang, W., Song, C., Deng, W., & Cui, X. (2017). Keratin 23 promotes telomerase reverse transcriptase expression and human colorectal cancer growth. *Cell Death & Disease*, *8*(7), e2961–e2961. https://doi.org/10.1038/cddis.2017.339

Zhao, H., Wu, L., Yan, G., Chen, Y., Zhou, M., Wu, Y., & Li, Y. (2021). Inflammation and tumor progression: Signaling pathways and targeted intervention. *Signal Transduction and Targeted Therapy*, *6*(1), 1–46. https://doi.org/10.1038/s41392-021-00658-5

Zhao, L., Li, Z., Vong, J. S. L., Chen, X., Lai, H.-M., Yan, L. Y. C., Huang, J., Sy, S. K. H., Tian, X., Huang, Y., Chan, H. Y. E., So, H.-C., Ng, W.-L., Tang, Y., Lin, W.-J., Mok, V. C. T., & Ko, H. (2020). Pharmacologically reversible zonation-dependent endothelial cell transcriptomic

changes with neurodegenerative disease associations in the aged brain. *Nature Communications*, *11*(1), 4413. https://doi.org/10.1038/s41467-020-18249-3

Zheng, G. X. Y., Terry, J. M., Belgrader, P., Ryvkin, P., Bent, Z. W., Wilson, R., Ziraldo, S. B., Wheeler, T. D., McDermott, G. P., Zhu, J., Gregory, M. T., Shuga, J., Montesclaros, L., Underwood, J. G., Masquelier, D. A., Nishimura, S. Y., Schnall-Levin, M., Wyatt, P. W., Hindson, C. M., … Bielas, J. H. (2017). Massively parallel digital transcriptional profiling of single cells. *Nature Communications*, *8*(1), 14049. https://doi.org/10.1038/ncomms14049

Zhou, Q., Zhou, S., Wang, H., Li, Y., Xiao, X., & Yang, J. (2020). Stable silencing of ROR1 regulates cell cycle, apoptosis, and autophagy in a lung adenocarcinoma cell line. *International Journal of Clinical and Experimental Pathology*, *13*(5), 1108–1120.

Zöller, M. (2011). CD44: Can a cancer-initiating cell profit from an abundantly expressed molecule? *Nature Reviews. Cancer*, *11*(4), 254–267. https://doi.org/10.1038/nrc3023