INAUGURAL DISSERTATION
FOR OBTAINING THE DOCTORAL
DEGREE OF THE COMBINED
FACULTY OF MATHEMATICS,
ENGINEERING AND NATURAL
SCIENCES OF THE
RUPRECHT-KARLS-UNIVERSITY
HEIDELBERG

Presented by

SPECIALIST, ZAIRA SEFERBEKOVA

born in:

MAKHACHKALA, RUSSIA

Oral examination: 21 JANUARY 2026

DECIPHERING THE CANCER ECOSYSTEM OF GLIOMAS USING SPATIAL TRANSCRIPTOMICS

Referees:

PROF. DR MORITZ GERSTUNG

PROF. DR OLIVER STEGLE

ABSTRACT

Recent efforts to incorporate a range of molecular characteristics into diagnostic guidelines underscore the necessity of comprehending the intricacies of tumour biology for better clinical decision-making. The role of environmental factors and cellular interactions is also increasingly recognised as a vital component in understanding cancer growth and progression. A complex cellular and spatial landscape is particularly evident in gliomas, which represent highly heterogeneous and plastic brain tumours. In this work, I develop a robust computational framework for the analysis of spatial single cell transcriptomics data and utilise it to conduct two studies on the spatial biology of gliomas.

In the first study, I conducted a comparative analysis of necrotic tissues in ten patients who were previously diagnosed with glioblastoma and subsequently presented with either tumour progression or radionecrosis. To this end, I generated a spatial single cell transcriptomic atlas consisting of over a million of cells and encompassing several brain-resident cell types and tumour states. The analysis of the annotated data revealed that radionecrotic samples contained abundant tumour cells with downregulated *EGFR* expression and were infiltrated by macrophages that contributed to gliosis. In contrast, samples with progression contained progenitor-like and cycling tumour cells that maintained high *EGFR* expression. The study offers invaluable insights into the spatial landscape and cellular interplay in radionecrosis and holds the potential to inform future research aimed at improving diagnostic and therapeutic strategies for glioblastoma patients.

In the second study, I designed a custom gene panel to enable a thorough investigation into the intricacies of cellular and spatial composition of over 300 samples from patients diagnosed with seven different glioma types. The examination of the annotated data set revealed both dis-

parities and commonalities in the tumour expression patterns among adult-type diffuse gliomas and ependymal tumours. The spatial resolution permitted systematic examination of the spatial neighbourhoods that were linked to individual tumour transcription programs. The study establishes the foundation for future research projects in the group that will employ tailored panels and offers a glimpse into the spatial organisation of gliomas. I believe that the generated atlas along with standardised clinical data can facilitate further attempts to identify clinically relevant associations.

To summarise, I posit that the research conducted within the scope of this dissertation stands to provide a valuable basis for future endeavours in the spatial field. Furthermore, the biological insights derived from the generated data can be used to inform a more focused exploration of the intricate biology of necrotic tissue and the broader spatial patterns characterising multiple gliomas in the future.

ZUSAMMENFASSUNG

Jüngste Bemühungen, eine Reihe molekularer Merkmale in Diagnoserichtlinien zu integrieren, unterstreichen die Notwendigkeit, die Feinheiten der Tumorbiologie zu verstehen, um bessere klinische Entscheidungen treffen zu können. Auch die Rolle von Umweltfaktoren und zellulären Interaktionen wird zunehmend als wichtiger Bestandteil für das Verständnis des Wachstums und Fortschreitens von Krebserkrankungen anerkannt. Eine komplexe räumliche Landschaft zeigt sich besonders deutlich bei Gliomen, die hochgradig heterogene und plastische Tumoren darstellen. Ein besseres Verständnis der Zusammensetzung des räumlichen Gliom-Ökosystems hat das Potenzial, die Entwicklung neuartiger Behandlungsstrategien zu beeinflussen. In dieser Arbeit entwickle ich einen robusten rechnerischen Rahmen für die Analyse räumlicher Einzelzell-Transkriptomikdaten und nutze ihn, um zwei Studien zur räumlichen Biologie von Gliomen durchzuführen.

In der ersten Studie führe ich eine vergleichende Untersuchung von nekrotischem Gewebe bei zehn Patienten durch, bei denen zuvor ein Glioblastom diagnostiziert worden war und die anschließend entweder eine Tumorprogression oder eine Strahlennekrose aufwiesen. Im Rahmen dieser Studie habe ich einen Einzelzell-Transkriptomik-Atlas erstellt, der aus über einer Million Zellen besteht und mehrere im Gehirn ansässige Zellen und Tumorzustände umfasst. Die Analyse der annotierten Daten ergab, dass die radionekrotischen Proben reichlich Tumorzellen mit heruntergeregelter *EGFR*-Expression enthielten und von grenznahen Makrophagen infiltriert waren, die zur Gliose beitrugen. Im Gegensatz dazu enthielten die Proben mit Progression progenitorähnliche und zyklische Tumorzellen, die eine hohe *EGFR*-Expression aufwiesen. Die Studie bietet wertvolle Einblicke in die räumliche Landschaft und die zellulären Wechselwirkungen bei der Radionekrose und hat das Potenzial, zukünftige Studien zur Verbes-

serung der Diagnose- und Therapiestrategien für Glioblastom-Patienten zu beeinflussen.

In der zweiten Studie entwerfe ich ein maßgeschneidertes Genpanel, um eine gründliche Untersuchung der Feinheiten der zellulären Landschaft und der räumlichen Zusammensetzung von über 300 Proben von Patienten zu ermöglichen, bei denen sieben Arten von Gliomen diagnostiziert wurden. Die Untersuchung des annotierten Datensatzes ergab sowohl Unterschiede als auch Gemeinsamkeiten in den Tumorexpressionsmustern zwischen diffusen Gliomen vom adulten Typ und ependymalen Tumoren. Die räumliche Auflösung ermöglichte eine systematische Untersuchung der räumlichen Nachbarschaften, die mit individuellen Tumortranskriptionsprogrammen in Verbindung standen. Die Studie legt den Grundstein für zukünftige Forschungsprojekte der Gruppe, die maßgeschneiderte Panels einsetzen werden, und bietet zudem einen Einblick in die räumliche Architektur von Gliomen. Ich bin davon überzeugt, dass der generierte Datensatz zusammen mit den zugehörigen klinischen Daten die Identifizierung klinisch relevanter Zusammenhänge erleichtern kann.

Ich gehe davon aus, dass die im Rahmen dieser Dissertation durchgeführten Forschungen eine wertvolle Grundlage für zukünftige Arbeiten im räumlichen Bereich bilden. Darüber hinaus können die aus den generierten Daten gewonnenen biologischen Erkenntnisse genutzt werden, um in Zukunft eine gezieltere Erforschung der komplexen Merkmale nekrotischen Gewebes und der breiteren räumlichen Muster, die multiple Gliome charakterisieren, zu ermöglichen.

CONTENTS

LIST OF FIGURES

LIST OF TABLES

NOMENCLATURE

AC	astrocyte
BAM	border-associated macrophage
CNS	central nervous system
CNV	copy number variation
CSF	cerebrospinal fluid
DCIS	ductal carcinoma *in situ*
DE	differential expression
EPN	ependymoma
EPN-MPE	myxopapillary ependymoma
EPN-SE	subependymoma
EPN-SP	spinal ependymoma
FFPE	formalin-fixed paraffin-embedded
GB	glioblastoma
GSC	glioma stem cell
H&E	haematoxylin and eosin
HGA	high-grade astrocytoma
HVG	highly variable gene
IDH	isocitrate dehydrogenase
IHC	immunohistochemistry
kNN	k-nearest neighbours
LGA	low-grade astrocytoma
MF	matrix factorisation
MES	mesenchymal
MRI	magnetic resonance imaging
NSC	neural stem cell
NMF	non-negative MF
NPC	neural progenitor cell

OC	oligodendrocyte
OLIGO	oligodendroglioma
OPC	oligodendrocyte progenitor cell
PCA	principal component analysis
PET	positron emission tomography
PVF	perivascular fibroblast
RN	radionecrosis
RNA-seq	RNA sequencing
RTK	receptor tyrosine kinase
scRNA-seq	single-cell RNA-seq
SNV	single nucleotide variant
sPCA	sparse PCA
TMA	tissue microarray
TME	tumour microenvironment
TMZ	temozolomide
VAF	variant allele frequency
VSMC	vascular smooth muscle cell
UMAP	uniform manifold approximation and projection
WHO	World Health Organisation

ACKNOWLEDGED CONTRIBUTORS

Abigail K. Suwala	Dr med. Abigail K Suwala, Department of Neuropathology, Heidelberg University Hospital, Heidelberg, Germany
Domenico Calafato	Domenico Calafato, Artificial Intelligence in Oncology, German Cancer Research Center, Heidelberg, Germany
Felix Sahm	Prof. Dr Dr med. Felix Sahm, Department of Neuropathology, Heidelberg University Hospital, Heidelberg, Germany
Gleb Rukhovich	Gleb Rukhovich, Artificial Intelligence in Oncology, German Cancer Research Center, Heidelberg, Germany
Michael Ritter	Dr rer. nat. Michael Ritter, Artificial Intelligence in Oncology, German Cancer Research Center, Heidelberg, Germany
scOpenLab	Single-cell Open Lab, German Cancer Research Center, Heidelberg, Germany

"We have no reason to mistrust our world, for it is not against us. Has it terrors, they are our terrors; has it abysses, those abysses belong to us; are dangers at hand, we must try to love them. And if only we arrange our life according to that principle which counsels us that we must always hold to the difficult, then that which now still seems to us the most alien will become what we most trust and find most faithful. How should we be able to forget those ancient myths about dragons that at the last moment turn into princesses; perhaps all the dragons of our lives are princesses who are only waiting to see us once beautiful and brave. Perhaps everything terrible is in its deepest being something helpless that wants help from us."

— RAINER M. RILKE

ACKNOWLEDGMENTS

Doing a PhD is rarely a solitary pursuit, even if it often feels like one, and I am deeply grateful to all the people who have shaped my journey and helped me reach this point.

First, I wish to thank *Prof. Dr Moritz Gerstung*, for giving me the chance to join his group and develop the projects from their earliest stages. I thank him for the opportunity to pursue a doctorate on some of the most exciting and rapidly evolving topics in translational research, and for trusting me with the freedom to explore them in my own way.

I would also like to thank *Prof. Dr Ana Martín-Villalba* and *Dr Pei-Chi Wei*, for taking the time to review and evaluate this work. I thank *Prof. Dr Oliver Stegle, Dr Lucy Yates*, and *Prof. Dr med. Stefan Pfister*, for their valuable input and constructive feedback throughout the years.

I wish to thank all *collaborators*, but particularly *Dr med. Abigail K. Suwala*, for being a source of both scientific and personal advice, and *Prof. Dr Dr med. Felix Sahm*, for allowing me to take part in the inspiring efforts to improve glioma patient care.

My sincere gratitude goes to *all colleagues* who became my dear friends. To *Domenico* and *Micha*, for their expertise in the experimental parts and constant encouragement and warmth. To *Stefan*, for making me feel included from the start and for always having my back. And to everyone in the lab, for the countless waffle sessions in the kitchen and lunch trips to Mensa. I thank you all for making our small group feel like a weird but wonderful family.

I thank *Mum* and *Dad*, for instilling in me the drive to study and persevere, and *my extended family*, for creating a sense of belonging that I continue to carry with me.

My warm gratitude also goes to my *chosen family*. To *Areeba* and *Eléonore*, for their endless support and ability to make even bad days brighter. To *Gleb* and *Artems*, for their

help with the computational analyses and Bayes theory, and for our tea breaks and a very short-lived book club. To *Nastya, Anya, Nadya, Misha*, and *Masha*, for staying close despite the distance and for reminding me of who I am beyond this PhD. To *Timothy*, for his love and care through all the late nights and moments of exhaustion. And finally, to *Feli*, who contributed absolutely nothing but brought me company and the best purring background noise one could ask for.

ABOUT THIS DISSERTATION

The present dissertation delineates my endeavour to establish a framework for the analysis of spatial data, with the ultimate objective of contributing to the field of cancer research. As a member of a family of medical professionals, my fascination with medicine and human diseases was kindled from a young age. I was therefore honoured to have the opportunity to contribute to that particular field of science through my work. Drawing upon both computational analysis and foundational biology, with this work I seek to optimise the balance between computational efficiency and biological plausibility.

The present dissertation is structured in the following manner. Chapter 1 serves as an introductory literature review, which underscores the significance of spatial biology in both general scientific research and the specific context of cancer studies. Subsequent to this, I present an overview of the biology of gliomas, which were selected for establishing the methodological framework to derive novel insights. The Chapter concludes with an overview of the spatial profiling technologies and computational methods for the analysis of the associated data.

The computational aspects of my work and underlying rationale are presented in Chapter 2. In this Chapter, I establish the fundamental principles of analysis, identify the most effective methods that align with the specific objectives of spatial studies, and discuss innovative approaches that have been developed within the scope of this work.

The results of the implementation of the derived methods are outlined in two chapters, with each devoted to a distinct project. In Chapter 3, I employ the devised framework for delineating the spatial architecture of necrotic tissue in glioblastoma patients. In Chapter 4, I present a large-scale pan-glioma project, to which I contributed at the early stages by developing the essential prerequisites, meticulously preparing and consolidating the data, and

formulating the initial findings based on my analysis of the data. The two Chapters each commence with a succinct overview of the background and methods relevant to the particular study, followed by the presentation of the analysis results and a discussion of the identified findings.

I conclude with Chapter 5, where I discuss the broader context of my work and outline potential avenues for future development and refinement, building upon the foundation that I have sought to establish.

DISCLOSURES

Hereby, I solemnly affirm that the contents of the present dissertation represent original work that has not been previously published in or submitted to, either in full or in parts, this or any other institution of higher education for the award of any degree or qualification. I hereby declare that the present dissertation is the product of my individual efforts and does not incorporate any contributions from collaborative work, with the exception of those explicitly stated within the text. I hereby state that I employed DeepL to enhance the readability of the text without altering its factual content. I hereby state that I used a combination of Python and the Adobe Illustrator software for the creation of the figures and visual aids. Where relevant, I incorporated and cited figures from previously published works with the publisher's consent.

ZAIRA SEFERBEKOVA

Heidelberg, 2026

1 INTRODUCTION

In this Chapter, I provide a comprehensive literature review to establish the context for subsequent parts of the thesis, which present the research results. The introduction starts with a rationale for studying the spatial aspects of cancer evolution. It then proceeds to delve into the intricate biology of gliomas, which have been selected as a primary subject to establish a framework for spatial analysis. The subsequent parts will address experimental approaches and consider existing computational methods that facilitate spatial analysis of tumours.

Section 1.1, Section 1.3 and Figures therein are reproduced from my first author publication in Nature Reviews Genetics [1]. For that publication, I conducted literature search, designed all figures, and contributed to the writing of the text. In comparison with the published review, the text has been updated and substantially rewritten in order to incorporate new evidence and to suit the presentation of this thesis. The scientific content remains similar. Section 1.2 and Section 1.4 present new contributions.

1.1 SPATIAL BIOLOGY OF CANCER

Cancer is a disease that evolves within the limits of the human body (Figure 1.1). Its growth is naturally restricted by the physical constraints of the organs and influenced by interactions with neighbouring cells. Consequently, it is essential that spatial context is considered in the study of cancer evolution to gain a comprehensive understanding of the disease. Here I examine in what ways spatial architecture influences the evolution of healthy and cancerous tissue.

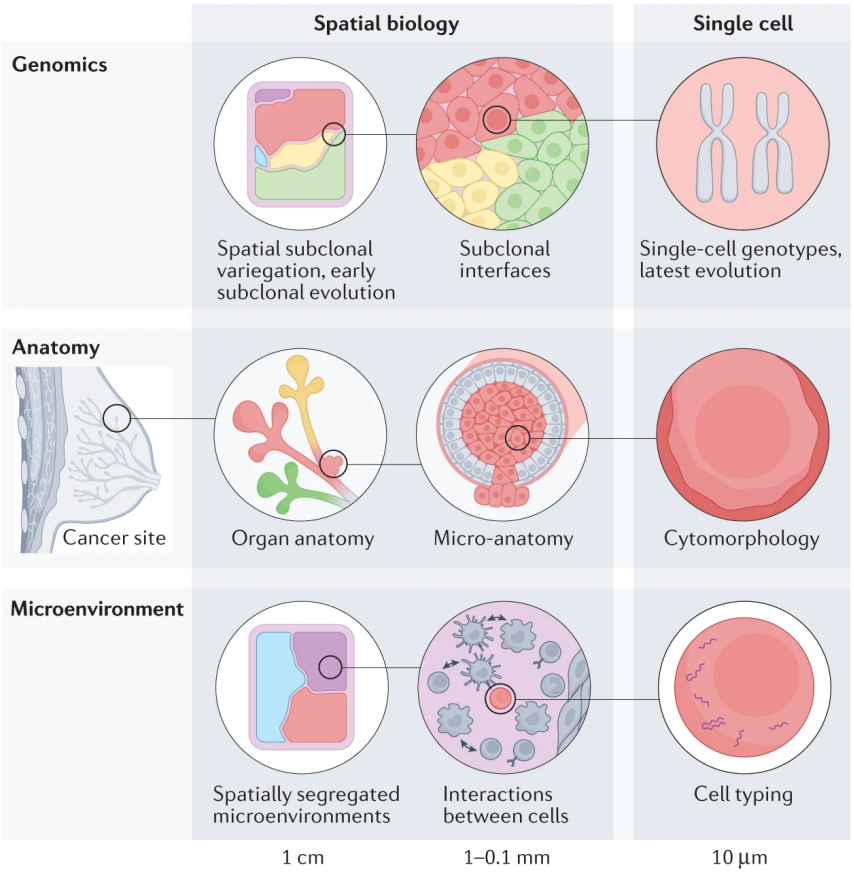

Figure 1.1: *Genetic, anatomical, and cellular contexts of cancer evolution.* Spatial analyses add a new level of complexity to single cell studies of cancer. Adapted with permission from [1].

1.1.1 *Tissue architecture and clonal evolution*

A CLONE *is the progeny of a single ancestral cell. Its increase in size within a cell population is called a* CLONAL EXPANSION.

It has long been established that cancers are mosaics of *clones* shaped by positive selection [2–5]. However, it was only in recent years that the presence of *clonal expansion* in morphologically normal tissues began to be recognised. Evidence suggests that normal cells too exhibit continuous accumulation of mutations in the course of the ageing process, which in turn leads to heterogeneity in healthy tissues [6]. It is conceivable that in the context of selective pressures these mosaics of normal subclones can transition into expanded clones. Spatial context, including tissue architecture and microscopic structure, plays a pivotal role in shaping the spatial evolution of mutant clones [7, 8].

Tissues exhibit a spectrum of microscopic architectures, with some demonstrating greater spatial constraints and others exhibiting less structural complexity (Figure 1.2). An exemplary illustration of the absence of anatomical barriers is the bloodstream, where clonal populations can disseminate throughout the entire blood system and engage in unrestricted competition with one another [9]. Consequently, progeny of a single mutated haematopoietic stem cell can reach a *variant allele frequency (VAF)* of up to 60% — a common phenomenon that is closely associated with ageing and is known as CLONAL HAEMATOPOIESIS [10, 11]. Mutations detected in clonal haematopoiesis include alterations in known cancer DRIVERS, which confer a selective advantage to the carrying clones and thereby facilitate their expansion with age [12].

