PhD Program in Biomolecular Sciences

Department of Cellular, Computational and Integrative Biology – CIBIO

36th Cycle

Deciphering Differentiation Trajectories in Neuroblastoma and Glioblastoma: Insights from Single-Cell and Spatial Transcriptomics

Prof. ALESSANDRO QUATTRONE

Department CIBIO, University of Trento

Ph.D. Thesis of

ENRICO SEBASTIANI

Department CIBIO, University of Trento

Academic Year 2023-2024

Declaration of original authorship

Declaration:
I, Enrico Sebastiani confirm that this is my own work and the use of all material from other sources has been properly and fully acknowledged.

Trento, 30/11/2024

Abstract

This thesis explores two highly aggressive tumors: neuroblastoma and glioblastoma. The first project focuses on neuroblastoma, a common extracranial pediatric tumor that typically arises in the adrenal gland during the development of the peripheral sympathetic nervous system. Through a meta-analysis that integrates different single-cell RNA sequencing datasets, I compared tumor cells with normal adrenal medulla cells to identify the differentiation stage that promotes the tumorigenic process. I found that neuroblastoma cells tend to recapitulate the regular differentiation program of the adrenal medulla and most of them exhibit features that resemble the characteristics of sympathoblasts and cycling sympathoblasts, but they do not complete their differentiation in sympathetic neurons. I then investigated which genes, chromosomal aberrations, and pathways block this terminal differentiation.

In the second project, I focused on glioblastoma, a highly aggressive primary brain tumor that affects adults. Through spatial transcriptomics, I analysed 12 tumor samples taken from three patients, each from four distinct areas, to verify if glioblastoma shows different characteristics depending on the sampling region. I observed that, unlike other regions, the outermost margin of the tumor has tumor cells resembling the OPC-like subtype. In contrast, the more aggressive MES-like and AC-like phenotypes are mostly represented in the central area and inner margins of the tumor. A more in-depth analysis of the interaction between tumor cells and tumor microenvironment cells has revealed that OPC-like tumor cells release immunosuppressive molecules, promoting growth and immune evasion. These results highlight the spatial complexity of glioblastoma and suggest possible therapeutic targets for this heterogeneous disease.

Table of Contents

Table of Figures

List of Abbreviations

5

5-ALA: AminoLevulinic Acid hydrochloride; 46

A

AIFA: Agenzia Italiana del Farmaco; 46

C

CCA: Canonical Correlation Analysis; 58
CNS: Central Nervous System; 3
CNVs: Copy Number Variations; 34
COG: Children Oncology Group; 10
CPCs: Connecting progenitor cells; 8

E

EGFR: Epidermal Growth Factor Receptor; 45
EMA: European Medicines Agency; 46

F

FLAIR: Fluid Attenuated Inversion Recovery; 48

G

GBM: Glioblastoma; 1
GEMs: Gel bead in EMulsions; 16
GO: Gene Ontology; 76
GRNs: Gene Regulatory Networks; 23
GSEA: Gene Set Enrichment Analysis; 76

H

H&E: Hematoxylin and Eosin; 64

I

INSS: International neuroblastoma staging system; 10

K

KNN: K-Nearest Neighbors; 58

M

MGMT: MethylGuanine-DNA-MethylTransferase; 45
MRI: Magnetic Resonance Imaging; 46; 48

N

NB: Neuroblastoma; 9
NCCs: Neural Crest Cells; 4
NMF: Non-Negative Matrix Factorization; 61

P

PCA: Principal Component Analysis; 21
PPIX: ProtoPorphyrin IX; 46

R

RPCA: Reciprocal PCA; 58

S

SAPs: Sympathoadrenal Progenitors; 6
SCPs: Schwann Cell Precursors; 6
scRNA-seq: Single Cell RNA Sequencing; 8; 15; 16; 17; 25; 48; 54; 68; 70; 71; 76; 78; 80
snRNA-seq: Single Nucleus RNA sequencing; 16
ST: Spatial Transcriptomics; 50; 51; 52; 54; 64; 65; 68; 76; 78; 80; 81; 82; 84; 88; 92

T

TAMs: Tumor Associated Macrophages; 52
TERT: TElomerase Reverse Transcriptase; 45
TMZ: TeMoZolomide; 45

W

WHO: World Health Organization; 44
wpc: Weeks post conception; 8
WPC: Weeks Post Conception; 8

X

Objectives of the Thesis

Cancer research has made remarkable progress in recent years thanks to the introduction of innovative technologies that allow the study and characterization of cancer cells with unprecedented resolution and detail. In particular, the advent of single-cell RNA sequencing (scRNA-seq) and spatial transcriptomics (ST) have significantly changed how we study cell transcriptomes. Indeed, using scRNA-seq, we can now investigate the transcriptomes at single cell level, making it possible to analyse tumor heterogeneity and identify the distinct cellular subpopulations in the tissue. On the other hand, ST adds a complementary dimension to scRNA-seq by preserving the spatial context of gene expression within the tissue architecture. In fact, unlike scRNA-seq, which involves tissue dissociation, ST maintains the positional information of RNA molecules. This allows us to study the interactions between the neighbouring cells within a given tissue and the localization of the different cell types.

In this doctoral thesis, I used these technologies to investigate the biology of two highly aggressive solid tumours: neuroblastoma (NB), a paediatric malignancy, and glioblastoma (GBM), an aggressive adult brain cancer. In the first project I studied the NB, the most common extracranial solid tumor in early childhood, accounting for approximately 7% of all paediatric cancers. This tumor typically arises in the adrenal gland during the early stages of the peripheral sympathetic nervous system development. However, despite recent advances, it remains unclear which cell type involved in the normal differentiation of the adrenal gland is responsible for the onset and progression of this tumor, as well as the mechanisms driving its development. The primary aim of this study was to characterize the differentiation trajectories of tumor cells and determine whether they can recapitulate the normal developmental processes of the adrenal medulla. I then tried to identify the differentiation stage at which tumor cells block their differentiation and promote tumor development. Moreover, I studied the genes, pathways, and transcription factors that influence the differentiation states of tumor cells, exploring their potential correlations with specific chromosomal aberrations.

In the second project I studied GBM, the most aggressive primary brain tumor in adults, characterized by high heterogeneity and resistance to standard therapies. Despite the recent advancements in the research, the biological complexity of GBM remains a significant challenge in oncology. In this work, I employed ST to investigate this disease by analyzing 12 samples from three patients collected from four distinct tumor regions: the core (the central tumor area), the inner edge (the area adjacent to the core), the outer

edge (a transitional zone found at the periphery of the tumor mass), and the distal area (healthy tissue distant from the cancer, serving as a reference). Using this approach, I built up a transcriptional atlas of GBM to study the spatial organization of this tumor and how cell identities and cell functions change across the different tumor regions. I then investigated the interactions between tumor cells and their surrounding microenvironment to examine potential crosstalk between different cell types and elucidate how these interactions influence tumor behavior and progression. These findings provide new insights into glioblastoma biology, highlighting the role of tumor spatial organization and interactions with the surrounding microenvironment.

Chapter 1

Studying differentiation trajectories in neuroblastoma

1.1 Introduction

1.1.1 The sympathetic nervous system

The autonomic nervous system, found in vertebrates, regulates the functionality of the internal organs without requiring conscious control. It can be divided into three different parts: the sympathetic nervous system, the parasympathetic nervous system, and the enteric nervous system. The sympathetic nervous system is crucial in the so-called fight-or-flight response and for maintaining body's homeostasis (Waxenbaum *et al*, 2023). It is formed by two chains of paravertebral ganglia localized ventrally and laterally to the spinal cord between the cervical and pelvic regions. These ganglia are formed by two types of neurons involved in signal transmission: preganglionic and postganglionic neurons. Preganglionic neurons have their soma localized in the spinal cord and they are cholinergic. They are responsible for sending the signals from the central nervous system (CNS) to the paravertebral ganglia when they make synapses with post ganglionic nerve fibers that send the signals to the final organs. Post ganglionic fibers, instead, use mainly the catecholamines norepinephrine and epinephrine as neurotransmitters with some exceptions for the kidney and sweat glands where they use dopamine and acetylcholine respectively (Cuevas *et al*, 2013; Hu *et al*, 2018). Neurons that release norepinephrine, epinephrine and/or dopamine are defined as noradrenergic.

1.1.2 Biology of the adrenal gland

The adrenal glands or suprarenal glands are small endocrine glands located on top of the kidney. These organs produce important hormones that are necessary for metabolism regulation, blood pressure and response to stress. It is possible to distinguish between two main parts: the adrenal cortex and the adrenal medulla. The adrenal cortex forms the largest portion of the gland, and it is located in the outer region. It can be divided into three distinct zones: the zona glomerulosa, zona fasciculata, and zona reticularis. Each one of those is responsible for producing specific hormones. The adrenal medulla instead is located at the center of the ground, within the cortex. Preganglionic nerves from the CNS connect to the cells in the adrenal medulla, stimulating the release of catecholamines. For this reason, the adrenal gland is commonly considered a modified ganglion of the sympathetic nervous system. The medullary region mainly contains secretory neuroendocrine cells named chromaffin cells. The morphology of chromaffin cells is like the one of other endocrine cells because they lack neurites and have large storing vesicles called chromaffin granules. These granules mainly contain hormones including catecholamines epinephrine and norepinephrine that are directly released in the blood circulatory system upon stimulation (Bornstein *et al*, 2012). Among their activities, the hormones mediate the increase of the hearth rate and the blood flow to the muscles and brain as well as the secretion of glucagon that mobilize energy stores in the form of glucose (**Figure 1.1 A-B**).

1.1.3 Biology of neural crest cells

Neural crest cells (NCCs) are a transient cell population uniquely found in vertebrates that originate during the early phases of embryonic development. NCCs can be considered a type of multipotent progenitor cell capable of differentiating in many different types of cells. They are formed during neurulation, a phase of the embryonic development characterized by the fusion of the neural ectoderm and the formation of the neural tube.

These cells detach from the most external region of the neural tube commonly referred to as neural plate border (Theveneau & Mayor, 2014a). The induction of the neural crest fate is mediated by signaling molecules such as bone morphogenetic proteins (Bmps), Wnt/b-catenin and fibroblast growth factors (Fgfs) (Labonne & Bronner-Fraser). NCCs detach from the neural tube shortly after its closure in a process called delamination mediated by the epithelial-to-mesenchymal transition (EMT). During EMT, cells lose their polarity and gain reduced adhesions allowing neural crest cells to delaminate (Theveneau & Mayor,

2014a). This process is activated by a group of regulators that belong to BMP, WNT, Notch and FGF signaling pathways (Ponzoni *et al*, 2022).

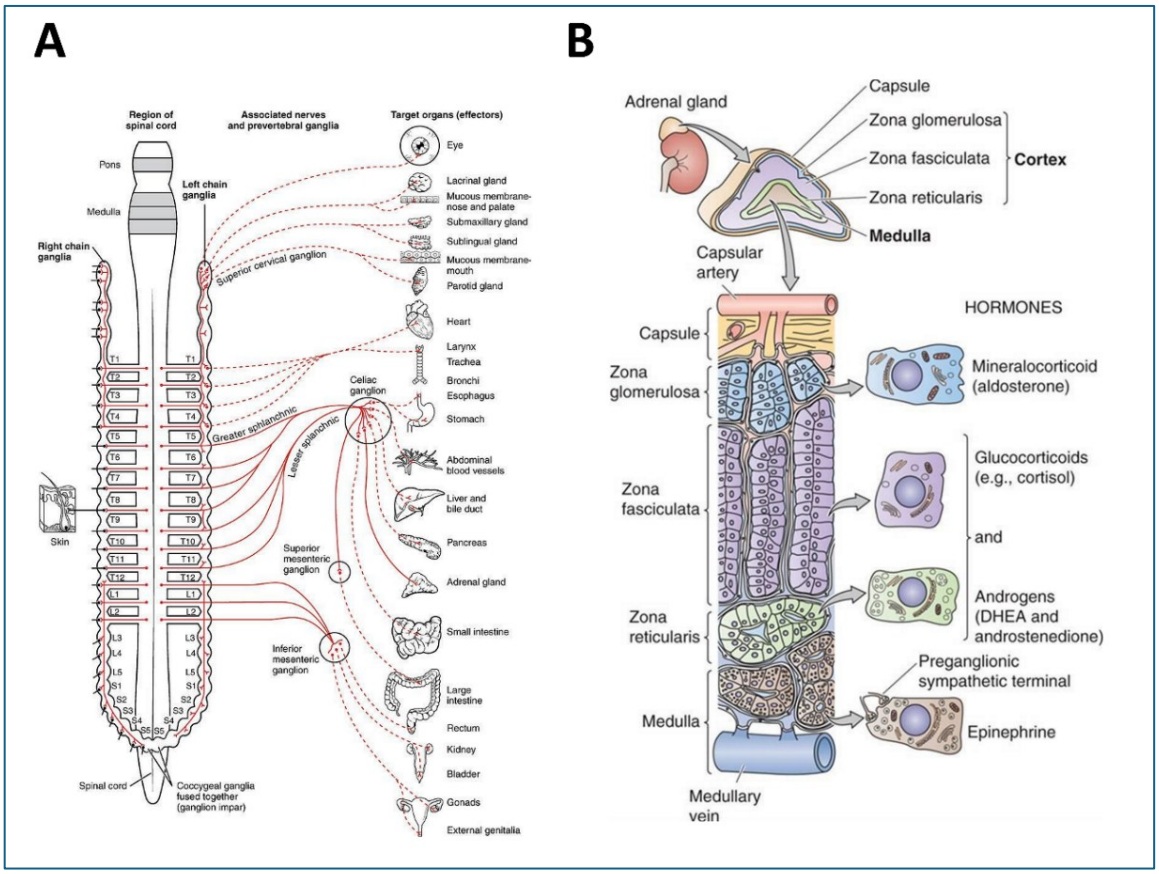

Figure 1.1 Adrenal gland anatomy and produced hormones.
(A) *Sympathetic nervous system diagram highlighting connections to target organs, including the adrenal gland.* ***(B)*** *Structure of the adrenal gland, showing the cortex (zona glomerulosa, fasciculata, reticularis) and medulla. The cortex produces hormones like aldosterone and cortisol, while the medulla releases catecholamines (epinephrine and norepinephrine) in response to stress, regulating heart rate and energy mobilization. Source of the images: https://doctorlib.info/physiology/medical-physiology-molecular/51.html, https://en.wikipedia.org/wiki/Sympathetic_nervous_system*

At this point, NCCs start to migrate along specific paths reaching numerous parts of the embryo, where they terminally differentiate in many different cell types that contribute to the formation of diverse anatomical structures. Along with the anterior-posterior neural axis of the embryo, NCCs can be divided into 4 different groups depending on their functions: cranial, cardiac, vagal, trunk and sacral neural crest (Bronner, 2012). Cranial NCCs migrate dorsolaterally to form the craniofacial mesenchyme, which later differentiates into cartilage and glial cells. Cardiac NCCs develop into melanocytes, neurons and connective tissues. Trunk NCCs instead follows one of three main pathways: they can migrate dorsolaterally to become pigment-producing melanocytes, ventrolaterally to form the dorsal root ganglia of the sensory neurons or take a more ventral path to develop into the

sympathetic nervous system and glial cells. Lastly, vagal and sacral NCCs contribute to form the parasympathetic ganglia within the gut. Delamination, migration, and post-migratory differentiation of NCCs are controlled by complex epigenetic and transcriptional programs that include histone modifications and the expression of specific transcription factors (Hu *et al*, 2014). It is well known that impaired development, differentiation and migration of NCCs can lead to different syndromes known as neurocristopathies but can also promote oncogenic processes in many pediatric malignancies including NB and craniofacial osteosarcoma (Ponzoni *et al*, 2022). Of note, not all the NCCs terminally differentiate and lose their multipotent capacity when they reach their post-migratory destination. Indeed, some subpopulations of NCCs are still present in different tissues at late embryonic stages and in the adult body. These cells that retain their stemness capacity may have been implied at different levels not only in tissue regeneration, but also in cancer promotion (Ponzoni *et al*, 2022).

1.1.4 Trunk neural crest cells

As mentioned before, after delamination from the neural tube, trunk NCCs can take three major routes. In the first one, they move dorsolaterally becoming melanocytes. These cells move through the dermis and enter the ectoderm through little holes in the basal lamina (Gilbert, 2000; Theveneau & Mayor, 2014b). The second route takes trunk NCCs ventrolaterally through the anterior half of each sclerotome. The sclerotomes are mesodermal cell structures derived from somites, which later develop into vertebral cartilage. Within the sclerotomes, NCCs give rise to the dorsal root ganglia, consisting of sensory neurons and glial cells. Finally, trunk NCCs can move ventrally and differentiate in sympatho-adrenal progenitors (SAPs) of the sympatho-adrenal lineage. At the level of the dorsal aorta, BMPs mediated signaling is directly involved in the migration of these progenitor cells and the subsequent differentiation of the primary sympathetic ganglia of the two paravertebral chains. Recent studies conducted in mice have also shown that part of these progenitor cells can migrate to para-aortic sites and differentiate to form secondary suprarenal sympathetic ganglia and chromaffin cells of the developing adrenal gland (Lumb & Schwarz, 2015; Lousado *et al*, 2017).

1.1.5 Schwann cell precursors

Following these three waves of early neural crest migration, a population of late migrating NCCs differentiate and become Schwann cells precursors (SCPs), a population of nerve associated multipotent cells. These cells use the axons of the preganglionic neurons to

migrate to their final destinations in contrast to the free chemoattractant-dependent migration of early migrating NCCs (Lousado *et al*, 2017; Ponzoni *et al*, 2022). For a long time, SCPs were believed to be primarily committed to Schwann cell differentiation, as their name implies. However, recent studies highlighted the potential role of these cells to differentiate in many different other cell types including parasympathetic and enteric neurons, melanocytes and odontoblasts (Kastriti *et al*, 2022).

Early developmental researchs suggested that adrenal chromaffin cells and sympathetic neurons share a common origin from a sympathoadrenal progenitor (Lousado et al., 2017b; Kholodenko et al., 2018a). This model was supported by the evidence that both sympathetic neurons and chromaffin cells express a common set of marker genes and the fact that chromaffin cells can be induced toward a sympathetic neuron-like phenotype under the stimulation of nerve growth factor (NGF) (Kholodenko *et al*, 2018a; Huber, 2006). However, in 2017 Furlan et al. surprisingly found that SCPs in mice generate about 80% of chromaffin cells in the adrenal gland (Furlan *et al*, 2017). In their work the authors showed that if they remove genetically the SCPs, chromaffin cells lose their ability to migrate toward the adrenal gland. They obtained the same result though the ablation of the preganglionic nerves demonstrating their importance of the migration of SCPs to the adrenal gland. Moreover, the authors demonstrated the complex mechanism underlying the differentiation of SCPs to chromaffin cells. At first, SCPs differentiate in an intermediate cell type named bridge cells that forms, as the name suggests, a continuous transient bridge state between SCPs and chromaffin cell cluster. In a second phase bridge cells suppress glial gene expression program and upregulate the genes involved in chromaffin cell differentiation. These results demonstrate that chromaffin cells of the adrenal medulla and sympathoblasts in the extra-adrenal ganglia originate from two different neural crest lineages. Experimental evidence showed that the separation between Schwann cell and sympathoadrenal lineages takes place very early during the embryonic development (**Figure 1.2**). Already at E10.5 it is possible to distinguish committed SAPs and SCPs. A panel of marker genes were characterized by studying the different transcriptional signatures. Among the others, Phox2b and Ascl1 while Plp1 and Sox10 define the neuronal and glial fate respectively (Furlan *et al*, 2017).

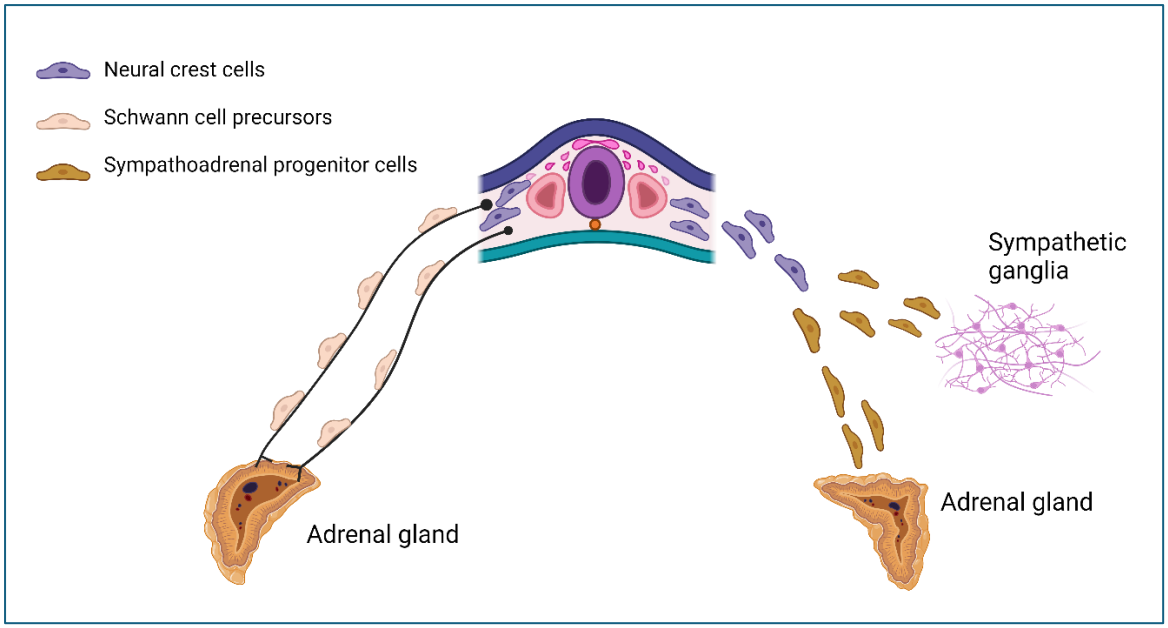

Figure 1.2 Migration Pathways of Trunk Neural Crest Cells
Trunk neural crest cells move ventrally and differentiate in sympatho-adrenal progenitors (SAPs) of the sympatho-adrenal lineage. They subsequently differentiate in primary sympathetic ganglia that form the two paravertebral chains or migrate to the adrenal gland to differentiate in chromaffin cells. Following this early neural crest cell migration, a population of late NCCs called SCPs use the axons of the preganglionic neurons to migrate to the adrenal medulla to differentiate in chromaffin cells and other cell types (Image created with Biorender software).

1.1.6 Cell populations of the adrenal medulla

In the past few years several studies have tried to characterize more in detail the different cell types of the human adrenal medulla. In 2020 Dong *et al.* used single cell transcriptomes to study both human embryonic and fetal adrenal glands identifying three cell populations in the adrenal medulla: SCPs, chromaffin cells and sympathoblasts (Dong *et al*, 2020). They also showed that in the first phase of adrenal medulla development, SCPs expand in number and differentiate in many quiescent chromaffin cells while the second phase is characterized by the expansion in number of cycling chromaffin cells. In 2021 Jansky et al. analyzed with scRNA-seq 17 fresh-frozen samples of human adrenal medulla spanning the first post conception week (wpc) identifying SPCs (SOX10+, PLP1+, ERBB3+), chromaffin cells (DBH+, TH+, CHGA+) and neuroblasts (another name of sympathoblasts) (ISL1+, ALK+) (Jansky *et al*, 2021). The authors identified a bridge population (ERBB4+, ASCL1+) that connects SCPs and chromaffin cells. Interestingly, at 8 wpc they identified a cell cluster of Connecting Progenitor Cells (CPCs) between bridge and chromaffin/neuroblast populations. These cells differentiate in chromaffin and neuroblast cells at 14 wpc. Using a higher cluster resolution, they also isolated two subpopulations of cycling cells in neuroblasts, and SCPs characterized by the high

expression of MKI67 and TOP2A that are not expressed in chromaffin and bridge cell clusters. This suggests that SCPs and neuroblasts can also have cycling characteristics. Moreover, the authors used trajectory inference coupled to RNA velocity to estimate the differential potential of all these cell types. According to this analysis, SCPs showed the highest stemness score, therefore it represents first cell population forming at the adrenal anlage during the early development (Jansky *et al*, 2021). Following the differentiation hierarchy, SCPs differentiate in bridge cells that differentiate in CPCs and subsequently bifurcate into neuroblasts and chromaffin cells. These results support the idea that SPCs are common progenitors of chromaffin and neuroblast populations. In 2021 Kameneva analyzed with more detail the different cell types of the developing human sympatho-adrenal regions from 6 to 14 wpc (Kameneva *et al*, 2021a). In this work she found that intra and extra adrenal sympathoblasts do not share the same progenitor cell. In fact, while intra adrenal sympathoblasts originate from the SCPs, the extra adrenal sympathoblasts directly originate from the migratory NCCs. Moreover, in her study, Kameneva analyzed the cell dynamics occurring in the adrenal medulla and discovered that intra adrenal sympathoblasts can differentiate in chromaffin cells and vice-versa.

1.1.7 Neuroblastoma

Neuroblastoma (NB) is a pediatric tumor of embryonic development arising from the peripheral sympathetic nervous system. It is known to be the most common extracranial solid tumor in children and according to recent data it accounts for 7% of all pediatric neoplasms in patients under 15 years and it is responsible for 15% of all pediatric deaths caused by cancers (Kholodenko *et al*, 2018b). Nevertheless, it is a relatively rare disease that affects 1 in 8000 live births with a median age at diagnosis spanning from 17-18 months with 40% of the patients younger than 1 year (Kholodenko *et al*, 2018a). NB can emerge at multiple sites within the sympathetic nervous system, but the most prevalent location is in the adrenal gland medullary region. However, it can also manifest everywhere along the sympathetic chain from pre- or paravertebral sympathetic ganglia in thorax, abdomen, neck or pelvis. NB can form metastasis to the bone marrow, bones, liver, lymph nodes or also the brain. The diagnosis is generally done using a combination of laboratory tests, radiographic imaging and pathology (Johnsen *et al*, 2019).

1.1.8 Staging and risk classification

NB is a complex tumor characterized by high biological heterogeneity which leads to a wide range of clinical manifestations. In some cases, NB regresses spontaneously leading

to full recovery without any medical intervention. However, in other patients NB can progress and form metastasis with highly unfavorable prognostic outcomes (Kholodenko *et al*, 2018a). Given this diversity of outcomes, in recent years there was a big effort to identify prognostic markers that can assist clinicians to classify patients in different risk groups to adjust the therapeutic treatment. The International Neuroblastoma Staging System (INSS) was the first standardized staging system introduced in 1988 to uniformly classify the extent of NB. The INSS staging system is based on the degree of surgical resection and the clinical evaluation of the primary tumor and metastases. It classifies NB into five stages (Irwin *et al*, 2021). Stage 1 is given to tumors with complete gross excision with representative ipsilateral lymph nodes that are negative for tumor microscopically. Stage 1 tumors are associated with the most favorable prognosis. Localized tumors with incomplete gross excision are classified as stage 2. Unresectable tumors infiltrating across the vertebral column with or without regional lymph nodes involvement are classified as stage 3. Stage 4 is assigned to primary tumors that disseminated to distant lymph nodes, bone, skin liver or other organs. These tumors are generally associated with poor prognosis. Finally, there is an additional stage referred as 4S that is given to localized tumors with dissemination limited to skin and bone marrow. Despite tumor dissemination, this stage is particularly interesting because it is associated with good prognosis and spontaneous regression. These tumors generally have a near triploid number of chromosomes, no MYCN amplification, no loss of chromosome 1p and lack telomerase expression.

In 2009 an additional risk classification system called International Neuroblastoma Risk Group (INRG) was introduced. Unlike the INSS, which is based on post-surgical data such as tumor extent and respectability, the INRG assesses risk prior to treatment by integrating clinical, radiological, and biological features, including MYCN amplification, histology, and chromosomal abnormalities (Monclair *et al*, 2009).This approach allows for more precise risk stratification and personalized management of patients, regardless of surgical outcomes. The INRG stage assessment together with a list of prognostic factors are used to define different risk classification systems to predict patient's risk at the time of diagnosis. In a recent clinical trial promoted by the Children's Oncology Group (COG) almost 5000 children with newly diagnosed NB were enrolled between 2007 and 2017 (Irwin *et al*, 2021). In this study, patients were assigned to one of four risk categories: very low, low, intermediate, or high risk based on INRG stage and a combination of easily measured clinical variables such as age, histological category, MYCN amplification and 1q and 11q aberration and ploidy status. The age of the patients seems to play a relevant role in risk assessment. Indeed, patients older than 18 months tend to show a poor

prognosis respect to younger ones which show a more favorable outcome. In this regard it is interesting to note that screening for catecholamine metabolites suggests that at least half of all NBs that arise in the first year of life are never detected due to the complete spontaneous regression (Matthay *et al*, 2016a). Finally, the amplification of the oncogene MYCN correlates with the worst prognosis also in patients staged as L1 and L2. Patients diagnosed with low- or intermediate-risk NB typically have a favorable prognosis, with a five-year survival rate exceeding 90%. On the other hand, high-risk NBs, which represent about 60% of cases, tend to show an unfavorable prognosis and despite aggressive treatments, the five-year survival rate remains under 50% (**Table 1**).

INRG.Stage	Age	MYCN.Status	11q aberration	Ploidy	Risk.Group
L1/L2		Not Amplified			Very Low
L1		Not Amplified			Very Low
		Amplified			High
L2	<18 mo	Not Amplified	No		Low
	≥18 mo	Not Amplified	Yes		Intermediate
		Not Amplified			Intermediate
M	<18 mo	Amplified	No	Hyperdiploid	High
	<12 mo	Not Amplified	Yes	Diploid	Low
	12-18 mo	Not Amplified		Diploid	Intermediate
MS	<18 mo	Amplified			Very Low
		Not Amplified	Yes		High

Table 1. NB Risk Stratification Based on INRG Classification.
The stratification of NB patients using key clinical and genetic features according to the International NB Risk Group (INRG) system. It categorizes patients based on INRG stage (L1, L2, M, MS), age, MYCN amplification status, presence of 11q aberration, and ploidy. The final column indicates the associated risk group (Very Low, Low, Intermediate, High). (Table adapted from Sokol & Desai, 2019)

1.1.9 Genetic aberrations in Neuroblastoma

Several studies demonstrated that pediatric cancers are characterized by less genomic alterations and mutations compared to adult tumors (Johnsen *et al*, 2019). Therefore, the identification of oncogenic drivers and the development of therapies is challenging for this malignancy (Jones et al., 2019). However, in the case of NB there are some recurrent genetic alterations that include chromosomal aberrations like gene amplifications,

deletions as well as somatic mutations but no one has ever identified a single mutation associated with the development of all NB cases. Whole chromosomal gains are typical of low-risk NB, in these cases tumor cells are frequently found hyperdiploid (triploid, penta or hexaploid). On the contrary, high-risk NBs are generally characterized by segmental chromosomal aberrations affecting only a part of a certain chromosome. The amplification of the MYCN gene, hemizygous deletions of 1p and 11q and the segmental gain of 17q are some of the most common chromosomal aberrations linked to poor prognosis in NB. Another common rearrangement in high-risk cases involves the chromosomal region 5p15.33 located next to the telomerase reverse transcriptase gene (TERT) which leads to an increased TERT expression (Matthay *et al*, 2016b). MYCN is an important master regulator transcription factor that can activate several genes that control important cancer hallmarks. MYCN is known to promote cell proliferation, growth, and modulate cell differentiation in the developing CNS. MYCN is highly expressed in migrating NCCs during the sympathoadrenal development in mice. On the other hand, in differentiating sympathetic neurons MYCN reduces its expression suggesting that this gene does not have any role in modulating sympathoadrenal maturation. The role of MYCN in the regulation of cell differentiation has been demonstrated in two recent studies where the overexpression on MYCN in sympathoadrenal precursor cells was shown to block the development of chromaffin cells promoting the formation of NB in zebrafish and mice. Moreover, MYCN has a role in the modulation of the apoptotic mechanisms. Indeed, during the development of sympathoadrenal a lot of precursor cells that do not differentiate die by induced apoptosis. An overexpression of MYCN during these maturation stages leads to the blockage of the apoptotic signaling resulting in NB development. For all these reasons, it is not surprising that MYCN oncogene is commonly found overexpressed through the chromosomal amplification in 22% of NBs (more than 10 copies for diploid genome) and is associated with high-risk disease and a highly aggressive phenotype with poor outcomes. Being such an important prognostic marker, it has been incorporated into current risk stratification systems. ALK gen is another important oncogenic driver in NB encoding for the anaplastic lymphoma receptor tyrosine kinase. The expression of ALK is important in the context of the sympathoadrenal lineage development. It is thought to regulate the balance between proliferation and differentiation of the cells modulating cellular pathway like MAPK and RAS-related protein 1 (RAP1). ALK is mutated in approximately 14% of high-risk NB cases and more than 35 different mutations have been identified. Most of them are point mutations detected in one of the three hotspot residues in the kinase domain F1174, F1245 and R1275 which lead to gene activation. The precise role of ALK is not completely understood, however it has been shown that an aberrant

activation of this gene leads to cellular transformation within the neural crest. Recent experiments showed that gain-of-function mutation of ALK can promote NB in mice under the control of dopamine-β-hydroxilase (Dbh) promoter, but it requires high expression of MYCN gene when this gene is controlled by Th promoter in mice and zebrafish. These results might suggest a cooperation between ALK and MYCN in driving tumorigenesis in NB. To this end, as these two genes reside on the same location 2p it is not surprising that these two genes are commonly found co-amplified (Johnsen *et al*, 2019).

Other important genomic rearrangements identified in NB include the loss of function alteration in ATRX gene which encodes for the RNA helicase, transcriptional regulator ATRX in 10% of the patients and rearrangements in the promoter of TERT gene encoding telomerase reverse transcriptase in 25% of patients. Interestingly, it was observed that ATRX and TERT are usually not mutated in cells with MYCN amplification. As TERT is one of the MYCN targets, this could suggest that overexpression of the TERT is essential in NB, and this could happen or by MYCN amplification or bypassing the checkpoint mutating ATRX gene. Finally, Polycomb complex genes ARID1A and ARID1B involved in chromatin remodeling are thought to have a role in NB as the haploinsufficiency for these genes are frequent in high-risk cases (Matthay *et al*, 2016b).

1.1.10 The problem of the cell of origin in neuroblastoma

In the recent years there has been a big effort to elucidate the similarities between the sympatho-adrenal cell populations found in humans and the molecular characteristics of NB. These analyses used scRNA-seq to dissect cell lineage trajectories of human adrenal medulla development.

In 2020 Dong et al., compared the single-cell transcriptomes of the adrenal glands from early embryos with adrenal NBs (Dong *et al*, 2020). This work showed that most of NB expression patterns reflect those of chromaffin cells and in some cases of sympathoblasts. Interestingly they found that the gene expression program expressed by most of adrenal NB tumor cells is associated with the one of the noradrenergic chromaffin cells. Furthermore, the undifferentiated chromaffin cell state appears to be closely linked to malignancy, suggesting that the heterogeneity observed in adrenal NB is driven by different levels of chromaffin cell proliferation and differentiation.

In 2021 Kildinsiute et al. made a comparison between human fetal adrenal gland and NB cells using scRNA-seq (Kildisiute *et al*, 2021). They discovered that the transcriptomic profile of tumor cells recapitulates the sympathoblasts gene expression despite the differences between the variability in cell cluster composition and individual patients.

These results were confirmed also by analysing bulk RNA seq data from TARGET and SEQC datasets.

In 2021 Jansky et al. studied in more detail the developmental programs recapitulated by NB cells in human adrenal medulla (Jansky *et al*, 2021). In their work, the authors analysed 7 adrenal medulla samples at different time points and different NB samples from patients belonging to different clinical subgroups. After the identification of malignant cells through the inference the copy number alteration state, the authors studied the similarity of the transcriptome between tumor cells and those of developing adrenal medulla populations. Differently from the work of Dong *et al*., they found that NB cells show the highest similarity with the adrenal neuroblasts. These inconsistences seem to originate from the different marker genes used to annotate the normal cell populations of the adrenal medulla. In particular, the expression of the gene CARTPT was widely debated. Many studies conducted in mice have shown that CARPT gene is highly expressed in sympathoblasts. However, in humans, CARPT is highly expressed in chromaffin cells. This mismatch led Dong et al. to classify human sympathoblasts as chromaffin cells and vice versa. These findings show the challenges of using mouse models to replicate human developmental processes and the development of NB. Moreover, the authors tried to correlate the cell transcriptomes of the normal cell populations with the NB risk identified by genetic aberrations such as MYCN amplification and/or TERT rearrangement (high risk) and the lack of telomere maintenance (intermediate/low risk). They found that low-risk NBs recapitulate the normal neuroblast population and exhibit the highest level of differentiation. On the other hand, NBs with MYCN amplification and those exhibiting mesenchymal characteristics which possess the greatest malignant potential show higher similarities with the undifferentiated cells showing characteristics of early neuroblasts.

In another work published in 2021, Kameneva and colleagues studied the developmental cell dynamics within human adrenal medulla and related sympatho-adrenal areas, comparing these to the diverse cell types found in NB (Kameneva *et al*, 2021b). They demonstrated that a significant number of adrenergic tumor cells show the highest similarity with the embryonic sympathoblasts. Additionally, a subset of these tumor cells showed a strong similarity with chromaffin cells, while mesenchymal tumor cells corresponded with the mesenchymal cells and the tumor cells with a Schwann cell-like signature were found to be like the SCP population.

In summary, all these findings could suggest that NB can originate from sympathoblasts at different stages of differentiation. Tumors with the characteristics of less differentiated sympathoblasts are typically associated with high-risk cases, while the ones with harboring features of more differentiated neuroblasts are typically associated with lower-risk tumors.

Moreover, dedifferentiation mechanisms may play a pivotal role in high-risk NBs, particularly in the presence of MYCN amplification. Furthermore, since the peripheral sympathetic nervous system is derived from neural crest precursors which migrate along various pathways and complete their differentiation in distinct locations, NB may potentially originate from any of these sites. Finally, genetic and epigenetic changes in extra adrenal sympathoblasts migrating to the sympathetic chain, similar to those in adrenal sympathoblasts, may drive NB development in paraspinal ganglia, reflecting genomic differences between adrenal and paraspinal NB tumors (Ponzoni *et al*, 2022).

1.1.11 Single cell RNA sequencing

Although all the cells in the human body share the same genome, their transcriptomes are different according to their function, their proliferation state and the interaction with the surrounding environment. Until ten years ago, the only way we could study the transcriptome of the cells was through bulk RNA-seq technology. However, the biggest limitation of this approach is that sequencing is performed on a mixture of cells. While it works well if we are studying cell lines of the same cell type, this becomes problematic if we are studying the transcriptomes of tissues formed by many different cell types because what we get is just an average gene expression signal of all these cells. This means that we obtain an average of cell identities, developmental states, responses and temporal positions in the cell cycle that basically doesn't represent none of these cells. The advent of scRNA-seq marked a breakthrough in the way we study cell transcriptomes and revolutionized the research. In fact, with scRNA-seq it is now possible to analyse the transcriptome at single-cell resolution for millions of cells in a single experiment. Although, scRNA-seq appeared in 2009 when people started to analyze the transcriptome of 9 cells individually, it got more attention starting from 2014 when technologies such 10X and Smart-seq2 came to the market allowing a higher sequencing throughput that has increased over the years reaching one million cells per experiment. Different methods have been developed over the years to study the transcriptome of single cells that differ for sensitivity and scalability. However, all these approaches can be broadly divided in two main categories: plate-based and droplet-based methods. The former ones like SMART-seq2 use FACS sorting to isolate individual cells into microfluidic chips like Fluidigm's C1 system. These methods usually have a low throughput allowing to sequence 300-500 cells per experiment. The latter ones rely on sophisticated microfluidic systems. By using tiny vessel-like tubes, the system can partition individual cells or nuclei into micro-scale water

droplets. These function as tiny reaction chambers in which the RNA molecules from each cell receive barcoded labels and prepare for sequencing.
The most famous droplet-based technology is 10X Genomics Chromium that allows to profile up to 10000 cells per run, hence the name 10X.

10X Library preparation

In this process, a microfluidic system separates individual cells into small water droplets, where specific reagents and barcoded gel beads are then added. These gel beads are covered with short DNA sequences also known as oligonucleotides. Each oligonucleotide consists of two barcoding elements and a component that binds to RNA. The RNA-binding section, which is a poly(T) sequence, attaches to mRNA molecules.
One of the barcoding elements consists of a unique barcode that is specific for each bead, so that it is possible to assign each mRNA molecule to its original cell. The second barcoding element is a unique molecular identifier (UMI), which is important for the precise measurement of gene expression for each RNA molecule.
These droplets, called Gel Beads in EMulsions (GEMs) at the end of the process contain one cell, a barcoded bead, and essential reagents. The reagents that include enzymes like reverse transcriptase and polymerase stimulate the cell to release its RNA, which then binds to the barcoded beads (**Figure 1.3**).

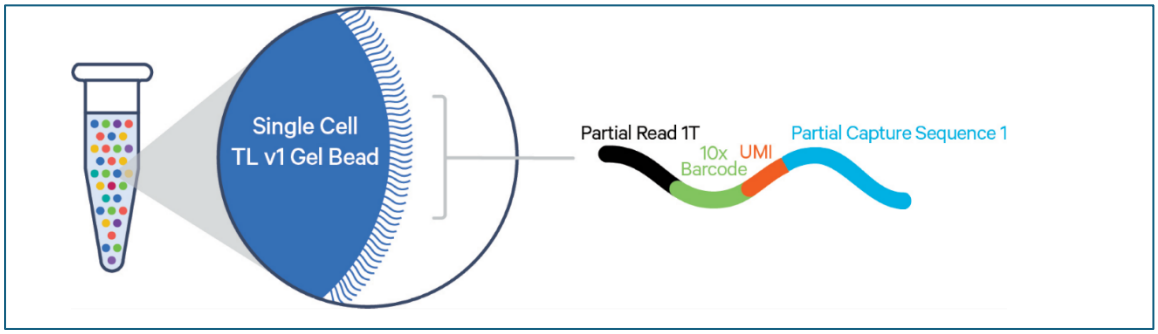

Figure 1.3 Single-Cell RNA Capture and Barcoding.
The beads carry oligonucleotides with a cell-specific barcode (10x Barcode), a unique molecular identifier (UMI), and a poly(T) sequence that binds to mRNA when it is released by the cell after the incorporation in the gel droplet. (Image from 10X Genomic website)

A variant of scRNA-seq is single nucleus RNA sequencing (snRNA-seq). Differently from scRNA-seq, this variant consists of isolating and sequencing only the cell nuclei. This technique is particularly suitable for fresh-frozen samples since the process of freezing may damage the cell membranes leading to mRNA leakage from the cytoplasm of the cells to the surrounding environment. Alternatively, it can be used for those tissues where the process of dissociation may lead to cell damage. SnRNA-seq presents some interesting

advantages over scRNA-seq. First, since most of the mRNA transcripts in the cytoplasm is of ribosomal and mitochondrial origin, these RNAs tend to saturate the sequencing reducing the detection of transcripts that are not mitochondrial nor ribosomal. Instead, since these transcripts cannot be found in the nucleus of the cells, using snRNA-seq it is possible to detect many more genes (about +60%) leading of a higher sequencing complexity. Secondly, snRNA-seq allows to detect more introns since mRNAs in the cytoplasm have been already spliced. However, snRNA-seq sequencing suffers more from the contamination of background noise generated ambient RNAs (Floriddia, 2022).

1.2 Aim of the study

Neuroblastoma is a tumor that arises during the embryonic development by the abnormal differentiation of neural crest cells. The identification of the developmental stage that promotes tumor formation is important to understand the molecular and cellular characteristics that drive tumor initiation and pathogenesis. In recent years, three main single-cell studies (*Jansky et al.*, 2021; *Kildisiute et al.*, 2021; *Dong et al.*, 2020) have tried to identify it by comparing the expression profile of tumor cells with those of healthy adrenal medulla cell populations. However, although these studies have given an important contribution to our understanding of NB, their results are affected by some important limitations:

1. **The small number of analysed samples**, which restricts the representation of the biological heterogeneity of this tumor, such as differences between tumor subtypes, disease stages, and rare cell populations.

2. **The presence of methodological differences** including variations in sequencing platforms (*10x Genomics*, *Smart-seq2*, *CEL-Seq2*) and data processing pipelines, which introduce batch effects and technical biases. Such differences complicate the direct comparisons of the results. Moreover, the three studies used reference datasets of normal adrenal gland that differ in their developmental stage. For instance, while *Dong et al.* and *Kildisiute et al.* focused on fetal adrenal gland tissue from later developmental stages, *Jansky et al.* examined a broader range of timepoints, leading to potential discrepancies in cell type representation.

3. **The use of three different methodologies** to evaluate the similarity between neuroblastoma tumour cells and normal adrenal medulla populations. Although each approach provides useful information, the absence of a unified method makes it difficult to directly compare the results between studies. Specifically:

- *Kildisiute et al.* used a logistic regression model with elastic net regularisation, trained on the reference dataset of the adrenal medulla to calculate a similarity score for each cell in the tumor dataset. This approach is limited as it strongly depends on the training set. Moreover, it excludes information on cellular transitions and may not capture tumour heterogeneity, as the model assigns a predefined similarity to already known populations without considering possible intermediate states.

- *Wang et al.* applied Non-negative Matrix Factorisation (NMF) to identify transcriptional meta-programs, some of which are associated with characteristics of chromaffin cells. However, this approach is more oriented towards deconvolution of tumour heterogeneity rather than a direct comparison with normal cells. Furthermore, the biological interpretation of signatures derived from NMF can be complex and strongly depend on the selection of the number of components, introducing possible biases in the analysis.
- *Westermann et al.* used SingleR, a method based on the correlation between expression profiles, to assign an identity to tumour cells based on reference datasets of normal medulla.

  The use of such different approaches therefore introduces variability in the results and limits the possibility of obtaining a coherent and comparable evaluation of the relationship between neuroblastoma and normal adrenal medulla precursors.

4. These studies do not sufficiently investigate **how the transcriptional profile and the genomic aberrations influence the differentiation** state of tumor cells. A better characterization of these aspects could help identify the signalling pathways and key genes involved in differentiation that are altered in NB. This, in turn, could explain why tumor cells do not complete the maturation process and block their differentiation at specific stages of development.

To overcome these limitations, in this project I integrated the datasets from these three studies into a single analysis, generating a comprehensive neuroblastoma dataset of single-cell and single-nucleus RNA-seq, which includes over 400,000 tumor cells. This approach allowed me to study the molecular characteristics of neuroblastoma on a high-resolution map, improving the ability to identify even rare populations that, in individual studies, could be absent or underrepresented.
In particular, I focused on the study of transcriptional characteristics and intra-tumour variability of neuroblastoma cells, trying to classify them into distinct groups based on their transcriptional profile and their degree of differentiation. In this regard, to evaluate the stage of differentiation of the tumour cells, I first integrated the datasets of the normal adrenal gland from the three studies, reconstructing a detailed cellular map of the adrenal medulla. By studying differentiation trajectories, I was able to reorder the adrenal medulla cells based on their differentiation potential, from the most stem-like to the most differentiated compartment. Using these results, I then analysed the degree of similarity between the tumour cells and those of the normal adrenal medulla, using a more accurate

approach based on the projection of the tumour cells on the reference dataset through Canonical Correlation Analysis (CCA). The adoption of this new method, more robust than those used in previous studies, allowed me to separate tumour cells into different subgroups based on different degrees of differentiation, which had not been identified previously.

Subsequently, I investigated the association between the transcriptional and genomic profiles of tumour cells, with particular attention to chromosomal alterations, to understand how these can influence cell differentiation. By integrating transcriptional data with copy number variation (CNV) analysis, I investigated whether specific genomic aberrations could affect the differentiation trajectory of neuroblastoma cells. This approach allowed me to identify key genes and altered signalling pathways that contribute to blocking the normal process of terminal differentiation of tumour cells. In particular, I analysed pathways involved in the regulation of the differentiation of sympathoblasts, which could have a central role in the differentiation block observed in high-risk neuroblastomas.

Overall, this analysis offers new perspectives on the transcriptional and genomic characteristics of neuroblastoma.

1.3 Materials and Methods

Single Cell RNA-Seq Analyses of neuroblastoma samples

Single-cell and single-nucleus RNA-seq datasets of NBs samples were obtained from three studies including 50 single-cell samples from the study of Dong et al, 2020, Kildisiute et al, 2021 and 36 single nucleus datasets from the study of Jansky et al, 2021. In addition, I analyzed control datasets of normal adrenal medulla samples: 17 from Jansky, 7 from Kildinsiute, and 9 from (Kameneva *et al*, 2021b). Downstream computational analyses were conducted using the Seurat R package (v4.4.0). I filtered out samples containing fewer than 1,000 cells to obtain better results from the integration analysis. Quality control steps were performed to remove low-quality or damaged cells. Firstly, I only retained cells with at least 800 Unique Molecular Identifiers (UMIs) and fewer than 40,000 UMIs to avoid cell doublets. Secondly, cells with fewer than 300 detected genes, a log10GenesPerUMI ratio below 0.80, or a mitochondrial UMI content exceeding 20% were excluded from the analysis. Finally, genes that were expressed in fewer than 10 cells were removed to keep only informative genes only. After applying these quality control measures, I retained 194,000 and 268,000 cells from single cell and single nuclei datasets respectively and 27000 cells for the control samples.

NB and control datasets were integrated separately. The datasets were divided by samples that were normalized, and the 2,000 most variable features were identified using the "vst" method. Integration features were selected, and anchors were identified using the FindIntegrationAnchors() function. The resulting integrated datasets were scaled, and Principal Component Analysis (PCA) were applied using the top 30 principal components. Clustering analyses were then performed on the integrated dataset. The FindNeighbors() function identified neighbor relationships using PCA dimensions 1 through 20, and clustering resolution was selected based on cluster stability, assessed with the Clustree package. Cluster identities were assigned using a reference panel of marker genes from literature.

Cell cycle analysis

I predicted cell proliferation state by computing cell cycle phase scores based on the expression of canonical markers of G2/M and S phase with the CellCycleScoring() function in Seurat. Phase scoring was performed on normalized counts prior to sample integration. To assess the impact of cell cycle variation in our data, we ran PCA analysis on the cells evaluating the expression of cell-cycle genes.

Identification of tumor cells

I used the inferCNV algorithm to predict the copy number chromosomal profile of the cells. As a normalization reference, I used normal sympathoblasts selected from the previously integrated datasets of normal adrenal medulla. InferCNV was run with the following parameters: cutoff = 0.1, denoise = TRUE, HMM = FALSE, k_obs_groups = 1, and write_expr_matrix = TRUE. Gene positions were obtained from Gencode human assembly release 44 (GRCh38.p14). A chromosome region was considered amplified or deleted if the associated inferCNV scoring was >1 or <1, respectively. Cells harbouring amplification of chromosome 17q were marked as tumor cells as it represents the most frequent genomic alteration in NB.

Evaluation of differentiation potential

The R package CytoTRACE2 was used to predict the differentiation state of tumor cells and the developmental potential from the scRNA-seq data. The highest score was associated with the cells with the highest stemness.

Mapping of Tumor Cells onto Controls

To better understand the transcriptional similarities of tumor cells and their relationship with normal developmental processes, tumor cells were mapped onto control cells. For this analysis both control and tumor Seurat objects were set to the "RNA" assay. The right number of principal components was determined using an Elbow plot and anchors between the tumor and control datasets were identified with the FindTransferAnchors() function, using dims = 1:15 and the "PCA" reduction as a reference.

The identified anchors were then used to transfer the control identities to the tumor cells via the TransferData() function. Tumor cells were then mapped on the UMAP embedding reference space of the control dataset using the function MapQuery() function with the parameters (reference.reduction = "pca", reduction.model = "umap").

Survival analysis

Survival analyses were performed using the GSE49711 dataset, downloaded from the public database Gene Expression Omnibus (GEO). This dataset includes a large cohort of NB patients. It includes gene expression data obtained through microarray technology together with detailed clinical information, such as MYCN amplification status, disease risk classification, age at diagnosis, and survival endpoints, including overall survival (OS) and event-free survival (EFS).

I then developed a custom pipeline to compute an aggregate score of the expression level for a given gene list. Samples were divided into two groups, High and Low, based on the 15th and 85th percentiles of the score distribution. Based on the survival scores, survival analyses were calculated using the survival R package using surv() function. The comparison between the High and Low groups was performed using the log-rank test, implemented with the survdiff() function, to determine statistical significance. Additionally, the hazard ratio (HR) was calculated directly from the log-rank test results.
Kaplan-Meier survival curves were generated using the survminer package and customized with the ggsurvplot() function. Finally, the statistical results of the analysis, including p-values, hazard ratios, and adjusted p-values (Benjamini-Hochberg method), were summarized in a heatmap created with the pheatmap package.

Pseudotime analysis

The Monocle3 R package (v.0.2.0) was used to compute pseudotime trajectories. The Seurat object of tumor cells was first converted into a CDS object, and the "RNA" assay was used for this analysis. Cells were ordered in pseudotime using the learn_graph() function, followed by the order_cells() function to assign their positions along the trajectory. The starting point of the trajectory was chosen based on the scoring calculated by CytoTRACE2, which predicts the differentiation state of cells from their transcriptome. Cells with the highest CytoTRACE2 scores, indicating the most stem-like and least differentiated state, were selected as the starting point. Genes differentially expressed along the pseudotime trajectory were identified using Moran's I, as implemented in the graph_test() function of the Monocle3 package. Genes were filtered based on a threshold of $P<0.00000001P < 0.00000001$ and Moran′s $I>0.15Moran's \, I > 0.15$, highlighting genes that are strongly correlated with the progression along the pseudotime.

Transcription factors analysis

Transcription factor analysis was performed using pySCENIC, a workflow designed to analyze gene regulatory networks (GRNs) at the single-cell level. The preprocessed and normalized Seurat object was filtered and converted into a loom file using the build_loom() function, including cell annotations for subsequent processing in pySCENIC. The workflow is divided in three main steps: (1) GRN inference using the pyscenic grn command which requires the a set of transcription factors (hs_hgnc_tfs.txt) to generate an adjacency matrix; (2) Motif enrichment analysis using the pyscenic ctx command which requires the specification of a motif database (motifs-v9-nr.hgnc-m0.001-o0.0.tbl) to identify target genes that are directly regulated by transcription factors; (3) Calculation of AUC scores for

the regulons in each cell using the pyscenic aucell command which generates a loom file containing regulon activity data. The AUC scores and their binarized versions were integrated into a Seurat object as separate assays for further analysis. Differentially regulons were identified with the function FindMarkers(), using a log-fold change threshold of 0.005. Heatmaps were generated with Seurat's DoHeatmap() function to visualize the scaled and centered AUC scores. Interaction networks between regulons were constructed using the R package igraph, based on adjacency matrices generated by pySCENIC, to highlight relevant connections ranked by importance.

Gene set and gene ontology enrichment analysis

I performed differential expression analyses between the samples using the Seurat function FindAllMarkers(). GSEA analysis was conducted using the R package fgsea (v1.16.0) with the following parameters: nperm=1000, minSize=15, and maxSize=500. Hallmark gene sets were retrieved from MSigDB (https://bioinf.wehi.edu.au/software/MSigDB/human_H_v5p2.rdata). Gene Ontology (GO) and Disease Ontology (DO) enrichment analyses were performed with the clusterProfiler R package (v3.18.1) using standard parameters, while pathway enrichment analysis was carried out with the ReactomePA R package (v1.34.0).

1.4 Results

I started the analysis by integrating the scRNA-seq and snRNA-seq datasets of normal adrenal gland from three different studies: Kildinsiute (scRNA-seq), Kameneva (scRNA-seq), Jansky (snRNA-seq) (Kildisiute *et al*, 2021; Kameneva *et al*, 2021b; Jansky *et al*, 2021). To preserve consistency and minimize technical variations, the samples were integrated separately for each study. To improve the quality of the data, a filtering step was applied to remove low-quality or damaged cells obtaining a total of 9298 cells from Kameneva, 58859 from Kildinsiute and 101826 from Jansky dataset. Subsequently, data reduction was performed with PCA analysis and cells were then visualized in a UMAP embedding space. Cell clusters were computed, and the identities were assigned based on the expression of a panel of reference markers from the literature. Adrenal cortex cells were characterized by the high expression of genes like CYP11A1 and HSD3B2 which are known to play a role in the biosynthesis of steroids typical of the adrenal cortical cells. Additionally immune populations were annotated based on the expression of the immune markers of myeloid cells such as CD68 and LYZ. Endothelial cells were identified by the expression of PECAM1 and VWF, and erythroblasts, by the expression of HBB and HBA1 that associated with biosynthesis of haemoglobin. I then identified different cell populations of the adrenal medulla: sympathoblasts (PHOX2B+ and TH+), chromaffin cells (CHGA+ and PNMT+), SCPs (SOX10+ and PLP1+), and bridge cells (ASCL+ and DBH+) (**Figure 1.4**). Since NB is a tumor that primarily originates in the adrenal medulla, I proceeded the analyses considering only the cell populations belonging to the adrenal medulla.

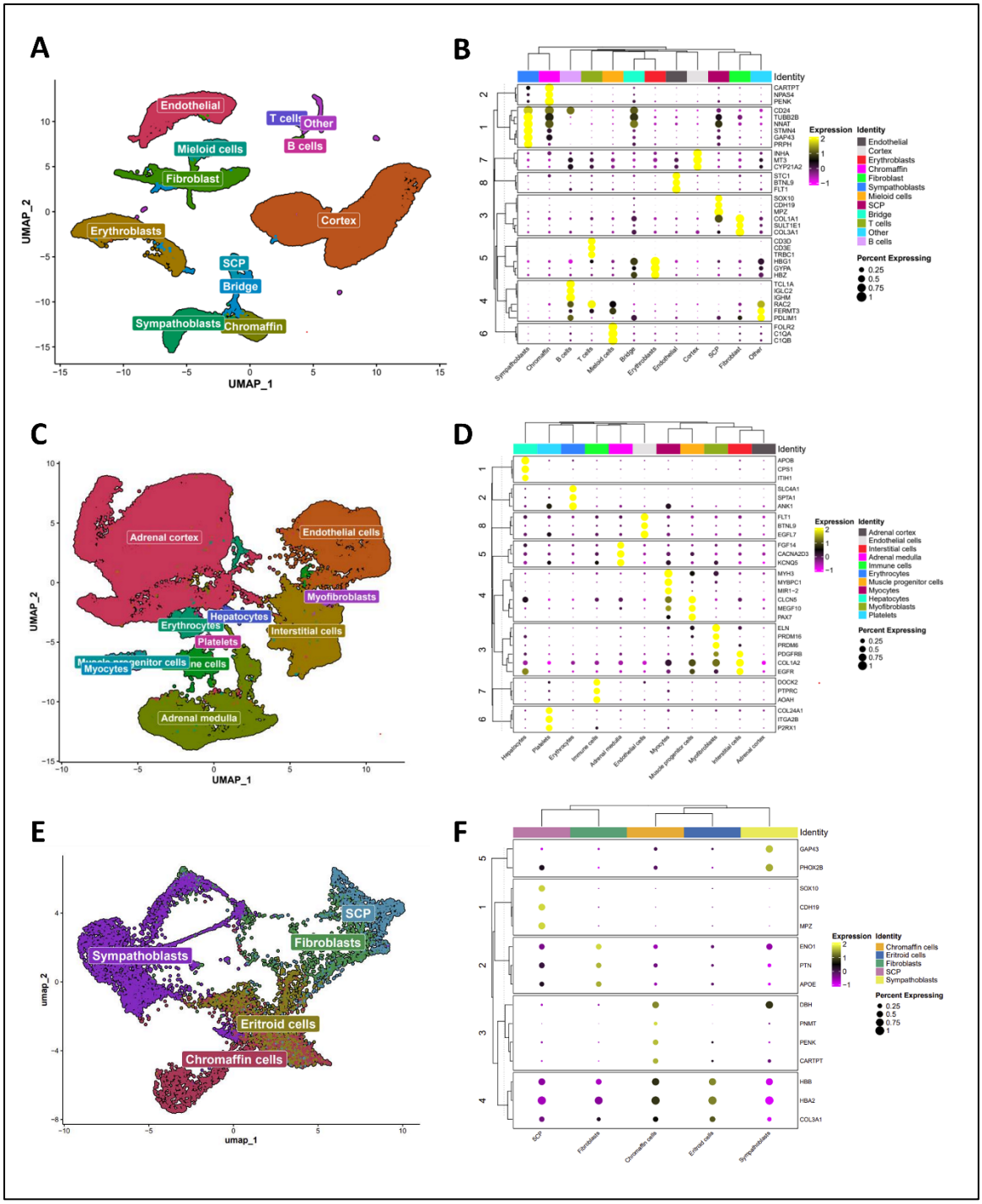

Figure 1.4 UMAP embeddings and marker expression for normal adrenal gland datasets.
UMAP visualizations of normal adrenal gland datasets from Kildinsiute ***(A)****, Jansky* ***(C)****, and Kameneva* ***(E)****. Cell clusters were identified based on marker expression: adrenal cortex (CYP11A1, HSD3B2), myeloid cells (CD68, LYZ), endothelial cells (PECAM1, VWF), erythroblasts (HBB, HBA1), sympathoblasts (PHOX2B, TH), chromaffin cells (CHGA, PNMT), SCPs (SOX10, PLP1), and bridge cells (ASCL1, DBH). Dot plots* ***(B, D, F)*** *show marker expression for annotated populations.*

From each dataset, I isolated the adrenal medulla cell populations, and I integrated the data of the three studies using Canonical Correlation Analysis to obtain an integrated map. The UMAP revealed the presence of six main cell populations. Clusters were annotated using the typical markers of the adrenal medulla cell populations reported in literature (Ponzoni *et al*, 2022). The SCP cluster was marked by the expression of SOX10, PLP1 and ERBB3, bridge cells by ERBB4 and ASCL1, chromaffin cells by CARTPT and PNMT, sympathoblasts by DBH and TH. A cluster of connecting cells was identified between the chromaffin and sympathoblasts cluster. A subset of cells belonging to the sympathoblasts cluster marked by the high expression of MKI67 and TOP2A were annotated as Cycling sympathoblasts (**Figure 1.5 A-B**). Correlation analysis using Pearson coefficient and hierarchical clustering showed a different degree of transcriptional similarity between the clusters that align with their developmental states. In fact, closely related cell populations show a positive correlation, such as Cycling Sympathoblasts and Sympathoblasts, indicating their close transcriptional states. Similarly Bridge cells and Cycling Sympathoblasts were positively correlated. On the other hand, progenitor populations, such as SCP, and fully differentiated cells, such as Chromaffin cells showed weaker or negative correlations reflecting their distinct developmental states (**Figure 1.5 C-D**).