VAF *measures mutant allele proportion in a population and approximates the clonal fraction.*

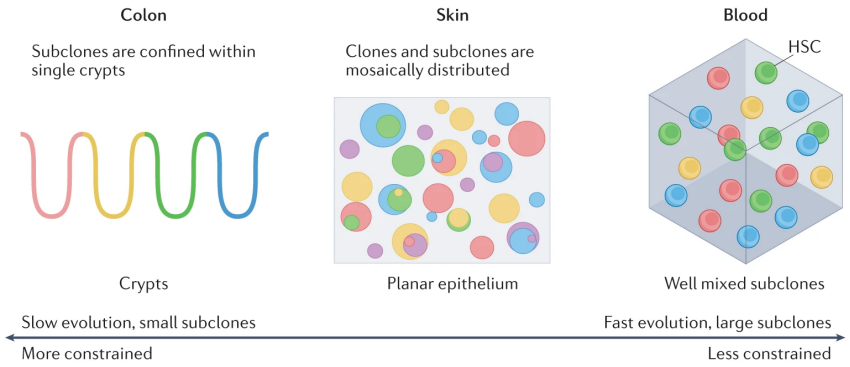

Figure 1.2: *A spectrum of physical architecture constraints in different tissues.* In highly organised tissues, such as the colonic epithelium, the spatial architecture constrains the clonal spread and impedes the evolutionary process. Conversely, the haematopoietic system exhibits no inherent physical limitations on clonal mixing, facilitating the most rapid evolution. Adapted with permission from [1].

In solid tissues, the presence of physical barriers can have a substantial impact on the competition of mutant clones, potentially impeding their growth and dissemination [13]. This is the case in skin, where mutant clones are able to compete and expand only in a two-dimensional manner, and their growth is additionally constrained by interactions with adjacent cells. As a result, aged sun-exposed

skin presents as a mosaic of numerous clones, with the majority exhibiting a VAF of less than 5% and a size below 1 mm², even in the presence of a driver mutation [14]. The transitional epithelium lining the urinary tract also comprises a patchwork of clones which have the capacity to proliferate extensively [15]. However, the majority of clones exhibit low VAFs of approximately 13% and are only a several hundred micrometers in diameter [16]. In the oesophagus, clonal growth occurs concurrently with age, ultimately resulting in the almost complete remodelling of the oesophageal epithelium. The median VAF of somatic mutations in this tissue is low, at approximately 1.6%, yet clones attain considerable sizes, reaching up to 12.8 mm² [17, 18]. Likewise, the liver is composed of a polyclonal community of small cell populations that exhibit minimal genetic similarity [19]. Finally, in the adult brain, where the number of stem cells is minimal, cortex regions also exhibit differences in clonal structure [20].

In tissues characterised by markedly constrained microanatomical features, a mutant clone is limited by the necessity of accommodating the complex spatial structure of the tissue. Consequently, clonal expansions are confined to microanatomical boundaries, resulting in a heterogeneous collection of monoclonal units [13]. For instance, in prostate, clonal expansions can be observed in both epithelial and stromal populations when exposed to selective pressure [21]. The main duct of the glandular system of prostate is constituted of embryonic clones. In the subsequent phase of puberty, new subclones populate novel components of ducts during their progressive branching. These results in a continuum of spatially constrained clonal populations, with a mixed structure in proximity to the urethra and almost pure clonality in distal branches [22]. In the colonic epithelium, where driver mutations are rare, a novel mutant clone will ultimately colonise the entire crypt [23]. The sole means by which a mutant clone can expand further is through crypt fission, which, on average, occurs only once in 27 years [24]. Consequently, all colonic crypts evolve as autonomous, fully clonal populations with only about 1% of the normal epithelium be-

ing occupied by clones with driver mutations [8]. In the stomach, glands are also frequently dominated by a single clone. However, the tubules of stomach glands branch towards the base, which distinguishes them from the straight tubular intestinal crypts and allows further colonisation of the epithelium [25]. Endometrial glands exhibit a complex branching structure similar to that in the stomach. In certain instances, two monoclonal lower ducts can coalesce, giving rise to an upper part with the two lower duct clones coexisting in subclonal states [26].

In certain conditions, such as inflammation or injury, there can be disruption to or remodelling of organ structure. The elimination of limits imposed by spatial constraints present in healthy tissue allows mutant clones to expand further [23]. A demonstrative example is the endometrium of the uterus, where the same clone can populate glands separated by hundreds of micrometers [27]. This phenomenon could be explained by monthly tissue breakdown, which allows new clones to colonise extensive zones of the endometrial lining. Moreover, in the case of endometriosis, an inflammatory condition that frequently results in the breakdown of local anatomical structures, somatic mutations can attain significantly higher VAFs, with some cases reaching up to 93% [28, 29]. In the cirrhotic liver, clones attain a diameter of millimetres, yet remain separated by fibrotic tissue [19]. In ulcerative colitis, as much as 83% of rectal epithelium can be repopulated by clones measuring up to 19 cm^2 [30]. In the instance of chronic inflammation of the stomach, clones in affected remodelled glands demonstrate higher levels of VAF compared to glands with normal gastric epithelium but remain localised in their expansion [25]. In conclusion, the accumulating corpus of evidence suggests a link between the spatial architecture of an organ and the manifestation of the complex clonality in normal tissue.

1.1.2 *Invasion and metastasis as an escape from spatial constraints*

DCIS *is a pre-malignant breast lesion that may progress to invasive ductal carcinoma.*

In the field of cancer research, the phenomenon of constrained growth within the spatial limitations of the tissue is called NON-INVASIVE. In the breast, non-invasive *ductal carcinoma in situ (DCIS)* populate intact ducts without breaching the myoepithelial layer and basement membrane [31]. In DCIS, at higher levels of organisation such as lobules, multiple clones frequently co-exist while pure clonality is exhibited only in microscopic acini and ducts [32]. A comparable pattern is observed in colorectal cancer, wherein a precursor lesion, referred to as monocryptal adenoma, is characterised by a clone populating the entire crypt and propagating solely through crypt fission [23].

Further clonal expansion typically necessitates a breakdown of tissue architecture and tumour infiltration into adjacent normal tissue – a transition to an INVASIVE form of disease [25, 27] (Figure 1.3). Invasive growth frequently coincides with the loss of normal morphological characteristics and is generally associated with more aggressive cancer according to many existing histopathological grading and clinical staging systems [33–36].

The existing subclones can give rise to invasive cancer in a parallel manner, thus demonstrating a branching pattern of evolution [37–39]. To date, coexisting spatial subclones have been identified in multiple solid cancers including, but not limited to, glioblastoma, skin, breast, colorectal, pancreatic, prostate, and renal cancers [37, 40, 41]. With respect to multiple myeloma, diverse patterns of cancer evolution co-exist or alternately emerge over time in conjunction with disease progression; in other instances, clonal expansions were observed to be contingent upon the anatomical location of the given subclone [42]. In general, the mode of tumour evolution can vary depending on numerous factors, such as the extent of cell dispersal [43, 44]. Regardless of the evolution mode, cancer subclonal driver alterations typically exhibit a greater degree of diversity in comparison to early clonal mutations [43, 45].

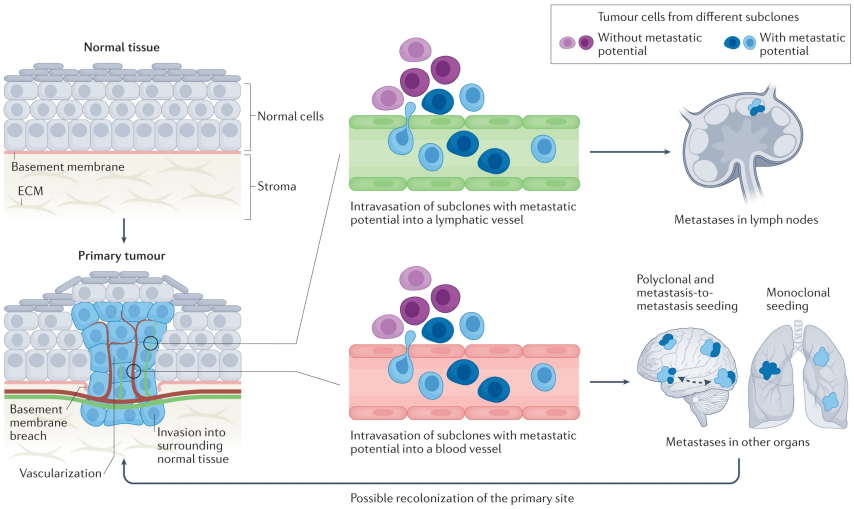

Figure 1.3: *Cancer progression is a spatial process.* In metastatic skin cancer, an invasive tumour disrupts the basement membrane to infiltrate the stroma and ultimately disseminate through the vasculature to the lymph nodes and distant anatomical sites. Adapted with permission from [1].

The ultimate triumph over spatial limitations in the context of cancer is METASTASIS, a process involving the dissemination of invasive tumour cells into the local blood or lymphatic vessels, subsequent circulation towards distant organs, and the seeding of a new metastatic site in them [46] (Figure 1.3). Not all locally invasive tumour cells have the capacity to metastasise. In certain instances, a clone not exhibiting dominance at the primary site can function as a seeding clone for a metastatic tumour [47, 48]. Pan-cancer studies have demonstrated that, for the majority of cancer types, the clones that gave rise to metastases acquired chromosomal aneuploidy already in the early stages of tumourigenesis [41, 49–52]. Following seeding, metastases continuously evolve, acquiring both private clonal and subclonal mutations [38, 53–55], and may undergo further expansions [32]. Nevertheless, the subclonal diversity in primary tumour sites is still significantly higher compared to metastases, which could be explained by different selective pressures acting in the primary and secondary sites [49].

The varying patterns of spread exhibited by different cancer types vary add another layer of complexity. In a linear model, a single seeding event from the primary tumour initiates the process, with subsequent secondary sites emerging through a series of metastasis-to-metastasis seeding events [48, 49, 56]. Another pattern, which is sometimes designated as "clonal diaspora", is distinguished by direct multifocal dissemination from a primary location to discrete metastatic sites [38, 54, 55]. Finally, in certain instances, cross-seeding may occur, when cells from one metastasis disseminate and seed a new metastatic tumour at a distinct location [48, 54, 56]. The precise factors that determine the metastatic spread pattern remain to be elucidated, although evidence indicates that treatment may impact this process. For instance, in colorectal, breast, and lung cancers, untreated metastases were characterised by clonal homogeneity. Conversely, administration of treatment prior to metastasis resulted in a substantial increase in the proportion of private clonal drivers [49]. Moreover, novel driver mutations are frequently associated with treatment resistance [50, 52]. These observations suggest that therapy may introduce new selective pressures, thus promoting further tumour evolution.

In summary, considerable evidence supports the hypothesis that tissue architecture profoundly affects cancer progression. However, it is imperative to take into account the cellular context, as the interactions of cancer cells with adjacent tumour and stromal cells are also poised to influence the trajectory of cancer evolution.

1.1.3 *Cellular context of cancer evolution*

A further significant factor contributing to tumour heterogeneity arises from its cellular context. The phenomenon of cross-talk between tumour cells has been a subject of research for a considerable period [57]. In the recent years, it is becoming increasingly recognised that stromal cells, which constitute the organ where the malignancy has manifested, can impact the growth and evolution of cancer.

Consequently, it is imperative to take the tumour microenvironment (TME) into consideration when studying cancer biology.

Cancer clones are engaged in a continuous battle for spatial dominance, not only with other cancer clones but also with stromal cells. In the oesophagus, early neoplastic lesions compete with mutant stromal clones located in the basal layer [58]. Consequently, if the mutant normal clone expands sufficiently to displace the nascent malignant clone, the entire cancer lesion is shed. A similar pattern can be observed in a colonic crypt, wherein the colonisation of the entire crypt by mutant cells is contingent upon the mutation arising in stem cells [23] (Figure 1.4). However, in certain instances, the development of cancer is facilitated by the colocalisation with other clones. This is evidenced by the findings in murine models of intestinal adenomas, wherein a driver mutation alone was insufficient for the formation of cancer [59]. The initiation of cancer required an increase in the density of mutant colonic crypts, thus demonstrating an example of cooperation.

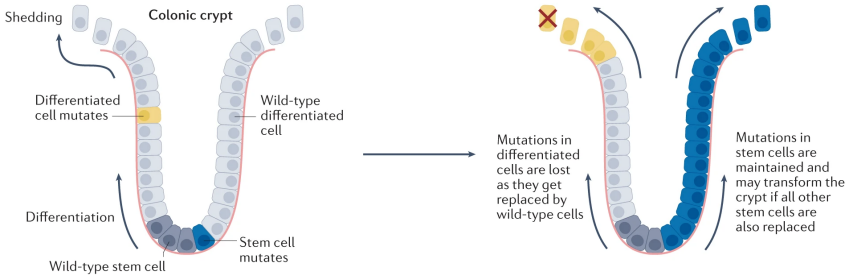

Figure 1.4: *Colonic crypt organisation defines clonal expansion and tumour evolution.* The anatomical structure can reduce the likelihood of malignant transformation by constraining the fixation capacity to mutations in few cells. Adapted with permission from [1].

It is evident that specific TMEs are frequently associated with distinct cancer subclones, transcriptional profiles, and clinical outcomes [37, 60–65]. For instance, colorectal tumours exhibiting chromosomal instability were found to be characterised by an immunosuppressed TME, whereas hypermutant lesions contained a high number of

cytotoxic T cells [39]. In cutaneous squamous cell carcinoma, basal tumour cells colocalised with keratinocytes at the leading edge [66]. In breast cancer, tumour cell expression correlated to the proximity of the cells to lesion edges [37]. Distinct phenotypes of tumour–immune interactions were found to be associated with differing survival rates amongst pancreatic [67] and breast [68] cancer patients. Furthermore, the reduction in thickness of the myoepithelial layer, alongside the replacement of normal fibroblasts with cancer-associated fibroblasts, was found to act as a protective factor against the progression of DCIS towards invasive breast cancer [31].

Recent advances in spatial profiling methodologies have enabled subcellular resolution, facilitating the examination of not only cellular colocalisation but also cell-to-cell interactions [65]. The capacity of tumour and stromal cells to exhibit different ligand–receptor interactions was shown to depend on the respective spatial neighbourhoods in cutaneous squamous cell carcinoma [66]. In the context of melanoma, the presence of a stem-like population of neural crest cells in close proximity to endothelial cells was demonstrated to be a contributing factor to the cancer growth. The underlying mechanism of this process depended on the *NOTCH3* receptor, which expression on neural crest cells preceded their *epithelial-mesenchymal transition (EMT)* [47].

In cancer, EMT describes the loss of polarity and adhesion alongside the gain of potency and invasiveness by epithelial cells.

1.2 BIOLOGY OF GLIOMAS IN ADULTS

The recent emphasis on incorporating a variety of molecular characteristics into diagnostic guidelines emphasises the necessity of understanding the intricacies of tumour biology to facilitate clinical decision-making. A notable illustration of this can be seen in the World Health Organisation (WHO)'s recent refinement of the classification of central nervous system (CNS) tumours, which introduced molecular parameters as a crucial component of the classification system, complementing the traditional histological characteristics [33]. Furthermore, DNA methylation pro-

filing has been utilised to distinguish clinically-relevant tumour entities that were previously considered homogeneous groups [69].

The present thesis investigates the spatial architecture of gliomas, which are the most common malignant primary CNS tumours [70]. In this Section, I explore the existing literature concerning the intricate biology of gliomas.

1.2.1 *A diverse group of CNS tumours*

For an extended period, clinicians have relied on histological findings, supported by tissue-based tests, for the diagnosis of CNS tumours [71] (Figure 1.5). Nonetheless, molecular biomarkers are increasingly being incorporated into diagnostic practice [33]. In the latest edition of WHO classification of CNS tumours, gliomas are classified into six families: paediatric-type diffuse low-grade gliomas, paediatric-type diffuse high-grade gliomas, adult-type diffuse gliomas, ependymal tumours, glioneuronal tumours, neuronal tumours, and circumscribed astrocytic gliomas [33]. Although gliomas can occur at any age, the vast majority of gliomas belong to the adult-type diffuse group [70]. Within each tumour type, the grading of CNS tumours is analogous to the grading of other tumours [72]. In essence, grade 1 tumours typically exhibit slow growth and clear margins, often resulting in a favourable prognosis if the tumour can be resected completely. Conversely, grade 4 tumours are characterised by invasive growth and high degree of malignancy and are associated with a poor prognosis.

In the classification of adult-type diffuse gliomas, the most significant molecular criteria are mutations of the *isocitrate dehydrogenase (IDH)* gene and the loss of the short arm of chromosome 1 together with the long arm of chromosome 19 (denoted as "1p/19q"). Accordingly, gliomas in adults are categorised into three distinct types: IDH-mutant astrocytoma, IDH-mutant and 1p/19q-codeleted oligodendroglioma, and IDH-wild type glioblastoma (GB) [33]. IDH-mutant gliomas exhibit increased 2-hydroxyglu-

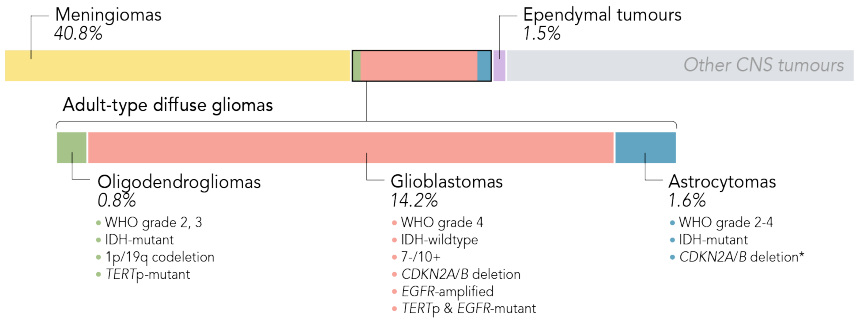

Figure 1.5: *Diverse CNS tumours and selected key characteristics.* Percentages indicate tumour frequency among all CNS tumours. *, intact in low-grade astrocytomas (LGAs). Based on [33, 70].

tarate levels, an accumulation of lipid clusters containing phospholipids and monounsaturated fatty acids, and a reduction in NADPH production [73]. The WHO grade 4 GBs represent the most aggressive glioma type that accounts for more than one half of all malignant CNS tumours [70]. The earliest copy number variation (CNV) event in the pathogenesis of GB is chromosome 10 loss followed by chromosome 7 gain (frequently designated as "7+/10-") [74, 75]. The most common gene alterations include *EGFR* and *PDGFRA* amplifications, recurrent mutations in *TERT* promoter and *EGFR*, *PTEN*, and *TP53* genes, and homozygous deletions of *CDKN2A/CDKN2B*. An aberrant exon 1–8 junction in the *EGFR* gene (known as "EGFR class III variant" or "EGFRvIII") represents a recurrent mutation in GB that is associated with elevated levels of triglycerides and polyunsaturated fatty acids [73]. In addition, the recent recognition of the clinical significance of DNA methylation profiling has led to the categorisation of GBs into three distinct groups: receptor tyrosine kinase (RTK)1, RTK2, and mesenchymal (MES) [69]. Patients diagnosed with either RTK class exhibit reduced survival rates and may benefit more from maximal resection in comparison with patients diagnosed with GB of the MES class [76, 77].

At present, no prevention strategies or screening protocols for early detection are available for adult-type diffuse gliomas. At the time of disease manifestation, patients typ-

ically exhibit non-specific neurocognitive symptoms [78]. Magnetic resonance imaging (MRI) is utili-sed as a pre-operative detection method, yet clinical decision-making continues to be dependent on tissue biopsy. Surgery is performed for two primary reasons: firstly, to diagnose the condition, and secondly, to remove as much of the tumour volume as is feasible. However, due to the location of diffuse gliomas and their highly infiltrative growth, total resection is often not possible. Following surgical removal, patients are subject to a combination of chemo- and radiotherapy, with the choice of detailed strategy informed by the specific molecular alterations present in a tumour [79]. For instance, alkylating agents such as temozolomide (TMZ) are frequently employed in the treatment of gliomas in patients undergoing chemotherapy. Nevertheless, it is now evident that, in the case of GB, its efficacy is predominantly beneficial to patients with tumours exhibiting *MGMT* promoter methylation [80]. Furthermore, the results of a phase III trial, known as INDIGO, demonstrated the potential benefits of IDH inhibitors for patients diagnosed with IDH–mutant grade 2 gliomas [81]. A plethora of innovative therapeutic modalities, encompassing immune checkpoint inhibitors and agents that target mutated proteins, are currently undergoing evaluation in the context of glioma management [82–84]. Nonetheless, the efficacy of these treatments is yet to be substantiated through definitive studies. Subsequent to the standard of care treatment, the ensuing follow-up entails a watch-and-wait approach, accompanied by frequent clinical examinations and MRI [78].