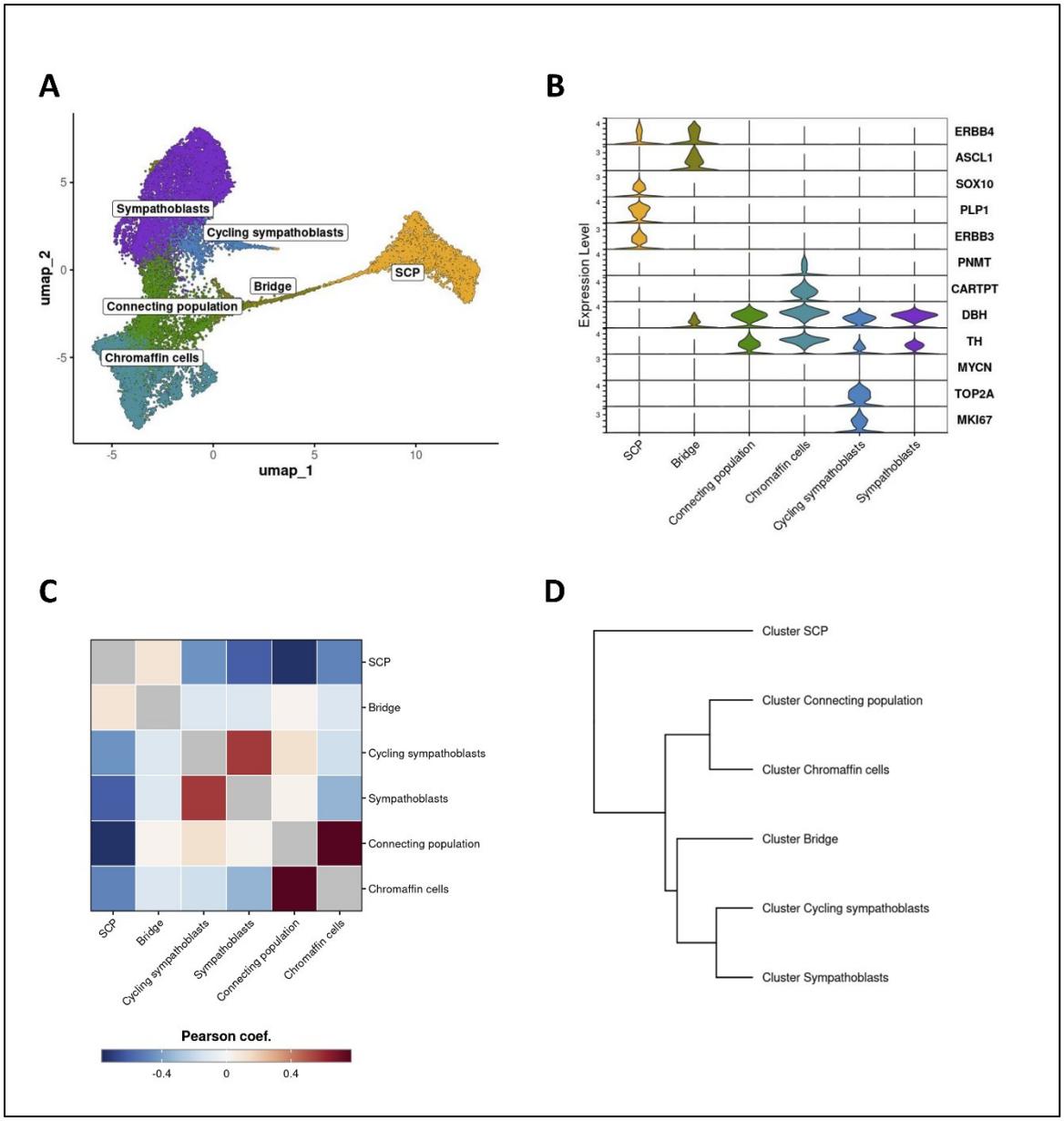

Figure 1.5 Transcriptional correlation and hierarchical clustering of adrenal medulla populations. ***(A)*** *UMAP embedding of adrenal medulla populations: SCPs, Bridge cells, Connecting population, Chromaffin cells, Cycling Sympathoblasts, and Sympathoblasts.* ***(B)*** *Violin plots showing expression of key marker genes for each population.* ***(C)*** *Heatmap of Pearson correlation coefficients between populations highlights transcriptional similarities, with positive correlations observed between related populations such as Cycling Sympathoblasts and Sympathoblasts or Bridge cells and Cycling Sympathoblasts.* ***(D)*** *Hierarchical clustering of populations reflects their developmental relationships, with progenitor populations (e.g., SCP) distinct from fully differentiated populations (e.g., Chromaffin cells).*

I then performed trajectory inference analysis to identify the differentiation trajectories of the populations of the adrenal medulla. First, I identified the cells with the higher stemness using CytoTRACE2. As expected, the SCP cell cluster showed the highest stemness scoring consistent with their role as primary progenitor cell population in the adrenal medulla. Interestingly, stemness signal was also observable within the cluster of cycling sympathoblasts which led me to the hypothesis that these cells might represent the Sympatho-Adrenal Progenitors (SAP), a population which is known to differentiate in both chromaffin and sympathoblasts cells (**Figure 1.6 A-B**). For this reason, I decided to choose both SCP and cycling sympathoblasts as starting point of the trajectory.

The analysis revealed a differentiation path originating from SCPs, transitioning through the Bridge population, and then bifurcating into two main branches: one leading to the chromaffin cells and the other to the sympathoblasts. Interestingly, the sympathoblasts cluster was characterized by a gradient of differentiation states with some cells on the tip of the cluster which seemed to be in a more advanced differentiation state (**Figure 1.6 C-D**). For this reason, I decided to perform the clustering on the predicted differentiation scores using k-means clustering approach (**Figure 1.6 E**). As expected, the algorithm identified a subcluster of sympathoblasts marked by significantly different differentiation score compared to the other sympathoblasts cells (**Figure 1.6 E-F**). I then identified the genes driving the differentiation toward this subcluster of differentiated sympathoblasts cells and to perform gene enrichment analysis. I found that most of these genes were enriched in gene ontologies like axogenesis, synapse organization and modulation of chemical synaptic transmission and other categories related with the terminal differentiation of the neurons, I called this new subcluster of cells sympathetic neurons. This supports the hypothesis that in the adrenal medulla there is a subset of cells of sympathetic neurons directly derived from sympathoblasts (**Figure 1.7 A-F**).

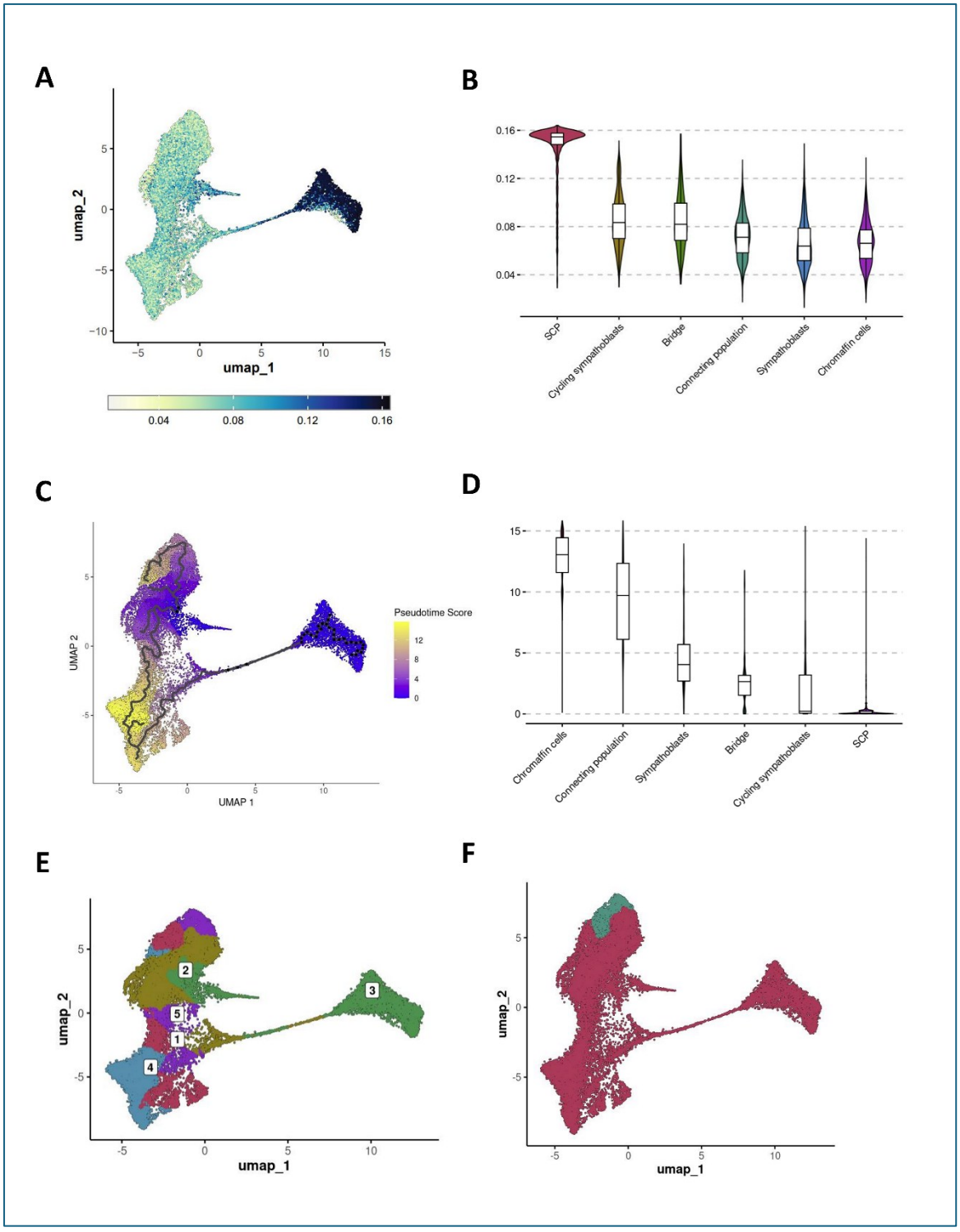

Figure 1.6 Trajectory inference and differentiation analysis of adrenal medulla populations.
(A) *UMAP embedding showing CytoTRACE stemness scores, with SCPs showing the highest stemness, followed by Cycling Sympathoblasts, indicating their potential role as Sympatho-Adrenal Progenitors (SAP).* ***(B)*** *Violin plots of stemness scores across populations, highlighting SCP and Cycling Sympathoblasts with the highest scores.* ***(C)*** *Pseudotime trajectory analysis showing a differentiation path originating from SCPs, transitioning through Bridge cells, and bifurcating into Chromaffin cells and Sympathoblasts.* ***(D)*** *Boxplots of pseudotime scores for each population, with Chromaffin cells and Sympathoblasts at terminal differentiation stages.* ***(E)*** *K-means clustering of differentiation scores identifies five subclusters, revealing heterogeneity*

within the Sympathoblast population. ***(F)*** *Highlight of cells within the Sympathoblast cluster with advanced differentiation states, supporting the hypothesis of a differentiation gradient.*

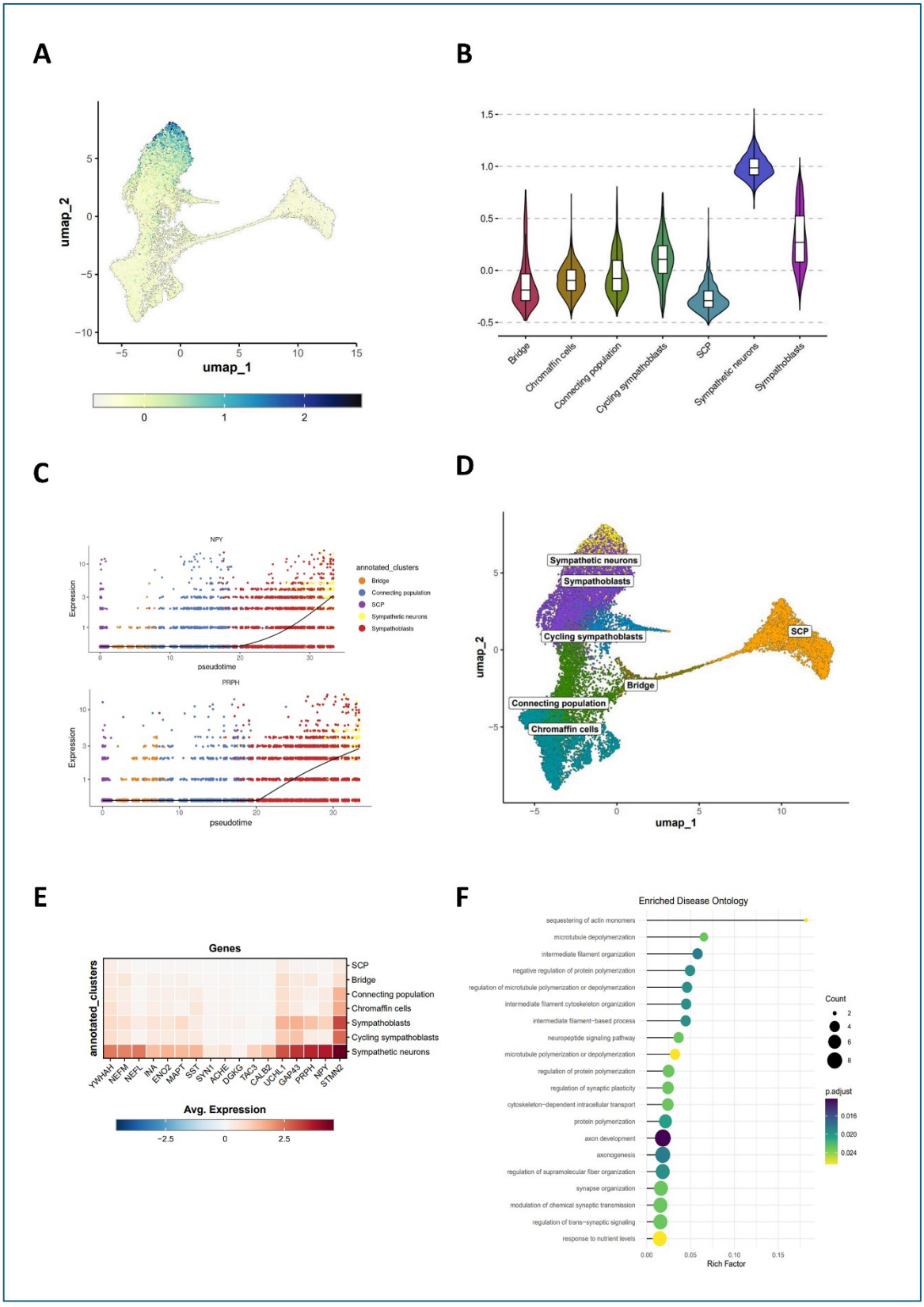

Figure 1.7 Identification and characterization of a subcluster annotated as Sympathetic Neurons. *(**A**) UMAP embedding showing the expression of the gene module identified during pseudotime analysis, highlighting its specific enrichment in the subcluster. (**B**) Violin plot showing the expression levels of the gene module in the identified subcluster, later annotated as Sympathetic Neurons. (**C**) Expression trends of two representative genes (NPY and PHOX2B) from the module, showing increased expression along the pseudotime trajectory leading to Sympathetic Neurons. (**D**) Updated UMAP embedding with the newly annotated Sympathetic Neurons subcluster. (**E**) Heatmap showing the average expression of selected genes from the module across annotated clusters, confirming the distinct transcriptional profile of the Sympathetic Neurons. (**F**) Disease ontology enrichment analysis of the gene module with terms associated with neuronal differentiation and function, supporting the annotation of this subcluster as Sympathetic Neurons.*

I then analysed the tumor datasets from the studies of (scRNA-seq) and Jansky (single-snRNA-seq). As I did for the control datasets, to reduce technical variation and ensure consistent representation of the cells, the single-cell and single-nucleus datasets were processed and integrated separately. After removing low-quality cells, I performed PCA analysis and visualized the cells in a UMAP embedding space to visualize the distribution of the different cell populations present in the tumors. I identified several clusters of stromal cell populations including fibroblasts expressing marker genes such as FN1 and COL1A1, endothelial cells expressing PECAM1 and VWF, and stromal Schwann cells identified by the expression of SOX10 and PLP1 positive and mesenchymal cells by the expression of ACTA2 and COL1A1 as well as a cluster of immune cells. In both datasets I then observed a dominant population of poorly characterized tumor cells that I called Unknown, which constitute most of the cells in the analysed datasets (**Figure 1.8 A-D**).

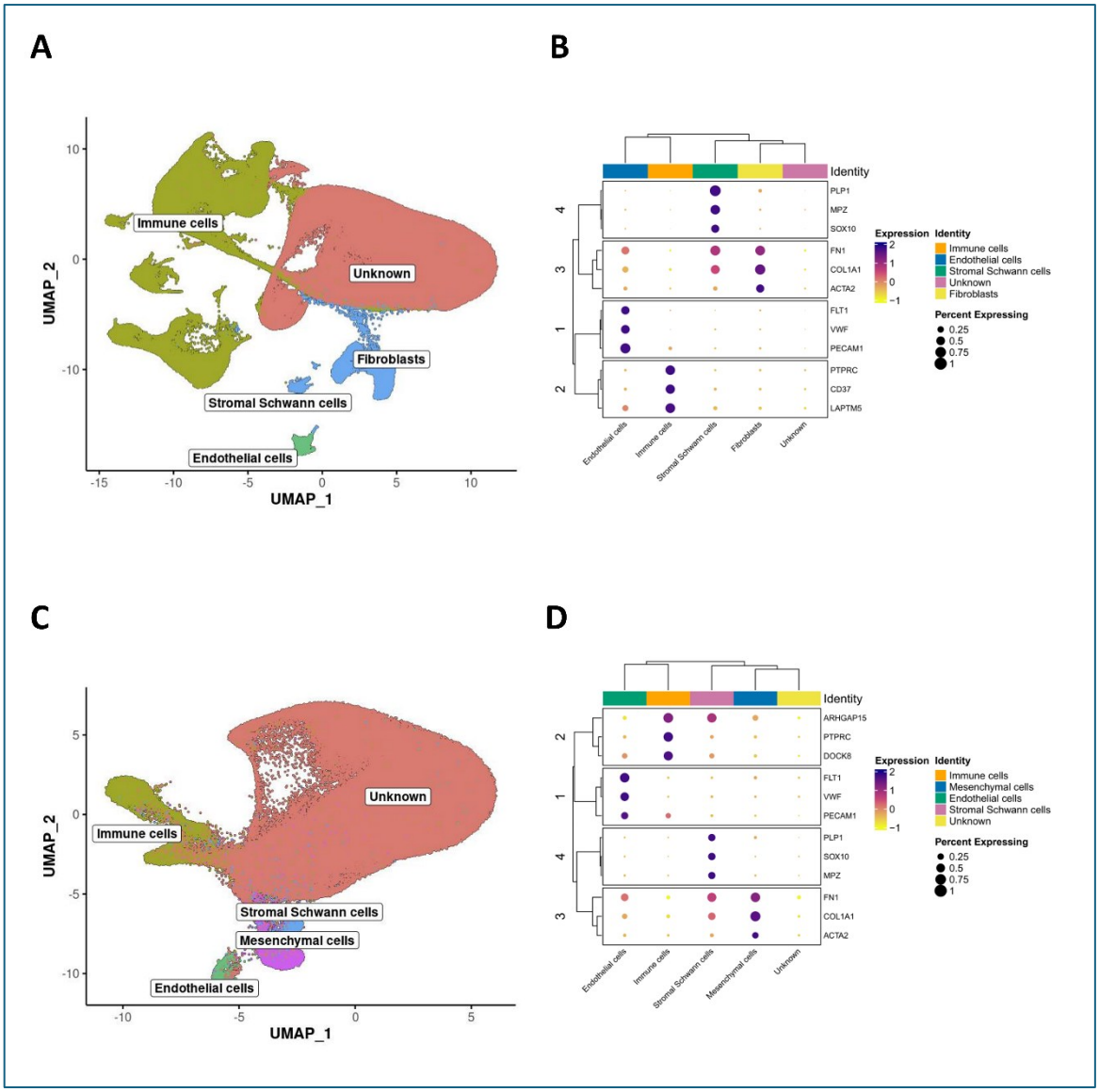

Figure 1.8 UMAP embeddings and marker expression of tumor datasets.
(A) *UMAP visualization of the single-cell dataset from Behjati, showing clusters of stromal populations, including fibroblasts (FN1, COL1A1), endothelial cells (PECAM1, VWF), stromal Schwann cells (SOX10, PLP1), and immune cells (PTPRC, CD7). A large cluster of tumor cells is labelled as "Unknown."* ***(B)*** *Dot plot showing marker expression for each identified population in the Kildinsiute dataset.* ***(C)*** *UMAP visualization of the single-nucleus dataset from Jansky, highlighting similar stromal populations and the dominant "Unknown" tumor cell cluster.* ***(D)*** *Dot plot showing marker expression for each population in the Jansky dataset. Both datasets reveal a prominent population of poorly characterized tumor cells, which constitutes the majority of the analysed cells.*

To identify tumor cells, I performed an analysis of copy number variations (CNVs). I performed the analysis separately for the integrated single cell and single nucleus data to maintain the specificities of each dataset and optimize the inference of CNVs. As reference population for signal normalization, I selected sympathoblasts derived from the control dataset of the adrenal medulla.

The most frequent identified mutations in the samples were the gain of 17q and 7q, and the loss of 1p and 19p. However, to reliably identify tumor cells, I focused on amplification of 17q and deletion of 1p, as these represent the characteristic aberrations of NB. Using these thresholds, I retained 194,000 and 268,000 cells from single cell and single nuclei datasets respectively. As I expected, the tumor cells were mainly located in the cluster I had previously defined as Unknown. In some samples I also observed additional aberrations, such as deletion of 6p, 2q, 4p, 11q, 14q and 3p (**Figure 1.9 A-B**).

I then calculated the correlations between these chromosomal aberrations and analysed them through PCA. PCA reduction allowed me to reduce the complexity of the data and visualize the relationships between the different CNVs. The first three principal components explain 25.87%, 14.04%, and 13.21% of the total variance, respectively, indicating that a good portion of the variability of chromosomal aberrations can be described by these components. Interestingly, I found that the gain of 7q and 17q tend to be highly correlated in the samples, suggesting that these aberrations may co-occur and share common molecular mechanisms in NB. In addition, I also observed a strong correlation between 6p loss and 1p loss, which appear to occur together, indicating a potential common synergy or predisposition in tumor processes in these regions (**Figure 1.9 C**).

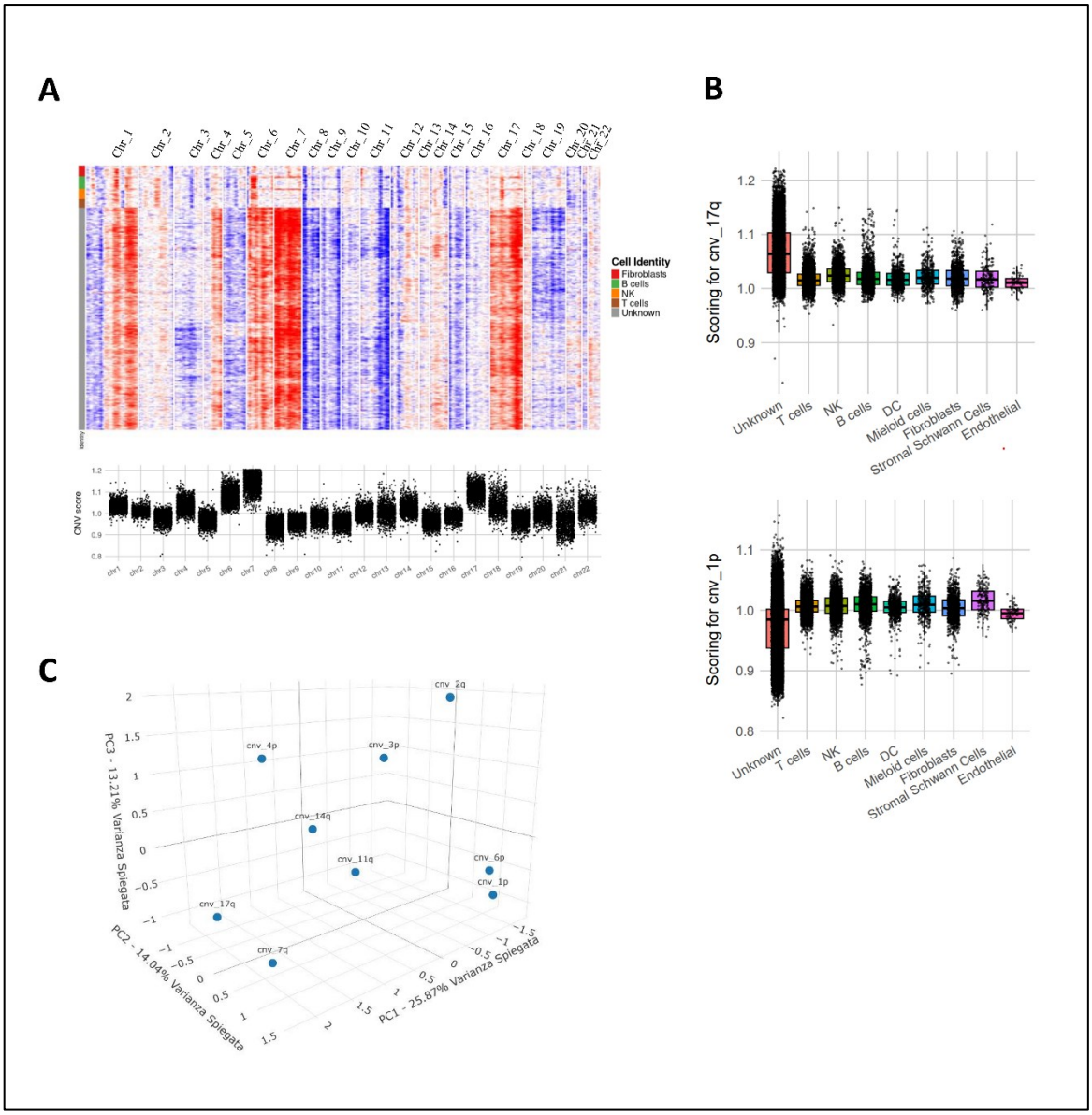

Figure 1.9 Copy number variation (CNV) analysis in NB.
(A) *Heatmap representing CNV profiles across the genome for different cell populations, with boxplots below showing CNV scores for each chromosome. Tumor cells, primarily located in the "Unknown" cluster, exhibit characteristic CNVs such as 17q gain and 1p loss.* ***(B)*** *Boxplots of CNV scores for 17q gain (top) and 1p loss (bottom) across cell populations. Tumor cells show the highest scores for these NB aberrations.* ***(C)*** *Principal Component Analysis (PCA) visualizing relationships among CNVs. The first three components explain 25.87%, 14.04%, and 13.21% of the variance, respectively, highlighting correlations such as 7q and 17q gain or 6p and 1p loss.*

To analyse the relationship between tumor and normal cells in terms of transcriptional similarities, I projected tumor cells onto the PCA structure of the previously analysed adrenal medulla cell populations. This analysis revealed that most of the tumor cells show the highest similarity with sympathoblasts suggesting that they likely interrupt their normal differentiation at this stage independently from the risk (**Figure 1.10 A-B**). Interestingly, about 20% of the high-risk tumor cells map onto the cycling sympathoblasts significantly more than both intermediate- and low-risk groups (nearly 10%). This result could suggest that cycling sympathoblasts may be important in tumor progression in high-risk NB cases, contributing to its more aggressive nature. On the other hand, intermediate-risk cases are more represented in the SCP and bridge populations compared to both low- and high-risk groups suggesting that intermediate risk could be characterized by a unique differentiation state that differs respect to high risk and low risk tumor cells.

I then analysed the chromosomal aberration profile of tumor cells for a specific set of genes belonging to the gene set “Positive Regulation of Neurogenesis” enriched in the sympathetic neurons cluster of the adrenal medulla (**Figure 1.10 C**). I found that these genes are predicted to be deleted in the tumor cells that map onto the cycling sympathoblast population, in particular in high-risk cases (**Figure 1.10 D**). To better investigate the clinical relevance of these findings, I also decided to perform a survival analysis using the expression levels of these genes. Interestingly, tumors characterized by a higher expression of these genes were associated with an improved overall survival (**Figure 1.10 E**). Finally, I identified the transcription factors that drive the terminal differentiation toward the sympathetic neurons. I found at least 18 transcription factors expressed at this stage, among them, GATA3, ISL1, and PHOX2A (**Figure 1.10 F**).

Given their pivotal function in normal development, these factors could be explored as potential therapeutic targets in pathological conditions such as NB, where differentiation is altered.

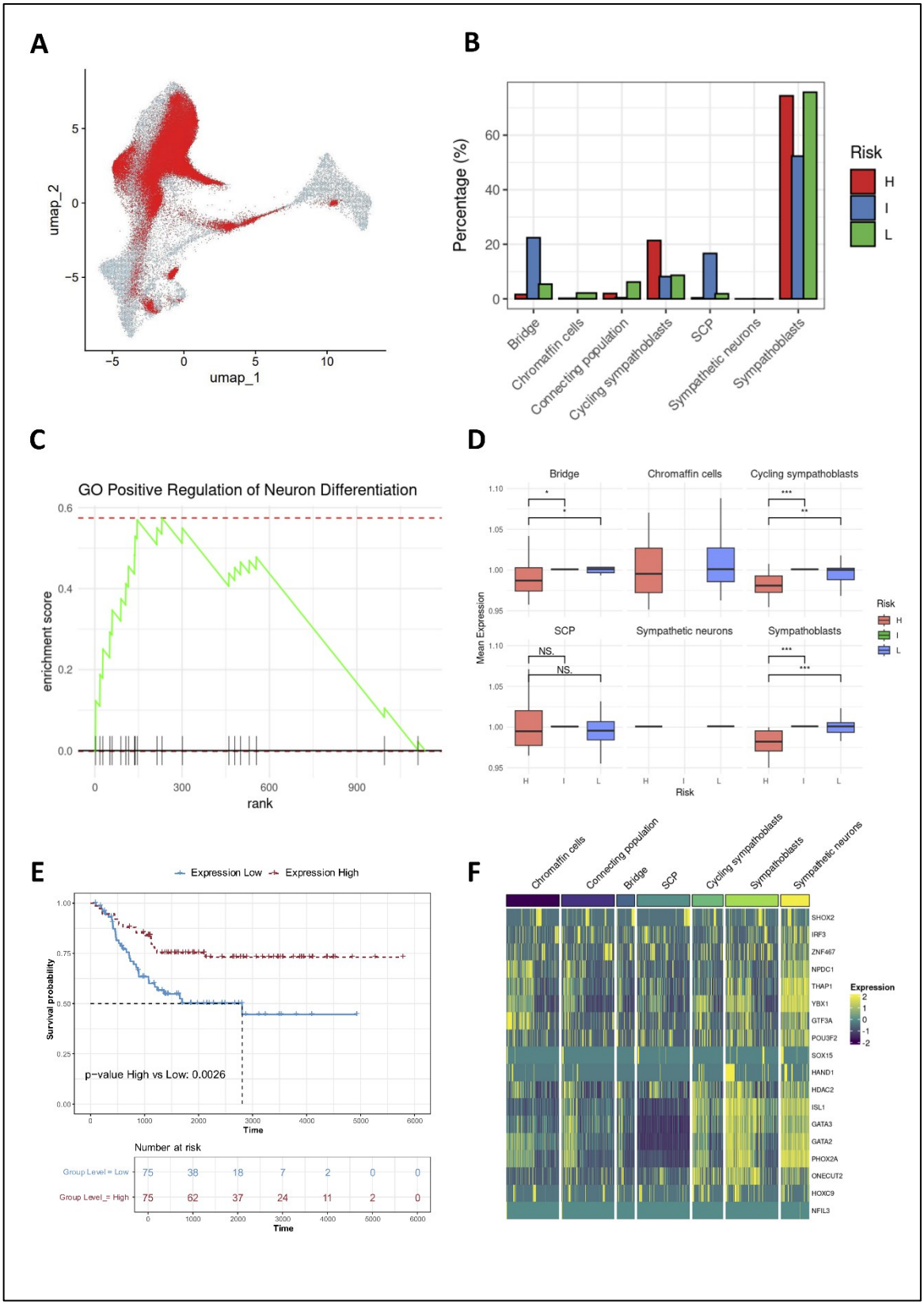

Figure 1.10 Analysis of tumor cell distribution and transcriptional regulation in adrenal medulla populations.

(A) *UMAP of tumor cells (red) mapped onto normal adrenal medulla populations (gray). Tumor cells predominantly map onto sympathoblasts, suggesting a differentiation block at this stage.* ***(B)*** *Distribution of tumor cells mapped to adrenal medulla populations, stratified by risk group (high [H], intermediate [I], and low*

[L] risk). ***(C)*** *Gene Set Enrichment Analysis (GSEA) for the "Positive Regulation of Neuron Differentiation" gene set.* ***(D)*** *Boxplots showing the expression levels of key neuronal differentiation genes. High-risk tumor cells exhibit significantly reduced expression levels (values < 1 indicate deletions), particularly in sympathoblasts and cycling sympathoblasts.* ***(E)*** *Kaplan-Meier survival analysis based on the expression of differentiation-associated genes. Patients with higher expression levels exhibit significantly better overall survival (p = 0.0026).* ***(F)*** *Heatmap displaying the expression of key transcription factors across normal adrenal medulla populations. Transcription factors such as GATA3, ISL1, and PHOX2A are highly expressed in sympathetic neurons.*

1.5 Discussion

In this work I conducted a meta-analysis that integrates the single cell and single nucleus RNA-seq datasets from three different studies to better investigate the underlying mechanisms of NB biology.