Ependymoma (EPN) is a rare CNS tumour of neuroepithelial origin [70] (Figure 1.5). According to the fifth edition of WHO classification [33], EPNs are classified based not only on histopathological and molecular features, but also on their anatomical location. Three distinct sites of EPN are recognised, namely supratentorial, posterior fossa, and spinal. Intracranial EPNs are rare in adults, and are more prevalent among paediatric and young adult populations. Conversely, spinal ependymomas (EPN-SPs) manifest more frequently in adult patients than in children

[85]. In addition, myxopapillary ependymoma (EPN-MPE) and subependymoma (EPN-SE) are recognised as distinct glioma types [33]. EPN-SEs are considered to be benign neoplasms that manifest in all three anatomical compartments [86], whereas EPN-MPEs constitute a specific glioma type affecting the lower spinal cord in adults [85]. EPN-SPs with *MYCN* amplification and posterior fossa EPN-SEs are known to exhibit an especially aggressive disease course. EPNs present a valuable research opportunity in the context of molecular biomarkers, given the recognised inadequacy of the WHO grading system in predicting clinical outcomes [87]. This underscores the potential for molecular analysis to provide novel insights into prognostic classification.

Nonetheless, despite extensive research conducted over many years [88–90], an ultimate cure for gliomas is yet to be discovered.

1.2.2 *High plasticity and heterogeneity of gliomas*

One potential explanation for the resistance of gliomas to therapeutic interventions is their plastic and heterogeneous nature. The heterogeneity of gliomas at the level of transcription was initially demonstrated in the The Cancer Genome Atlas bulk RNA sequencing (RNA-seq) data of GB [74]. In that study, four transcriptional subtypes were described: classical, mesenchymal, proneural, and neural. It is important to note that further research revealed the latter subtype to be non-tumour specific and likely a result of sample contamination with normal brain tissue [91]. Nonetheless, the subtypes were associated with differing survival rates, and delineated distinct spatial regions of the same tumour. Furthermore, their proportions shifted at recurrence, demonstrating a clinical relevance [74].

Subsequent single-cell RNA-seq (scRNA-seq) studies yielded congruent transcription states in adult-type gliomas, with three (namely, astrocyte (AC)-, oligodendrocyte progenitor cell (OPC)-, and neural progenitor cell (NPC)-like) resembling neurodevelopmental cell types, and an addi-

tional MES-like state [92–94] (Figure 1.6a). Of these, NPC- and OPC-like states were characterised by increased proliferation and reduced differentiation, while AC- and MES-like states represented a more differentiated progeny [95]. Similarly to what has been observed previously, a tumour may comprise cells of multiple states, and, due to high plasticity, these cells may undergo state transitions [92].

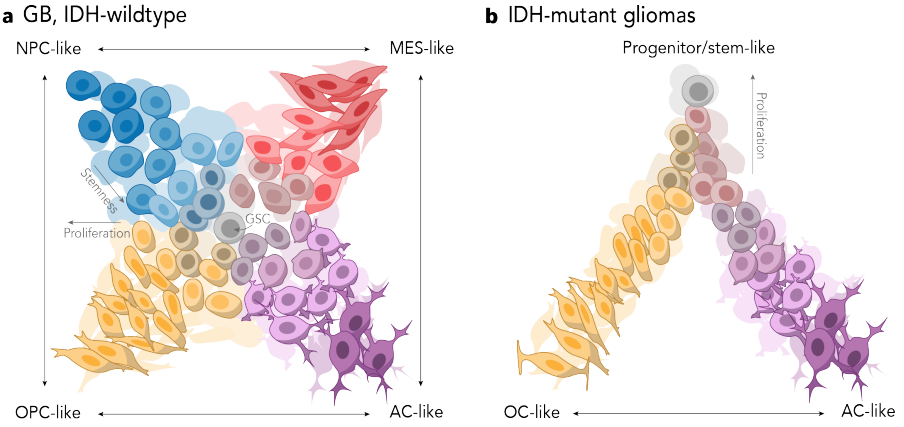

Figure 1.6: *Glioma states.* Four transcription states of GB *(a)* and IDH-mutant gliomas *(b)*. GSC, glioma stem cell. Illustration based on [96].

Similar transcription states were found in IDH-mutant gliomas [97–99] (Figure 1.6b). However, these tumours comprised primarily three states: non-proliferating and more differentiated AC- and oligodendrocyte (OC)-like states, and a cycling state resembling a progenitor or stem population. While the proportions of the states remained unchanged at recurrence, they did vary across grades. In addition, IDH-mutant gliomas demonstrated a generally lower degree of cellular plasticity and a more stable transcription state hierarchy [100].

Transcriptional heterogeneity of EPNs has not been as extensively studied, yet some understanding is emerging. It has been observed that EPN cells manifest in a variety of transcriptional states, encompassing cycling, undifferentiated, and differentiated populations as well as metabolic programs [101, 102]. The proportion of undifferentiated and cycling states was associated with more aggressive EPN groups, whilst the more benign EPNs were charac-

terised by a higher prevalence of differentiated programs [101]. Additionally, three hierarchical trajectories were identified, all beginning with the undifferentiated neural stem cell (NSC)-like state and each one progressing towards either undifferentiated glial progenitor-like, undifferentiated NPC-like, or differentiated ependymal-like cells (through an intermediate astroependymal-like state) [101]. Cycling and differentiated programs were shared amongst diverse EPNs, while undifferentiated cell states were more distinctive of a specific type [101] or existed in a mutually exclusive manner and followed divergent differentiation trajectories depending on the EPN type [102]. Interestingly, some of EPN programs resembled those identified in adult-type diffuse gliomas [101]. In particular, the undifferentiated NPC-like program was similar to the diffuse glioma NPC-like program, one of the metabolic programs that was enriched in glycolytic signature resembled the MES2-like glioma state, and the intermediate astroependymal-like program correlated with the AC-like program in GB. In contrast, the undifferentiated NSC-like and differentiated ependymal-like EPN programs did not correlate with any known adult-type diffuse glioma states, likely indicating cancer cell populations that were unique to EPNs.

Despite the long-standing recognition of the heterogeneity of gliomas, the precise origin of their transcriptional variability remains to be elucidated. It is conceivable that distinct progenitor cells could give rise to separate populations of glioma cells, thus generating the observed cellular complexity. In the adult human brain, only a small number of cells have retained the capacity for self-renewal and the generation of new cells. Consequently, NSCs and OPCs have long been hypothesised to be the cells of origin for gliomas [103]. Although the latter are established to be restricted to the oligodendrocyte lineage, under certain conditions, they can undergo dedifferentiation into a multipotent stem-like state [104]. The aforementioned glioma transcription states bear a resemblance to these neurodevelopmental cell types, thus lending further credence to the cell of origin hypothesis. A recent study of GB using scRNA-seq reported a fifth glioma state resembling

progenitor-like cells [93]. These progenitor-like cells exhibited a high rate of proliferation and had a capacity to differentiate into all glioma states. Nevertheless, other studies do not appear to support the notion of a defined progenitor population [92, 105]. In fact, a xenograft study suggests that all four states may in fact be generated by any glioma cell [92].

A conceivable explanation could be that the glial progenitor-like cancer cells are not a distinct cell type, but alternatively, a plastic cellular state guided by external TME cues [106]. This is substantiated by the observation that the discrepancy between glioma samples from disparate anatomical locations is more pronounced than that between samples from distinct tumours [107, 108]. Furthermore, in EPNs, brain-resident TME cells were proposed to have the capacity to promote the differentiation of progenitor cells towards a particular differentiated state [102]. The advent of novel technologies has led to a proliferation of research exploring spatial glioma heterogeneity, a subject that will be addressed in the following Subsection.

1.2.3 *Spatial architecture of gliomas*

In gliomas, the concept of spatial niches has been recognised for many years, with the initial classification being based primarily on histological observations [109]. In recent years, a growing body of evidence derived from single cell studies has begun to provide a more profound understanding of the spatial organisation of gliomas [110–113].

A necrotic core has been recognised as an intrinsic feature of GB, characterised by the presence of macrophages and glioma cells that are exposed to nutrient-deprived conditions under hypoxia [95, 114] (Figure 1.7). The perinecrotic glioma cells were shown to express a quiescent and wound response gene signatures [95]. Furthermore, hypoxia was demonstrated to facilitate the organisation of gliomas in a spatial manner, whereas non-hypoxic regions generally exhibited a lower degree of organisation [102, 115]. Chapter 3 is devoted to a comprehensive examina-

tion of the spatial architecture of necrotic tissues in individuals previously diagnosed with GB. Vessels (which are frequently abnormal in gliomas), pericytes, astrocytes, and various types of immune cells are typically categorised into a peri-vascular niche. Endothelial cells within the perivascular niche are thought to contribute to glioma stem cell (GSC) maintenance, invasion, and tumour growth [116]. The invasive edge located at tumour periphery is thought to contribute to tumour recurrence and is comprised of invasive, stem-like cycling glioma cells as well as microglia and mature AC-like glioma state [95, 116, 117]. Recent studies identified an additional proliferative niche composed of glioma cells expressing OPC- and NPC-like signatures [110].

A growing body of research employing spatial profiling suggests a potential association between spatial neighbourhoods and clinical outcomes in glioma patients. For instance, in GB, increased survival was found to be associated with macrophage-enriched neighbourhoods as well as neighbourhoods composed of AC-like, NPC-like, and MES-like glioma states [111, 113]. Conversely, a colocalisation of AC-like, OPC-like, and MES-like states was identified as a negative prognostic factor [113]. In Chapter 4, I discuss the findings of a study examining spatial architecture across multiple glioma types.

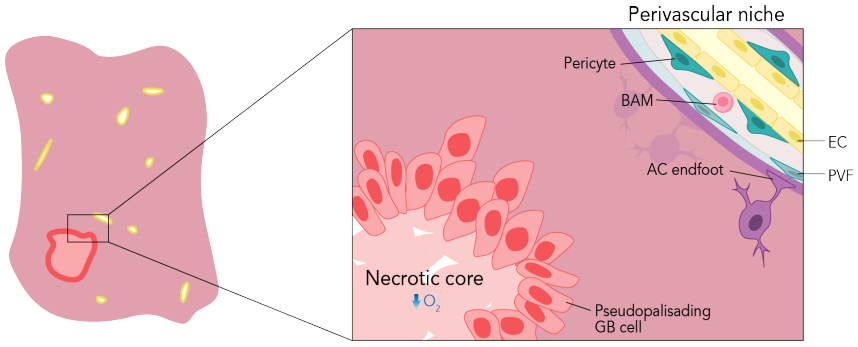

Figure 1.7: *Spatial niches in GB.* Pseudopalisading necrosis and a perivascular niche comprised of vascular and immune cells. BAM, border-associated macrophage; EC, endothelial cell; PVF, perivascular fibroblast. Illustration based on [109, 118].

Finally, there has been a recent surge of interest in so-called glioma tumour networks. Specifically, glioma cells were demonstrated to establish connections with each other and with astrocytes via gap junctions, thereby forming a functional, communicating synthicium [117]. The invasive edge was characterised by a predominance of unconnected glioma cells, whereas connected glioma cells remained stationary within the bulk of the tumour. Furthermore, a small proportion of tumour cells, which act like periodic cells and express high levels of the potassium channel gene, were observed to induce calcium oscillations within the tumour network [119]. The dynamic nature of these periodic glioma cells' identity was evident. Additionally, up to ten percent of both connected and unconnected glioma cells were observed to form heterogeneous glutamatergic synapses with neurons. It was hypothesised that through these synapses neurons may induce calcium oscillations within glioma networks [120].

In light of the evidence presented, it can be concluded that the spatial context is of paramount importance in the investigation of cancer evolution. As computational and experimental methodologies evolve rapidly, they are enabling the generation of increasingly extensive atlases with superior resolution. The review of technologies and analytical methods will be the focus of the subsequent Sections.

1.3 EXPERIMENTAL TECHNOLOGIES FOR SPATIAL PROFILING

The advent of numerous single cell methodologies has engendered a paradigm shift in our understanding of cellular properties and variability [121–124]. Nevertheless, the initial stage involves dissociating solid tissues and organs to yield a single cell sample for subsequent profiling. This step consequently results in the loss of crucial spatial information, thereby hindering the capacity to evaluate the spatial organisation of single cell architectures. A plethora of technological innovations has facilitated the spatial genomic, epigenomic, transcriptomic, and proteomic profil-

ing. The choice of the most appropriate platform depends on the specific scientific problem to be addressed (Figure 1.8a). The key factors to be considered are resolution, throughput, field of view size, and sensitivity.

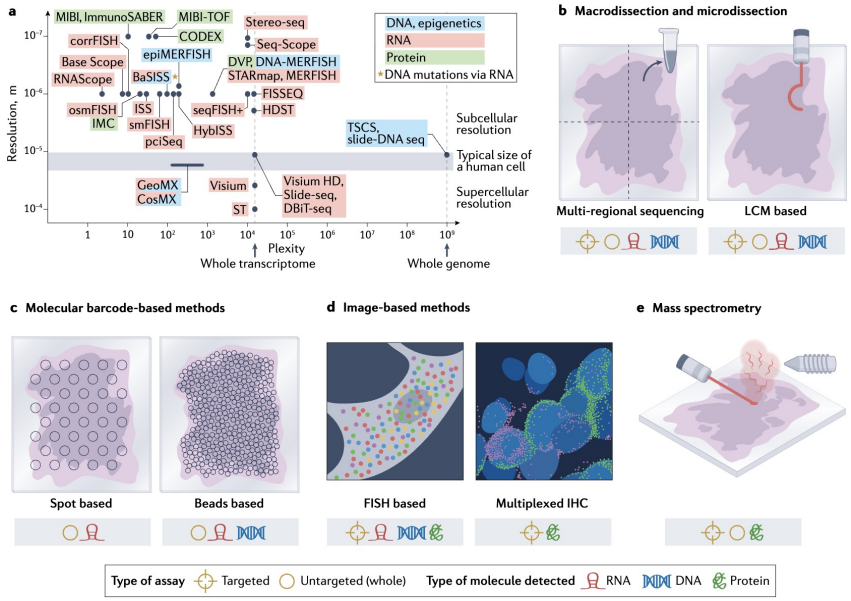

Figure 1.8: *Experimental approaches to study spatial cancer ecosystem.* Higher resolution typically implies lower multiplexity *(a)*. Spatial methods can be grouped into macro- and microdissection-based *(b)*, barcode-based *(c)*, imaging-based *(d)* methods or methods employing mass spectrometry *(e)*. Re-printed with permission from [1].

1.3.1 *Sequencing-based methods*

The foundation for sequencing-based approaches was laid by tissue dissection techniques, which entail the extraction of numerous sample regions that researchers subsequently subject to bulk sequencing [125–129] (Figure 1.8b). The resolution of tissue dissection methods depends on the cutting technology employed. In some cases, the resolution can reach microscopic levels, or even the level of individual cells. The known position of the extracted samples provides the spatial resolution for the resulting bulk data sets. Furthermore, the capacity of tissue dissection methodolo-

gies to integrate with any subsequent high-throughput sequencing renders them remarkably versatile and applicable to a wide range of tasks [130]. A frequently employed tissue dissection method is laser-capture microdissection, which utilises laser ablation to isolate minute regions of interest [131–133]. An alternative approach typically employed in the analysis of large anatomical structures entails the acquisition of discrete tissue cryosections through serial microtomy [134, 135].

The recent advancement of spatially-resolved sequencing-based methods entails the use of spatial molecular barcodes [136–142] (Figure 1.8c). In essence, spatial molecular barcodes define unique sequences for each location, which enables the subsequent mapping of the captured molecules to their original spatial coordinates. The primary advantage of the sequencing-based methods is their capacity to generate unbiased high-throughput data. Nevertheless, the resolution of such approaches is inherently constrained by the dimensions of the designated spots, which typically fall below the size of a single cell. Despite ongoing development of enhanced-resolution approaches [143–146], currently existing sequencing-based methods generate bulk cell profiles within each spot, necessitating additional analytical procedures to comprehensively assess cell-level composition.

1.3.2 *Image-based methods*

Over a span of many years, *in situ* hybridisation of fluorescently labelled probes has played a pivotal role in elucidating the spatial distribution of DNA and RNA with subcellular resolution [147, 148]. This traditional approach has later undergone a quantitative evolution, characterised by an increase in *multiplexity* [149–153] (Figure 1.8d). Notable advances were achieved through the development of sequential barcoding, wherein each molecule is identified through several cycles of hybridisation, with the order of fluorescent probes detected defining the final sequence [154, 155]. Several methodologies employ a similar

MULTIPLEXITY *denotes the capacity to simultaneously profile multiple molecules within a given sample.*

approach, sequential antibody staining, for spatial protein profiling [156–158]. However, sequential antibody staining approach is constrained by the necessity for a considerable period of incubation. Sequential antibody staining, utilised by several methodologies for spatial protein profiling [156–158], has limited appeal due to the significant time investment required for antibody incubation. An alternative method, termed *in situ* sequencing, is based on the use of fluorescently labelled padlock probes and rolling circle amplification [159–163]. Given its high specificity, *in situ* sequencing was furthermore adapted for the detection of mutations [32]. In the case of multiplexed protein analysis, it is possible to visualise DNA-barcoded antibodies using *in situ* polymerisation with fluorescently labelled nucleotides [164].

Image-based methods are capable of providing subcellular resolution, yielding readouts originating from individual cells. Nevertheless, the present limitation is a highly restricted number of targeted molecules, typically up to hundreds. Furthermore, the necessity of defining the panel in advance confines the investigation to the study of a priori defined phenomena. Novel approaches promise to offer larger target panels, with some already reaching up to several thousands of targets [165].

1.3.3 *Mass spectrometry-based methods*

The spatial distribution of proteins is frequently studied by means of mass spectrometry (Figure 1.8e). In this class of methods, the target proteins are firstly labelled using metal-conjugated antibodies, and then the tissue sections are subjected to laser [166] or ion beam ablation [167–169]. The application of mass spectrometry can be integrated with other methodologies, for example, with laser capture microdissection [170] or combinatorial tagging [171]. In addition to proteins, mass spectrometry-based methods are capable of analysing even smaller molecules, lipids and metabolites in particular. Nevertheless, the efficacy of these methodologies depends on the specificity of the an-

tibodies employed, and their utility is constrained by the size of the panel, which typically enables the profiling of only a few dozen proteins. The development of methodologies that facilitate the spatial profiling of the complete proteome at the level of individual cells remains a subject of ongoing research.

To summarise, a plethora of innovative experimental approaches permit spatial proteomic, transcriptomic, and metabolic characterisation of tissue sections. Nonetheless, the advent of novel technologies that facilitate profiling of extensive fields of view, yield exhaustive readout (e. g.whole transcriptome) with subcellular resolution, and enable concurrent spatial profiling of multiple modalities from a single sample is required to facilitate a nuanced comprehension of the intricacies of cancer biology.

1.4 COMPUTATIONAL TOOLS FOR SPATIAL DATA ANALYSIS

Given the nascent nature of spatial profiling technologies, there is currently no established set of best practices for computational analysis of these data. The following Section will examine the computational challenges in spatial transcriptomics data analysis and evaluate potential approaches to address these challenges. The primary focus of this Section will be data that can be generated using 10x Xenium, as this is the protocol that has been implemented in the projects in this dissertation.

1.4.1 *Gene panel design*

In the context of conventional scRNA-seq data analysis, the existence of whole transcriptomes provides comprehensive cell expression profiles, enabling detailed analysis of cellular states and the identification of transcriptomic signatures. Conversely, in image-based spatial transcriptomics, the limitation of panel size gives rise to a combinatorial problem, necessitating a balance between the identification of all possible cell types and the ability to explore

greater expression diversity within these cell types. The optimal balance between these two objectives remains an open challenge and is contingent on the specific objectives of the study in question.