The integration of the adrenal medulla datasets revealed the presence of four main cell population: SCP, bridge, chromaffin cells, connecting population and sympathoblasts. As previously identified by Jansky et al, 2021, I also identified a population of cells that forms a subcluster of the sympathoblasts cluster, characterized by the high expression of proliferation-associated genes and a high stemness. These cells are unlikely to be part of the same differentiation lineage of the SCP, as this would imply that SCPs, by differentiating first into bridge cells and then into sympathoblasts, would dedifferentiate and regress into a more stem-like state of cycling sympathoblasts. This would have little biological meaning, and it is difficult to explain from a biological perspective. Therefore, it is much more plausible to assume that these cells are the population of sympathoadrenal precursors (SAPs) that originate from the neural crest cells. These cells are well known to play a key role in the formation of the entire sympathoadrenal system, including intra- and extra-adrenal sympathetic neurons and chromaffin cells in the adrenal medulla (Ponzoni *et al*, 2022). For this reason, it might be more appropriate to consider these cells as part of a separate differentiation lineage from SCPs. Because of that, differently from the previous studies, I decided to calculate the differentiation trajectory, using both the SCP cluster and the cycling sympathoblasts cluster as the starting points of the differentiation trajectory. Using this approach, I found the presence of a subset of cells in the sympathoblasts cluster that has not been detected before characterized by the expression of a set of genes involved in the neuronal terminal differentiation. This suggests that the sympathoblasts of the adrenal medulla show different degrees of differentiation and a subset of them have the characteristics of fully matured neurons.

The direct mapping of tumor cells with normal adrenal medulla populations showed that tumor cells can be grouped in different subgroups depending on their differentiation state. In particular, I found that most of the cells resemble the sympathoblasts cell population, independently form the associated risk and at least 20% of high-risk tumor cells showed high similarities to cycling sympathoblasts. This finding suggests that high-risk tumors are characterized by the presence of a subgroup of cells that maintains a more immature and proliferative state, typical of cycling sympathoblasts. This could contribute to increase tumour aggressiveness, influencing both disease progression and response to therapy. These results are in line with previous studies; however, unlike what has been observed

in the other works, it is interesting to note that although some tumour cells also map to more differentiated cell types, such as chromaffin cells, none map to the sympathetic neuron cluster. This aspect is particularly relevant and introduces an innovative aspect in this type of analysis. Previous studies have limited themselves to highlighting a strong expression of the sympathoblasts signature in tumour cells but have not made any distinction between sympathoblasts and sympathetic neurons. In fact, as confirmed by the trajectory analyses, the latter derive directly from the sympathoblasts and are present in the adrenal medulla.

Another interesting result, not highlighted in previous studies, concerns the fact that intermediate-risk tumour cells tend to be more represented in SCP and bridge populations than in low- and high-risk groups. This suggests that intermediate risk could be characterised by a peculiar differentiative state, distinct from both low-risk and high-risk tumour cells.

In the second part of this work, I investigated the role of chromosomal aberrations in determining the degree of differentiation of tumour cells. To this end, I used *InferCNV*, a tool that allows to infer the genomic status of cells from transcriptomic data. Previous studies have mainly used *InferCNV* for a qualitative evaluation of genomic aberrations, based on the graphical output in the form of a heatmap. However, to date, no study has fully exploited the potential of this tool by analysing its numerical output, which provides a relative scoring for each gene, classifying specific gene regions as amplified or deleted. Starting from this quantitative data, through further analyses, I reconstructed the genomic profile of each previously identified tumour cell subgroup, trying to identify the deregulated gene programmes in relation to the chromosomal aberrations.

By investigating the impact of the genomic aberration profile on the differentiation state of the cells, I found that genes involved in neurogenesis are frequently deleted in tumor cells of high-risk tumors that recapitulate the differentiation state of cycling sympathoblasts. The importance of these genes in the prognosis of NB is supported by survival analyses as the high expression of these genes seem to be associated with better survival outcomes. Additionally, the analysis of the transcription factors that are differentially expressed across the populations of the adrenal medulla identified a group of transcription factors that regulate the terminal differentiation of sympathoblasts. Among these, GATA3, ISL1, and PHOX2A have crucial roles in promoting neuronal differentiation. GATA3 is important in the early stages of neuronal lineage commitment, regulating genes involved in the differentiation and maturation of sympathetic neurons (Tsarovina *et al*, 2004). ISL1 functions as a master regulator in terminal differentiation, driving the expression of genes required for neuronal growth, axonal guidance, and synapse formation (Zhang *et al*, 2018).

PHOX2A is crucial for the specification and survival of sympathetic neurons, controlling genes involved in neurotransmitter synthesis and vesicle transport, and guiding the differentiation of progenitor cells into specific neuronal subtypes (Lo *et al*, 1999). Together, these transcription factors form a regulatory network that coordinates the progression of cells from progenitor states to fully differentiated neurons, leading to the terminal development of the sympathetic nervous system. This analysis has interesting clinical implications, as these transcription factors could be considered as potential targets for treating NB. In fact, the modulation of their activity might help restore the disrupted differentiation programs in tumors, opening new possibilities for therapies.

Overall, while this study has provided new insights into the biology of NB, further investigations are still necessary to better elucidate the underling mechanisms that drive NB progression. In this study I focused on the analysis of chromosomal aberrations in a specific subgroup of tumour cells. However, it would be interesting to extend these analyses to the other tumor subgroups to better understand which genes and regulatory programmes modulate the identity of these cells, determining their degree of differentiation.

Another interesting analysis could involve validating these findings using bulk RNA-seq data available in cohorts such as SEQC and/or TARGET. In particular, it would be useful to verify if the signatures identified in the tumour subgroups through single-cell data can be also found in bulk data. Subsequently, the relationships between the expression of these signatures and specific clinical characteristics could be explored, with the aim of developing predictive models through machine learning approaches.

Finally, it would be interesting to investigate how tumor cells interact with the surrounding tumor microenvironment analysing the ligand-receptor interactions. The characterization of these mechanisms could improve our knowledge of tumor biology and help the development of new therapies for patients affected by NB.

Chapter 2

Spatial Transcriptomics reveals oligodendrocyte differentiation in glioblastoma infiltrative regions

2.1 Introduction

2.1.1 Brain tumors

According to the American brain tumor association, brain tumors represent the fifth most common form of cancer with an incidence rate of 7 cases per 100000 individuals annually making them the most diffuse adult solid tumor. Even the patients who survive will be affected by long term consequences of exposing the brain to surgery, radiotherapy and/or chemotherapy (Aldape *et al*, 2019). It is reported that there are about 150 different brain tumors that can be classified as primary or metastatic, but most of them are non-malignant (85%) and are diagnosed in people aged 20 years old (Ostrom *et al*, 2019).

Primary brain tumors develop from brain tissues or its surrounding microenvironment and can be broadly divided into glial or non-glial. Glial tumors originate from glial cells, which provide support and isolation within the brain, while non-glial tumors arise from structures such as nerves and blood vessels. However, at late stages they can form metastasis in other parts of the body (such as the lungs or breast) and spread in other parts of the brain, often through bloodstream. About 78% of all malignant tumors are represented by gliomas. They originate from various types of glial cells, including astrocytes, ependymal cells, and oligodendrocytes. For this reason, glial tumors can be classified as astrocytomas, ependymomas, GBMs, medulloblastomas, and oligodendrogliomas (Zeng *et al*, 2015).

2.1.2 Glioblastoma

Glioblastoma (GBM), IDH wild type is the most aggressive form of malignant glioma that has an incidence rate of 3-6 cases per 100,000 people annually (Grech *et al*, 2020) . Life expectancy remains at approximately 12–18 months despite the improvements in diagnosis and therapeutic treatments with only about 5.5% of patients surviving five years after diagnosis (Shah *et al*, 2022; Sharma *et al*, 2023). This is because GMB tends to be highly resistant to current treatments due to its heterogeneity, immune adaptation and invasive capabilities (Sharma *et al*, 2023). GBM cells can easily invade the neighboring healthy brain tissue but in some few cases they can also migrate to other areas such as the cerebrospinal fluid, the brain's outer membrane, or bones. Tumor infiltration from the primary site makes the recurrence inevitable. Indeed, most of the patients experience tumor regrowth within two years from the surgery. On the other hand, metastasis outside the brain is uncommon and it generally occurs only in 0.4-0.5% of cases and involves bones, lymph nodes, lungs, and liver (Zhang *et al*, 2021). Symptoms of GBM can include persistent weakness, vision loss and other manifestations (Alexander & Cloughesy, 2017). The progression of neurological signs usually occurs slowly but a small number of patients show these symptoms very fast, often due to intracranial hemorrhage. That’s why some individuals may experience headaches, nausea, and vomiting episodes. Cognitive function changes are also common, especially among the elderly. For many patients, the time from the appearance of symptoms to diagnosis is less than three months and 84% of diagnoses are made within the first six months (Ohgaki & Kleihues, 2007).

GBM is classified as a diffuse astrocytic glioma IDH- wildtype. IDH is the isocitrate dehydrogenase and it plays a crucial role in cellular metabolism. IDH is involved in the citric acid cycle, also known as the Krebs cycle, that is important for the generation of energy through the oxidation of acetyl-CoA from carbohydrates, fats, and proteins into carbon dioxide and chemical energy in the form of ATP (Tang *et al*, 2015). In 2016 the World Health Organization (WHO) revised the classification system for glioblastoma (GBM), introducing a distinction based on IDH mutation status, categorizing tumors as either IDH-wildtype or IDH-mutant. However, the updated 2021 WHO guidelines specified that GBM should be identified only as IDH wild type (WHO Classification of Tumors Editorial Board 2022; (Louis *et al*, 2021)). In contrast, IDH mutant neoplasms are classified as astrocytomas or oligodendrogliomas, dependent on the concurrent existence of a 1p/19q codeletion (Louis et al, 2021).

2.1.3 Genetic landscape of glioblastoma

Glioblastomas typically exhibit many genetic or molecular aberrations (Aldape *et al*, 2019). Diagnosis often relies on the mutation of the TElomerase Reverse Transcriptase (TERT) promoter, as well as amplification, mutation, rearrangement, or splicing alterations of the Epidermal Growth Factor Receptor (EGFR) gene. Additionally, GBM is usually characterized by the simultaneous gains of chromosome 7 and losses of chromosome 10 (Louis et al, 2021; WHO Classification of Tumors Editorial Board 2022). These chromosomal changes frequently appear in tandem (Stichel et al. 2018). In some cases, amplification may occur only on the arm q of chromosomes 7 and 10 (chr.1q and chr.10q gains). These genetic markers, together with histopathological features like microvascular proliferation and necrosis, form the basis for diagnosing IDH wild-type GBM (Louis et al, 2021).

2.1.4 Therapeutic treatment

The conventional approach for treating GBM is the surgical removal of the tumor together with a margin of safety. However, the ability of GBM cells to invade the surrounding tissues limits the therapeutic efficacy (Seker-Polat *et al*, 2022). In fact, cell migration in the brain parenchyma poses a significant obstacle to tumor excision as infiltrative cells are difficult to detect and if they escape from the surgery, they induce the regrowth of tumor. Surgical removal is commonly combined with adjuvant radiotherapy and administration of the chemotherapeutic agent TeMoZolomide (TMZ). TMZ is an alkylating compound that alters the structure of DNA or RNA molecules, specifically targeting N7 and O6 positions on guanine and N3 on adenine, adding methyl groups (Lee, 2016). These methylations can lead to mutations if they are not corrected by the DNA mismatch repair mechanism, or alternatively they can be reverted through the action of O6-methylguanine-DNA methyltransferase (MGMT), an enzyme that repairs DNA (Lee, 2016). MGMT is an enzyme that repairs alkylated guanine nucleotides by transferring the methyl group from guanine's O6 position to its own cysteine residues (Yi *et al*, 2019). This mechanism is crucial to prevent genetic mutations, cellular apoptosis, and the subsequent cancer formation induced by alkylating agents (Yi et al, 2019). Consequently, it is not surprising that a reduced activity of this enzyme correlates with an enhanced survival prognosis for GBM patients. Furthermore, the MGMT gene expression is directly influenced by the methylation status of CpG islands within the MGMT promoter region; Because of that, methylation of the MGMT promoter that leads to a decrease of MGMT expression, is associated with an increase efficacy of TMZ therapy.

2.1.5 The role of Gliolan in surgical resections

Gliolan is a drug approved by the European Medicines Agency (EMA) and by Agenzia Italiana del Farmaco (AIFA) for the use during neurosurgical resections of malignant gliomas. It is an intraoperative optical image drug that allows the identification of the tumor mass by the emitted fluorescence of cancer cells (Hadjipanayis & Stummer, 2019). Aminolevulinic acid hydrochloride (5-ALA) is the principal component of Gliolan and it is the biochemical precursor of hemoglobin that promotes the production of fluorescent protoporphyrin IX (PPIX) in many epithelial and neoplastic tissues, among them malignant gliomas (Regula et al. 1995). This induced fluorescence can be observed using a neurosurgical microscope mounted with a fluorescence module that switches from standard white xenon light to violet-blue excitation lighting (Stummer *et al*, 2006). In recent years, many studies demonstrated the pronounced PPIX fluorescence for malignant glioma cells, while normal brain tissues do not exhibit PPIX accumulation following 5-ALA administration (Suero Molina, Schipmann, and Stummer 2019). The absence of PPIX in healthy tissues is attributed to the insufficiency of ferrochelatase, the key enzyme responsible for the last step of heme biosynthesis, in GBM cells, which leads to the accumulation of PPIX only in neoplastic tissues, making it detectable under light excitation in the 400-nm spectrum (Stummer *et al*, 2006). Intraoperative fluorescence visualization with Gliolan facilitates the discrimination of distinct regions within the neoplasm. Strong Gliolan fluorescence (ALA positive) is associated with the tumor's proliferative zone, while weak fluorescence marks the invasive margin at the tumor's periphery. In contrast, adjacent non-neoplastic brain tissue, or necrotic tumor areas, do not exhibit fluorescence (ALA negative). In 2006, Dr. Walter Stummer conducted a study showing that the use of 5-ALA could lead more efficiently to complete tumor resections leading to better patient outcomes compared to traditional methods (Stummer et al. 2006). The application of Gliolan presents significant advantages for patients in terms of complete tumor excision and progression-free survival, without raising any toxicological issues (Stummer et al, 2006).

2.1.6 Histology

GBM is primary composed of astrocytic tumor cells, which often show poor differentiation and have atypical nuclei and a high variability in cell morphology (pleomorphism). Magnetic Resonance Imaging (MRI) of Glioblastoma often reveals a pattern characterized by an infiltrative and heterogeneous appearance which resembles the structure of a ring

made by a central necrosis zone and an adjacent peritumoral edema (**Figure 2.1**) (Alexander & Cloughesy, 2017).

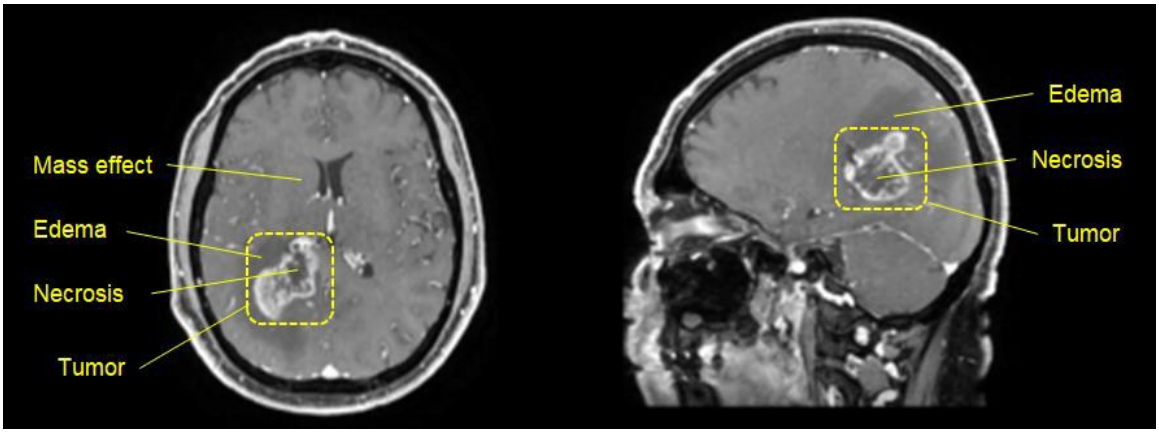

Figure 2.1 Primary IDH-WT Glioblastoma (GBM).
Axial (left) and sagittal (right) T1-weighted contrast-enhanced MRI images show a GBM near the right ventricular trigone. The mass exhibits peripheral enhancement, central necrosis, surrounding edema, and a mass effect. Courtesy of A. Prof Frank Gai.

From a histological point of view, GBM presents unclear margins and a variety of colors on its cross-section; the outer areas often show grey to pink colors, whereas the central parts are characterized by yellowish necrotic areas. Necrosis is the direct consequence of the abnormal proliferation of tumor cells which leads to nutrients and oxygen depletion. This creates an environment characterized by strong acidic and hypoxic conditions inducing cells to necrosis. Necrotic tissue is typically found close to brain structures without a clear transitional tumor zone. It may constitute up to 80% of the total tumor mass. Ravi *et al*. has recently studied the impact of the necrotic microenvironment on tumor cells. Studying the spatial transcriptomes of 20 patients affected by GBM, the authors found that tumor necrotic areas induce the malignant cells to substantially change their gene expression profile (Ravi *et al*, 2022). The most important changes involve the switching toward glycolytic metabolism and the arrest of cell proliferation. Because of that, the authors observed that these cells accumulate genomic instability making necrotic areas a driver for microevolution and favoring the evolution of therapy resistant clones. Interestingly, while most of the cells undergo apoptosis due to the harsh environmental conditions, some cells seem to activate a transcriptional program (“GO” program) that allow them to escape and invade the surrounding healthy tissues. These cells with migratory capabilities show mesenchymal-like features. They called this model “Go or Grow” (Ravi *et al*, 2022). This aggressive invasive behavior is a common feature of all diffuse gliomas, in particular in GBM (Law et al, 2008). Infiltration involves the brain white

matter region, although it can also be found in cortical and deep grey matter areas. Such expansion may lead to the development of a bilateral, mirror-image lesion, referred to as a butterfly glioma (WHO Classification of Tumors Editorial Board 2022). In GBM, infiltrating cells are identified in T2 hyperintense Magnetic Resonance Imaging (MRI) areas and in the so-called Fluid Attenuated Inversion Recovery (FLAIR) regions as well as beyond glioma border and they drive local recurrence following initial treatment. Typically, these cells exhibit a gradient of decreased density that correlates with the distance from the central mass of the tumor. Individual infiltrating tumor cells can be detected histologically several centimeters away from the main tumor site. Notably, these infiltrating cells escape from the surgical resection and are not targeted by the maximum doses of radiotherapy (WHO Classification of Tumors Editorial Board 2022).

2.1.7 Glioblastoma heterogeneity

In 2010 a study of Verhaak *et al*. demonstrated that GBM is characterized by at least three transcriptional signatures, also called programs: proneural, classical and mesenchymal. These programs are characterized by specific gene expression profiles and specific genetic events. For example, alterations in the PDGFRA gene are common in the proneural subtype while EGFR is commonly mutated in the classical subtype. These subtypes are variable also within the same tumor. In recent years, the neural subtype has been later removed (Wang *et al*, 2017). With the advent of single cell technologies, there has been a big effort to better characterize at single cell level these different subpopulations (Darmanis *et al*, 2017; Wang *et al*, 2017; Neftel *et al*, 2019). In 2019 Neftel performed scRNA-seq analyses in 20 adult and 8 pediatric GBM samples. In this study they identified 6 meta-modules, each one characterized by a unique gene expression signature that resemble the normal brain developmental. The first two modules were called MES-1 and MES-2 and they are identified by the expression of genes of the mesenchymal phenotype (Neftel et al. 2019). The difference between MES-like 1 and MES-like 2 is related with the expression of genes in MES-2 involved in the hypoxia response (HILPDA), stress (DDIT3), and glycolysis (ENO2, LDHA) (Neftel et al. 2019). The other meta-module is the NPC-like signature that is further subdivided into two sub meta-modules: NPC1 and NPC2 like. NPC1-like module expresses OPC-related genes (OLIG1, TNR) while NPC-2 like module expresses genes of the neuronal lineage which suggests the tendency to differentiate toward either OPCs or neurons (Neftel *et al*, 2019). Finally, they identified OPC-like and AC-like modules with oligodendrocyte and astrocytic like signatures respectively. Each module was also characterized by additional features that are not

expressed in the normal cell type counterpart, causing a deviation respect to the normal cell type program. Moreover, the comparison of the signatures of these meta-modules with TCGA atlas revealed that the classical subtype overlaps with the AC-like signature, the mesenchymal subtype with MES-2 like signature and finally the proneural with a mixture signature of OPC and NPC-like cells. Interestingly, they also found that 15% of tumor cells express two distinct meta-modules mainly AC-MES like and NPC, OPC-like. They defined this state as “hybrid state”. More recently in 2020 the group of Couturier performed a direct comparison between GBM tumor cells and cells from normal brain development (Couturier *et al*, 2020). They demonstrated that GBM cells are organized into programs that reflect the normal brain development. Indeed, GBM cells can be hierarchically organized into four different lineages: oligodendrocytic, neuronal, astrocytic and mesenchymal that resemble the ones previously identified (Neftel *et al*, 2019). However, differently from the work of Neftel they also identified a fifth cellular state found at the intersection of these four lineages in the diffusion embeddings that transcriptionally corresponds to a bipotent glial progenitor cell in the normal developmental brain. Bipotent glial progenitor cells are a type of CD133-positive cells characterized by a mixture signature of oligodendrocytes, astrocytes and progenitor cells. These cells exhibit a significantly higher proliferation rate compared to other differentiated cancer cells and for this reason, the authors have been proposed these as the potential cells of origin for GBM. This observation suggests that new clones may emerge from this progenitor population and propagate as their progeny undergo differentiation. In support of this, they demonstrated that these cancer stem cells are the most tumorigenic in xenograft models and they drive cancer growth (Couturier *et al*, 2020).

All these studies that use scRNA-seq have provided a great improvement in our knowledge of GBM heterogeneity. However, they do not provide any insight into the interactions occurring between tumor cells and the surrounding microenvironment which has been shown to play an important role in cancer promotion and progression (Sharma *et al*, 2023).

2.1.8 Spatial transcriptomics

ScRNA-seq represents a significant advancement in the field of transcriptomics, in particular for immunologists since it is very simple to extract immune cells such as T and B lymphocytes from a variety of sources including blood circulation, lymphoid tissues, peripheral body parts, and tumors (Williams *et al*, 2022). These cells are not generally fixed within tissues. In contrast, cells like neurons require specific protocols developed for

tissue dissociation because they are very fragile and are strongly integrated with tissue matrices. Therefore, the process of cell extraction from intact tissues without causing stress or cellular deaths is the main limitation of this technique.

Spatial transcriptomics can provide information on the transcriptome of these cells together with their context, without the need for tissue dissociation. This technology allows scientists to connect the information coming from the gene expression profile of the cells with their organization in the tissue by retaining the spatial localisation of RNA molecules within tissues. The spatial context in which this gene expression occurs can be important for understanding the phenotypic properties, state and function of the cells. Indeed, the position of the cell in relation to the surrounding microenvironment can provide several clues like the type of signals that cell might receive. ST offers new possibilities for studying the interactions between cell surface proteins and their ligands at mRNA levels. This becomes especially important when studying complex biological tissues like tumors or neural tissues because in this context the interactions between the cells that make up the immune responses and the dynamics within the tumor microenvironment are extremely important. The size of the tissue can range from a small (<1mm²) piece to whole organ sections and the number of genes detected can range from tens, thousands or even all genome (Williams *et al*, 2022). The methods used to profile genes in spatial context can be broadly divided in two categories: by imaging and by sequencing.

Imaging techniques can be further subdivided into in situ hybridization and in situ sequencing. The first technique uses fluorescence labelled gene specific probes to bind mRNAs (typically via hybridization). The other is the in-situ sequencing (ISS) of amplified mRNAs, where the transcripts are sequenced directly inside a tissue block or section via sequencing by ligation.

The second, much broader approach is by sequencing. There are two sequencing techniques: array based and microdissection. The most famous array-based technique is the Visium commercialized by 10X Genomics. Visium uses spatially barcoded probes in tissue sections to capture mRNAs. Visium offers a spatial resolution of 55 µm (as opposed to previous methods which can only reach 100 um) and it is compatible with fresh-frozen or FFPE samples, making it more versatile across clinical as well as preserved tissues (**Figure 2.2**). The main issue of Visium is that each sequenced spot can contain more than one cell, which means that it detects lots of mRNA from different cells at the same spatial spot. To come over this problem there are some bioinformatic methods based on spot deconvolution that allows to deconvolve the cellular composition within each spot. Finally, microdissection method is based on the dissection of tissue regions from a larger sample followed by subsequent downstream analysis like RNA-sequencing. A well-known

example is laser capture microdissection (LCM), in which cells or groups of cells are captured from a tissue slice by cutting them using a laser beam.

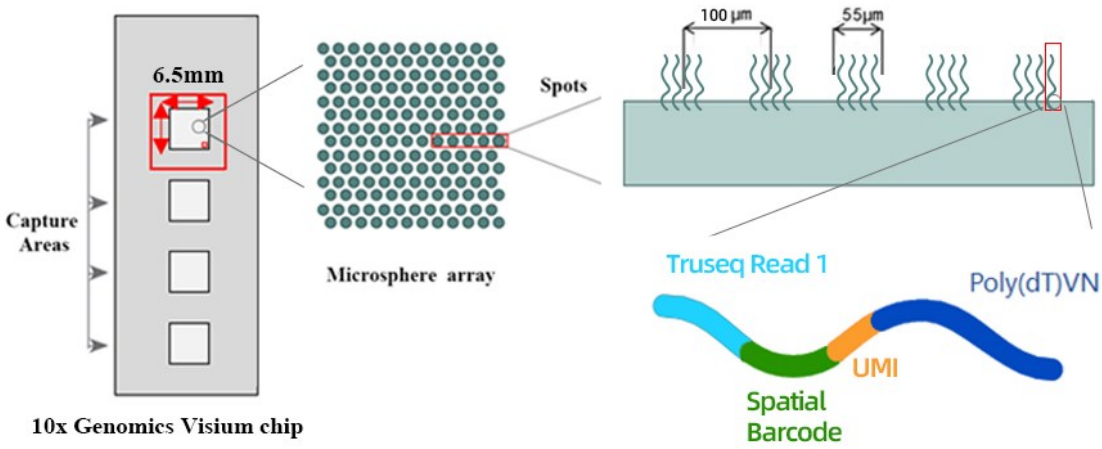

Figure 2.2 Overview of the 10x Genomics Visium platform for ST.
The 10x Genomics Visium chip consists of multiple capture areas, each measuring 6.5 mm, containing a high-density microsphere array. Each spot within the array is spaced 100 µm apart and has a diameter of 55 µm. Spots are coated with capture oligonucleotides containing a unique spatial barcode, a unique molecular identifier (UMI), and a poly(dT) sequence for mRNA capture. During library preparation, mRNA binds to the poly(dT) tail, enabling spatially resolved transcriptomic analysis. The resulting spatial barcodes and UMIs are sequenced using the TruSeq Read 1 workflow, enabling the reconstruction of gene expression patterns in tissue sections.

2.1.9 Insight from spatial analysis

In 2022 Ravi *et al*. performed an extensive ST analysis on GBM samples. In this study the authors tried to link the identity of the tumor cells with their spatial localization. In their study, the authors analyzed 20 GBM patients, and they identified 5 different spatial gene signatures: radial glia, OPC, spatial neuronal development, reactive hypoxia, and reactive immune. The first three programs overlap with the corresponding gene signatures previously identified by Neftel while reactive hypoxia is characterized by the expression of genes of glycolysis and hypoxia. This program was identified in the necrotic tumor margin area and was associated with non-cycling cells that show a mesenchymal-like phenotype. Integrating ST with metabolomics data, Ravi found in hypoxic regions of GBM an important metabolic shift between glycolysis and the pentose phosphate pathway. The extremely low oxygen conditions probably promote this metabolic reprogramming in the necrotic tumor areas, where cells must adapt themselves to maintain energy production and biosynthesis. This metabolic adaptation is strongly correlated with an increased genomic

aberration suggesting that such regions act as “bioreactors” for genomic instability. Indeed, cells found in these regions of chronic hypoxia are more likely to develop mutations and structural alterations in their DNA that contribute to tumor heterogeneity and promote the emergence of therapy-resistant subclones. These alterations seem to be important for tumor cells to allow the switch between proliferative and migratory behaviours depending on their microenvironment. This theory was called "go-or-grow hypothesis”, where cells either focus on proliferating or migrating in response to environmental cues. For all these reasons, the hypoxic microenvironment seems to have an important role in the selective pressure on the surrounding cells, promoting their transformation into a more aggressive, therapy-resistant phenotype.

Finally, the reactive immune program is characterized by the expression of glial genes such as AQP4, GFAP, CD44 but differently from the radial glial cells, they show a high expression of genes of the immunity like HLA-DRA and CCL4. This state signature overlaps with the hybrid signature AC-like and MES-like previously identified by Neftel (Neftel *et al*, 2019). Moreover, spatial analysis revealed that these areas present an increased immune infiltration of exhausted T cells and CD163+ tumor associated macrophages (TAMs) that promote an immunosuppressive environment and tumor cell migration in distant areas. Overall, this demonstrates the impact of the host microenvironment on tumor cell differentiation and transcriptional heterogeneity (Ravi *et al*, 2022) .

More recently, a study of (Greenwald *et al*, 2024) tried to dissect more deeply the complexity of the spatial organization of GBM by combining ST and spatial proteomics. One of the principal findings of this work was the discovery that GBM is organized into distinct layers that are enriched for specific cellular states. This non-random organization suggests a high level of structural complexity within the tumor. Moreover, they found that a key factor that drives this organization was hypoxia, which seems to play a crucial role in the organization of the spatial structure of the tumor. Indeed, hypoxic regions, that are often located in necrotic tumor cores, influence the organization of surrounding cells of the invasive margins of the tumor. In this regard, they found that in the proximity of hypoxic regions, there are specific pairs of cellular states that tend to cluster together across different tumors. This finding supports the hypothesis that hypoxia is not just a marker of aggressive tumor regions but also an essential organizer of the tumor cellular architecture. Together, the findings from Ravi and Greenwald offer new insights into the spatial organization of GBM and the influence of factors like the role of the microenvironment and the hypoxia on tumor heterogeneity. However, despite these advancements, there is still

the need to investigate with more details the correlation between the transcriptional programs and their spatial localization within the tumor. The way the different transcriptional programs are organized across the GBM microenvironment still needs to be explored. Finally, while Greenwald gave an important contribution in the identification of the tumor cellular states that cluster together and the interactions between them, their influence on tumor progression remain to be investigated. This aspect could be important for understanding the mechanisms of therapeutic resistance and tumor evolution.

2.2 Aim of the study

Over the last few years, research using scRNA-seq has substantially improved our knowledge of GBM showing that tumor cells can be divided into four distinct cell identities. These four cell states correspond to three neurodevelopmental lineages and an inflammatory, mesenchymal-like state. Moreover, adding another layer of complexity, it has been demonstrated that these four GBM states are plastic and interchangeable.

However, despite the valuable contributions of these scRNA-seq studies, they are still incapable of providing a complete description of GBM heterogeneity integrating the information coming from the gene expression profile of the cells with their spatial context. This could be important for an improved understanding of the interactions between different cell subpopulations and their relation to both tumor as well as neuronal microenvironment. In recent years, spatially resolved approaches have been used in several studies focusing on GBM. These studies have advanced our comprehension of GBM, but they largely focused on the study the tumor locally rather than the overall organization and complexity.

Among the features that most distinguish GBM is its invasive nature into surrounding brain tissues. These infiltrating GBM cells are in T2 hyperintense magnetic resonance imaging (MRI) areas and in the so-called Fluid Attenuated Inversion Recovery (FLAIR) regions as well as beyond glioma border and they drive local recurrence following initial treatment. This is mainly because they evade surgical removal and do not receive the highest doses of radiotherapy. Research is evolving toward the study of cellular crosstalk among both tumor cells and components within their microenvironment, specifically concerning immune systems. Nevertheless, the properties of invading tumor cells are still largely unknown, and we need a comprehensive molecular analysis to clarify what constitutes invasive GBM cells.