CELL TYPE RECOVERY. The ability to successfully identify and categorise cell types is fundamental to all subsequent analyses. In spatial contexts, the failure to annotate a cell introduces an additional difficulty — artificially missing parts of the sample. This complicates later spatial pattern and interaction discovery and may even lead to false conclusions. A natural approach for marker gene selection is to use known literature *markers* [172]. To this end, curated cell type marker databases could be used, such as The Human Protein Atlas [173], the Human Cell Atlas [174], PanglaoDB [175], and CellMarker [176]. The evolution of scRNA-seq technologies has provided a new way to study cell type-specific expression in detail. In the presence of an annotated scRNA-seq atlas for the chosen tissue and condition, cell type marker selection amounts to a statistical comparison of the gene expression of each cell cluster against the rest [177]. A multitude of methodologies exists for such analyses, with the majority relying on existing annotations [178–182], but some offering cluster-free approaches [183–186].

A good MARKER *is a gene with high expression in only one cell type.*

EXPRESSION DIVERSITY. To facilitate a comprehensive understanding of sample biology that extends beyond the spatial localisation of individual cell types, it is essential to include genes that drive expression heterogeneity across cell populations. One potential strategy involves incorporating genes that are components of signalling pathways, metabolic networks, or biological processes, which are of particular interest in this context. To this end, several existing databases of conserved pathways could be utilised, the Molecular Signatures Database [187] being one of the commonly used ones.

An alternative approach that does not require prior knowledge of gene function is to include top *highly variable genes (HVGs)* as their variable expression is believed to un-

derlie specific functional diversity of cells within a given tissue [188]. A different mathematically-driven approach is to apply a dimensionality reduction algorithm to preserve as much information within the data as possible while reducing the overall number of dimensions. One common linear approach, known as principal component analysis (PCA), is based on the ranking of linear combinations of features (i. e.genes) according to the amount of variation they capture [189]. The main advantages of this method are its computational efficiency and interpretability, as PCA enables gene importance evaluation through the computation of *loadings*. PCA-based methods outperform alternative approaches in terms of variation recovery [190]. Nevertheless, a potential constraint of PCA arises from the exclusive density of all components. In the context of gene expression, sparse vectors may be a more appropriate choice, as they clearly delineate individual gene contributions. To this end, a variant of PCA termed sparse PCA (sPCA) that is designed to identify sparse components for optimal reconstruction of initial data could be utilised [191].

HIGHLY VARIABLE GENES *are defined and ranked by their variance-to-mean expression ratios.*

PCA LOADINGS *correspond to the correlation between original features and principal components.*

STUDY-SPECIFIC CONSIDERATIONS. To design a panel, it is also necessary to consider the objectives of a specific study. While it is challenging to propose a strictly outlined set of steps to follow, there are a number of considerations that can be made in advance, based on the study question. For instance, additional markers can be incorporated for a particular cell type that is known to be difficult to recover, or alternatively, markers recapitulating different levels of granularity (e. g.neurons and inhibitory neurons) can be introduced [192]. Furthermore, in the context of cancer, it is imperative to differentiate between malignant and non-malignant cells. To this effect, the inclusion of well-known tumour drivers that are anticipated to be homozygously lost or overexpressed specifically in cancer cells can be considered [193]. An alternative strategy is to select genes uniformly across all chromosomes, thus enabling CNV analysis to be performed at a later stage to identify malignant cells [194].

1.4.2 *Spatial signal assignment*

Prior to the analysis of spatial single cell data, detected molecules need to be assigned to individual cells. This step assumes that cellular boundaries are known, thus necessitating prior cell segmentation. Despite numerous attempts to develop a generalised algorithm for all tissues [195–198], there is still no universally recommended tool that works well in all settings.

The raw output of image-based spatial transcriptomic profiling typically comprises nuclear boundaries, as determined by a segmentation run on the nuclei-stained (DAPI) image, and cell boundaries, as determined by expanding the corresponding nuclei [199]. However, in a recent benchmark [200], alternative approaches based on Bayesian mixture models [195] performed best, although nuclear transcripts were sufficient to recover all cell types. Given the high level of dependency of downstream analyses on the selected segmentation algorithm, a more prudent approach would be to define cell types based on nuclear transcripts and thereafter incorporate cytoplasmic transcripts to investigate further expression diversity.

Nonetheless, nuclear masks may be a suboptimal solution for segmenting cells with irregular morphology [201]. In such cases, it is possible to rely upon grouping transcripts into signatures without delineating cell boundaries. Several such segmentation-free approaches have been developed [201–204], which, in addition to their primary function, may also assist in the identification of subcellular and extracellular compartments.

1.4.3 *Spatial single cell data analysis*

Once the spatial transcriptomic signal has been assigned to cells, the computational workflow is similar to a streamlined scRNA-seq analysis. Although there are many publications that describe current best practices (for example, see [172]), there is a range of important considerations for

working with spatial data that will be outlined in this Subsection.

CELL ANNOTATION. A typical scRNA-seq analysis includes cell filtering and count normalisation, which may not be optimal for spatial data [205]. As discussed above, in a spatial context, filtered cells pose additional problems (e. g.non-biological areas of absent tissue) that complicate further analysis. Furthermore, spatial patterns in the distribution of transcript counts may reflect a biologically relevant sample structure [206] and therefore any raw count transformations should be made with care. To overcome these challenges and to recover as many cells as possible, an alternative strategy of label transfer can be employed [207].

BIOLOGICAL VARIABILITY. Following initial processing, transcriptomics data can still contain unwanted variability that may mask interesting signals. A common example of this is the cell cycle, which can dominate cell clustering and therefore compromise the detection of less prominent cellular states [172]. In addition, tests assessing differentially expressed genes are sensitive to technical confounding factors [208]. To remove unwanted variability, a variety of batch correction methods can be applied [209]. However, overcorrection may lead to loss of heterogeneity — a crucial characteristic of many tumours [210]. Furthermore, in the case of cancer, cycling cells often represent a biologically-relevant population [211]. Therefore, whether or not to batch correct depends entirely on the specific research question.

TRANSCRIPTIONAL PROGRAMS. Manual annotation relies on pre-existing knowledge and cannot identify novel transcriptionally distinct populations. To discover new transcriptional patterns hidden in the data, unsupervised learning tools should be used [210]. An intuitive choice is *matrix factorisation (MF)*, as genes encoding parts of the same biological pathway should be co-regulated across samples and can therefore be represented as latent components of

MF *decomposes a matrix into two matrices of lower rank.*

the count matrix [212]. A popular algorithm commonly used in scRNA-seq is non-negative MF (NMF), owing to the strict non-negativity of transcriptional data [213–215]. NMF finds interdependent, additive and equally important components that can be interpreted as transcriptional programs. Since NMF handles dropout events by imputing missing signal, it is well suited to sparse single cell data. However, like other unsupervised algorithms, the number of components must be specified in advance, and the choice of an optimal value remains an open problem in the field. In addition, initialisation can greatly affect the convergence rate as well as the result, since the objective functions for NMF are non-convex [216]. Random initialisation has a low computational cost; however, it does not generally generate reproducible results. Hence, to select the best number of components and the optimal local minimum, the algorithm should be run a number of times with a different starting point. This drastically increases the computation time and reduces the advantage of the approach. A more sophisticated starting point can be calculated by using a low-dimensional representation of the input matrix. Theoretically, this should result in a deterministic model that gives meaningful results and only needs to be run once [216]. Nonetheless, in this approach, additional considerations such as biological relevance have to be taken into account when choosing the optimal number of components.

1.4.4 *Spatial exploration*

The ultimate objective of applying spatially resolved approaches is to gain insight into the spatial heterogeneity of tissue. Spatial information can be utilised in numerous ways, and biological hypotheses should inform the method selection. This Subsection will delineate a number of general techniques, extending from visual observations to quantitive associations defined by local neighbourhoods.

VISUAL ASSESSMENT. When a single cell atlas consists of only tens of cell types; therefore, spatial exploration may begin simply by examining the location of individual cell types [95]. Furthermore, when the number of targeted genes is low, it is feasible to visually explore the spatial expression of the selected genes of interest. The identification of spatial patterns is already facilitated by such straightforward approaches, especially in tissues that exhibit clearly defined cell zonation [110, 217].

SPATIAL STATISTICS. The calculation of spatial metrics typically involves the construction of a SPATIAL NEIGHBOURHOOD GRAPH, wherein nodes represent cells and edges connect neighbouring cells. The most common methods for constructing a neighbourhood graph include radius-based, k-nearest neighbours (kNN), and Delaunay triangulation. The selection of the algorithm is dictated by the nature of the task at hand, given that graphs defined by the three approaches differ in their characteristics [218].

In particular, a graph based on a radial distance implies direct biological meaning of spatial scale. Nonetheless, this approach can result in irregular communities when cell densities are not uniform across the sample tissue. Moreover, this approach does not take into account neighbourhood topology, as cells on opposite sides of cavities and cells on the same side are treated equivalently, provided they are within the same distance. The kNN method guarantees that all cells are assigned an equal number of neighbours regardless of the local cell density. Nevertheless, it may result in neighbourhoods with little biological reasoning, for instance, including cells separated by spatial obstacles. Finally, kNN approach will produce neighbourhoods biased towards more prevalent cell types. Hence, in the context of biological systems, Delaunay triangulation may yield a balanced solution. Its performance is not contingent on a predefined radius or number of neighbours, thereby ensuring greater adaptability to local cell densities. Additionally, the Delaunay triangulation is topology-aware and discourages connections across gaps, as the algorithm maximises minimum angles and therefore encour-

ages near-equiangular triangles [219]. Consequently, if a triangle is formed by connecting cells across a gap, it is likely to be rejected unless there is no alternative. Finally, to achieve a greater degree of control, Delaunay method can be combined with post-filtering of biologically unrealistic long edges.

Following the construction of the neighbour graph, the quantitative measurement of observed patterns can be facilitated by the application of spatial statistics [220, 221]. Metrics originating from graph theory, including centrality and clustering coefficients, Ripley's statistics, and autocorrelation scores have been implemented in `squidpy` [222] Python package and can be employed directly. Moreover, novel spatial statistics and other measures of the degree of spatial tissue organisation are being constantly developed [113, 115].

NEIGHBOURHOOD ANALYSIS. Despite their simplicity, visual assessment and spatial statistics calculation do not exploit the breadth of information contained in high-dimensional transcriptomic data. It is conceivable, however, that certain patterns can become apparent only when relationships between different cell populations are taken into consideration [223]. Therefore, for comprehensive exploration of tissue spatial complexity, it is necessary to define recurrent spatial communities or SPATIAL NEIGHBOURHOODS.

One common method of defining a neighbourhood is to conceptualise it as an aggregation of cells of particular types [95, 111, 224]. To illustrate this principle, one may consider the perivascular niche in the brain, wherein endothelial cells forming a vessel are surrounded by vascular smooth muscle cells (VSMCs) or pericytes, fibroblasts, astrocytic endpoints, and various immune cells including border-associated macrophages (BAMs). Despite its simplicity, the cell type-based approach heavily relies on the cell annotation and its granularity. An alternative approach to defining a neighbourhood is to consider local conditions that manifest as altered transcript counts. An example of a such neighbourhood could be a hypoxic region characterised by the expression of genes responsible for survival

in low oxygen conditions. A plethora of cell-type-agnostic methods that utilise transcript data directly have been developed for the annotation of expression-driven neighbourhoods [110, 225, 226].

Although simple methods were found to be advantageous [111, 200, 224], both approaches are valid, and the comparison between them can assist in the identification of novel spatial biomarkers [227].

AIMS OF THIS DISSERTATION

To address the knowledge gaps in the field of glioma research and to tackle the computational challenges presented in Chapter 1, I set the following objectives within the scope of this dissertation:

1. To establish a robust and efficient computational pipeline to enable the analysis of spatial single cell transcriptomics data;
2. To decipher the complex ecosystem of necrotic tissue in GB patients;
3. To launch a large-scale pan-glioma study and elucidate the discrepancies in cellular landscape and spatial organisation across seven gliomas.

2

COMPUTATIONAL METHODS FOR SPATIAL ANALYSES

As aforementioned, owing to the recent emergence of spatial technologies, a consensus on best practices for computational analysis of spatial data remains to be established. In this Chapter, I present a comparison of the methods that were available at the beginning of my doctorate a few years ago. Moreover, due to the paucity of tools readily available for analysis, I was required to devise novel approaches to spatial analysis. These methods are also discussed in this Chapter. Given that both projects included in this dissertation employed the 10x Xenium technology, the following text will address only the analysis of the associated data.

2.1 DESIGNING A CUSTOM GENE PANEL

While commercial spatial platforms provide predesigned gene panels, a custom panel that is tailored to specific needs may be required when the research question is more narrow or involves rare cancer types. The study discussed in Chapter 4 encompasses adult-type diffuse gliomas and multiple EPNs. To develop a custom panel of 350 target genes for a detailed spatial profiling of these gliomas, I formulated the following recommendations:

1. The panel should recover all cell types expected to be found in gliomas;
2. The panel should capture the diversity of gene expression beyond that determined by a cell type;
3. The panel should incorporate genes of particular interest for glioma research.

The following text discusses the approaches I developed to guarantee the fulfilment of each of the three requirements.

A SILHOUETTE SCORE *measures clustering quality and ranges from -1 (worst) to 1 (best).*

CELL TYPE RECOVERY. To conduct cell type marker selection, I used a published GB data set [228] and a cosine similarity-based method named `COSG` [178] (Figure 2.1). To evaluate the capacity of the gene set to recover all cell types of interest, I compared the *silhouette scores* of cell type clusters within two distinct data sets: one consisting of the whole transcriptome (Figure 2.1a), and the other containing only the selected markers (Figure 2.1b). The average silhouette score for clusters in whole transcriptome space was -0.1022, and for the marker space it reached 0.02242 (Figure 2.1c). Considering that result, I concluded that the selected markers were efficient at recovering cell types.

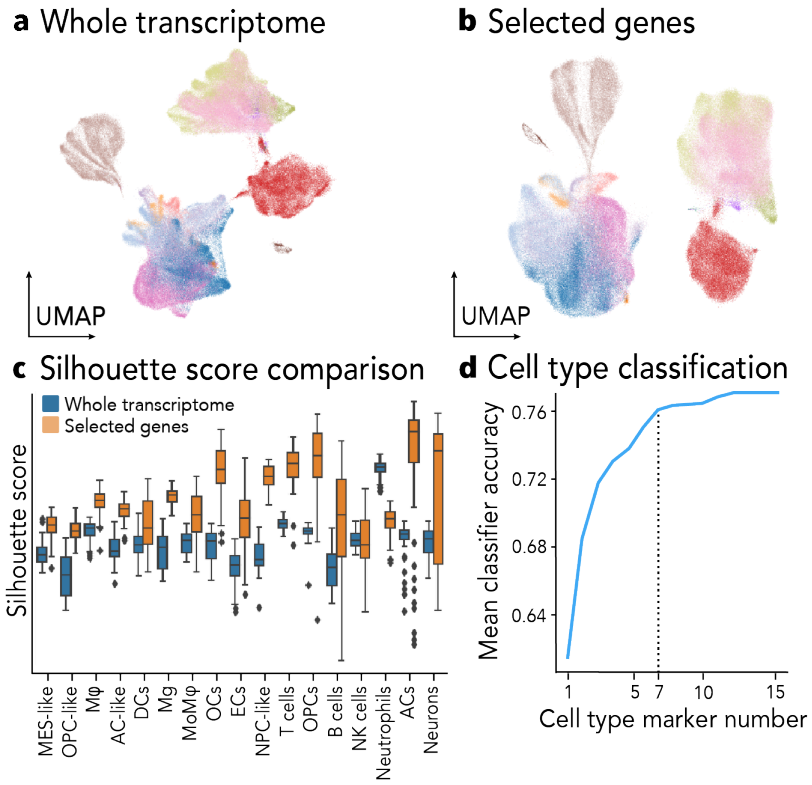

Figure 2.1: *Cell type recovery.* UMAP of the GB data set using the whole transcriptome *(a)* and only markers *(b)*. Silhouette score per cell type *(c)*. Classification accuracy depending on the number of markers *(d)*. DC, dendritic cell; EC, endothelial cell; Mg, microglia; Mφ, macrophage; MoMφ, monocyte-derived Mφ; NK, natural killer.

Due to the limited size of the panel, it is necessary to define the smallest number of markers per cell type sufficient to recover all cell clusters. To this end, I calculated the cell classification accuracy using the Python package `spapros` [190]. As, on average, the accuracy curve reached a plateau at seven markers (Figure 2.1d), I decided to proceed with seven markers per cell type.

To inform the selection of seven genes out of 50 markers per cell type, I calculated the expression correlation within each cell type (Figure 2.2). Certain cell types, such as neurons, exhibited a high positive correlation between their markers, while other cell types, for example neutrophils, demonstrated minimal to no correlation. The correlation coefficients were instrumental in optimising the recovery of expression information while reducing the number of redundant markers for each cell type. To this end, I incorporated at least one marker per individual cluster of correlated markers. In instances where the number of markers was still below seven, I opted to include genes exhibiting low correlation with other markers or additional markers for each cluster.

Since the study in Chapter 4 focuses on multiple glioma types, I additionally evaluated the performance of the cell type markers on publicly available astrocytoma [97] (Figure 2.3a) and oligodendroglioma [98] (Figure 2.3b) data sets. As demonstrated by the high level of accuracy of the classifier calculated using `spapros` [190], the selected cell type markers were effective in accurately capturing major cell types present in both data sets.

EXPRESSION DIVERSITY. Next, I employed both PCA and sPCA on all three data sets in order to identify genes that accounted for the expression variation in these tumours (Figure 2.4a-b). To evaluate the performance of the methods, I selected genes with a minimum of 0.5 loading values across all components and calculated the panel quality metrics implemented in `spapros`. In the case of astrocytoma, both methods demonstrated comparable performance (Figure 2.4c). Conversely, for oligodendroglioma, sPCA was more effective in selecting uncorrelated mark-

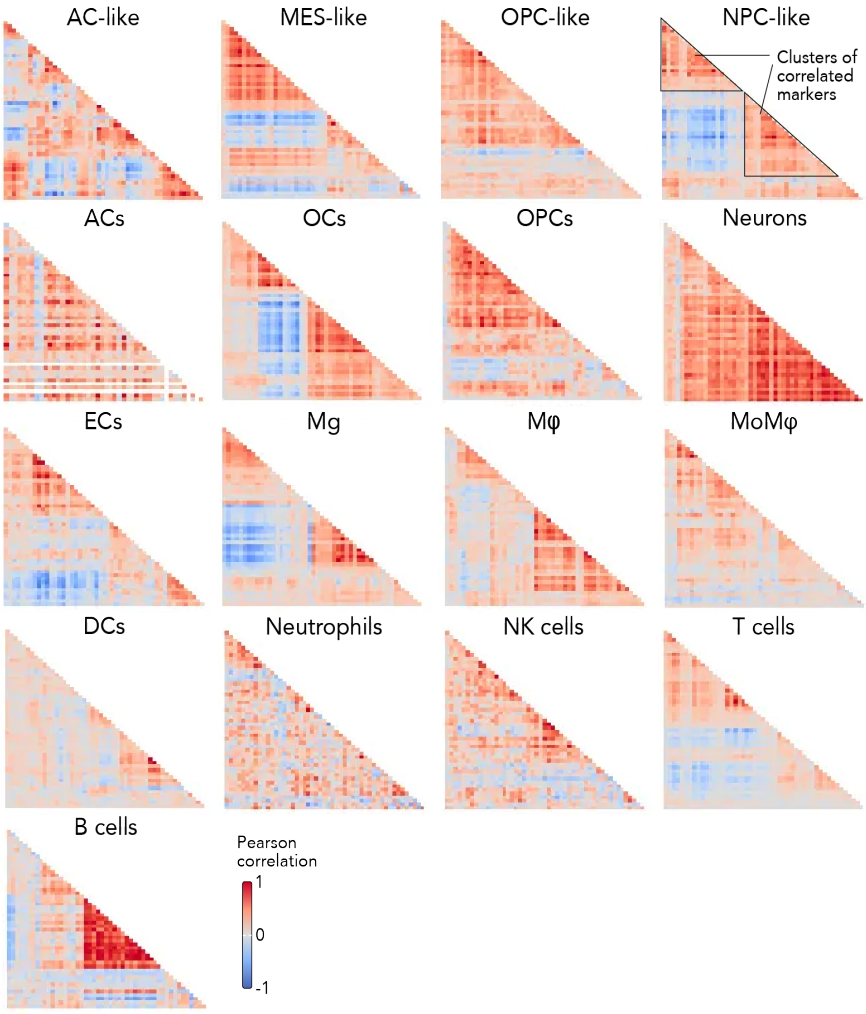

Figure 2.2: *The correlation of marker expression.* Clustering highlights redundant markers.

ers and preserving local cell neighbourhoods (Figure 2.4d). Therefore, I opted to employ sPCA to incorporate genes reflecting expression diversity within the glioma types under study.