By using ST, the primary objective of this study was to characterize the phenotype and the molecular features of infiltrating tumor cells in GBM and to understand how these cells interact with the surrounding environment.

2.3 Materials and Methods

Patient samples

Tumor samples were obtained from three patients affected by GBM after surgical resection at Santa Chiara Hospital in Trento. The local ethics committee of the Santa Chiara Hospital in Trento approved the using the patient samples. All enrolled patients were fully informed and signed the preoperative informed consent form. Clinical characteristics are summarized in **Table 1**.

Histopathological analyses of tumor sections were performed by the Pathology Department of Santa Chiara Hospital. A team of neuropathologists reviewed tissue sections and tumor types were diagnosed according to the WHO classification of CNS tumors (WHO Classification of Tumors Editorial Board 2022).

Tissue collection and handling

After resection, tissue samples were snap-frozen in isopentane pre-cooled with liquid nitrogen. Frozen samples were transported to the laboratory and stored at -80°C. Frozen samples were embedded in OCT and stored at -80°C until further processing.

RNA quality control and tissue optimization

For each area, 10-15 sections were pooled and total RNA was extracted using Trizol. RNA quality was assessed using the Bioanalyzer (Agilent) according to the manufacturer's instructions. The correct permeabilization time for the Visium Spatial Gene Expression protocol was determined for each area with the 10X Visium spatial tissue optimization kit (PN-1000193).

Visium Spatial Gene Expression Protocol

Spatial gene expression experiments were performed using the fresh-frozen 10X Visium Spatial Gene Expression Kit (PN-1000184). Briefly, tissues embedded in OCT were sectioned on a cryostat at -20°C. 10 µm thick tissue sections were placed on the Spatial Gene Expression slides. One tissue section per area was placed on a Visum Spatial Gene Expression capture area. Specifically, the four sampled areas from each patient were placed on the four capture areas of the same slide. Visium slides were fixed in pre-cooled 100% methanol and H&E staining was performed according to the 10X Genomics protocol. Mosaic brightfield images were captured using a Zeiss Axio Imager M2 microscope at 10x magnification. The brightfield configuration includes the use of a color camera (3x8 bit,

2424x2424 pixel resolution), white balance functionality, 2.18 µm/pixel capture resolution, and 2-10 millisecond exposure times. Permeabilization was performed according to the manufacturer's instructions using the predefined time for each area as determined by the Tissue Optimization Protocol. The RNA was then reverse transcribed, a second strand was produced, and cDNA was amplified using the appropriate number of PCR cycles as determined by qPCR. cDNA quality has been evaluated using an Agilent Bioanalyzer high-sensitivity chip. cDNA was amplified for 18 cycles and purified using AMPure beads. Enzymatic fragmentation and two-sided size selection were performed. P5, P7, i7, and i5 samples and TruSeq Read 2 were added by end repair, A-tailing, adapter ligation, and PCR. AMPure beads were used for post-ligation clean-up. P5 and P7 indexes were added by PCR and the final library was size selected using AMPure beads. The libraries passed the quality control and were sequenced on a NovaSeq 6000 with an S1 flow cell.

Single Cell RNA-Seq Analyses: Quality Control and Preprocessing

ScRNA-seq datasets from IDH wildtype GBM samples were retrieved from three studies:(Wang *et al*, 2022; Neftel *et al*, 2019; Couturier *et al*, 2020). To ensure consistency in the analyses, we included only the 3' tagged samples processed using the 10X Genomics platform, resulting in a total of 38 samples. Downstream computational analyses were conducted using the Seurat R package (v4.4.0). We filtered out samples containing fewer than 1,000 cells to prevent issues during integration steps. Additional quality control measures included the removal of damaged or low-informative cells by excluding those with fewer than 800 UMIs, fewer than 300 detected genes or a mitochondrial UMI content exceeding 30% of total UMI counts. Cells with more than 40,000 UMI counts, likely to be doublets, were also excluded. Genes with no counts were disregarded. Low complexity cells, identified as those with less than 0.8 log10GenesPerUMI, were excluded due to their generally low informativeness. Cell doublets were imputed using the DoubletFinder R package (v2.0.3), and background contamination was calculated with SoupX (v1.6.0). Each sample was processed separately following the standard Seurat pipeline. Data were normalized for sequencing depth scaling with a factor of 10,000 per cell. Highly variable genes were identified using the variance stabilizing transformation method with the function FindVariableGenes(), selecting 2,000 features per dataset. These variable genes were used for principal component analysis (PCA). Dimensionality reduction visualization was performed using Uniform Manifold Approximation and Projection (UMAP), selecting the first 30 principal components.

Integration, Dimensionality Reduction, and Cluster Identification

Sample integration was performed using the Canonical Correlation Analysis (CCA) algorithm implemented in Seurat. The percentage of mitochondrial genes in cells was regressed out. PCA and UMAP reductions were applied to the “integrated” assay, choosing the first 30 principal components as determined by the elbow plot. Clusters were identified using the graph-based clustering approach. Initially, we computed the K-Nearest Neighbors (KNN) graph based on the Euclidean distance in PCA space with the FindNeighbors() function, selecting 30 principal components. Cells were then clustered using the Louvain algorithm with the FindClusters() function, setting the resolution parameter to 1. Cluster identities were assigned based on a reference panel of markers from the literature.

Identification of tumor cells

We employed the inferCNV algorithm to predict the CNV profile of the cells using normal macrophage cells as normalization reference. To avoid the possibility that some tumor cells could also be detected in the macrophage cluster, we decided to choose as reference only the cells where the expression of macrophage markers was above the cluster mean (we called them Ref_Macrophages). Infercnv was run with the following parameters: cutoff=0.1, denoise=TRUE, HMM=F and k_obs_groups=1, write_expr_matrix = T. Doing so, we increased our confidence that these cells were truly macrophages. Gene positions were taken from Gencode human assembly release 44 (GRCh38.p14). We considered a chromosome region amplified or deleted if the associated Infercnv scoring for that specific region was >1 or <1 respectively. Cells harbouring the amplification of chromosome 7 or deletion of chromosome 10 were marked as tumor cells.

Tumor cells integration and identity assignment

After the identification of the tumor cells, samples were integrated using Reciprocal PCA (RPCA). This method was chosen over the Canonical Correlation Analysis (CCA) because tumor cells tend to exhibit high variability and poorly conserved gene expression profiles across different datasets. The high variability makes it likely that many cells do not have corresponding matching types in other datasets which is a requirement for CCA. On the other hand, RPCA provides a more conservative approach to integration because it ensures that cells with distinct biological states are less likely to be forced into alignment. This helps in preserving the inherent biological differences and avoids overcorrection that could arise with CCA. Each dataset was independently normalized using the

`NormalizeData()` function to ensure that the gene expression levels were comparable across cells. We identified the variable features of each dataset using the `FindVariableFeatures()` function with the vst selection method, selecting the top 2000 variable genes. To identify the features that are consistently variable across both datasets, we used the function SelectIntegrationFeatures(). These features were subsequently used for scaling and performing PCA. We identified integration anchors using the FindIntegrationAnchors() function with the RPCA method by setting the reduction parameter to rpca. The datasets were integrated into a single Seurat object using the IntegrateData() function, which makes use of the previously identified anchors. Standard workflow steps for visualization and clustering were performed on the “integrated” assay. These steps included data scaling, principal component analysis (PCA) and the visualization on a UMAP space. We finally identified the nearest neighbours and clustered the cells using the `FindNeighbors()` and `FindClusters()` functions, respectively, with a resolution parameter set to 1.0.

Evaluation of differentiation potential

The R package CytoTRACE2 was used to predict the differentiation state of the tumor cells and the developmental potential from the scRNA-seq data.

Cell cycle analysis

We predicted the cell cycle state by computing the cell cycle phase scores based on the expression of a list of canonical markers of G2/M and S phase with the function CellCycleScoring() implemented in Seurat. Phase scoring was calculated on normalized counts prior to sample integration. To assess the impact of cell cycle variation in our data, we ran PCA analysis on the cells to evaluate the expression of cell-cycle genes.

Pseudotime analysis

Pseudotime trajectory analysis was perfomed using the Monocle3 R package (v.0.2.0). The Seurat object was first converted into a CDS object using the “RNA” assay for the analysis. Cells were ordered along the pseudotime trajectory using the learn_graph() function, followed by cell ordering with the function order_cell(). A node from the cluster identified as tumor stem cells, was chosen as the starting point for the trajectory. Differentially expressed genes along the pseudotime were identified using the function graph_test() based on Moran's I statistic. Significant genes were selected applying a filtering threshold of $P < 1e\text{-}8$ and Moran’s $I > 0.15$.

RNA velocity

RNA velocity analysis was run to infer the future state of individual cells in terms of differentiation state based on spliced and unspliced mRNA dynamics. Starting from BAM alignment files, RNA velocity was computed using the Velocyto pipeline, which estimates the direction and speed of cellular transitions. At first, the spliced and unspliced RNA counts were extracted from the Seurat object and then processed using the scVelo package (v. 0.2.4) to estimate the RNA velocity vectors, which were projected onto the UMAP embedding. Using this method, we could study the trajectories and infer the dynamic cellular states, providing insights into the potential future cell states along the differentiation paths.

Transcription factor activity analysis

Transcription factor activities were inferred using a Bayesian approach using the R package BITFAM. Raw counts from the integrated Seurat object were normalized using the function BITFAM_preprocess() and the TFs activities were inferred with the function BITFAM_activities(). We then used a random forest model to identify the cell-type specific TF activity profile. Cluster heatmap of TFs was generated using Pheatmap R package.

Spatial data processing

Raw sequencing reads from spatial experiments were processed using the standard pipeline for fresh frozen tissues provided by Space Ranger software (v.2.0.1) specifically developed for Visium spatial gene expression data processing. Alignment was performed using human transcriptome reference GRCh38. The obtained gene expression matrices were analysed with the R package Seurat (v.4.4.0). For each sample we obtained an average of 7000 UMI counts and 2500 detected genes. Sample quality controls included the evaluation of UMI and gene counts together with the cell mitochondrial read content. Sample complexity was assessed computing the log10GenesPerUMI metric. Mitochondrial read content mean was found to be lower than 15% of the total gene counts in each sample indicating a low percentage of damaged or lysed cells. Log10genesPerUMI values ranged from 0.90 – 0.95 suggesting a good complexity of our datasets. Data were normalized using the SCTranform() function implemented in Seurat. We then used CCA to perform multi-sample integration. We ran PCA analysis and UMAP projection of the integrated samples selecting the first 15 principal components. The number of PCA component threshold was selected according to the elbow plot. Non-Negative Matrix Factorization (NMF) dimensional reduction analysis was performed using the R package Singlet with the function RunNMF() while we used RunLNMF() function on the merged

spatial object to account for batch integration effects. To evaluate similarity between the different tissue sections we computed for each spot the similarity scores based on Spearman correlation using R package SingleR (v1.0.5) (de.method="wilcox").

Identification of tumor spots

We used inferCNV algorithm to study the copy number alteration profile of the tissue areas. We ran the analysis separately for each patient to mitigate any possible batch effects in the analysis. Outer edge areas were used as normalization reference. Infercnv was ran with the same parameters used to analyse single cell data. We identified tumor spots by the amplification of chromosome 7 or the deletion of chromosome 10. We assigned a chromosome region amplified or deleted if the associated infercnv scoring for that specific region was >1 or <1 respectively. For patient 1 we decided to be more stringent in selecting these thresholds since the scoring of the spots was markedly above or below these thresholds (> 1.02 and < 0.9 respectively).

Projection and deconvolution analyses

Tumor cells were projected onto the integrated GBM single cell data with the Seurat FindTransferAnchors() and TransferData() functions using the first 40 PCA components of the reference dataset. Projection results were orthogonally validated with SingleR (v.1.8.0). Deconvolution analysis was performed on tissue sections with the R package CARD (v. 2.1.0) using as reference the single cell datasets of GBM samples from Bhaduri (Bhaduri *et al*, 2020). We re-processed the scRNA-seq data using the standard Seurat workflow and assigned cluster identities with the markers provided by Bhaduri's work. The analysis was performed using the raw counts from each spatial sample. Moreover, to be more accurate, previously identified tumoral and normal spots were deconvoluted using respectively the tumoral and normal cells of the single cell dataset. Statistical t-tests were conducted to evaluate differences between deconvolution results using the function stat_compare_means() from the R package Ggpubr (v. 0.4.0).

Cell-Cell Communication Analysis

We used CellChat R package (v.1.6.0) to analyse ligand-receptor interactions among the identified cell types. Cellchat predicts the biological signalling communication between the cells by assigning a probability value to each interaction and performing a permutation test. The likelihood of cell-cell communication is computed combining gene expression and spatial information. As suggested by CellChat documentation, given the low sensitivity of spatial imaging technologies, we calculated the average gene expression per cell group

using 10% truncated mean with the function computeAveExpr(type = truncatedMeanm trim = 0.1). In ComputeCommunProb() we set scale.distance parameter.

Gene set and gene ontology enrichment analysis

We performed differential expression analyses between the samples using the Seurat function FindAllMarkers(). GSEA analysis was run using the R package fgsea with the following parameters: nperm=1000, minSize = 15, maxSize=500. Hallmarks gene sets were retrieved by https://bioinf.wehi.edu.au/software/MSigDB/human_H_v5p2.rdata. Gene Ontology and Disease Ontology enrichment was performed with ClusterProfile R package using parameters (v. 4.8.1) while for pathway enrichment we used ReactomePA R package (v. 1.42.0).

2.4 Results

Spatial resolution atlas of different areas of glioblastoma

To study how the spatial localization influences the phenotype of GBM cells, we analysed different tumor biopsies taken from 3 patients affected by GBM. The samples were obtained during surgery using navigation-guided sampling in combination with Gliolan fluorescence. For each patient we analysed four different tumor regions: a central tumor area with visible necrosis that we called “Core” region, an internal tumor area marked by a strong Gliolan fluorescence that we called “Inner edge”, a transition zone between the tumor and healthy tissue, marked by a weak Gliolan fluorescence signal and FLAIR hypersensitivity, which we called the “Outer edge" and finally, a distant area in the healthy tissue defined as the "Distal area" (**Figure 2.3 A**). The resected tumor samples were embedded in Optimal Cutting Temperature compound, cryosectioned, and stained with Hematoxylin and Eosin (H&E) (**Figure 2.3 B**). We obtained an atlas of 12 samples that were processed using ST. The detailed patient characteristics and molecular features of the respective samples are shown in **Table 1**.

The samples were processed according to the ST Visium protocol following the standard Seurat workflow (see Materials and Methods). Briefly, we performed a quality control step to remove low quality spots that likely contained dead cells (**Figure S1)**. We then integrated all the sample areas and applied NMF to capture the main differences. The cells were then visualized using a UMAP embedding and we identified the clusters to explore the transcriptional similarities between the areas. Our results showed that the Inner edge and Core regions tend to cluster together within each patient suggesting a strong transcriptional similarity. On the other hand, the Distal area and Outer edge regions cluster together, independently from the patient of origin (**Figure 2.3 C**). This finding was in line with our expectations, as the outer edge is likely characterized by low infiltration of tumor cells. Therefore, it is reasonable that these areas resemble the distal regions more closely.

It is interesting to note that the Core and Inner edge tumor regions show significant differences between the patients, suggesting a strong heterogeneity across the different patients even between the same tumor regions. On the other hand, this inter-patient heterogeneity was not observed in the distal and outer areas, which are characterized by a lower tumor content suggesting low inter-patient variability for these regions.

Finally, we evaluated the transcriptional similarity between the different areas by performing correlation analyses of the samples using Pearson coefficient. Doing this, we confirmed the previously observed results. Indeed, the outer edge of the tumor and the

distal areas were highly correlated independently from the patients (Pearson coefficient approximately 1), while on the other hand, Inner edge and Core areas were highly correlated only between the same patients (**Figure 2.3 D**).

To identify the tumor cells in our datasets, we inferred the CNV profile from the transcriptome of each spot, using the distal areas as reference. As expected, the Core and the Inner edge regions of all the patients showed a strong signal of chromosome 7 amplification and the chromosome 10 deletion that are the most common CNVs of GBM cells. On the contrary, the outer edge regions showed a limited number of spots with CNVs, confirming that this area is made by a mixture of tumor and normal cells (**Figure 2.4 A-C left panels**). Of note, we also observed less common CNVs that are patient-specific, such as the deletion of chromosome 9p in patient 1, and the deletion of chromosome 13 and 19 in patient 2. We retrieved from this analysis the associated scores of chromosomal aberrations and classified the spots as tumorigenic those with values higher than 1 in the region 7q and values lower than 1 in the region 1p (**Figure 2.4 A, B, C right panels and D and Figure S3 and Figure S4**).

SAMPLE NUMBER	SEX	Gliolan	HISTOPATHOLOGY	IDH	MGMT	Ki67	ATRX	p53
PZ 1	M	Yes	GBM grade IV	WT	NT	25%	WT	+/-
PZ 2	M	Yes	GBM grade IV	WT	Unmet	25%	NT	80%
PZ 3	M	No	Giant-cell GBM grade IV	WT	Unmet	20%	NT	90%

Table 1. Patient demographics and clinical information for ST cohort.
In this table, we collect some relevant information about the patients belonging to the ST cohort. Specifically, we collected some demographic information such as sex. We recorded the intraoperative use of Gliolan, the confirmed histopathology of the tumor and some molecular information of the samples (e.g. IDH mutation status, MGMT methylation status, ki67 percentage of positive cells, ATRX mutation status and p53 percentage of mutated cells). WT= Wild type; NT= Non-tested; Unmet= Unmethylated.

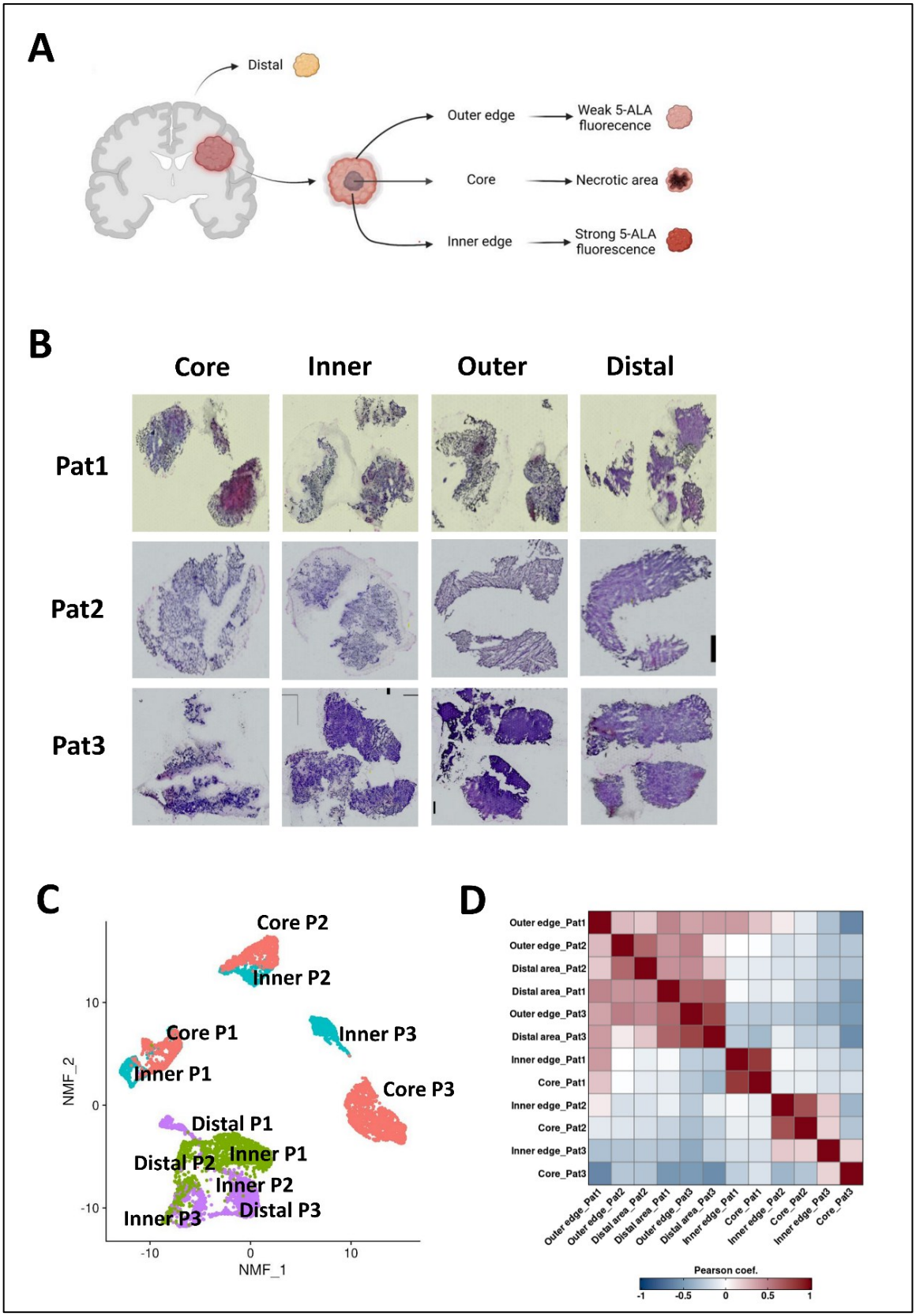

Figure 2.3 Overview of tumor region sampling and analysis in GBM patients.
(A) *Scheme of the four tumor sampled regions: the Core, the Inner edge, the Outer edge and the Distal area.* ***(B)*** *H&E staining images of the Core, Inner edge, Outer edge, and Distal area for three different patients.* ***(C)*** *UMAP embedding showing the clustering of the samples based on transcriptional similarities. Core and Inner*

edge samples cluster together within patients, while Outer edge and Distal area samples group together across patients. ***(D)*** *Heatmap showing Pearson correlation analysis.*

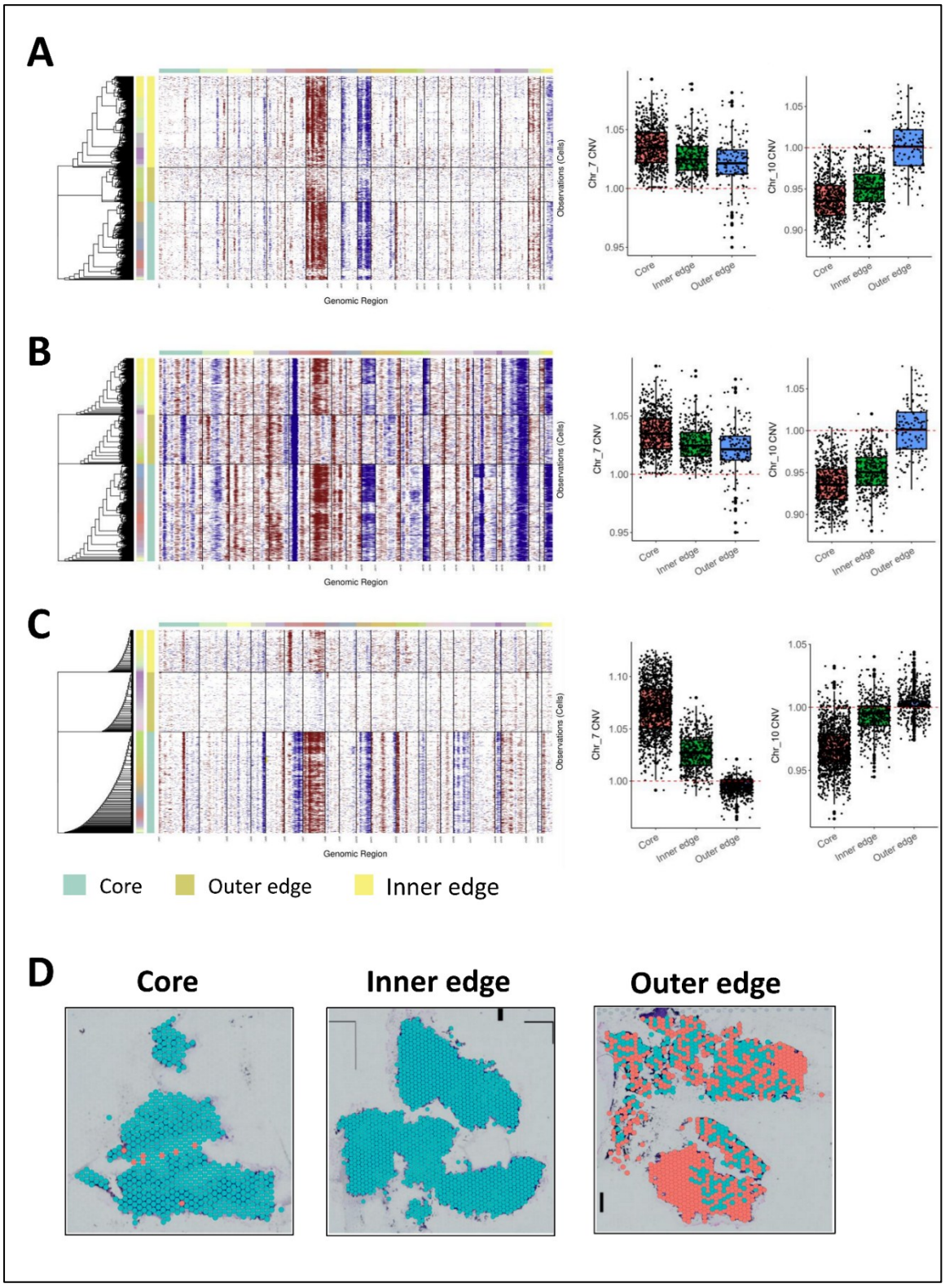

Figure 2.4 CNV profiling and spatial localization of tumor cells of the different tumor regions. ***(A-C)*** *Heatmaps representing inferred CNVs for chromosome 7 amplification (7amp) and chromosome 10 deletion (10del) in three GBM patients. Each panel shows the CNV profiles of the core, inner edge, and outer*

edge regions of the tumor. The right panels display box plots of the CNV scores across the three regions, with values above 1 indicating amplifications and values below 1 indicating deletions. ***(D)*** *Spatial mapping of tumor spots (in red) of the core, inner edge, and outer edge areas (from left to right). The core and inner edge are predominantly composed of tumor cells with 7amp and 10del, while the outer edge shows a mixture of normal and tumor cells, confirming the heterogeneous nature of this region.*

Next, we attempted to characterize the identity of tumor spots in our tissues. However, ST experiments are limited by their intrinsic low resolution because each spot can contain up to 8 cells. This implies that studying gene expression directly on tissues can be challenging as the signal could be confounded by the presence of different cells within the same spot. To overcome this limitation imposed by the technique, we decided to perform a deconvolution analysis of the spot signals using an external single-cell dataset to identify the different cell populations. For this analysis we used a recently published scRNA-seq dataset made by 11 GBM tumor samples (Bhaduri *et al*, 2020). To improve the accuracy of the analysis, deconvolution of the tumor spots was performed using tumor cells and the one of the normal spots using normal cells from Bahduri's dataset.

The most abundant cell types identified in the normal spots included radial glia, glycolytic progenitors, mixed progenitor neurons, endothelial cells, and dividing progenitors (**Figure 2.5 A**). Moreover, we observed a high presence of endothelial cells in the outer edge of the tumor, indicating an increased vascularization in the peripheral areas. This could suggest that these areas can better support the tumor by providing nutrients and oxygen.

On the other hand, the tumor spots showed different cell composition depending on the tumor sampling region. In the tumor core and inner edge areas, we observed the predominance of AC-like (astrocytic-like) and MES-like (mesenchymal-like) cells, which are often associated with increased tumor aggressiveness and resistance to therapy. In contrast, in the outer edge region we observed a strong enrichment of OPC-like tumor population (**Figure 2.5 B-C and Figure S5**).

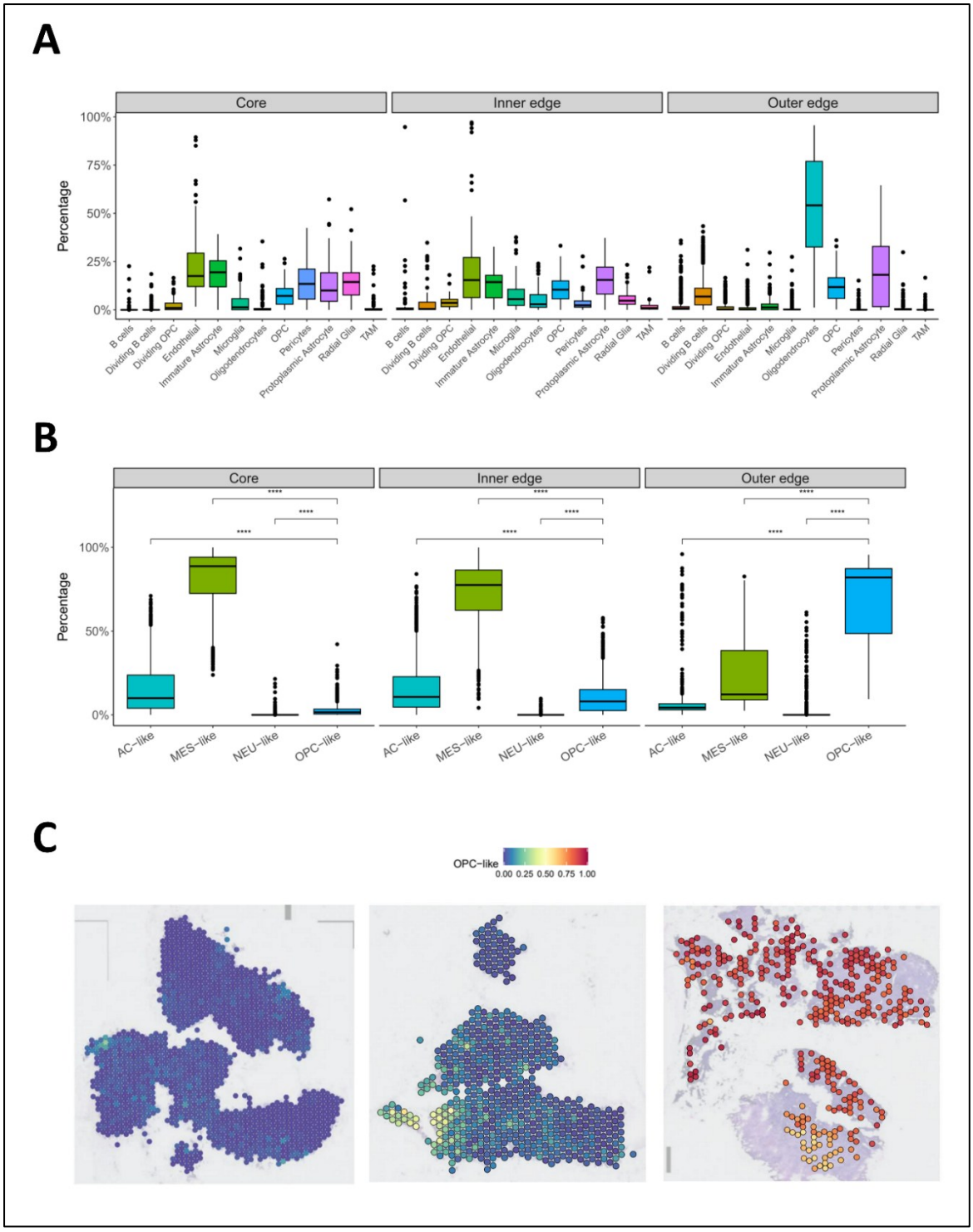

Figure 2.5 Deconvolution of cell types in GBM tissues.
(A) *Deconvolution of normal spots reveals the presence of radial glia, glycolytic progenitors, and endothelial cells, with higher endothelial cell abundance in the outer edge area, suggesting an increased vascularization.* ***(B)*** *Deconvolution of tumor spots showing AC-like and MES-like cell signatures dominating the core, while OPC-like signature is enriched in the outer and inner edges, suggesting a role in tumor proliferation.* ***(C)*** *OPC-like cells signature distribution (from left to right) in the core, inner edge and outer edge regions. The figure highlights the higher abundance of OPC-like cells in the outer edge region.*

To further validate our results and provide a detailed characterization of GBM tumor cell states, which has been shown to be an important feature of GBM, we built a reference dataset of GBM tumor cells integrating 38 scRNA-seq GBM samples from three different studies (Wang *et al*, 2022; Couturier *et al*, 2020; Neftel *et al*, 2019). After standard quality control preprocessing (**Figure S6**), data were filtered and integrated using CCA to mitigate technical variability between samples and exclude any batch effects (see Materials and Methods). We reduced the data dimensionality using PCA followed by projection onto a two-dimensional embedding UMAP. Cluster identities were assigned based on the expression profiles of a reference panel of marker genes. We identified a cluster of T cells (IL7R+, CD52+), macrophages (CD68+, CD14+), fibroblasts (COL1A1+, DCN+), oligodendrocytes (APOD+, PLP1+). We then identified a large cluster of cells with uncertain identity that we named Unknown that exhibited a mix of gene signatures typical of mesenchymal cells, oligodendrocytes, and astrocytes signatures (**Figure 2.6 A**). Based on this, we hypothesized that it consisted of tumor cells. To validate our hypothesis, we inferred the CNV profile from the expression data (**Figure 2.6 B**). As done previously for our spatial dataset, we annotated tumor cells based on the presence of chromosome 7 amplification and chromosome 10 deletion (**Figure 2.6 C**). As expected, tumor cells identified through CNV analysis were predominantly found within the “Unknown” cluster.
After the isolation of the tumor cells, we reprocessed them by performing PCA and UMAP reductions to study more deeply the different cellular subtypes. The stem cell compartment was identified as the one showing the highest predicted stemness score and by the expression of the typical stemness markers (**Figure S7**)**.** Moreover, the other identified branches were assigned to each GBM state: OPC-like, MES-like, NPC-like, and AC-like using the gene signatures provided by Neftel (**Figure 2.7 A-B and Figure S8**) (Neftel *et al*, 2019). We then performed pseudotime analysis to confirm that the differentiation trajectory originates from the stem cell compartment and branches into the four major GBM subtypes and we isolated the set of genes that drive the differentiation into each of these four branches (**Figure 2.7 C**). These results were also supported by RNA velocity analysis that provides information on the future states of individual cells. The velocity vectors show differentiation moving outward from the stem cell population toward the four subtypes, supporting the idea that the stem cell compartment is the starting point for GBM cell differentiation (**Figure 2.7 D**). To further confirm the differentiation of GBM cells, we generated cellular states plot where cells are represented on a transcriptional space built on the key gene signatures of the associated GBM states. This plot shows how cells were positioned in relation to the expression of these four signatures. Stem cells, shown in black, are centrally located in line with their undifferentiated nature. On the contrary, the

differentiated cell types are located on the four corners, reflecting their strong enrichment in one of the four key GBM signatures (**Figure 2.7 E**).