To finalise the gene panel, I included additional markers for specific challenges that may arise in the glioma study. In particular, since glioma cells express markers of brain-resident cells [104], it can be difficult to distinguish surrounding tissue from tumour. In order to take this into account and facilitate glioma cell identification, I included additional astrocytic markers into the panel. Furthermore, the gene list was further augmented with genes involved

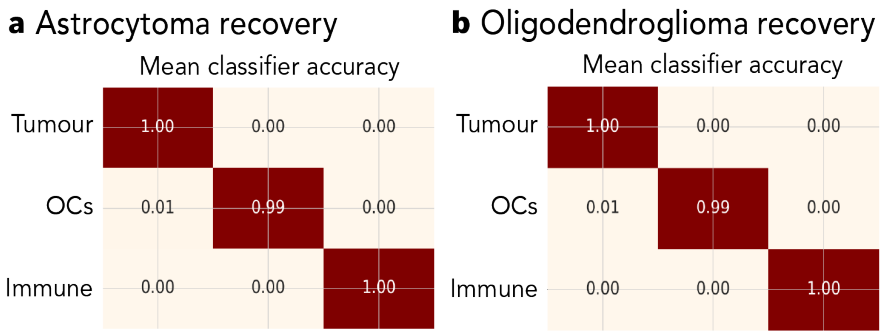

Figure 2.3: *Cell type recovery in IDH-mutant gliomas.* Classification accuracy for astrocytoma *(a)* and oligodendroglioma *(b)*.

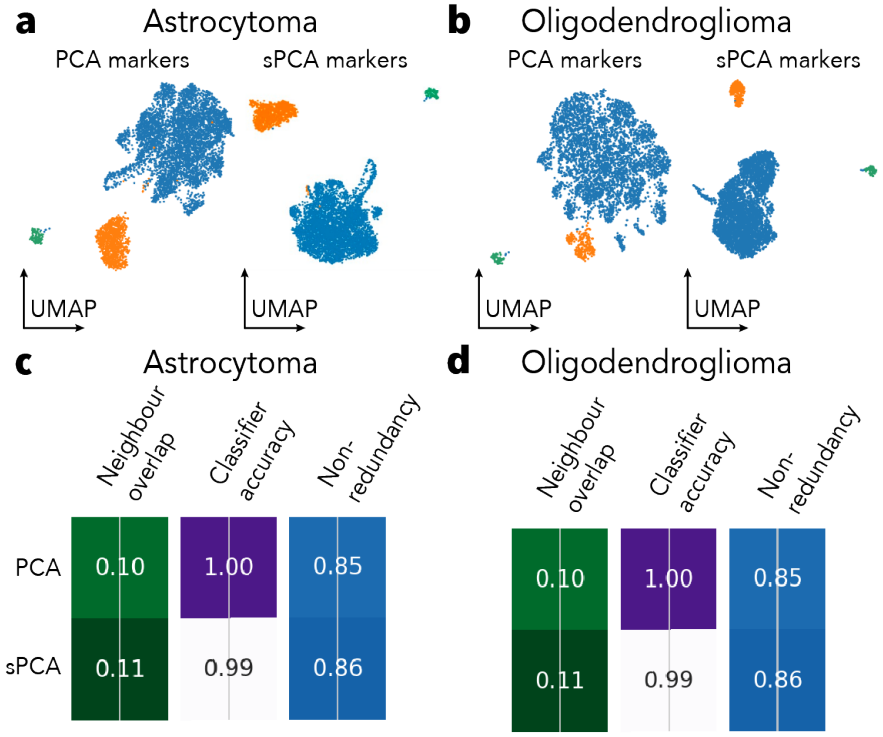

Figure 2.4: *Gene selection using PCA.* UMAP of the marker spaces for astrocytoma *(a)* and oligodendroglioma *(b)*. Panel performance on astrocytoma *(c)* and oligodendroglioma *(d)* data.

in tumour networks [229], therapeutic targets, genes with common mutations and CNVs, and previously established principal markers of glioma states not covered by the transcriptomic atlas (e. g.NPC1- and NPC2-like programs) [92]. Finally, I expanded the panel by adding genes implicated in the hallmark functional pathways, namely proliferation, angiogenesis, hypoxia, glycolysis, apoptosis, necrosis, ferroptosis, growth, senescence, and invasion.

2.2 CHOOSING A CELL SEGMENTATION METHOD

To select the best segmentation approach, I used a data set produced in the study that will be discussed in Chapter 4 to compare three methods: distance-based segmentation, nuclear segmentation, and `Baysor` [195] (Figure 2.5). The raw output of the 10x Xenium platform comprises nuclear boundaries, as determined by a segmentation run on the nuclei-stained (DAPI) image, and cell boundaries, as determined by expanding the corresponding nuclear masks [199]. I utilised these outputs in the comparison, the first one representing nuclear segmentation and the latter representing a distance-based approach.

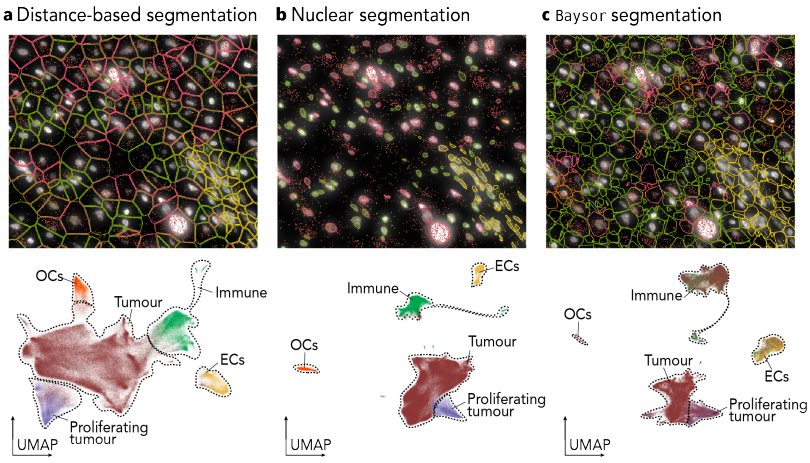

Figure 2.5: *Segmentation approaches.* Spatial masks *(top)* and UMAP embedding *(bottom)* of the distance-based *(a)*, nuclear *(b)*, and `Baysor` *(c)* segmentation. Cell colours in the top panel correspond to the cell types in the bottom panel; white staining shows nuclei; dots show transcripts.

Distance-based segmentation led to considerable signal mixing among cells of differing types, a finding that may be attributable to cell spatial colocalisation (Figure 2.5a). It is evident that the highly heterogeneous morphology of brain cells, characterised by dendrites and irregular shapes, may necessitate a more sophisticated approach than merely extending masks by a pre-defined universal radius.

Conversely, the exclusive use of nuclear transcripts produced well-defined cell type clusters with little overlap (Figure 2.5b). Concurrently, the available signal was sufficient to annotate all cell types present in the data. However, a considerable proportion of transcripts located outside the nuclei, which may offer crucial information regarding the cell expression, was discarded.

Finally, `Baysor` presented a balanced solution that produced well-defined cell type clusters while capturing the maximum amount of transcript information (Figure 2.5c). Nonetheless, due to the persisting complexity of delineating cell boundaries away from the nuclei, a certain degree of signal leakage might be inevitable. Therefore, the selection of the appropriate segmentation algorithm may be contingent upon the specific requirements of the task at hand.

2.3 ADAPTIVE CELL ANNOTATION

CELL ANNOTATION. In a spatial context, it is imperative to maximise the recovery of cells. This necessity will be particularly apparent in Chapter 3, where I discuss expansive necrotic regions that exhibit reduced transcript levels despite intact nuclei (Figure 2.6). In this case, a conventional universal filtering procedure, based on raw transcript counts, would have resulted in the exclusion of all cells from these regions.

To overcome the challenge, I used an alternative, more adaptable strategy:

1. Initial annotation is performed on a subset of "high quality" cells;

2. The cell type labels are transferred to all segmented cells with a certain confidence score. Then, depending on the task at hand, cells can be filtered to maximise either the annotation confidence or the number of the preserved cells.

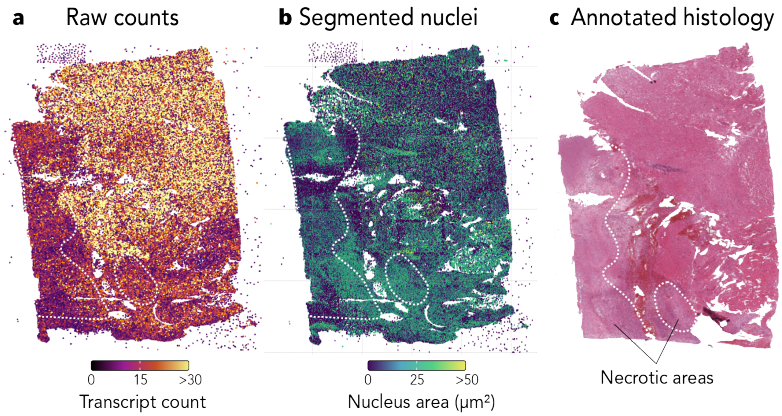

Figure 2.6: *Transcript counts reflecting sample histology.* Raw transcript counts *(a)*, nuclear area *(b)*, and the corresponding haematoxylin and eosin (H&E) staining *(c)*.

In the following text I discuss each step in more detail.

INITIAL ANNOTATION. The initial annotation included removal of cells with low counts, count normalisation by a universal size factor, log 1p-transformation, and gene count scaling. The next step was to perform a PCA and to build a kNN-graph on the basis of the first principal components. The generated kNN-graph was further clustered using the Leiden algorithm to identify cell types and visualised using uniform manifold approximation and projection (UMAP).

At this stage, a decision must be made regarding the necessity of batch correction prior to the definition of cell type clusters. In order to make a decision, I analysed the data from multiple glioma types (discussed in detail in Chapter 4). Upon examination of the UMAP and cell clustering (Figure 2.7), it became apparent that TME clusters were defined by their respective cell types, while tumour cells were grouped according to glioma type and/or sample origin.

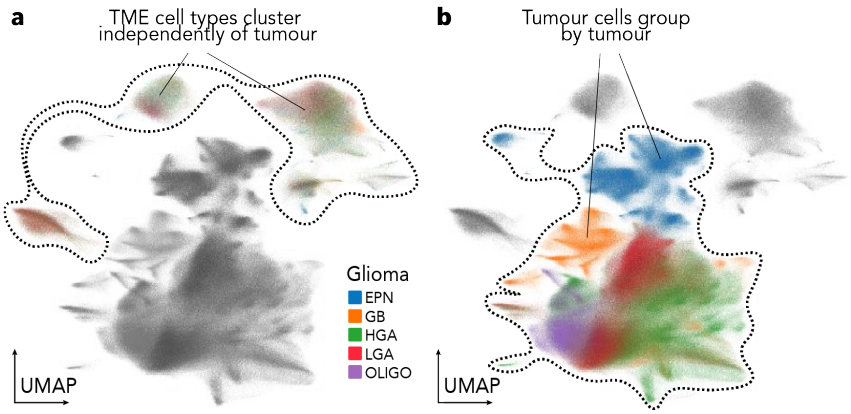

Figure 2.7: *Batch effect during cell annotation.* TME cell clusters defined by cell type *(a)*. Batch effect for tumour cells *(b)*.

Based on these observations, I concluded that batch correction was not required at the level of broader cell groups, since the difference in expression observed between cell types exceeded the diversity present within cell type clusters, thereby preventing their mixture.

In a standard scRNA-seq analysis, it is common practice to further examine each broad cell type cluster in isolation, aiming to generate fine cell subtype annotations. In this case, another question that must be addressed concerns the necessity of re-scaling the transcript counts prior to the re-clustering. As evidenced by an investigation into T cells (Figure 2.8), no re-scaling impeded the identification of proliferating T cells, potentially due to the greater expression diversity between cell types that exceeded the subtle variations between cell subtypes (Figure 2.8a). Conversely, re-scaling of gene expression resulted in the aggregation of proliferating T cells into a separate cluster (Figure 2.8b). Therefore, I concluded that re-scaling represented a better strategy to facilitate the annotation of subtle cellular states.

An alternative approach exists, however, which is not dependent on the subjective elements of data preparation and annotation. In this approach, cell states are defined by means of matrix factorisation (see Section 2.4).

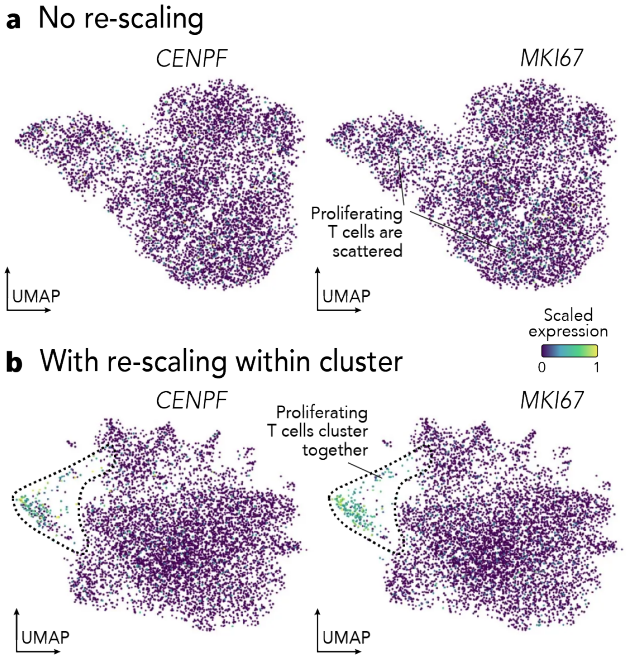

Figure 2.8: *Re-scaling during cell annotation.* UMAP embedding of T cells with no re-scaling *(a)* and after re-scaling *(b)*.

LABEL TRANSFER. In order to annotate the filtered cells and to refine the previous annotations, I employed models using variational inference (scVI) that are part of the `scvi-tools` Python package [207] (Figure 2.9). I trained a scVI model [230] with a Poisson count distribution on a random sample from the annotated data set to compute the latent space. In the next step, I used the weights of the trained scVI model to initialise and train a scANVI model [231] to transfer the annotations to all cells (Figure 2.9a).

Such approach facilitated greater adaptability and customisation, enabling either the prioritisation of cells with high confidence predictions or threshold adjustment according to the task at hand (Figure 2.9b). Furthermore, in Chapter 3 and Chapter 4, the approach described above results in the discovery of a new cell type.

2.4 DESCRIBING TRANSCRIPTION PROGRAMS

As mentioned above, an alternative to manual annotation of fine cell types is MF. Using a GB data set (discussed

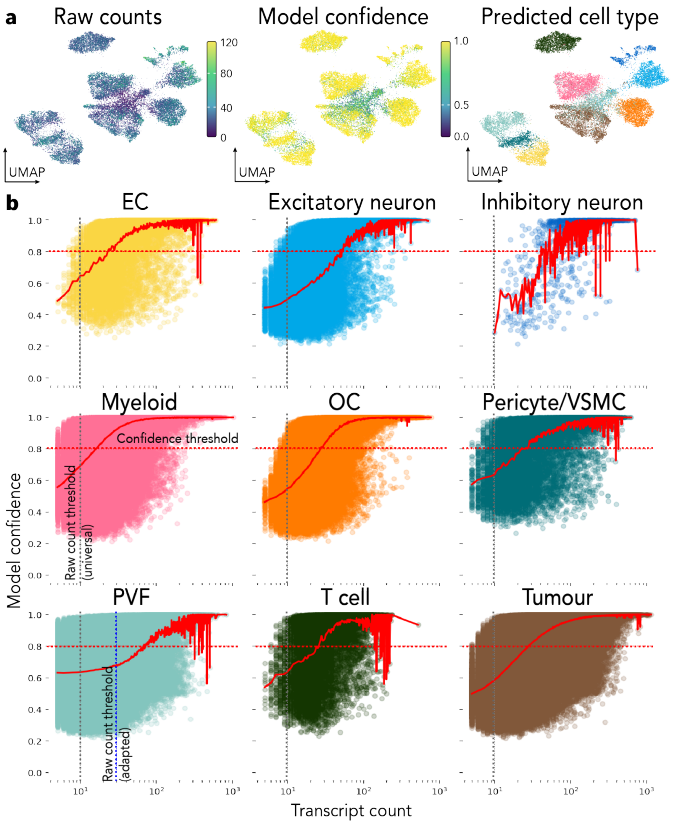

Figure 2.9: *Adaptive thresholds for cell quality control.* UMAP of sampled cells coloured by count, prediction confidence, and label *(a)*. Model confidence depending on the cell type *(b)*. Cell type colours in *(a)* match the colours used in *(b)*. Red curves show averaged values.

in Chapter 3), I tested two NMF initialisation methods for finding transcriptional programs for tumour cells: random (implemented in `consensus NMF` [232] Python package, further referred to as rNMF) and low-rank approximation (non-negative double singular value decomposition used by default in `scikit-learn` [233] Python package, further referred to as lrNMF) (Figure 2.10). For rNMF, I chose ten components, as this solution corresponded to the highest stability and a low error rate. In the case of lrNMF, the solution was comparatively stable regardless of the number of components. Since numbers higher than ten produced programs consisting of one gene only, I concluded than ten components would be optimal.

The rNMF produced programs that were mutually uncorrelated (Figure 2.10a) and could be clearly annotated based on the contributing genes. In contrast, the lrNMF approach yielded highly correlated programs that clustered into groups defined by similar functional signatures clusters (e. g.a hypoxic mesenchymal signature) (Figure 2.10b). Furthermore, one of the lrNMF programs was attributed to a single gene, *EGFR* (Figure 2.10b), which undermined the very objective of identifying metaprograms. Such factor redundancy was effectively addressed by the rNMF algorithm which consolidated the correlated lrNMF programs into single uncorrelated programs (Figure 2.10c).

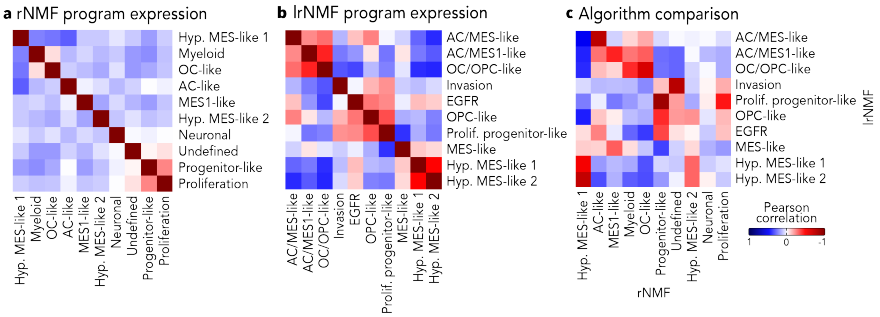

Figure 2.10: *Comparison of rNMF and lrNMF solutions.* Expression correlation for rNMF *(a)* and lrNMF *(b)* programs. Correlation between rNMF and lrNMF programs *(c)*.

In addition, the expression of rNMF programs was observed within discrete populations of tumour cells (Figure 2.11a), highlighting the heterogeneity of glioma cells. Conversely, most lrNMF programs were distributed over cells across all patients, exhibiting no discernible patterns. Moreover, rNMF programs were confined to specific spatial areas (Figure 2.11b), whereas in the case of lrNMF, the redundancy of programs was furthermore evident as their spatial expression considerably overlapped (Figure 2.11c).

Consequently, in both studies discussed in this dissertation, I employ the rNMF method (further referred to by the corresponding package name, `consensus NMF` [232]), in view of its ability to capture information about heterogeneous cell populations in a spatial and sample-based manner demonstrated here.