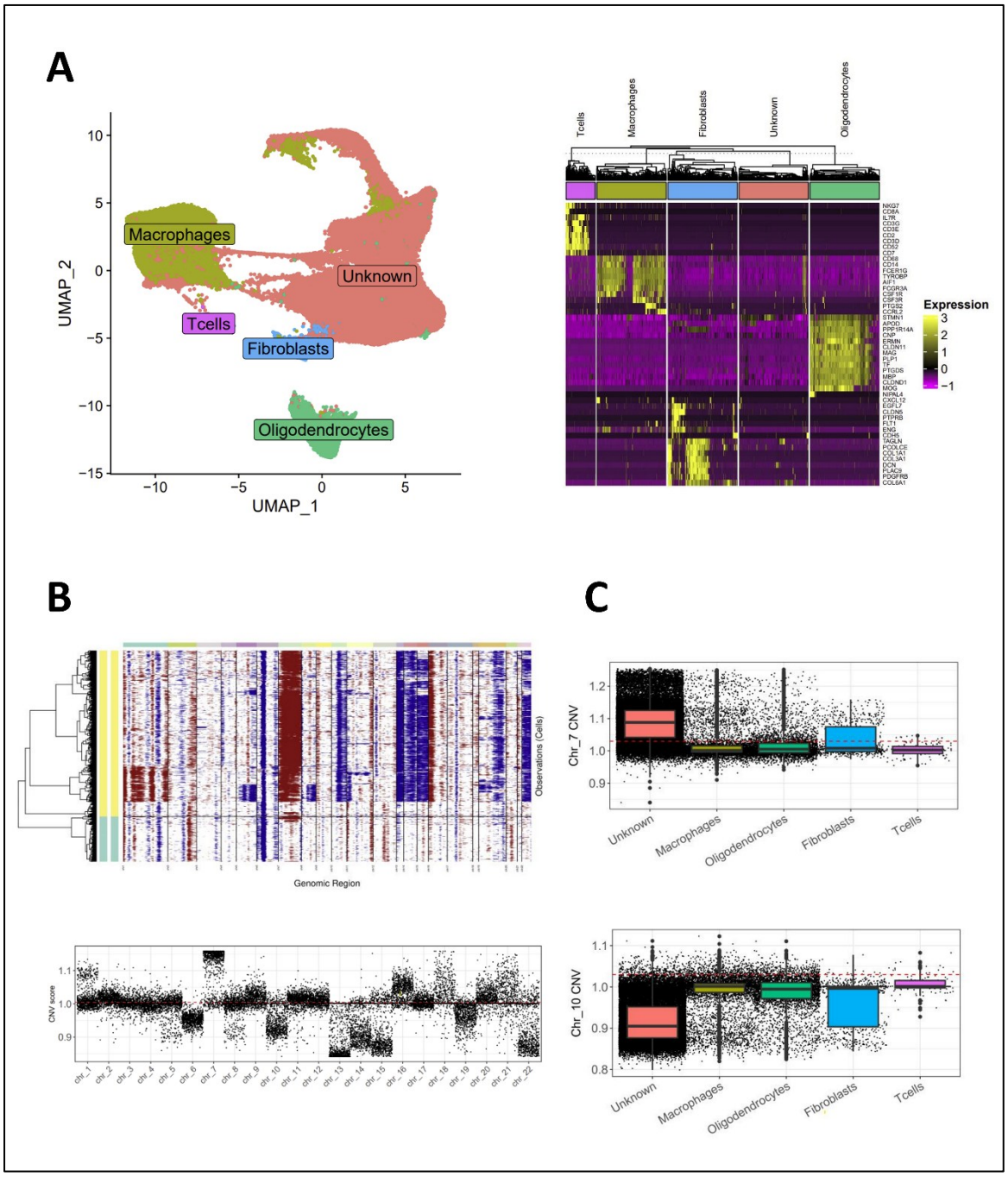

Figure 2.6 Characterization of Glioblastoma Tumor Cell Clusters through scRNA-seq Integration and CNV Inference.

(A) *Left: UMAP of GBM tumor cells from the integration of 38 scRNA-seq samples, Right: Heatmap showing marker genes used to annotate each cluster.* ***(B)*** *CNV profiles of the different clusters inferred from expression data highlighting the main chromosomal aberrations found in each chromosomal region (red is used for amplification, blue for deletion).* ***(C)*** *The associated scoring for Chr7 amplification and Chr10 deletion confirms the tumorigenic nature of cells in the "Unknown" cluster.*

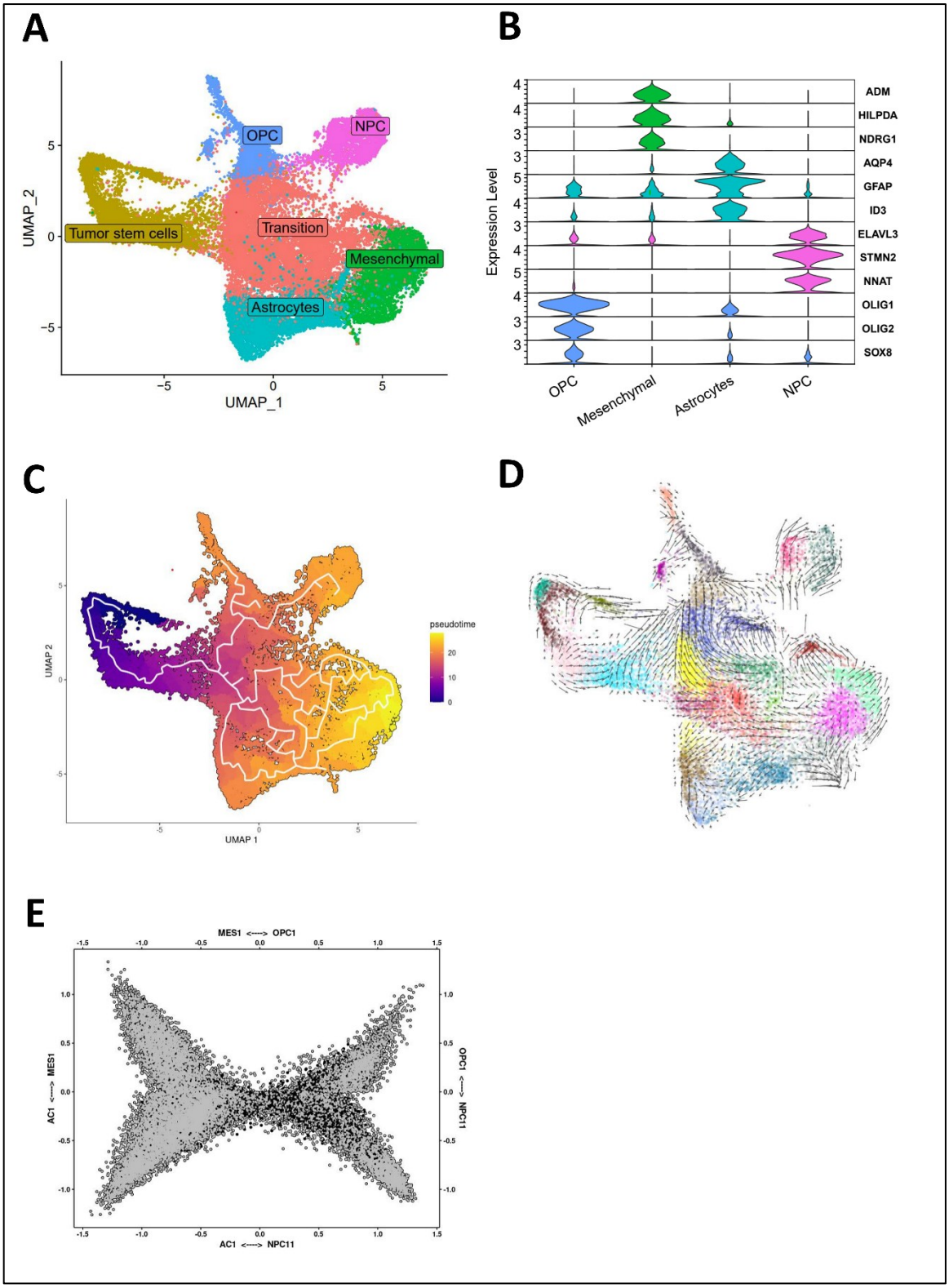

Figure 2.7 Tumor cell subtype distribution in GBM single cell datasets.
(A) *UMAP plot showing five main branches corresponding to the stem cell compartment and the 4 GBM subtypes: OPC-like, MES-like, NPC-like, AC-like.* ***(B)*** *Annotated clusters confirming the identity of these subtypes based on the expression of known markers.* ***(C)*** *Pseudotime analysis illustrates the differentiation trajectory from stem cells into the four subtypes.* ***(D)*** *RNA velocity plot with arrows indicating the predicted future states of cells, showing differentiation from the stem cell population towards the four subtypes.* ***(E)*** *Cellular States Plot representing GBM signatures (OPC-like, MES-like, NEU-like, AC-like). Stem cells, shown*

in black, are centrally located, underscoring their undifferentiated nature and potential to differentiate into the surrounding subtypes. This plot further confirms the distinct identities and differentiation states of the GBM cell populations.

At this point, we tried to analyze the distribution of cancer cell states in our spatial atlas. To do this, we projected the tumor spatial spots identified in each tumor area onto the PCA structure of our GBM reference single-cell dataset. In accordance with our previous results, we found that cancer spots in the core and inner edge mostly resembled the AC-like and MES-like cell states, whereas the OPC-like state was enriched in cancer spots found in the outer edge region (**Figure 2.8**). These results were also confirmed by correlation analyses using an external single cell reference dataset from the study of Nowakowski et al, 2017 (**Figure S9**).These findings suggest that cancer cells within the infiltrative area seem to differentiate towards the oligodendrocyte lineage.

We then decided to investigate the ligand-receptor interactions between the different identified cell types in the outer edge regions. For this analysis, we selected the outer edge tissue from patient 3. To assign the identities to the spots, we used the previously deconvolution results obtained from the Bhaduri dataset. Since the deconvolution analysis provides a probability score for each cell type, we assigned the identity to each spot based on the highest score. To obtain better results, for the normal spots we focused only on the five most relevant cell types that can be found in brain tissue: astrocytes, macrophages, oligodendrocytes and microglia cells, while for the tumor spots we considered all the four tumor signatures (OPC-like, AC-like, NEU-like, and MES-like) (**Figure 2.9 A-C**).

We then analyzed the interactions between all the analyzed cell types in terms of receptor-ligand expression focusing our attention on the interactions between OPC-like tumor cells and the normal cells in the microenvironment. Interestingly, we found that tumor cells release important ligands to the surrounding microglia cells belonging to the GRN, TGFβ, and MK signalling pathways. These pathways are known for playing an important role in the modulation of the tumor microenvironment response mediating immunosuppressive mechanisms (**Figure 2.9 D-G**). TGFβ is well-known for its ability to suppress the immune system by inhibiting cytotoxic T cells, promoting regulatory T cells, and reducing the infiltration of immune cells in the tumor area (Kim *et al*, 2021; Ma *et al*, 2024). Similarly, GRN (Granulin) contributes to inflammation and tumor progression by creating a favourable environment for tumor growth and modulating immune responses (Elkabets *et al*, 2011). MK (Midkine) plays a role in cell survival and angiogenesis, promoting tumor development and supporting immune evasion (Hu *et al*, 2021). Together, these pathways are important to allow cancer cells to influence the surrounding microenvironment to evade immune detection and promote their own survival.

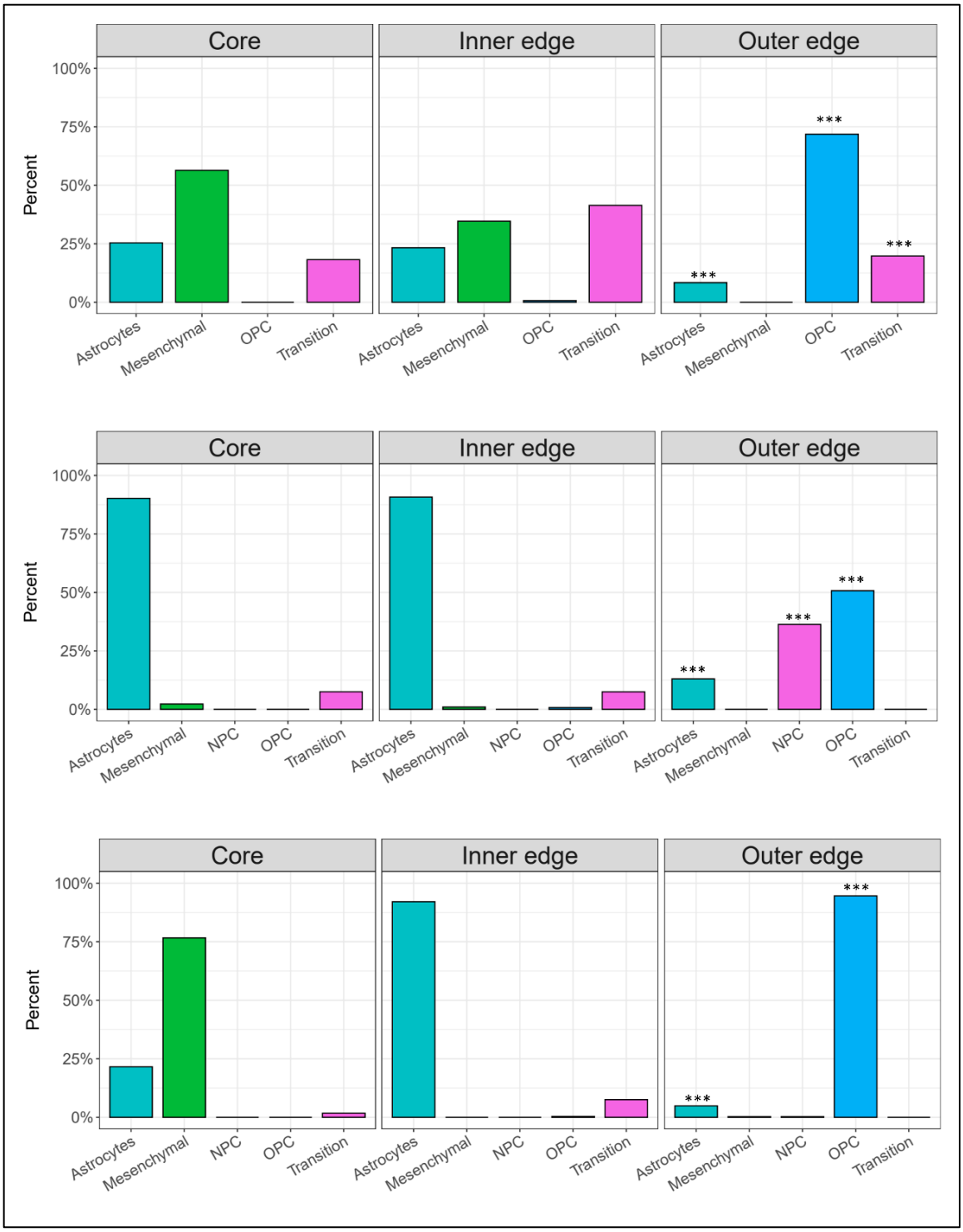

Figure 2.8 Distribution of the different tumor cell populations in the different tumor areas.
Each row represents a different patient (from the top to the bottom), patient 1, 2 and 3. Enrichment of MES-like and AC-like populations in the core and inner edge, with a significant prevalence of OPC-like cells in the outer edge.

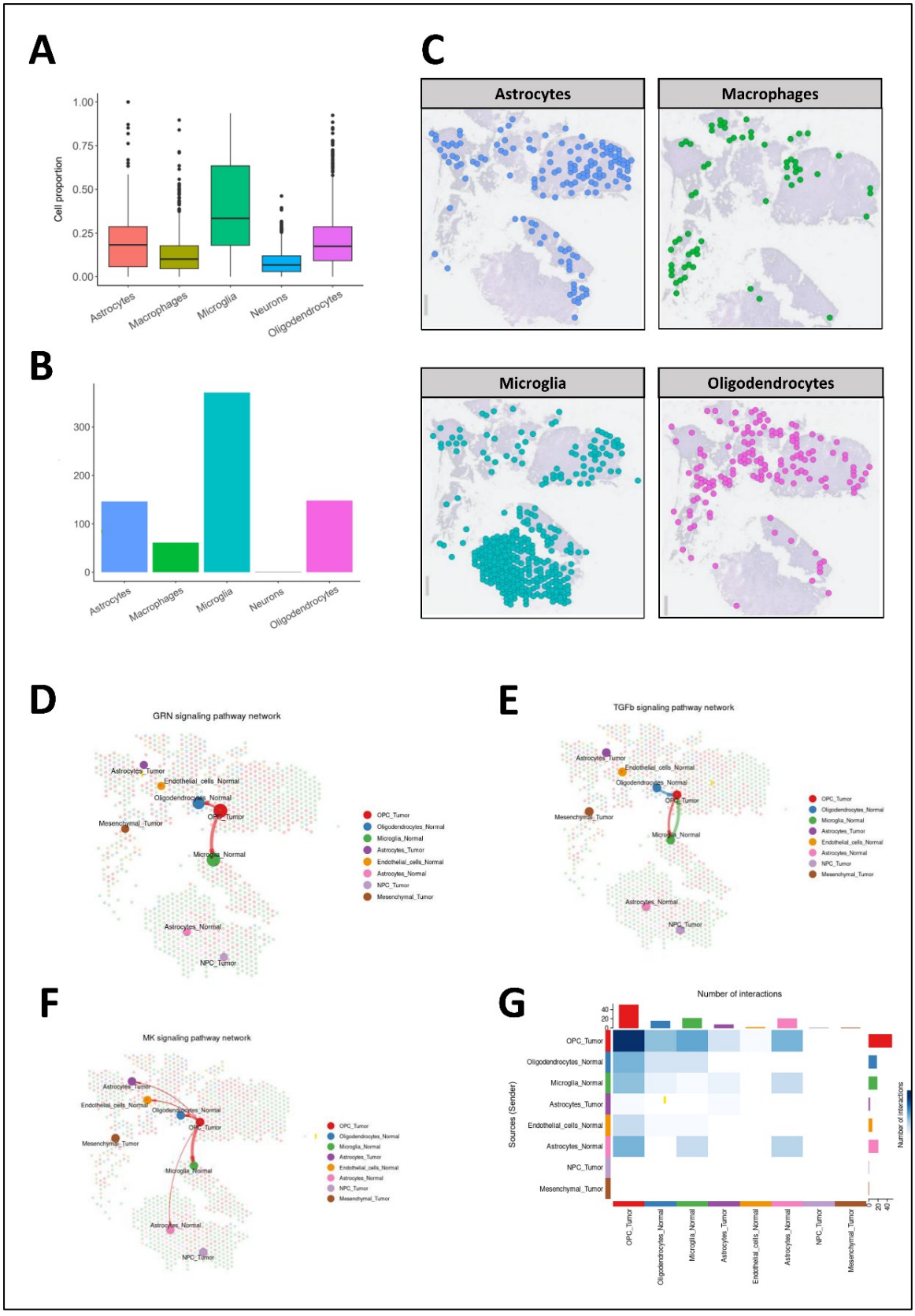

Figure 2.9 Cell type interactions and signalling pathways in the outer edge tissue of GBM.
(A) *Boxplot representing the scoring of the deconvoluted five key brain cell types (astrocytes, macrophages, microglia, neurons, and oligodendrocytes) identified in the outer edge tissue sample from patient 3.* ***(B)*** *Bar chart shows the total number of identified cells for each of the five cell types in the outer edge region.* ***(C)*** *Spatial distribution of astrocytes, macrophages, microglia, and oligodendrocytes in the outer edge tissue.* ***(D-***

***F)**: Network representations of key ligand-receptor interactions between OPC-like tumor cells and microglia. The GRN (D), TGFβ (E), and MK (F) signalling pathways are shown, highlighting their roles in immune suppression (TGFβ), cell survival, and angiogenesis (MK). **(G)** Heatmap illustrating the number of ligand-receptor interactions across various normal and tumor cell types, with a focus on the key signalling pathways (GRN, TGFβ, and MK).*

To better confirm our findings, we decided to perform gene enrichment analysis across the different tumor areas. GSEA analysis performed on the upregulated genes (fold change > 1) of the outer region compared to the core and inner edge areas showed a significant enrichment of signatures related to oligodendrocyte development and differentiation. In particular the enriched gene sets "oligodendrocyte differentiation" and "neuron projection development" support the presence of oligodendrocyte-like tumor cells in the outer regions (**Figure 2.10 A-C**). We also performed GSEA analysis on the upregulated genes in the core with respect to the inner edge area. GSEA results showed enrichment of pathways associated with "hypoxia," "apoptosis," and "inflammatory response," suggesting that the core region is under significant metabolic and environmental stress (**Figure 2.10 D**). Moreover, gene ontology (GO) terms showed an upregulation of processes such as "oxidative phosphorylation," "ATP metabolic processes," and "response to hypoxia" in the core, confirming the hypoxic nature of this region and its impact on the cellular microenvironment (**Figure 2.10 E**).

Finally, we decided to validate our results using two external datasets from two studies: a ST dataset from Greenwald and a scRNA-seq dataset from Darmanis (Greenwald *et al*, 2024; Darmanis *et al*, 2017). Greenwald dataset includes samples from various tumor regions: the tumor, the necrotic region, the T1 region and from the infiltrative area. We performed cell spots deconvolution using the previously defined signatures in our reference scRNA-seq GBM dataset and confirmed that bulk and T1 areas are enriched of mesenchymal and astrocytic signatures whereas the infiltrative area showed a significant enrichment of OPCs, consistent with our findings (**Figure 2.11 A**). We also decided to validate our results using a scRNA-seq dataset from Darmanis (Darmanis *et al*, 2017). This dataset includes samples from the tumor region, the periphery and the distal areas. In this dataset, the tumor region showed a predominance of AC-like and MES-like subpopulations. On the contrary, the OPC-like subpopulation was mainly found in the infiltrative area (**Figure 2.11 B**). By combining data from both ST and scRNA-seq datasets, we could confirm the enrichment of the OPC cluster in the infiltrative areas of GBM.

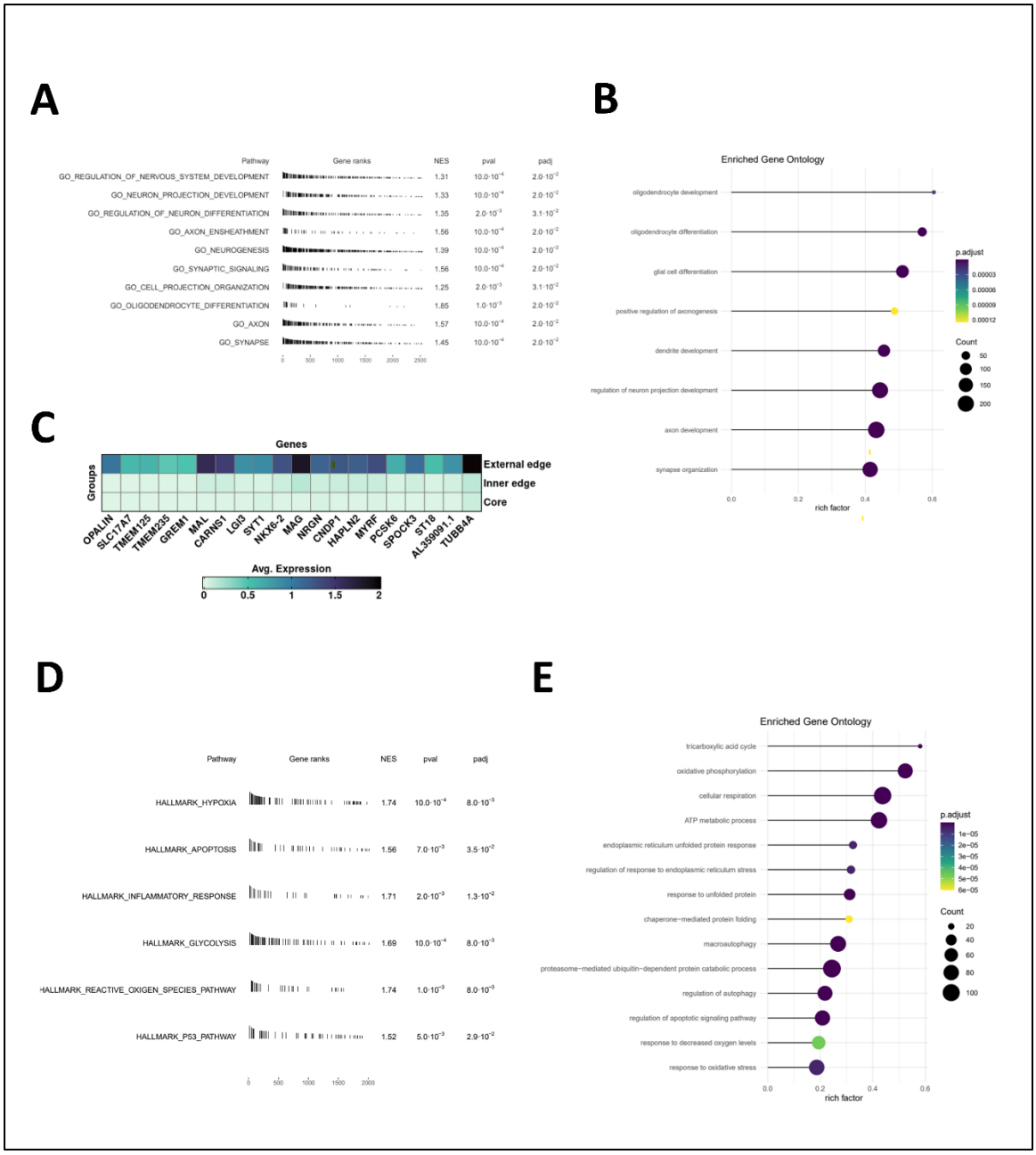

Figure 2.10 Gene enrichment analysis reveals distinct biological processes across different tumor regions.

(A) *Gene Set Enrichment Analysis (GSEA) performed on the outer edge versus core and inner edge regions highlights significant enrichment in pathways related to oligodendrocyte development and differentiation, including "oligodendrocyte differentiation" and "neuron projection development."* ***(B)*** *Gene Ontology (GO) analysis confirms enrichment of processes such as "oligodendrocyte differentiation" and "synapse organization" in the outer edge, supporting the presence of oligodendrocyte-like tumor cells in this region.* ***(C)*** *Heatmap showing the expression of representative genes enriched in the outer edge compared to core and inner edge regions.* ***(D)*** *GSEA of the core region versus inner edge reveals significant enrichment of hypoxia-associated signatures, including pathways like "hypoxia," "apoptosis," and "inflammatory response," indicating metabolic and environmental stress in the core.* ***(E)*** *GO analysis of the core highlights processes such as "oxidative phosphorylation," "ATP metabolic processes," and "response to hypoxia," confirming the hypoxic microenvironment and its influence on cellular functions.*

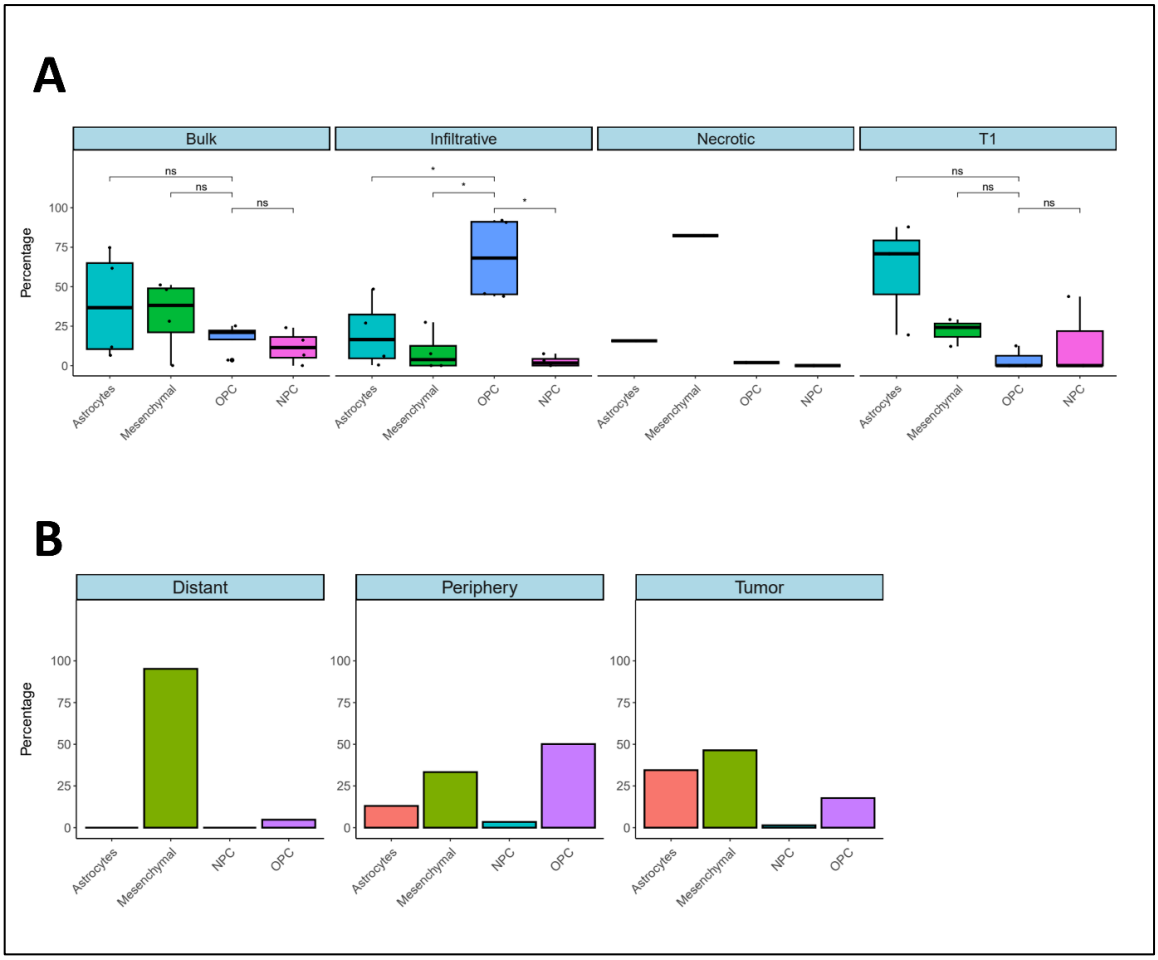

Figure 2.11 Validation of OPC enrichment in infiltrative areas using external ST and scRNA-seq datasets.
(A) *Deconvolution analysis using GBM reference scRNA-seq signatures from Greenwald ST dataset revealed a predominance of mesenchymal and astrocytic clusters in the bulk and T1 areas, while the infiltrative region showed a significant enrichment of OPC-like clusters, consistent with our results.* ***(B)*** *Validation using the single-cell RNA sequencing dataset from Darmanis showed a predominance of AC-like and MES-like subpopulations in the distant and tumor areas. In contrast, the OPC-like cluster was predominantly observed in the infiltrative area at the tumor's periphery.*

2.5 Discussion

In this study, we explored the complexity of GBM using ST. We analyzed GBMs samples from three different patients, and for each patient, we analyzed four different tumor areas to investigate how spatial organization influences the nature and identity of tumor cells. The sampled regions included the "Core" (central necrotic area), the "Inner edge" (characterized by strong Gliolan fluorescence), the "Outer edge" (a transitional zone with low fluorescence and FLAIR sensitivity), and the "Distal area" (healthy tissue distant from the tumor). This approach allowed us to create a high-resolution ST atlas of the different tumor regions.

ST represents a cutting-edge technology that allows the study of the gene expression directly on the tissue, preserving the spatial organization. This gives the possibility to analyze the dynamics of the cells in their natural context. In recent years, this technique has been widely used to explore the spatial heterogeneity of GBM allowing researchers to study the expression of specific gene sets or molecular signatures in different tumor regions. This has significantly contributed to our understanding of the structural and functional complexity of this malignancy. However, ST suffers from an intrinsic limitation of low resolution. Indeed, each spot can contain up to 8 cells, making it challenging to analyze gene expression at the single-cell level. This limitation makes more complicated the accurate identification of cellular populations within the samples which potentially leads to a loss of detail and introduces a significant bias in the characterization of the cellular composition in complex regions like GBM. To overcome the limitations of ST, we applied a deconvolution technique that uses an external scRNA-seq dataset to infer the cellular composition of each individual spot. This approach gave us the possibility to have a clearer and more accurate view of the cellular composition of the tumor and its interactions with the surrounding microenvironment, which wouldn’t have been possible with ST data alone. This analysis revealed a significant presence of OPC-like cells (oligodendrocyte progenitor cells) in the outer edge region, where they make up most tumor cells. These cells are found in the areas where the tumor invades healthy tissue, suggesting that they play a key role in GBM infiltration by supporting tumor spread into surrounding tissue. Previous studies have shown that OPC-like cells are one of the main cell populations in GBM known for their strong cellular plasticity (Neftel *et al*, 2019). Similarly, *Bhaduri et al. (2020)* and *Couturier et al. (2020)* have shown the resilience and adaptability of OPC-like cells in GBM. This plasticity allows OPC-like cells to adapt to different tumor environments and contribute to tumor progression which makes them particularly dangerous in terms of tutor aggressiveness and invasiveness (Neftel *et al*, 2019). For these reasons, we suppose that

OPC-like cells, due to their plasticity and stem cell-like properties, may play a key role in driving tumor recurrence and resistance to treatment. This hypothesis is reasonable if we consider that these cells are located in the most distal regions of the tumor which makes them less likely to be fully removed during surgical interventions. Moreover, this location can allow them to evade therapies and can contribute to tumor regrowth. These findings are in accordance with some studies which support the role of these cells in promoting tumor progression and recurrence (Filbin *et al.*, 2018).

The strength of our results is supported by external validation using both ST datasets from Greenwald and scRNA-seq datasets from Darmanis (Greenwald *et al*, 2024; Darmanis *et al*, 2017). Indeed, both studies confirmed the higher presence of OPC-like cells in the outer edge of GBM.

The results from Gene Ontology (GO) enrichment analysis further support our data, showing that tumor cells present in the outer edge are significantly enriched for genes involved in oligodendrocyte differentiation. Among these, OPALIN, MAG, and TNR. OPALIN is essential for oligodendrocyte differentiation and supports the processes of myelination and neuronal maturation. MAG (Myelin-Associated Glycoprotein) stabilizes the interactions between oligodendrocytes and axons, while TNR (Tenascin R) modulates the extracellular environment, promoting cellular adhesion and migration.

In the core and inner edge of GBM, deconvolution data showed a predominance of mesenchymal (MES-like) and astrocytic (AC-like) cells in these areas with a similar distribution. These cell populations, known for their aggressiveness and therapy resistance, are in line with the findings of Greenwald et al. (2023) and Couturier et al. (2020) that highlight the importance of mesenchymal cells under hypoxic conditions.

Our Gene Set Enrichment Analysis (GSEA) confirmed in these areas the activation of pathways such as hypoxia response, apoptosis, and inflammation. This result is in line with previous findings that hypoxia in GBM involves poor vascularization and metabolic stress, driving tumor aggressiveness and therapy resistance (Neftel *et al*., 2019); This result supports the idea that the GBM core is under significant environmental stress. This could explain the reason why cells in this region activate compensatory pathways, including oxidative phosphorylation and ATP-related metabolic processes, that support tumor adaptation and progression.

An additional interesting result of our study regards the interactions occurring between OPC-like tumor cells and the cells in the surrounding environment, especially microglial cells. According to our data, tumor cells seem to release molecules that can activate specific signaling pathways in the microglial cells, including the GRN (Granulin), TGFβ

(transforming growth factor beta), and MK (Midkine) pathways. These signaling pathways are known to play a key role in the modulation of the immune environment by suppressing immune responses and creating favorable conditions for tumor growth. The TGFβ pathway is well-known for its ability to counteract the anti-tumor immunity. Indeed, it has been shown that TGFβ can reduce immune cell infiltration into the tumor mass by inhibiting cytotoxic T cells and promoting regulatory T cells (Batlle & Massagué, 2019). Doing so, it contributes to tumor progression and immune resistance. Similarly, Granulin (GRN) pathway plays an important role in enhancing inflammation and in the modulation of the immune responses, creating a pro-inflammatory tumor environment that helps GBM progression (Saeedi-Boroujeni *et al*, 2023). Another key pathway we identified in our analysis is the Midkine (MK) pathway that is associated with a protein involved in cell survival and blood vessel formation. MK supports tumor development by promoting the growth of new blood vessels and enabling immune evasion, creating conditions that allow tumor proliferation and resistance to targeted therapies (Filippou *et al*, 2019).

All these pathways highlight the complex relationship between OPC-like tumor cells and the immune environment in GBM. The tumor cells not only evade immune activity but also actively manipulate the surrounding environment to support their growth and survival. Specifically, the ability of OPC-like cells to interact with microglia through the release of immunosuppressive signals suggests that this cell population could be crucial for the tumor's progression and invasiveness in peripheral areas.

In conclusion, this study provides an advancement in understanding the spatial differences in GBM. By combining ST with scRNA-seq, we could overcome the resolution limits imposed by ST and we could map with more accuracy the cell identities in the different tumor areas. This innovative approach revealed OPC-like cells as the dominant population in the outer edge, suggesting their key role in GBM progression and recurrence.

In the future it would be interesting to investigate more deeply the role of OPC-like cells, especially in tumor recurrence, characterizing these cells at the molecular level. Moreover, it would be important to investigate the interactions occurring between these cells and the surrounding environment, focusing in particular on the interactions with the immune cells. This could lead to the development of new therapeutic options able to target the tumor cells or modulate their interactions with immune cells of the microenvironment.

2.6 Supplementary figures

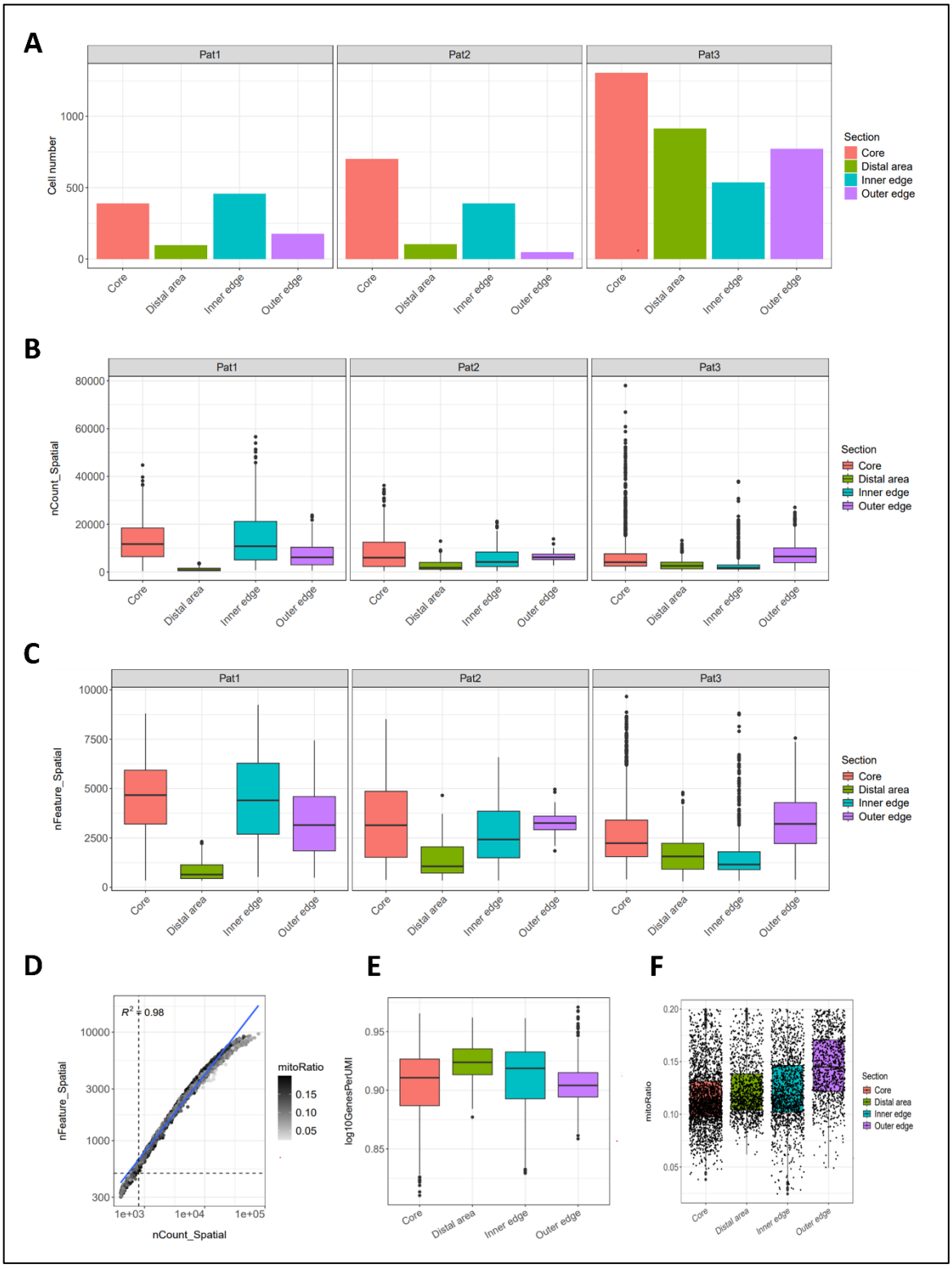

Figure S1 Quality control metrics for ST data
The figure shows standard metrics to access the quality of our spatial samples: distribution of spot counts ***(A)****, transcript counts (nCount_Spatial)* ***(B)****, features counts (nFeature_Spatial)* ***(C)****, correlations between*

nFeature_Spatial and nCount_Spatial ***(D)****, log-transformed genes per UMI (log10GenesPerUMI) across tumor sections* ***(E)****. mitochondrial RNA content (mitoRatio)* ***(F).***

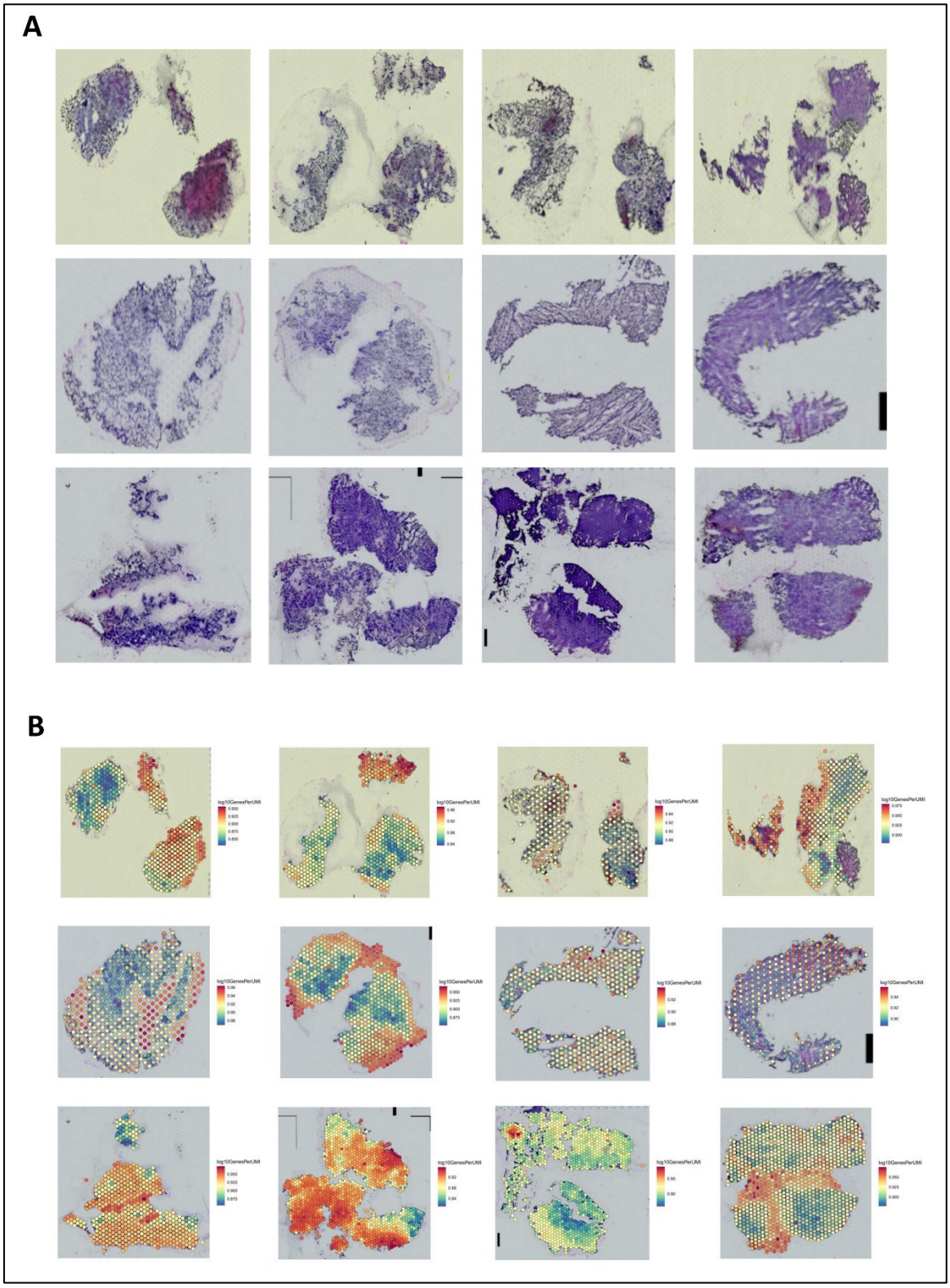

Figure S2 Histological Analysis and Gene per UMI ratio.
(A) *Histological sections of tumor samples stained with hematoxylin and eosin (H&E).* ***(B)*** *Spatial mapping of transcriptional activity on the tissues using the log10GenesPerUMI metric which reveals spatial heterogeneity.*

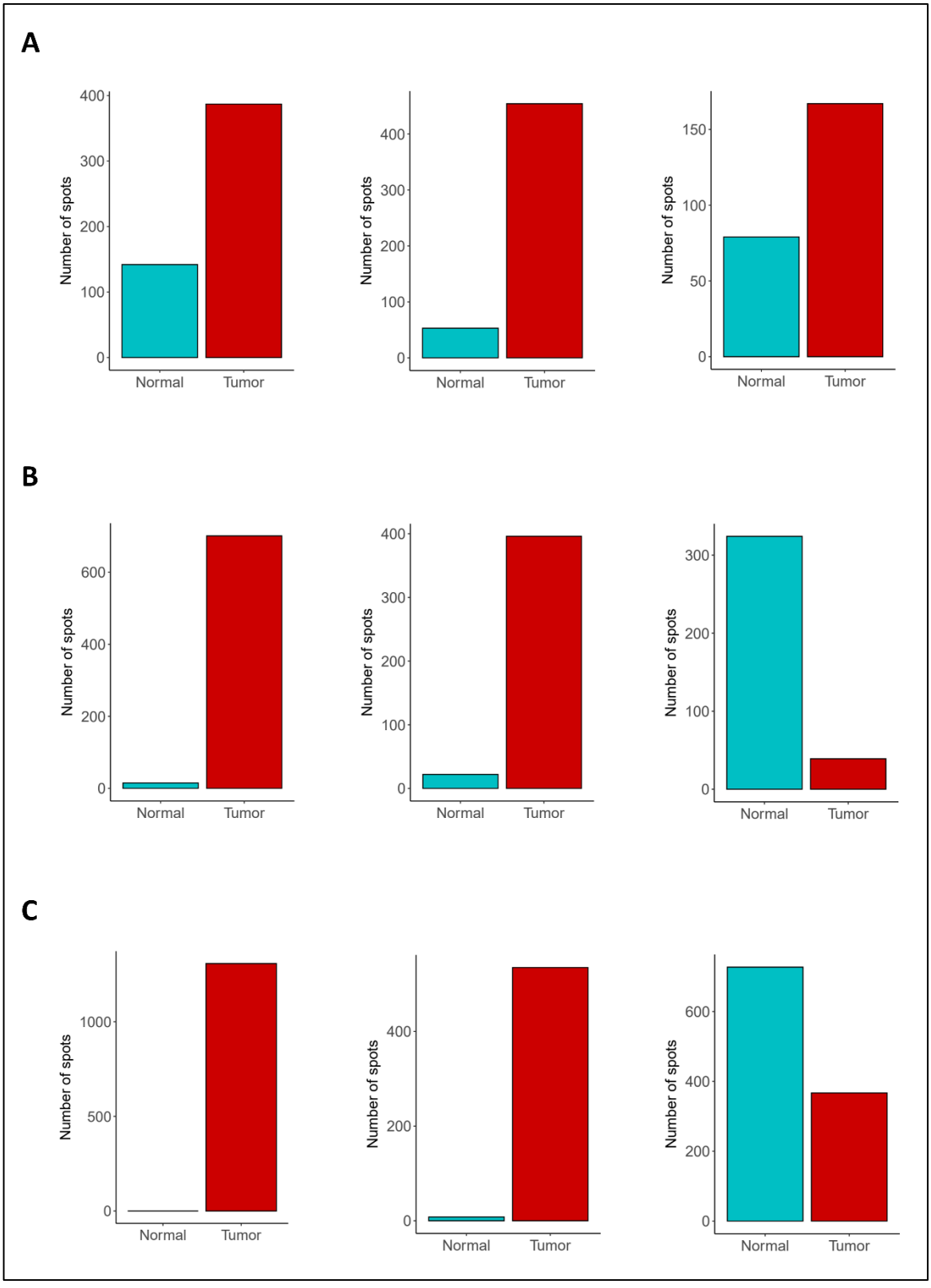

Figure S3 Number of Spots Identified as Tumor or Normal Across Tumor Sections.
The figure shows the distribution of spots classified as tumor or normal for three patients, with each row corresponding to one patient: ***(A)*** *distribution in patient 1,* ***(B)*** *distribution in patient 2,* ***(C)*** *distribution in patient 3. For each patient, the columns represent different tumor sections: Core (left), Inner edge (center), and Outer edge (right). The bar plots show the number of spots identified in each section.*

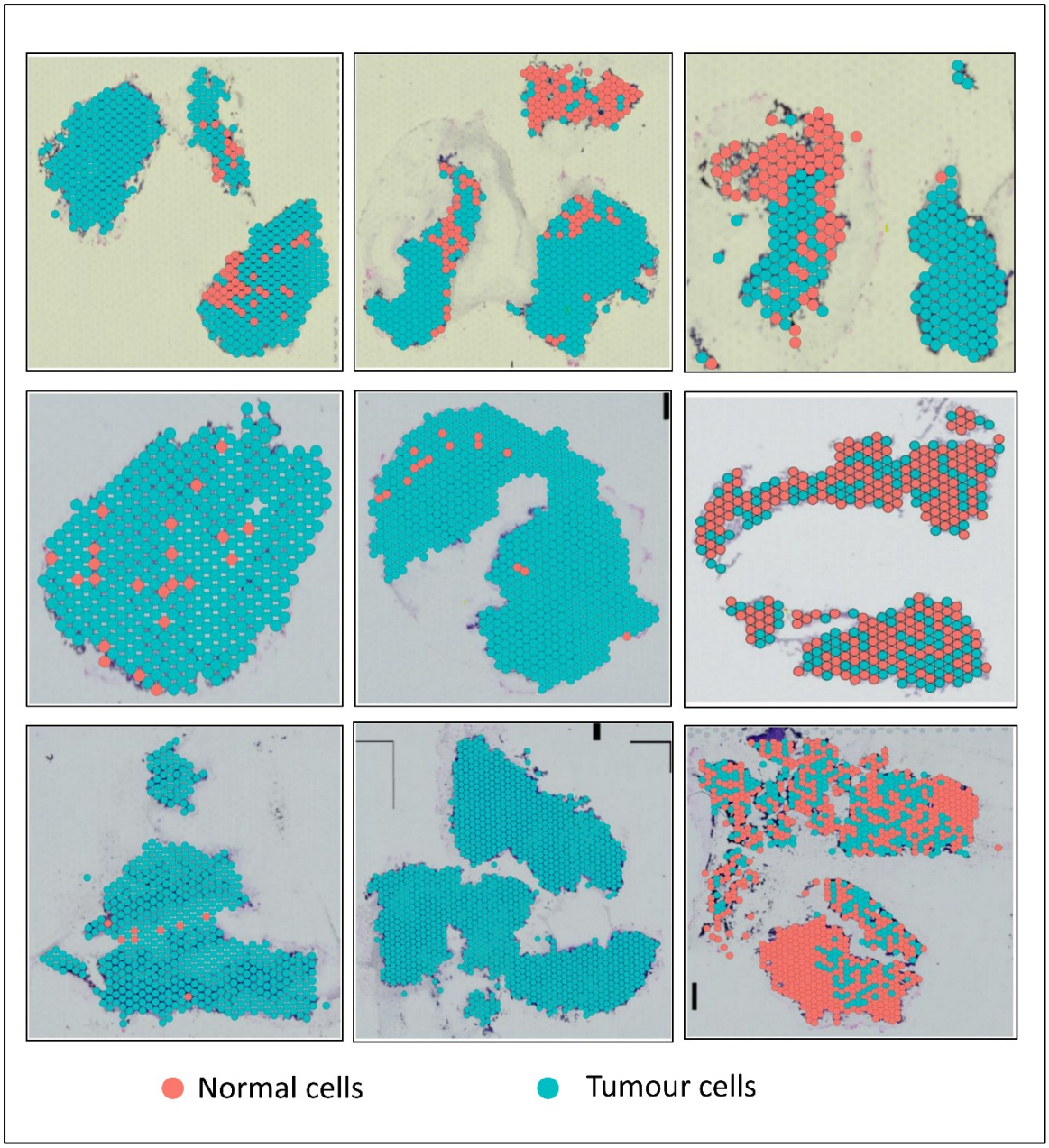

Figure S4 Spatial distribution of tumor and normal cells across tissue sections.
The figure illustrates the spatial mapping of spots classified as tumor (blue) or normal (red) across multiple tissue sections from three patients. Each row corresponds to one patient, with the columns representing different tumor sections: Core (left), Inner edge (center), and Outer edge (right). The spatial distribution highlights the heterogeneity in the proportion and localization of tumors and normal.

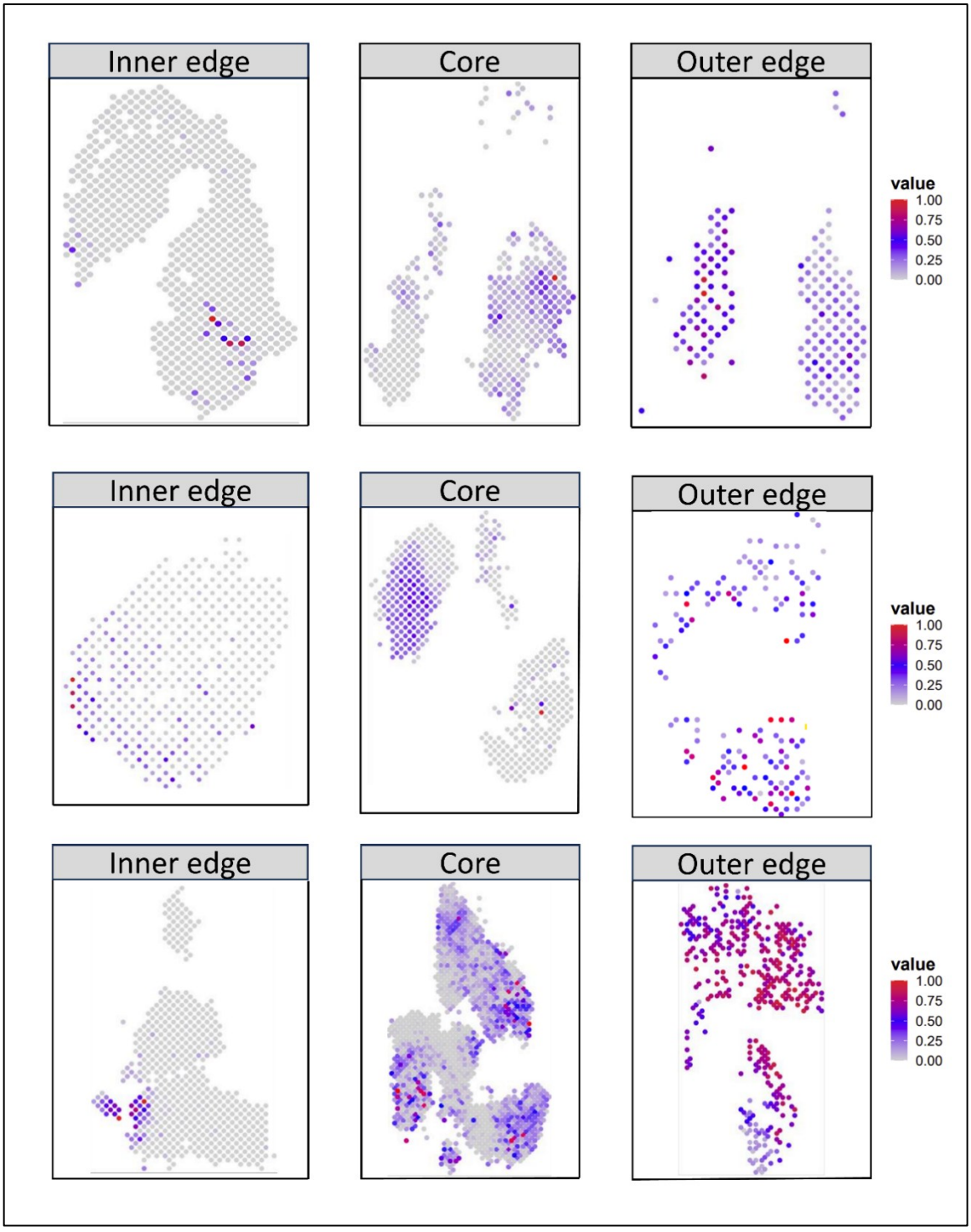

Figure S5 ST results of deconvolution score for OPC-like cells.
The figure illustrates the distribution of signal intensity for the OPC-like signature derived from the Bhaduri dataset, which includes 11 GBM tumor samples. Each row corresponds to a patient (from top to bottom: patient 1 to patient 3), with regions—Inner edge, Core, and Outer edge—depicted sequentially within each row. The color gradient represents deconvolution score intensity, where red indicates higher scores (stronger resemblance to the OPC-like signature) and blue indicates lower scores (weaker resemblance).

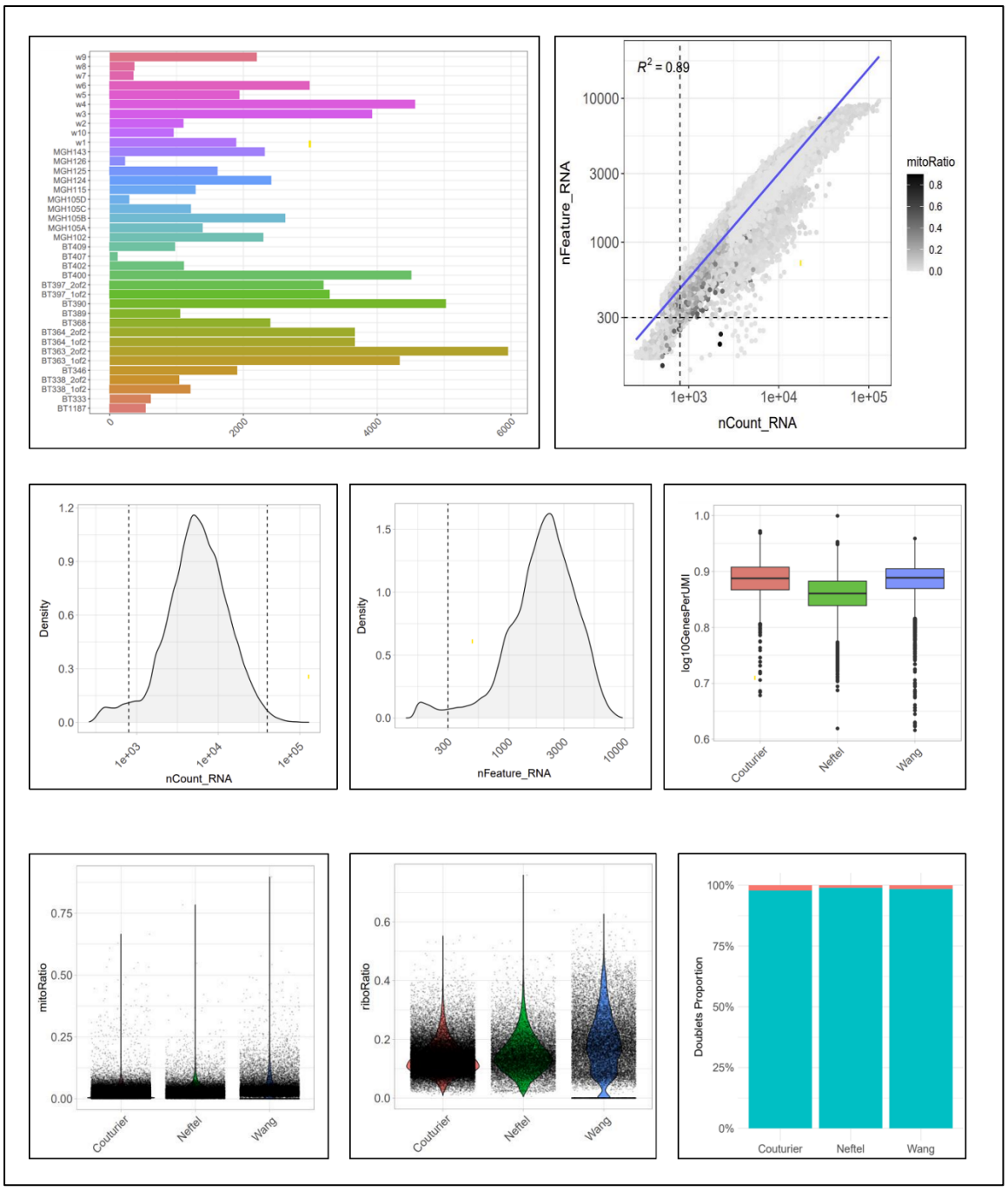

Figure S6 Quality Metrics for Single-Cell Datasets.
The figure summarizes the quality control metrics for single-cell RNA-sequencing datasets from three studies (Couturier, Neftel, Wang). The bar plot displays the number of cells for each sample. The scatter plot shows the correlation between the total number of transcripts detected (nCount_RNA) and the number of unique features (nFeature_RNA), with mitochondrial RNA content (mitoRatio). Density plots illustrate the distribution of nCount_RNA and nFeature_RNA, with filtering thresholds highlighted. The box plot visualizes the log-transformed number of genes per UMI (log10GenesPerUMI) across studies. Violin plots depict the distribution of mitochondrial RNA content (mitoRatio) for each study. The final bar plot shows the proportion of doublets detected in each dataset.