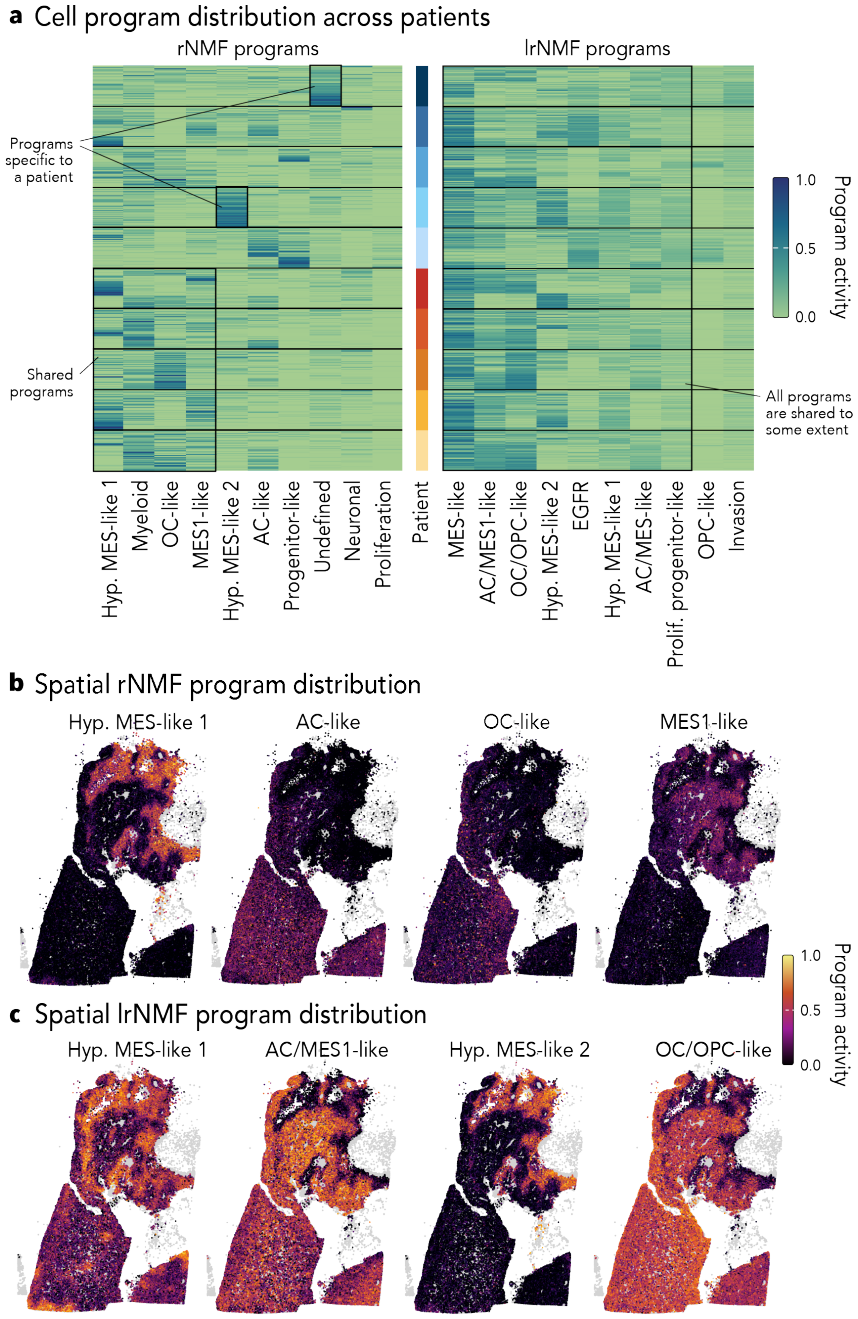

Figure 2.11: *Comparison of heterogeneity captured by rNMF and lrNMF programs.* Program distribution across cells for two algorithms *(a)*. Spatial localisation of rNMF *(b)* and lrNMF *(c)* programs.

2.5 DESCRIBING SPATIAL PATTERNS

VISUAL EXAMINATION. When a study involves a small number of samples and cell types (as in Chapter 3), spatial exploration could be started with the visual examination of the location of individual cell types. In certain cases, cell type distribution within anatomical regions may offer insight into the nature of the region itself (Figure 2.12).

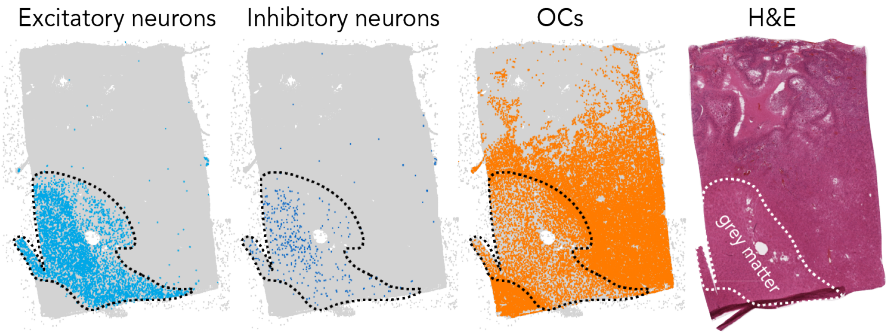

Figure 2.12: *Visual assessment of cell type location.* The majority of neuronal signals are derived from the grey matter, where the somas are situated.

SPATIAL NEIGHBOURHOODS. To identify recurring spatial patterns in a quantitative manner, it is essential to compute a neighbourhood graph. Despite the existence of numerous algorithms for the construction of such a graph (see Subsection 1.4.4), for a general task of identifying spatial neighbourhoods, I selected the Delaunay-based approach as it produced neighbourhoods of similar size, that were independent of cell type or sample region (Figure 2.13a), whereas the results of the radius-based approach depended on the selected radius (Figure 2.13b-c).

The resulting neighbourhood graph can then be visualised in order to identify recurrent spatial connections among cell types. The capacity for customisation of such visualisation is noteworthy, suggesting its suitability for a range of analytical applications. To illustrate, I visualised spatial connections existing in samples from two groups from Chapter 3 (Figure 2.14). In the given example, the size of the nodes is proportional to the frequency of neighbours

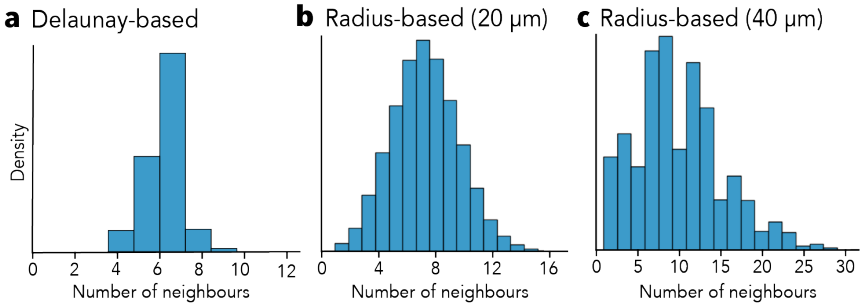

Figure 2.13: *Comparison of neighbour distribution for three graphs.* Distribution of the number of neighbours per cell for the Delaunay-based *(a)*, 20µm-radius-based *(b)*, and 40µm-radius-based *(c)* spatial graphs.

of the same type, and the width of an edge is proportional to the neighbouring frequency of the corresponding two cell types. Samples in the first group demonstrated high interconnectivity (Figure 2.14a), whereas the second group exhibited the loss of interconnectivity, with only three cell types maintaining spatial connections (Figure 2.14b).

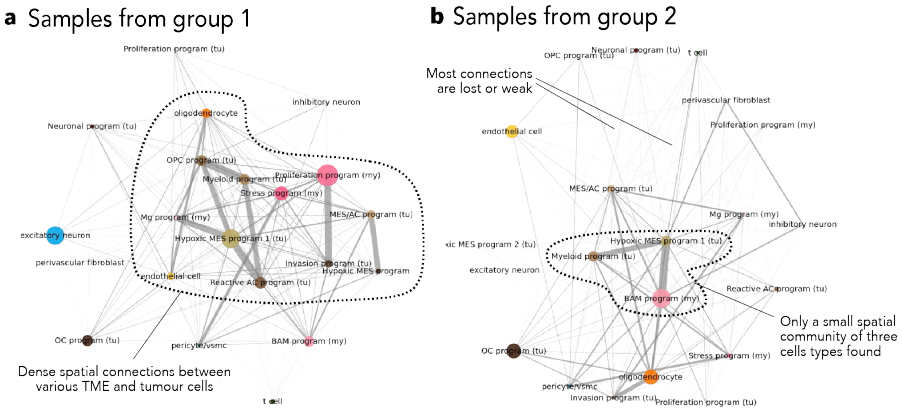

Figure 2.14: *Graph visualisation for spatial analysis.* Spatial connections in samples from group 1 *(a)* and group 2 *(b)*.

In order to facilitate a quantitative calculation of neighbourhoods, which would be independent of subjective visual evaluation, I employed a method that defined neighbourhoods based on their cell type composition, similarly to previously described approaches [111, 224]. The initial step involved the construction of a matrix, where the rows corresponded to each cell in the data set, and the columns denoted the number of neighbours of each cell type. I

subsequently clustered the matrix using the k-means algorithm to obtain cellular neighbourhoods.

LOCATION-DEPENDENT EXPRESSION CHANGES. The analysis of spatial single cell data provides a valuable opportunity to investigate context-dependent expression patterns. In the two studies discussed in this dissertation, the number of profiled genes was insufficient for the differential expression (DE) tools that are frequently applied in the scRNA-seq studies, as the data was too sparse to distinguish noise and signal leakage from true DE expression.

To address this challenge, I adopted a different approach (Figure 2.15). To this end, for all cells belonging to a specific group, I calculated the correlation coefficients between gene expression and the distance of these cells to a target spatial location. To account for inter-sample variation, the correlation coefficients were first calculated on a per-sample basis and then averaged across samples. As the distribution of averaged correlation coefficients should approach a normal distribution even in cases where sample correlations are not normally distributed, I made an assertion that genes located at the extremes of the empirical correlation distribution (i. e.below the 5th or above the 95th percentile) could be interpreted as statistically and biologically meaningful.

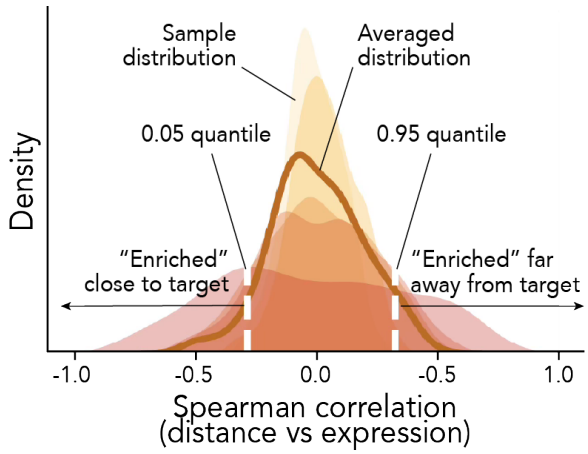

Figure 2.15: *A custom method for gene enrichment analysis.* Correlation between expression and a spatial distance used to define gene enrichment.

3

SPATIAL PROFILING OF NECROSIS IN GLIOBLASTOMA

3.1 BACKGROUND AND CONTEXTUAL FRAMEWORK

The current standard of care for GB involves radiotherapy, a procedure that frequently results in complications, such as radionecrosis (RN) [234]. RN manifests in approximately 20% of GB patients, typically within several months to several years after radiation exposure [235, 236]. In routine MRI scanning, which is frequently employed to assess disease status, RN can be misinterpreted as GB progression, as both are characterised by contrast enhancement and perilesional oedema. Nonetheless, progressive GB and RN demand different therapeutic strategies, and incorrect diagnosis may lead to the worsening of symptoms [78]. Despite the potential of perfusion MRI and amino acid positron emission tomography (PET) to facilitate differential diagnosis, these techniques have yet to achieve wide usage [235]. Consequently, the differential diagnosis of GB progression and post-treatment complications is currently based on histopathological evaluation [237]. Histologically, the manifestation of RN is characterised by coagulative necrosis, vessel wall hyalination, and marked vasculature damage [238]. Conversely, in GB, a necrotic core is encircled by hypercellular zones comprised of elongated tumour cells, known as PSEUDOPALISADES [239]. Despite extensive research, the fundamental causes of the disparities between RN and pseudopalisading necrosis in GB remain to be fully elucidated.

In the study discussed in the present Chapter, I employed spatial single cell transcriptomic profiling of 266 genes to generate the first spatial atlas of progressive GB and RN encompassing eight brain-resident cell types and ten tumour transcription programs. In RN samples, large numbers of infiltrating BAMs and the prevalence of mu-

tant cells with low *EGFR* expression were identified. Conversely, progressive GB contained progenitor-like and proliferating tumour cells that exhibited high *EGFR* expression. The present study offers invaluable insights into the intricate architectural patterns and cellular interplay in RN, which may assist in improving GB patient outcomes in the future.

The text and Figures herein are reproduced from the manuscript that has been published on medRxiv (Seferbekova, Ritter, Ruckhovich, Schinkewitsch, Köberer, Grassl, Kessler, Goidts, Ratliff, Herold-Mende, Krieg, Etminan, Platten, Wick, Reuss, von Deimling, Sahm, Gerstung & Suwala [240]) and is currently in revision at Neuro-Oncology. For that manuscript, I analysed all data, interpreted the results, prepared all figures, and wrote all the text. In comparison with the published manuscript, some parts were omitted and the text has been rewritten to follow the overarching narrative of this dissertation. The Figures were updated to fit the general style of this work. The scientific content remains similar.

3.2 MATERIALS AND METHODS

3.2.1 *Specimen collection*

Abigail K. Suwala and Felix Sahm provided formalin-fixed paraffin-embedded (FFPE) material from nine GB patients, who previously underwent standard of care treatment consisting of chemo- and radiotherapy. At the time material was collected, MRI scans of all patients revealed signs of tumour progression in accordance with the RANO guidelines [241]. Nevertheless, upon histological examination, four patients were confirmed to have tumour progression, four exhibited radionecrotic changes, and one patient was found to have both. Abigail K. Suwala assembled the sample cohort and performed neuropathological annotations on consecutive H&E slides.

3.2.2 *Spatial single cell transcriptomic experiment*

Abigail K. Suwala selected areas of interest, encompassing both necrosis and the adjacent tissue. Michael Ritter carried out the Xenium sample preparation, involving cutting tissue sections (5 µm thick), sample mounting onto Xenium slides, incubation, and air-drying. Michael Ritter in collaboration with scOpenLab processed the samples using the Xenium Sample Preparation Kit and the off-the-shelf Xenium Human Brain Panel, comprising 266 genes (PN-1000599), in accordance with the manufacturer's protocols.

3.2.3 *Spatial single cell transcriptomics data analysis*

CELL SEGMENTATION. Gleb Rukhovich utilised `Baysor v0.6.1` [195] to segment whole cells. The nuclear segmentation, informed by DAPI staining, was used as a prior, with a confidence threshold set at 0.1. The minimal number of transcripts required for a cell to be classified as real was set to 5. After segmentation, I combined the generated per sample expression matrices into one data set.

CELL ANNOTATION. For computational data analysis I used `Scanpy v1.10` [242] and `Python v3.9`. Initially, I filtered out 1,360,821 (56% of all segmented cells) cells that contained less than 30 transcripts. Next, I normalised, `log1p`-transformed, and scaled the transcript counts for the remaining 1,065,178 (44%) cells. I performed PCA on the scaled data with default parameters, and generated a kNN graph using 50 principal components and 15 neighbours. For visualisation purposes, I embedded the graph using UMAP with default parameters. Finally, to annotate cells, I clustered them using the Leiden algorithm with a resolution of 0.5. I annotated six major cell types (Oligodendrocytes, Neurons, Vascular, Myeloid, Lymphoid, and Tumour cells) based on cell type marker expression (Figure 3.1). In the final step, counts for Vascular cells, Lymphoid cells, and Neurons were re-scaled and re-analysed

separately to produce perivascular fibroblasts (PVFs), Endothelial cells, T cells, and Inhibitory and Excitatory neurons. I labelled 14,185 cells from the Lymphoid cluster as "Unknown" due to the lack of specific expression.

REFINING CELL ANNOTATION. To annotate filtered out cells and refine the generated annotations, I used models implemented in `scVI-tools v1.2.0` [207]. First, to obtain the latent representation of the transcript space, I trained an scVI model [230] on a random sample from the annotated data set (115,165 cells stratified by patient and cell type, excluding "Unknown") with Poisson gene likelihood distribution and 16 latent dimensions. Cell clustering employing the latent space yielded two clusters for Endothelial cells. I re-labelled one cluster as Pericytes/VSMCs due to its high expression of *NR2F2*, *CDH6*, and *SLIT3* and retrained the scVI model with this cell type added to the labels. Subsequently, I used the new model to generate decoded transcript fractions per cell. To transfer annotations and obtain cell type probabilities for the entire data set, I used the same training sample to train a scANVI model [231] initialised with weights from the scVI model. Cells with at least 10 transcripts and 0.8 cell type probability were assigned the predicted label. As PVFs presented the greatest challenge and exhibited the lowest model confidence (Figure 2.9b), I assigned this label to cells that had at least 30 transcripts. The remaining cells were annotated as "Uninformative".

TRANSCRIPTION PROGRAMS. To identify transcription programs within the Tumour and Myeloid clusters, I employed the `consensus NMF` [232]. Ten and four programs were found to achieve the highest stability-to-error ratio for the Tumour and Myeloid clusters correspondingly (Figure A.1). For the Tumour cluster, I excluded from the subsequent analysis the minor (less than 20% of total tumour signal) Myeloid and Neuronal programs as well as sample-specific Invasion and Hypoxic MES2b-like programs. For the Myeloid cluster, although the optimal stability-to-error ratio was also achieved for the solution with ten programs,

the additional six programs mostly corresponded to brain-resident cell transcriptional profiles. Hence, I hypothesised that the extra programs could be explained by an over-training and proceeded with four Myeloid programs. All programs were annotated based on the top contributing genes.

3.2.4 *Spatial analysis*

SPATIAL COMMUNITY IDENTIFICATION. For the spatial computations, I used functions implemented in the `squidpy` [222] package. First, I constructed a spatial graph based on Delaunay triangulation, filtering out neighbours at the distance more than 100 µm. Next, for a given pair of cell types, a metric of neighbourhood enrichment was defined as a z-score standardising the observed number of colocalisation events compared to the number of colocalisation events in 1,000 random permutations of cell labels.

OVEREXPRESSED GENES PRIORITISATION. To identify genes that were overexpressed in cells of a target cell type in close proximity to cells of another cell type, I implemented a prioritisation strategy (see Chapter 2). In short, for the two cell types, I computed Spearman correlation coefficients between cell distances and target cell gene expression for each gene and sample and calculated median values within diagnosis groups. Genes with a correlation coefficient below the 0.05 quantile of the diagnosis-level distribution were considered "overexpressed" in the target cell type in close proximity to the second cell type. Similarly, genes with a correlation coefficient above the 0.95 quantile were considered "overexpressed" away from the second cell type.

3.3 RESULTS

3.3.1 *Spatially-resolved transcriptomics facilitates RN and GB comparison*

The objective of the present study was to describe the spatial cellular landscape across the RN and progressive GB histologies. To this end, ten samples from a cohort of nine patients who had undergone the current standard of treatment were collected (Table 3.1). According to the histological assessment, four cases met the criteria for progressive GB, four demonstrated features of RN, and one patient presented with both. The samples were analysed using the Xenium *In Situ* platform and the Human Brain Panel (Figure 3.1a).

Patient	Histology	Age	Days after radiation*	Status
Patient 1	GB	58	231	Deceased
Patient 2	GB	66	252	Alive
Patient 3	GB	66	998	Alive
Patient 4	GB	57	58	Alive
Patient 5	RN	52	351	Deceased
Patient 6	RN	60	*No data*	*No data*
Patient 7	RN	50	226	Alive
Patient 8	RN	79	53	Deceased
Patient 9	RN & GB	59	141	Alive

Table 3.1: *Patient cohort.* *, time of material collection in days after radiation exposure.

In general, GB samples displayed higher cell densities in comparison with samples exhibiting RN histology (Figure 3.1b, $P = 1.29 \times 10^{-5}$, two-sided Mann-Whitney U test). Following the processing of raw data (see Section 3.2), I generated a spatial single cell transcriptomics atlas of RN and progressive GB, that encompassed 1,189,460 cells and

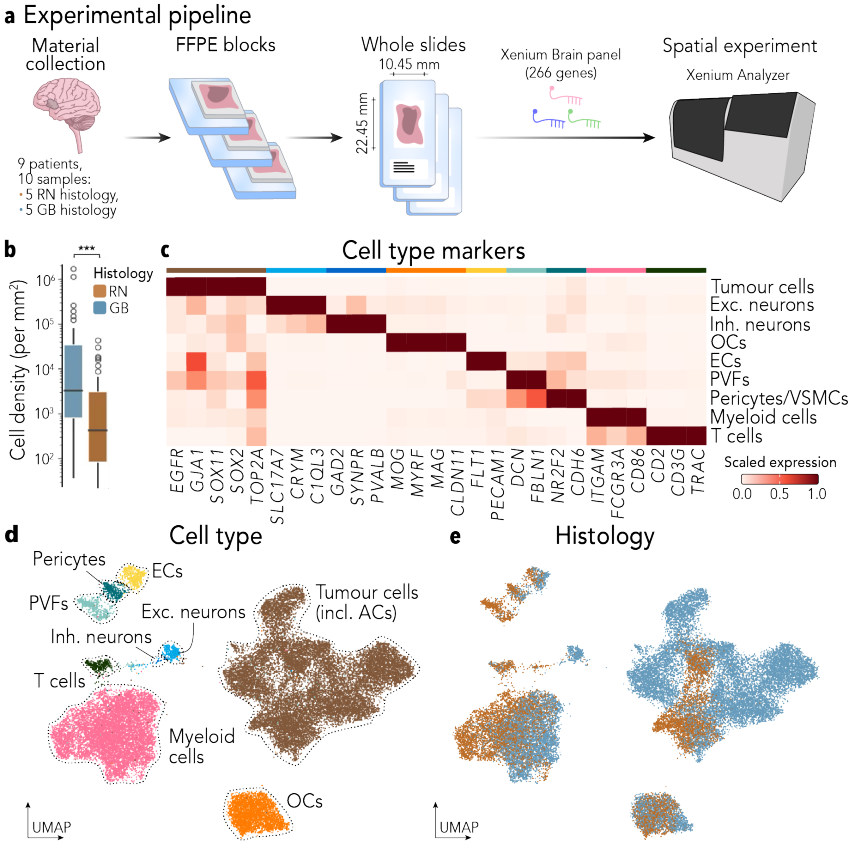

Figure 3.1: *Atlas of necrotic tissue in GB patients.* Experimental pipeline *(a)*. Cell density comparison *(b)*. Cell type markers *(c)*. UMAP of annotated cell coloured by type *(d)* and histology *(e)*. Adapted with permission from [240].

112,423,484 transcripts. The annotated cell types included Tumour cells and eight TME cell types, namely Myeloid cells, Oligodendrocytes, PVFs, Pericytes/VSMCs (further as "Pericytes"), Endothelial cells, T cells, and Excitatory and Inhibitory neurons (Figure 3.1c). The presence of all nine cell types, inclusive of Tumour cells, was identified in both GB and RN samples (Figure 3.1d-e).

During the annotation process, one major cluster showed expression of both glioma and astrocytic marker genes (Figure A.2). Concurrently, no individual cluster was found to exclusively express astrocyte-specific markers. The distinction between malignant and TME cells is a frequently encountered challenge in scRNA-seq and spatial transcriptomics analysis and is traditionally solved through the anal-

ysis of CNV profiles [243]. However, due to the insufficient number of target genes in the panel, CNV analysis could not be conducted in the present study. Furthermore, GB has been shown to contain cell populations exhibiting transcriptional similarities to brain-resident astrocytes [92], which additionally complicated the distinction. To address the problem, I annotated the cluster as "Tumour cells (including astrocytes)" and subsequently identified cell populations via transcription program inspection (see Subsection 3.3.2). To ascertain the validity of this approach, I compared the annotated Tumour cell percentages with the expected tumour cell counts, the latter being devised from VAFs of *TERT* promoter mutations, a hallmark alteration in GB [244] (Figure A.3). The surprising conclusion was that the established annotations may in fact be an underestimation of the number of glioma cells in samples regardless of histology. On the basis of this evidence, I considered the proposed strategy adequate and employed the same approach for the Myeloid cluster. In this particular case, the strategy was further driven by the observation that, in the brain, myeloid populations form a transcriptional continuum, as opposed to individual cell type clusters [245]. Consequently, I delineated myeloid cell states through the identification of transcription programs (see Subsection 3.3.3).

The comparison of cell proportions revealed that there was no statistically significant difference between Uninformative cell numbers in the two histologies (Figure 3.2a, $P > 0.05$, two-sided Mann-Whitney U test), indicating that the employed filtering procedure did not have a bias towards one particular histology. In addition, since irradiation is known to be associated with RN changes in the white matter, a significant proportion of which consists of axons [246], I hypothesised that the low transcript number within axons would not be captured efficiently in a spatial single cell experiment. This, in turn, should result in a reduced number of neurons annotated in samples with RN histology. Indeed, I found a statistically significant difference in the proportions of Inhibitory neurons (Figure 3.2b, $P = 0.028571$, two-sided Mann-Whitney U

test). However, after multiple test correction, this difference was rendered insignificant. The other seven cell types did not show significant differences between GB and RN samples (Figure 3.2b, $P > 0.05$, two-sided Mann-Whitney U test with Benjamini–Hochberg correction). In both sample groups, Tumour cells constituted the predominant cell type, while Inhibitory neurons were the least prevalent cell population.

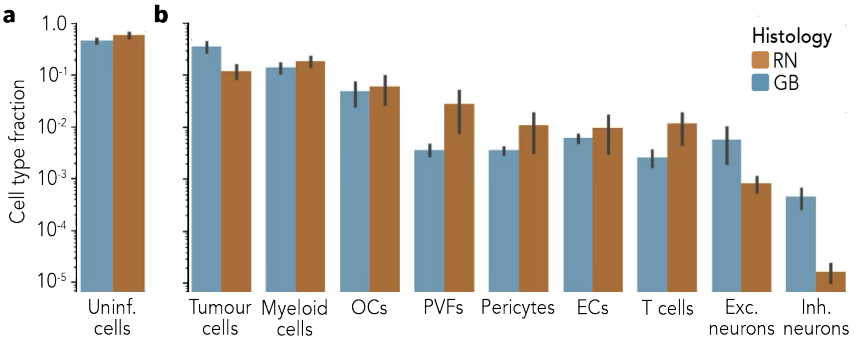

Figure 3.2: *Cell type proportion comparisons.* GB and RN samples did not differ in proportions of filtered-out cells *(a)* or any annotated *(b)* cell types. Adapted with permission from [240].

Taken together, I generated a spatial single cell transcriptomic atlas of nine cell types across ten samples with RN changes and GB, which enabled a nuanced comparison of the two histologies.

3.3.2 *Progenitor-like tumour state is downregulated in RN samples*

To characterise the transcriptomic heterogeneity concealed within the Tumour cluster, I annotated ten tumour transcription programs using `consensus NMF` [232] (see Section 3.2). Subsequently, given that four tumour transcription programs accounted for either sample-specific or minor signals in the Tumour cluster (Figure A.4), I focused on the remaining six programs for downstream analyses (Figure 3.3).

In alignment with previously reported findings [247], I identified two distinct transcription programs associated

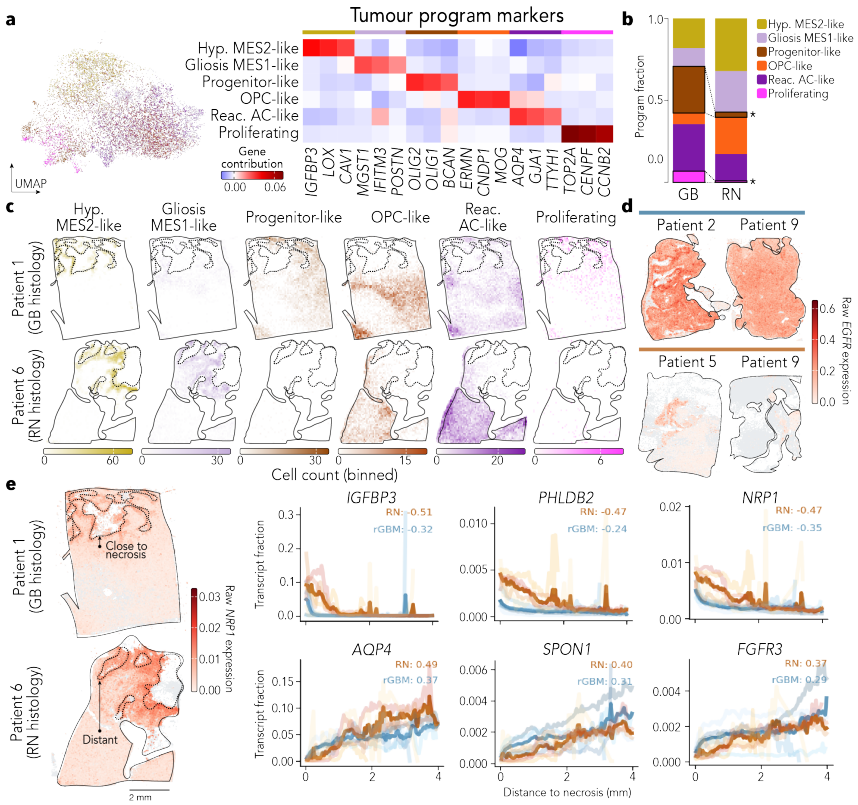

Figure 3.3: *Tumour transcription programs.* Program markers *(a)*. Program distribution between the two histologies *(b)*. Spatial distribution of programs *(c)*. *EGFR* expression in samples where the gene is amplified *(d)*. Genes enriched perinecrotically (*top*) and away from necrosis (*bottoms*) *(e)*. Hyp., hypoxic; Reac., reactive. Adapted with permission from [240].

with the MES-like glioma state. The first program constituted the largest proportion of the Tumour cluster and was characterised by the presence of both MES2-like (*IGFBP3*) and hypoxia-response (*LOX*, *CAV1*, *HILPDA*) markers (Figure 3.3a). The distribution of these tumour cells aligned with perinecrotic areas (Figure 3.3c, Figure A.5a), indicating the potential stress environment. Consequently, I annotated that transcription program as Hypoxic MES2-like program.

The second MES program was enriched in genes previously assigned to the MES1-like program (*MGST1*, *SERPINA3*, *IFITM3*) (Figure 3.3a), which are involved in injury response. Spatially, the program was expressed in re-

gions in close proximity to Hypoxic MES2-like cells, but distant from areas of necrosis (Figure 3.3c, Figure A.5a). I hypothesised that this program may represent cells in the direct vicinity of stress-affected tissue that facilitate wound-healing. The program was therefore annotated as Gliosis MES1-like program.

The third program was associated with glial (*OLIG2*, *OLIG1*) and neuron (*SOX4*) progenitor markers, as well as general stem (*BCAN*, *PTPRZ1*, *NOTCH1*) markers (Figure 3.3a), therefore it was annotated as Progenitor-like tumour program. The proportions of cells expressing this particular program were different in RN samples in comparison to GB samples (Figure 3.3b, $P = 0.047619$, two-sided Mann-Whitney U test with Benjamini-Hochberg correction). Notably, *EGFR*, a pivotal driver in GB biology, was among the genes associated with this program. Since *EGFR* amplification is common in GB, it is therefore conceivable that all samples with the amplification should have high *EGFR* expression. However, upon comparing *EGFR* expression among the three patients with confirmed *EGFR* amplification (one with GB histology, one with RN histology, and one with both samples), it was revealed that the expression levels of *EGFR* in RN samples were low, whereas GB samples demonstrated high expression (Figure 3.3d). This discrepancy was particularly pronounced in two samples from the same patient (Patient 9). Furthermore, an assessment of DNA methylation array data revealed that the *EGFR* amplification was not genetically lost (Table A.1).

The fourth program was enriched in oligodendocytic (*CNDP1*, *ERMN*) and OPC (*MAG*) markers (Figure 3.3a) and was therefore termed OPC-like. Tumour cells expressing this program could represent OPCs differentiating towards mature oligodendrocytes. In a similar manner, the fifth program was characterised by the markers of brain-resident astrocytes (*AQP4*, *FGFR3*, *SOX9*, *SPON1*) and was annotated as Reactive AC-like program. Finally, the Proliferating program was expressed in the lowest fraction of Tumour cells was associated with cell cycle-related genes. Similarly to Progenitor-like tumour cells, Proliferating tu-

mour cells were present in different proportions in GB and RN samples (Figure 3.3b, $P = 0.047619$, two-sided Mann-Whitney U test with Benjamini-Hochberg correction).

Finally, to facilitate the analysis of potential associations between Tumour gene expression and the distance to necrosis, I employed a gene prioritisation procedure (see Subsection 3.2.4). A negative correlation was observed between distance to necrosis and the expression of genes associated with cell migration (*NRP1*, *IGFBP3*, *PHLDB2*). It has been previously observed that pseudopalisading cells may exhibit migratory behaviour directed away from hypoxic conditions [114]. The present study posits that, despite distinct histopathological representation (i. e.the absence of an elongated shape), perinecrotic cells in RN exhibit a similar migratory transcription signature as pseudopalisading glioma cells.

Taken together, the findings outlined here demonstrated that GB samples contained larger fractions of Progenitor-like and Proliferating tumour cells compared to RN. Furthermore, Tumour cells in GB samples maintained high *EGFR* expression, whilst in RN, they expressed low levels of *EGFR* even in the presence of *EGFR* amplification.

3.3.3 *Border-associated macrophages infiltrate RN samples*

In a similar manner to the Tumour cluster, I employed `consensus NMF` [232] to describe four transcription states for cells in the Myeloid cluster (Figure 3.4, Figure A.5b, see Section 3.2). The first program was associated with *P2RY12*, *CX3CR1*, and *GPR34* expression and corresponded to brain-resident Microglia (Figure 3.4a). The second program was enriched in *LYVE1*, *THBS1*, and *CD163*, markers of BAMs. Finally, I identified Proliferation and Hypoxia programs, which resembled the tumour transcription programs but were defined by different genes (Figure 3.4a). Furthermore, Proliferating myeloid cells had significantly higher fractions in GB samples in comparison to RN samples (Figure 3.4b, $P = 0.0317$, two-sided Mann-Whitney U test with Benjamini-Hochberg correction).

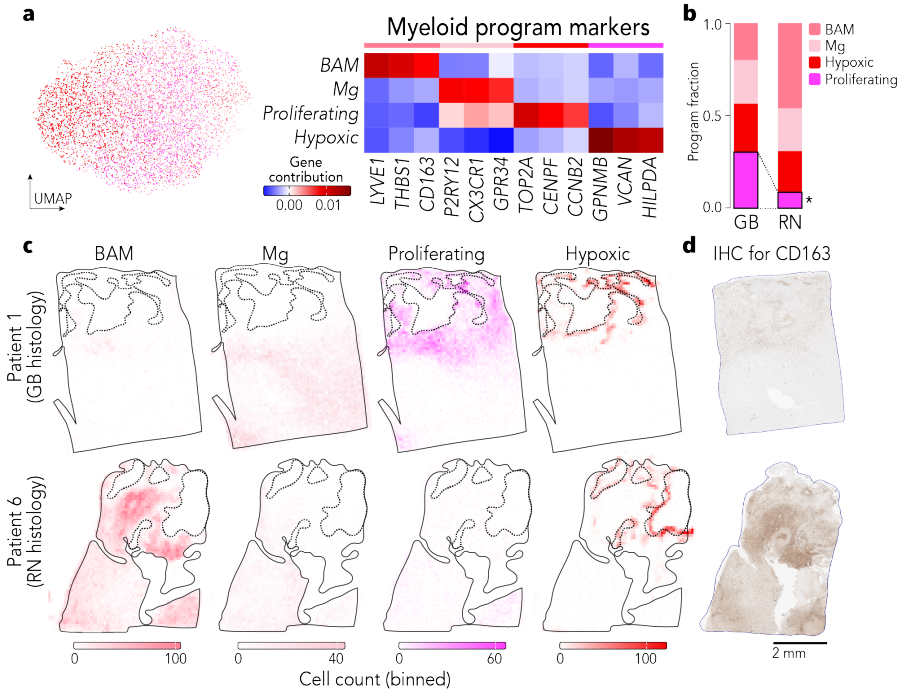

Figure 3.4: *Myeloid transcription programs.* Program markers *(a)*. Program distribution between the two histologies *(b)*. Spatial distribution of programs *(c)*. immunohistochemistry (IHC) staining for a BAM marker *(d)*. Adapted with permission from [240].

On average, BAMs constituted the largest proportion of the Myeloid cells in samples with RN histology — four RN samples exhibited higher BAMs fractions than all GB samples. Surprisingly, I observed that, in GB, BAMs were restricted to smaller areas, while in RN samples, they demonstrated a diffuse pattern (Figure 3.4c, Figure A.5b). This observation was corroborated by IHC staining for CD163, which showed pronounced positive staining in the sample with RN histology and not in the GB sample (Figure 3.4d).

The results of the analysis outlined here indicated that the Proliferation myeloid transcription program is more prevalent in samples with GB histology than in RN samples. Furthermore, in RN, the most prevalent myeloid population, BAMs, was enriched across the whole sample area, whereas in GB samples BAMs were observed in few small clusters.

3.3.4 *RN and GB histologies exhibit different spatial architectures*

To investigate the differences in spatial architectures underlying GB and RN histologies, I utilised a permutation test (see Section 4.2). I assessed whether particular cell types demonstrated a preference for colocalisation, forming recurrent spatial communities (Figure 3.5, Figure A.6).

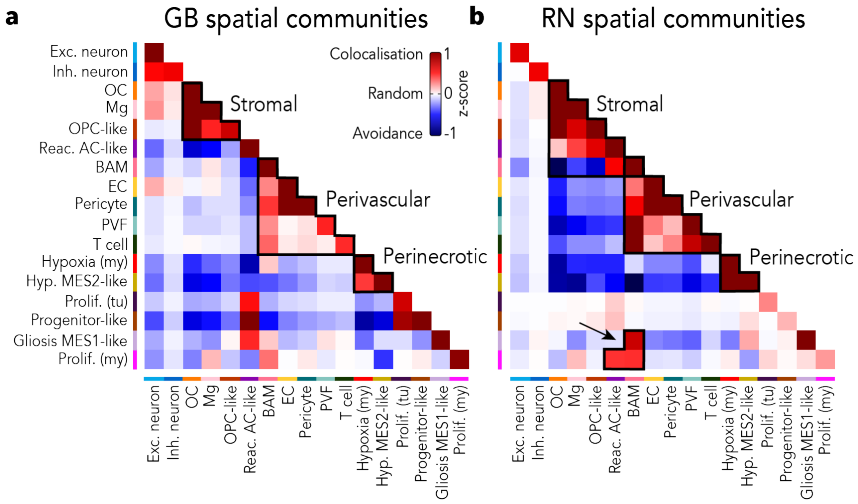

Figure 3.5: *Recurrent spatial communities.* Three recurrent communities in GB *(a)* and RN *(b)* samples. Adapted with permission from [240].

The first spatial community was formed by Hypoxic myeloid and Hypoxic MES2-like tumour cells (Figure 3.6), delineating a potential area with the conditions of severe hypoxic stress. Furthermore, BAMs, Pericytes, Endothelial cells, and T cells formed a second spatial community, the Perivascular community, which was shared by both histologies (Figure 3.5). It is noteworthy that, despite the marked severity of vessel damage observed after brain irradiation, this finding suggests that the Perivascular niche is preserved in RN samples. Finally, I observed a colocalisation of brain-resident cell types, namely Neurons, Oligodendrocytes, and Microglia and defined this spatial community as "Stromal".