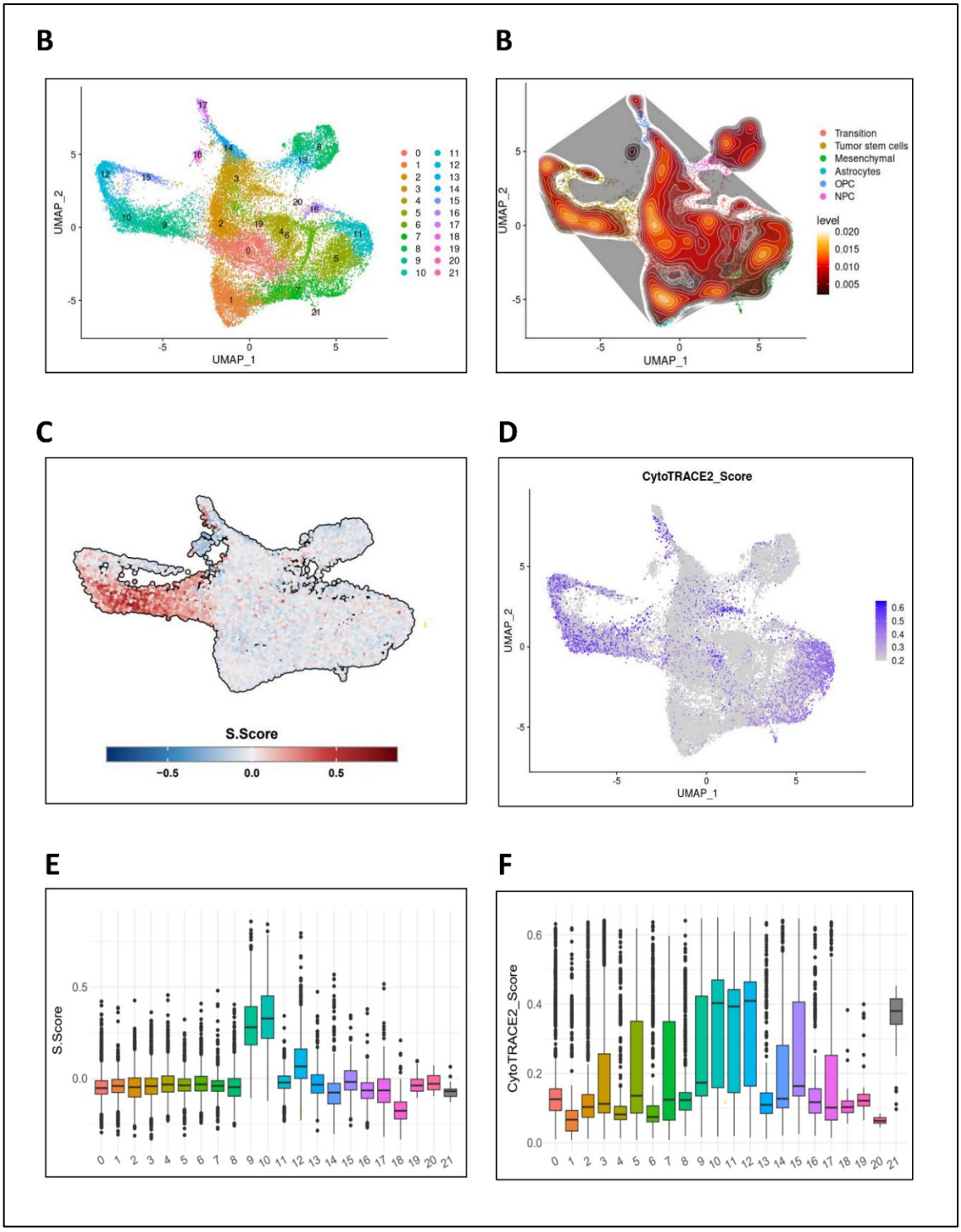

Figure S7 Integrated UMAP of GBM tumor cells from Couturier, Neftel and Dong studies.
(A) *UMAP representation of tumor cells, with the identified clusters.* ***(B)*** *UMAP of cell densities.* ***(C)*** *UMAP colored by S.Score, representing the cell cycle S phase activity. Higher scores (red) indicate cells actively proliferating.* ***(D)*** *UMAP colored by CytoTRACE2 score, reflecting the differentiation potential of the cells, with higher scores (purple) indicating a more stem-like state.* ***(E)*** *Boxplot displaying the distribution of S.Scores across the identified clusters, highlighting variability in proliferative activity.* ***(F)*** *Boxplot displaying the distribution of CytoTRACE2 scores across the identified clusters, showing the differentiation potential of each cluster.*

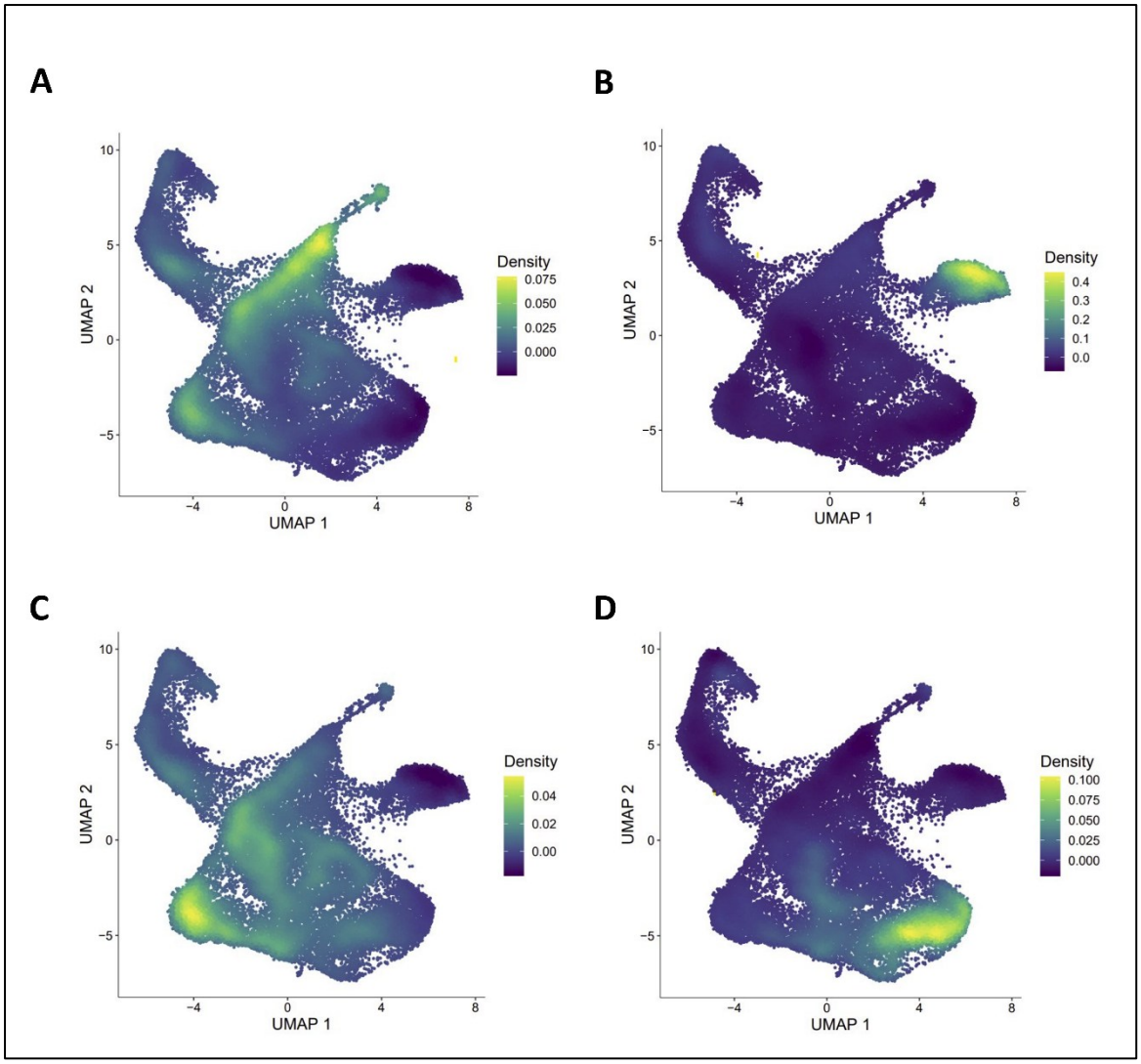

Figure S8 Distribution of GBM Molecular Signatures.
UMAP showing the density distribution of the four Neftel signatures of GBM: ***(A)*** *OPC-like,* ***(B)*** *NEU-like,* ***(C)*** *AC-like,* ***(D)*** *MES-like. The panels show the distribution of cells for each signature, highlighting the transcriptional heterogeneity of tumor populations.*

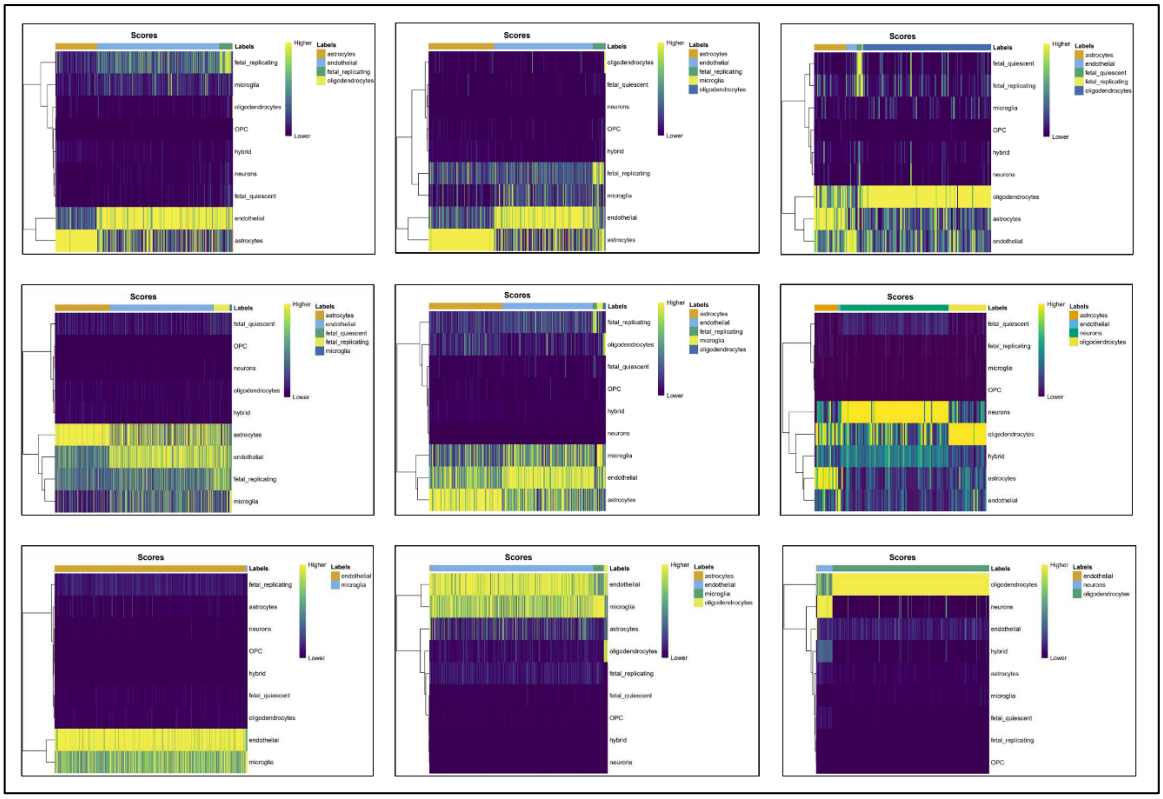

Figure S9 Correlation analyses of ST spots performed using the single-cell RNA-seq dataset from Nowakowski et al. (2017) on the human cortex.

The columns represent the spatial regions of the tumor: (from left to right) Core, Inner edge, and Outer edge, while the rows correspond to the three patients (Patient 1, Patient 2, and Patient 3). The heatmaps show the scores (Pearson correlations) between the tissue spots and the various cell types, with higher scores (yellow) indicating stronger similarity. Annotated cell types include OPC, astrocytes, neurons, oligodendrocytes, microglia, and others.

Conclusions

In this thesis, I studied the biological characteristics of two tumors distinguished by origin, affected population, and clinical behaviour: neuroblastoma and glioblastoma. Neuroblastoma is a rare and highly heterogeneous paediatric tumor that originates from the progenitor cells of the sympathetic nervous system. On the other hand, glioblastoma is the most common malignant brain tumor in adults, known for its extreme aggressiveness and biological complexity.

In the first project, I investigated the cellular and transcriptional dynamics underlying neuroblastoma by performing a meta-analysis of different single-cell and single-nucleus RNA sequencing datasets. In this study, I found that cancer cells mainly resemble the state of sympathoblasts and cycling sympathoblasts of the normal adrenal gland. The absence of cancer cells resembling the more differentiated sympathetic neurons suggests that these cells are somehow blocked in an undifferentiated state unable to complete the terminal differentiation. This behaviour appears to be driven by the aberrant regulation of the genes involved in neurogenesis combined with the presence of specific chromosomal aberrations. Moreover, survival analyses indicate that the downregulation of genes driving terminal differentiation is associated with poorer patient outcomes.
Although further studies are needed to better understand the role of these genes in regulating the differentiation state of neuroblastoma cells, the framework employed in this study is innovative. Combining transcriptomic analysis with chromosomal aberration inference offers a powerful approach to uncover mechanisms underlying neuroblastoma cell differentiation and identify new therapeutic targets for future interventions.

In the second project, I explored the biological complexity of glioblastoma by integrating spatial and single-cell transcriptomic techniques. I analysed the different cell populations within the tumor and studied how the spatial localization affects the cell identity in the tissue. Specifically, I observed a predominance of AC-like and MES-like subtypes in the central regions and the inner areas of the tumors. In contrast, the outer edge seemed to be predominantly composed of tumor cells exhibiting OPC-like characteristics. Moreover, I investigated the molecular interactions occurring between OPC-like tumor cells of the outer edge and the surrounding microenvironment. I found that tumor cells can influence and modulate the behaviour of the immune cells downregulating their anti-cancer activity. These results not only improve our understanding of GBM heterogeneity, but they give us new opportunities for the development of targeted therapeutic strategies.

In conclusion, this work demonstrates the potentiality given by the application of scRNA-seq and ST in cancer research. The possibility given by scRNA-seq to analyze tumor tissues at an unprecedented high resolution, down to individual cells, allows to study how tumors evolve, develop resistance to therapies, and interact with their surrounding environment. Meanwhile, spatial transcriptomics allows to capture the spatial organization within the tumor microenvironment and map cell-to-cell interactions.

Integrating these technologies with advanced computational tools offers exciting opportunities to better understand treatment resistance, improve outcome predictions, and design more effective, targeted therapies for patients.

Bibliography

Aldape K, Brindle KM, Chesler L, Chopra R, Gajjar A, Gilbert MR, Gottardo N, Gutmann DH, Hargrave D, Holland EC, *et al* (2019) Challenges to curing primary brain tumours. *Nature Reviews Clinical Oncology 2019 16:8* 16: 509–520

Alexander BM & Cloughesy TF (2017) Adult Glioblastoma. *J Clin Oncol* 35: 2402–2409

Batlle E & Massagué J (2019) Transforming Growth Factor-β Signaling in Immunity and Cancer. *Immunity* 50: 924–940

Bhaduri A, Di Lullo E, Jung D, Müller S, Crouch EE, Espinosa CS, Ozawa T, Alvarado B, Spatazza J, Cadwell CR, *et al* (2020) Outer Radial Glia-like Cancer Stem Cells Contribute to Heterogeneity of Glioblastoma. *Cell Stem Cell* 26: 48-63.e6

Bornstein SR, Ehrhart-Bornstein M, Androutsellis-Theotokis A, Eisenhofer G, Vukicevic V, Licinio J, Wong ML, Calissano P, Nisticò G, Preziosi P, *et al* (2012) Chromaffin cells: the peripheral brain. *Mol Psychiatry* 17: 354–358

Bronner ME (2012) Formation and migration of neural crest cells in the vertebrate embryo. *Histochem Cell Biol* 138: 179–186

Couturier CP, Ayyadhury S, Le PU, Nadaf J, Monlong J, Riva G, Allache R, Baig S, Yan X, Bourgey M, *et al* (2020) Single-cell RNA-seq reveals that glioblastoma recapitulates a normal neurodevelopmental hierarchy. *Nature Communications 2020 11:1* 11: 1–19

Cuevas S, Villar VA, Jose PA & Armando I (2013) Renal Dopamine Receptors, Oxidative Stress, and Hypertension. *Int J Mol Sci* 14: 17553

Darmanis S, Sloan SA, Croote D, Mignardi M, Chernikova S, Samghababi P, Zhang Y, Neff N, Kowarsky M, Caneda C, *et al* (2017) Single-Cell RNA-Seq Analysis of Infiltrating Neoplastic Cells at the Migrating Front of Human Glioblastoma. *Cell Rep* 21: 1399–1410

Dong R, Yang R, Zhan Y, Lai HD, Ye CJ, Yao XY, Luo WQ, Cheng XM, Miao JJ, Wang JF, *et al* (2020) Single-Cell Characterization of Malignant Phenotypes and Developmental Trajectories of Adrenal Neuroblastoma. *Cancer Cell* 38: 716-733.e6

Elkabets M, Gifford AM, Scheel C, Nilsson B, Reinhardt F, Bray MA, Carpenter AE, Jirström K, Magnusson K, Ebert BL, *et al* (2011) Human tumors instigate granulin-expressing hematopoietic cells that promote malignancy by activating stromal fibroblasts in mice. *J Clin Invest* 121: 784–799

Filippou PS, Karagiannis GS & Constantinidou A (2019) Midkine (MDK) growth factor: a key player in cancer progression and a promising therapeutic target. *Oncogene 2019 39:10* 39: 2040–2054

Floriddia E (2022) The impact of ambient RNA. *Nature Neuroscience 2022 25:12* 25: 1583–1583

Furlan A, Dyachuk V, Kastriti ME, Calvo-Enrique L, Abdo H, Hadjab S, Chontorotzea T, Akkuratova N, Usoskin D, Kamenev D, *et al* (2017) Multipotent peripheral glial cells generate neuroendocrine cells of the adrenal medulla. *Science* 357

Gilbert SF (2000) The Neural Crest.

Grech N, Dalli T, Mizzi S, Meilak L, Calleja N & Zrinzo A (2020) Rising Incidence of Glioblastoma Multiforme in a Well-Defined Population. *Cureus* 12

Greenwald AC, Darnell NG, Hoefflin R, Simkin D, Mount CW, Gonzalez Castro LN, Harnik Y, Dumont S, Hirsch D, Nomura M, *et al* (2024) Integrative spatial analysis reveals a multi-layered organization of glioblastoma. *Cell* 187: 2485-2501.e26

Hadjipanayis CG & Stummer W (2019) 5-ALA and FDA approval for glioma surgery. *J Neurooncol* 141: 479–486

Hu B, Qin C, Li L, Wei L, Mo X, Fan H, Lei Y, Wei F & Zou D (2021) Midkine promotes glioblastoma progression via PI3K-Akt signaling. *Cancer Cell Int* 21: 1–13

Hu N, Strobl-Mazzulla PH & Bronner ME (2014) Epigenetic Regulation in Neural Crest Development. *Dev Biol* 396: 159

Hu Y, Converse C, Lyons MC & Hsu WH (2018) Neural control of sweat secretion: a review. *British Journal of Dermatology* 178: 1246–1256

Huber K (2006) The sympathoadrenal cell lineage: Specification, diversification, and new perspectives. *Dev Biol* 298: 335–343

Irwin MS, Naranjo A, Zhang FF, Cohn SL, London WB, Gastier-Foster JM, Ramirez NC, Pfau R, Reshmi S, Wagner E, *et al* (2021) Revised Neuroblastoma Risk Classification System: A Report From the Children's Oncology Group. *Journal of Clinical Oncology* 39: 3229–3241

Jansky S, Sharma AK, Körber V, Quintero A, Toprak UH, Wecht EM, Gartlgruber M, Greco A, Chomsky E, Grünewald TGP, *et al* (2021) Single-cell transcriptomic analyses provide insights into the developmental origins of neuroblastoma. *Nat Genet* 53

Johnsen JI, Dyberg C & Wickström M (2019) Neuroblastoma—A Neural Crest Derived Embryonal Malignancy. *Front Mol Neurosci* 12

Kameneva P, Artemov A V., Kastriti ME, Faure L, Olsen TK, Otte J, Erickson A, Semsch B, Andersson ER, Ratz M, *et al* (2021a) Single-cell transcriptomics of human embryos identifies multiple sympathoblast lineages with potential implications for neuroblastoma origin. *Nat Genet* 53: 694–706

Kameneva P, Kastriti ME & Adameyko I (2021b) Neuronal lineages derived from the nerve-associated Schwann cell precursors. *Cellular and Molecular Life Sciences* 78 doi:10.1007/s00018-020-03609-5 [PREPRINT]

Kastriti ME, Faure L, Von Ahsen D, Bouderlique TG, Boström J, Solovieva T, Jackson C, Bronner M, Meijer D, Hadjab S, *et al* (2022) Schwann cell precursors represent a neural crest-like state with biased multipotency. *EMBO J* 41

Kholodenko I V., Kalinovsky D V., Doronin II, Deyev SM & Kholodenko R V. (2018a) Neuroblastoma Origin and Therapeutic Targets for Immunotherapy. *J Immunol Res* 2018

Kholodenko I V., Kalinovsky D V., Doronin II, Deyev SM & Kholodenko R V. (2018b) Neuroblastoma origin and therapeutic targets for immunotherapy. *J Immunol Res* 2018 doi:10.1155/2018/7394268 [PREPRINT]

Kildisiute G, Kholosy WM, Young MD, Roberts K, Elmentaite R, van Hooff SR, Pacyna CN, Khabirova E, Piapi A, Thevanesan C, *et al* (2021) Tumor to normal single-cell mRNA comparisons reveal a pan-neuroblastoma cancer cell. *Sci Adv* 7

Kim BG, Malek E, Choi SH, Ignatz-Hoover JJ & Driscoll JJ (2021) Novel therapies emerging in oncology to target the TGF-β pathway. *Journal of Hematology & Oncology 2021 14:1* 14: 1–20

Labonne C & Bronner-Fraser M Induction and Patterning of the Neural Crest, a Stem Cell-Like Precursor Population.

Law M, Young RJ, Babb JS, Peccerelli N, Chheang S, Gruber ML, Miller DC, Golfinos JG, Zagzag D & Johnson G (2008) Gliomas: predicting time to progression or survival with cerebral blood volume measurements at dynamic susceptibility-weighted contrast-enhanced perfusion MR imaging. *Radiology* 247: 490–498

Lee SY (2016) Temozolomide resistance in glioblastoma multiforme. *Genes Dis* 3: 198–210

Lo L, Morin X, Brunet JF & Anderson DJ (1999) Specification of Neurotransmitter Identity by Phox2 Proteins in Neural Crest Stem Cells. *Neuron* 22: 693–705

Louis DN, Perry A, Wesseling P, Brat DJ, Cree IA, Figarella-Branger D, Hawkins C, Ng HK, Pfister SM, Reifenberger G, *et al* (2021) The 2021 WHO Classification of Tumors of the Central Nervous System: a summary. *Neuro Oncol* 23: 1231

Lousado L, Prazeres PHDM, Andreotti JP, Paiva AE, Azevedo PO, Santos GSP, Filev R, Mintz A & Birbrair A (2017) Schwann cell precursors as a source for adrenal gland chromaffin cells. *Cell Death Dis* 8: e3072

Lumb R & Schwarz Q (2015) Sympathoadrenal neural crest cells: the known, unknown and forgotten? *Dev Growth Differ* 57: 146–157

Ma R, Sun JH & Wang YY (2024) The role of transforming growth factor-β (TGF-β) in the formation of exhausted CD8 + T cells. *Clin Exp Med* 24: 1–9

Matthay KK, Maris JM, Schleiermacher G, Nakagawara A, Mackall CL, Diller L & Weiss WA (2016a) Neuroblastoma. *Nat Rev Dis Primers* 2

Matthay KK, Maris JM, Schleiermacher G, Nakagawara A, Mackall CL, Diller L & Weiss WA (2016b) Neuroblastoma. *Nature Reviews Disease Primers 2016 2:1* 2: 1–21

Monclair T, Brodeur GM, Ambros PF, Brisse HJ, Cecchetto G, Holmes K, Kaneko M, London WB, Matthay KK, Nuchtern JG, *et al* (2009) The International Neuroblastoma Risk Group (INRG) Staging System: An INRG Task Force Report. *Journal of Clinical Oncology* 27: 298

Neftel C, Laffy J, Filbin MG, Hara T, Shore ME, Rahme GJ, Richman AR, Silverbush D, Shaw ML, Hebert CM, *et al* (2019) An Integrative Model of Cellular States, Plasticity, and Genetics for Glioblastoma. *Cell* 178: 835-849.e21

Nowakowski TJ, Bhaduri A, Pollen AA, Alvarado B, Mostajo-Radji MA, Di Lullo E, Haeussler M, Sandoval-Espinosa C, Liu SJ, Velmeshev D, *et al* (2017) Spatiotemporal gene expression trajectories reveal developmental hierarchies of the human cortex. *Science* 358: 1318

Ohgaki H & Kleihues P (2007) Genetic pathways to primary and secondary glioblastoma. *Am J Pathol* 170: 1445–1453

Ostrom QT, Cioffi G, Gittleman H, Patil N, Waite K, Kruchko C & Barnholtz-Sloan JS (2019) CBTRUS Statistical Report: Primary Brain and Other Central Nervous System Tumors Diagnosed in the United States in 2012-2016. *Neuro Oncol* 21: 1–100

Ponzoni M, Bachetti T, Corrias MV, Brignole C, Pastorino F, Calarco E, Bensa V, Giusto E, Ceccherini I & Perri P (2022) Recent advances in the developmental origin of neuroblastoma: an overview. *Journal of Experimental and Clinical Cancer Research* 41: 1–28

Ravi VM, Will P, Kueckelhaus J, Sun N, Joseph K, Salié H, Vollmer L, Kuliesiute U, von Ehr J, Benotmane JK, *et al* (2022) Spatially resolved multi-omics deciphers bidirectional tumor-host interdependence in glioblastoma. *Cancer Cell* 40: 639-655.e13

Saeedi-Boroujeni A, Purrahman D, Shojaeian A, Poniatowski ŁA, Rafiee F & Mahmoudian-Sani MR (2023) Progranulin (PGRN) as a regulator of inflammation and a critical factor in the immunopathogenesis of cardiovascular diseases. *Journal of Inflammation 2023 20:1* 20: 1–14

Seker-Polat F, Degirmenci NP, Solaroglu I & Bagci-Onder T (2022) Tumor Cell Infiltration into the Brain in Glioblastoma: From Mechanisms to Clinical Perspectives. *Cancers (Basel)* 14

Shah JL, Li GH & Soltys SG (2022) Glioblastoma Multiforme. *CyberKnife Stereotactic Radiosurgery: Brain* 1: 85–98

Sharma P, Aaroe A, Liang J & Puduvalli VK (2023) Tumor microenvironment in glioblastoma: Current and emerging concepts. *Neurooncol Adv* 5: 1–16

Sokol E & Desai A V. (2019) The Evolution of Risk Classification for Neuroblastoma. *Children 2019, Vol 6, Page 27* 6: 27

Stummer W, Pichlmeier U, Meinel T, Wiestler OD, Zanella F & Reulen HJ (2006) Fluorescence-guided surgery with 5-aminolevulinic acid for resection of malignant glioma: a randomised controlled multicentre phase III trial. *Lancet Oncol* 7: 392–401

Tang M, Luo H & Lu J (2015) Genetically altered cancer epigenome. *Epigenetic Gene Expression and Regulation*: 265–289

Theveneau E & Mayor R (2014a) Neural Crest Cell Migration: Guidance, Pathways, and Cell-Cell Interactions. *Neural Crest Cells: Evolution, Development and Disease*: 73–88

Theveneau E & Mayor R (2014b) Neural Crest Cell Migration: Guidance, Pathways, and Cell-Cell Interactions. *Neural Crest Cells: Evolution, Development and Disease*: 73–88

Tsarovina K, Pattyn A, Stubbusch J, Müller F, van der Wees J, Schneider C, Brunet JF & Rohrer H (2004) Essential role of Gata transcription factors in sympathetic neuron development. *Development* 131: 4775–4786

Wang L, Jung J, Babikir H, Shamardani K, Jain S, Feng X, Gupta N, Rosi S, Chang S, Raleigh D, *et al* (2022) A single-cell atlas of glioblastoma evolution under therapy reveals cell-intrinsic and cell-extrinsic therapeutic targets. *Nat Cancer* 3: 1534–1552

Wang Q, Hu B, Hu X, Kim H, Squatrito M, Scarpace L, deCarvalho AC, Lyu S, Li P, Li Y, *et al* (2017) Tumor Evolution of Glioma-Intrinsic Gene Expression Subtypes Associates with Immunological Changes in the Microenvironment. *Cancer Cell* 32: 42-56.e6

Waxenbaum JA, Reddy V & Varacallo M (2023) Anatomy, Autonomic Nervous System. *StatPearls*

Williams CG, Lee HJ, Asatsuma T, Vento-Tormo R & Haque A (2022) An introduction to spatial transcriptomics for biomedical research. *Genome Med* 14: 1–18

Yi GZ, Huang G, Guo M, Zhang X, Wang H, Deng S, Li Y, Xiang W, Chen Z, Pan J, *et al* (2019) Acquired temozolomide resistance in MGMT-deficient glioblastoma cells is associated with regulation of DNA repair by DHC2. *Brain* 142: 2352–2366

Zeng T, Cui D & Gao L (2015) Glioma: an overview of current classifications, characteristics, molecular biology and target therapies. *Front Biosci (Landmark Ed)* 20: 1104–1115

Zhang Q, Huang R, Ye Y, Guo X, Lu J, Zhu F, Gong X, Zhang Q, Yan J, Luo L, *et al* (2018) Temporal requirements for ISL1 in sympathetic neuron proliferation, differentiation, and diversification. *Cell Death Dis* 9

Zhang W, Cai YY, Wang XL, Wang XX, Li Y, Han GY, Chu YJ, Zhang YX & Hao FR (2021) Bone Metastases of Glioblastoma: A Case Report and Review of the Literature. *Front Oncol* 11

Acknowledgements

I would like to thank Professor Alessandro Quattrone, my thesis supervisor, for giving me the opportunity to conduct these research projects in his laboratory, I am very grateful for his guidance and support. During my PhD, I had the opportunity to significantly improve my knowledge and skills in computational biology by learning how to process data from these advanced cutting-edge techniques.

I also want to thank Professor Dr. Frank Westermann and his team at the DKFZ (German Cancer Research Center) in Heidelberg, where I had the opportunity to conduct part of my research abroad. Their expertise offered valuable insights for my projects.

Lastly, I want to thank Angelo for his collaboration on the neuroblastoma project, as well as Chiara Argento and Emanuele for their support and assistance with the glioblastoma project. This work was also made possible thanks to their contributions, I have sincerely appreciated their help and support.