Additionally, I observed multiple instances of colocalisation between only two cell types. In particular, in GB sam-

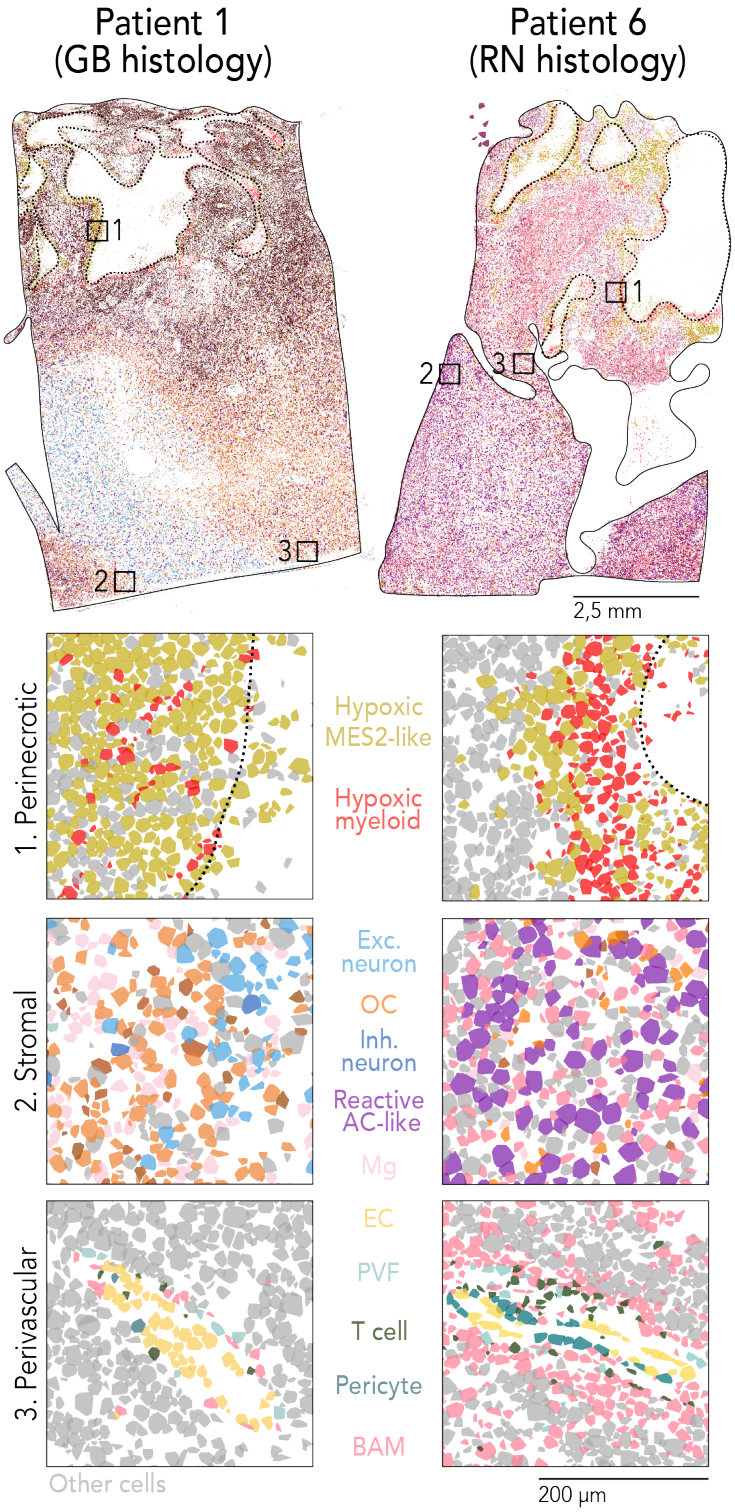

Figure 3.6: *Recurrent spatial communities in RN and GB samples.* Cell composition of the three recurrent spatial communities. Adapted with permission from [240].

ples, Proliferating tumour cells colocalised with Progenitor-like tumour cells (Figure 3.5a). Moreover, I observed a colocalisation of Reactive AC-like cells with BAMs within the stromal community only in samples displaying RN histology (Figure 3.5b, Figure A.6b, showed with an arrow). Furthermore, BAMs recurrently colocalised with Gliosis MES1-like tumour cells in RN samples (Figure 3.5b). To assess whether such colocalisations could be associated with a functional interplay between the cell types, I evaluated how gene expression correlated with spatial distances. In the case of BAMs in proximity to Gliosis MES1-like tumour cells, the colocalisation was associated with an upregulation of cytokine production in BAMs, particularly *TNF* (Figure A.7).

To summarise, the current study posits that necrotic tissue in GB patients is characterised by an intricate spatial architecture, displaying variations in minor cellular communities depending on the exact histology of the sample. In cases exhibiting RN histology, I observed the colocalisation of BAMs with Reactive AC-like cells and Gliosis MES1-like tumour cells. Concurrently, the upregulation of BAM cytokine production was evident when they were located close to Gliosis MES1-like tumour cells.

3.4 DISCUSSION

The study presented in this Chapter is the first to provide a comprehensive analysis of RN changes in comparison to GB histology using spatially-resolved single cell transcriptomics. According to the results of the study, samples which were histologically diagnosed with RN changes and no GB progression (i. e.no evidence of pseudopalisades) contained a high number of cells with characteristic CNVs and *TERT*p mutations. Despite the frequent detection of sparse tumour cells in pseudoprogression following treatment [248], it is yet to be elucidated why only in progressive GB and not in RN samples these tumour cells organise into pseudopalisades.

The study additionally revealed that GB samples exhibited higher fractions of Progenitor-like and Proliferating tumour cells, while in RN samples, these tumour cells were rare. The discrepancy was particularly pronounced in two samples collected from the same patient. A key GB driver, *EGFR*, was one of the genes associated with the Progenitor-like transcription program. Interestingly, *EGFR* expression differed between the two histologies in a similar manner. In particular, in the presence of *EGFR* amplification, tumour cells in samples with GB histology overexpressed *EGFR*, whilst in RN samples, *EGFR* expression was markedly low. In a study on postnatal rats, the expression of *EGFR* in progenitor cells was demonstrated to play a pivotal role in maintaining their proliferative capacity [249]. In the current study, Proliferating tumour cells were found to colocalise with Progenitor-like cells in GB samples. Therefore, despite the limited number of samples and low Proliferating tumour cell number, it could be hypothesised that, in GB samples, the proliferating population is replenished from cells in the progenitor pool.

As demonstrated in earlier studies [246], brain irradiation can have an adverse impact on the NPC and OPC populations. Given that the Progenitor-like transcription program was associated with markers of brain-resident progenitors, it could be hypothesised that these tumour cells may be similarly sensitive to radiotherapy. Moreover, Proliferating cells are expected to be more radiosensitive [250]. Nevertheless, all patients in the cohort underwent the same course of treatment that included radiation therapy. Furthermore, RN typically manifests from months to years after the exposure [236], and all patient material used in this study was collected after comparable time periods in both groups (Table 3.1). Hence, it is difficult to attribute the observed discrepancies to irradiation alone.

It is also notable that an amplification signal for the *EGFR* was detected in the DNA methylation data in both GB and RN samples; yet, at the transcriptomic level, *EGFR* expression was dramatically low in samples with RN histology. One potential explanation for this could be the death of tumour cells with high *EGFR* expression. This

follows from the fact that the rate of RNA degradation is faster than that of DNA degradation, hence DNA signal might provide enough evidence to detect the amplification even if RNA has already degraded and no signal is present in scRNA-seq data. However, unlike DNA, DNA methylation can undergo substantial alterations when stored for a duration exceeding three days [251].

Therefore, I speculate here that tumour cells in RN samples downregulate their *EGFR* expression as well as the Progenitor-like and Proliferation transcription programs. Theoretically, such transcriptional plasticity could enable a substantial population of glioma cells to transition into a resistant state of quiescence by reducing their proliferation. One possible driver of such plasticity could be the TME and/or spatial architecture, as the discrepancies were also observed in two samples that had different histologies but were collected from the same patient.

In support of that hypothesis, I observed distinct spatial patterns between GB and RN samples. In cases of GB histology, BAMs were confined to the small areas surrounding blood vessels, a pattern that mirrors the distribution of BAMs in the healthy brain [252]. In contrast, in RN cases, BAMs represented the largest myeloid population and infiltrated the parenchyma. Interestingly, BAM markers were previously demonstrated to be expressed in brain parenchyma in several other brain pathologies [253, 254]. Within parenchyma, BAMs colocalised with Reactive AC-like cells and Gliosis MES1-like tumour cells. The proximity of the latter was associated with increased expression of cytokine-related genes, particularly *TNF*, in BAMs. The upregulation of cytokines has been linked to neuroinflammation [255], which may provide a rationale for the cognitive impairment and epileptic seizures observed in patients with RN [235]. Furthermore, TNF has been identified as a key factor in the development of gliosis, which can result in the formation of glial scar tissue [256]. A study that used transgenic rodent models demonstrated that glial scar tissue could function as a form of protective environment for glioma cells, thus contributing to GB resistance [257]. The increased presence of BAMs in RN

samples also has the potential to offer a novel diagnostic modality. Despite the fact that the presence of glioma cells in the cerebrospinal fluid (CSF) is infrequent [258], certain BAM markers, including CD163, have soluble protein forms [259] and have therefore the potential to be detected in blood plasma or CSF biopsies. The identification of a biomarker which could be evaluated non-invasively to differentiate RN from GB progression would considerably enhance clinical decision-making.

In order to provide a comprehensive discussion of the findings, it is imperative to consider the limitations of this study. Firstly, a fundamental limitation inherent to any study of necrosis is the presence of numerous dead cells that lack molecular signals suitable for investigation. Secondly, the cohort was confined to patients who satisfied the RANO criteria for tumour progression [241], as determined by MRI scanning. This inherent limitation is difficult to overcome, as the second surgery that provides tissue material is performed solely in the event of a suspected tumour recurrence. Consequently, it would be prohibitively difficult to obtain RN samples in cases where there is no radiological progression diagnosis. This limitation renders it challenging to determine the direct extrapolation of current findings to cases of radiologically diagnosed RN. Nevertheless, it is important to acknowledge that, despite the term "radionecrosis" being predominantly utilised in the field of imaging, it is histology that remains the accepted gold standard for making the definitive diagnosis [237].

Finally, the current study is limited by the number of targeted genes. Despite a select number of markers specific to GB, the commercially available Brain Expression Panel, utilised in the present study, exhibited a limited capacity to describe tumour cell expression diversity, given its primary focus on neuron subsets. The advent of rapidly advancing spatially-resolved profiling technologies, offering simultaneous exploration of thousands of genes, might enable the identification of finer spatial architectures shaped by necrosis in GB patients.

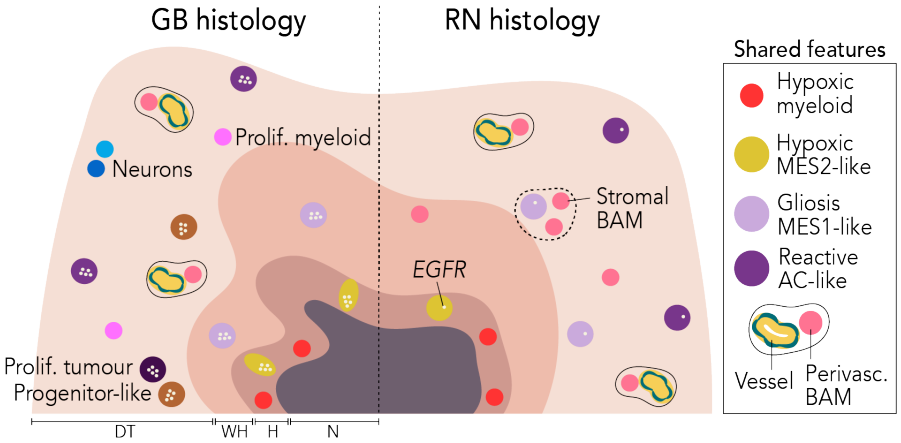

Figure 3.7: *A schematic representation of the findings.* DT, distant tissue; H, hypoxia; N, necrosis; perivasc., perivascular; prolif., proliferating; WH, wound-healing. Adapted with permission from [240].

To summarise, the present work elucidates the spatial characteristics of tissue with RN changes, thereby establishing a foundation for future research aimed at enhancing the treatment of GB patients.

4

A PAN-GLIOMA SPATIAL TRANSCRIPTOMIC ATLAS

4.1 BACKGROUND AND CONTEXTUAL FRAMEWORK

Adult gliomas are a diverse group of highly plastic and heterogeneous tumours [33]. The tumour tissue is comprised of subpopulations of shared stem-like cells and various more differentiated progenies, the presence of which is associated with varying degrees of tumour resistance and disease progression [92, 97, 98, 115, 260]. Gliomas are also found to contain non-malignant cells that constitute the TME, which include neurons, astrocytes, oligodendrocytes, vascular, and myeloid cells [111, 112, 261, 262]. Collectively, malignant cells and the TME form complex communities that are associated with survival and therapy response [95, 110, 113, 115]. As tumours are generally studied in isolation, there is still a limited understanding of the recurrent patterns in spatial organisation and cell interactions across different entities.

Within the study discussed in this Chapter, I designed a custom gene panel and subsequently utilised it for the spatial profiling of single cells from seven gliomas, namely adult-type diffuse gliomas and ependymal tumours. The panel demonstrated efficacy in its depiction of the complex cellular landscape across all examined glioma types. The use of the annotated data set enabled the investigation of tumour expression programs and identification of recurrent spatial neighbourhoods. The constructed comprehensive atlas of seven adult gliomas could serve as a valuable resource for the exploration of spatial biomarkers and their correlation with clinical data.

4.2 MATERIALS AND METHODS

The thorough discussion of computational approaches and the underlying rationale can be found in Chapter 2. The following summary is provided to outline the specifics and deviations from the aforementioned workflow.

4.2.1 *Custom panel design*

The cell markers were selected on the basis of publicly available scRNA-seq data sets of gliomas. In short, a GB data set with 1,135,677 cells [228] was utilised. The marker selection process involved 680,156 cells (60% of the total data). I utilised `COSG` [178] for the identification of 50 markers for the 17 distinct cell types present in the data set. Mast cells, radial glia, and plasma B cells were excluded. The remaining 453,437 cells (40%) were used to construct a kNN graph and compute UMAP for visualisation. Furthermore, I used 3,400 cells (0.3%) from the latter set to calculate the silhouette score and metrics for classification quality (implemented in the `Python` package `spapros`[190], see Figure 2.1). The average silhouette score and random forest classifier accuracy demonstrated that a set of seven marker genes per cell type was sufficient to successfully identify all the cell types analysed (Figure 2.1d). In order to select seven genes out of 50, I computed expression correlation within cell types and prioritised genes that were uncorrelated or negatively correlated (Figure 2.2). Following this, I expanded the data set to include cell types not present in the GB scRNA-seq data set, by incorporating known markers for other immune cells and genes of interest in GB research. Finally, to complement the panel and cover expression diversity within and across cell populations, I included genes with the highest loadings for top 30 sparse principal components computed for the aforementioned GB data, as well as scRNA-seq astrocytoma and oligodendroglioma data sets [97, 98]. Following multiple filtration procedures and rigorous evaluations on the 10x

Genomics side, the final panel was composed of 343 genes (Figure 4.1a).

4.2.2 *Sample collection & experimental procedures*

Abigail K. Suwala and Felix Sahm provided archival FFPE samples from 283 patients who were diagnosed with one of seven glioma types: oligodendroglioma (OLIGO), LGA, high-grade astrocytoma (HGA), GB, EPN-SE, EPN-MPE, or EPN-SP. Using tissue microarrays (TMAs) with 313 cores of 1.5 mm size, Domenico Calafato in collaboration with scOpenLab performed spatial transcriptomic profiling using the 10x Xenium platform and the designed panel (Figure 4.1b).

4.2.3 *Spatial single cell transcriptomics data analysis*

SEGMENTATION. I used the assignment the segmentation of DAPI staining to define nuclear boarders. Only transcripts within the nuclear masks were used for further analysis.

CELL ANNOTATION. Utilising `Scanpy v1.10` [242] and `Python v3.9`, I filtered out 867,569 (31.01% of total) cells (nuclei) with less than 30 nuclear transcripts. Counts for the remaining 1,929,837 (68.99%) cells were `log1p`-transformed and scaled. Next, I computed PCA on the scaled data with default parameters, and constructed a kNN graph using 50 principal components and 15 neighbours. UMAP with default parameters was used for visualisation purposes. For cell annotation, I performed Leiden clustering with a resolution of 0.2. Four of the clusters were pooled and annotated as "Tumour cells". The remaining clusters were annotated as T cells, Oligodendrocytes, Neurons, Vascular, Myeloid/B cells, and Mast cells based on cell type marker expression. In the final step, counts for Vascular cells, Myeloid/B cells, and Neurons were re-scaled and re-analysed separately to produce PVFs, Endothelial cells,

Myeloid cells, B cells, and Inhibitory and Excitatory neurons.

REFINING CELL ANNOTATION. Cell annotations were refined and transferred to the filtered cells using `scVI-tools v1.2.0` [207] according to the pipeline outlined in detail in Section 3.2. The final atlas comprised 2,797,406 cells.

TRANSCRIPTION PROGRAMS. I used the `consensus NMF` [232] to identify nine tumour transcription programs (Figure 4.4, Figure A.9). I excluded from the subsequent analysis Immune and Interferon-response programs that accounted for less than 20% of total signal (Figure A.10). All programs were annotated based on the top contributing genes.

4.2.4 *Neighbourhood detection*

To characterise recurrent spatial neighbourhoods, I utilised functions from the package `squidpy` [222]. Initially, a cell neighbourhood graph was constructed from Delaunay triangulation, with all edges exceeding 50 µm being removed. Subsequently, I utilised the graph connectivity matrix to calculate the number of neighbours of each cell type per cell. This procedure produced a $N \times N$ matrix, where N denotes the number of cell types. Subsequently, the matrix was clustered using the K-means algorithm with the number of clusters ranging from 5 to 30. The final choice of five clusters was determined by its highest silhouette score (Figure A.12). The final step included neighbourhood annotation according to the relative proportions of different cell types present within the neighbourhoods.

4.2.5 *Metadata standardisation*

The manual assignment of sample IDs to each core was performed using TMA slide images in collaboration with Abigail K. Suwala. Next, I aggregated and standardised clinical data collected by Abigail K. Suwala separately for

each TMA. To this end, I assigned standardised English nomenclature to all columns. In instances where a single category was represented by multiple entries, I aligned all entries with the standardised category names. Subsequently, I parsed the columns representing CNVs and single nucleotide variants (SNVs) and represented each significant diagnostic alteration as a separate identifier column. The remaining non-diagnostic alterations were pooled and saved to a new column. To standardise survival time, I converted dates to the more standardised units of days. Finally, columns containing irrelevant metadata were removed, and the remaining columns were used for all subsequent analysis requiring clinical information.

4.3 RESULTS

4.3.1 *Custom panel captures complex cellular landscapes across gliomas*

The aim of the present study was to characterise the spatial cellular landscapes across seven glioma types, namely LGA, HGA, OLIGO, GB, EPN-SE, EPN-MPE, and EPN-SP (Figure 4.1). In order to ensure the capture of all cell types, as well as the expression heterogeneity observed across distinct tumours, a custom gene panel was designed (Figure 4.1a). The panel encompassed 56 brain-resident and 61 tumour cell markers, 108 markers of finer subsets of immune cells, 40 markers that covered glioma expression heterogeneity, and 78 genes that are representative of hallmark biological pathways or active research in gliomas. Utilising the panel, the spatial transcriptomic profiling of 313 cores from 283 patients was conducted using the Xenium *In Situ* platform (Figure 4.1b). Furthermore, clinical data (e. g.survival, grade, and progression status) as well as molecular data (e. g.mutational and copy number profiles) were collected from discrete sources. The data underwent a rigorous processing and standardisation procedures, and was matched with the single cell data, generating a complete and integrated data set (Figure 4.1c).

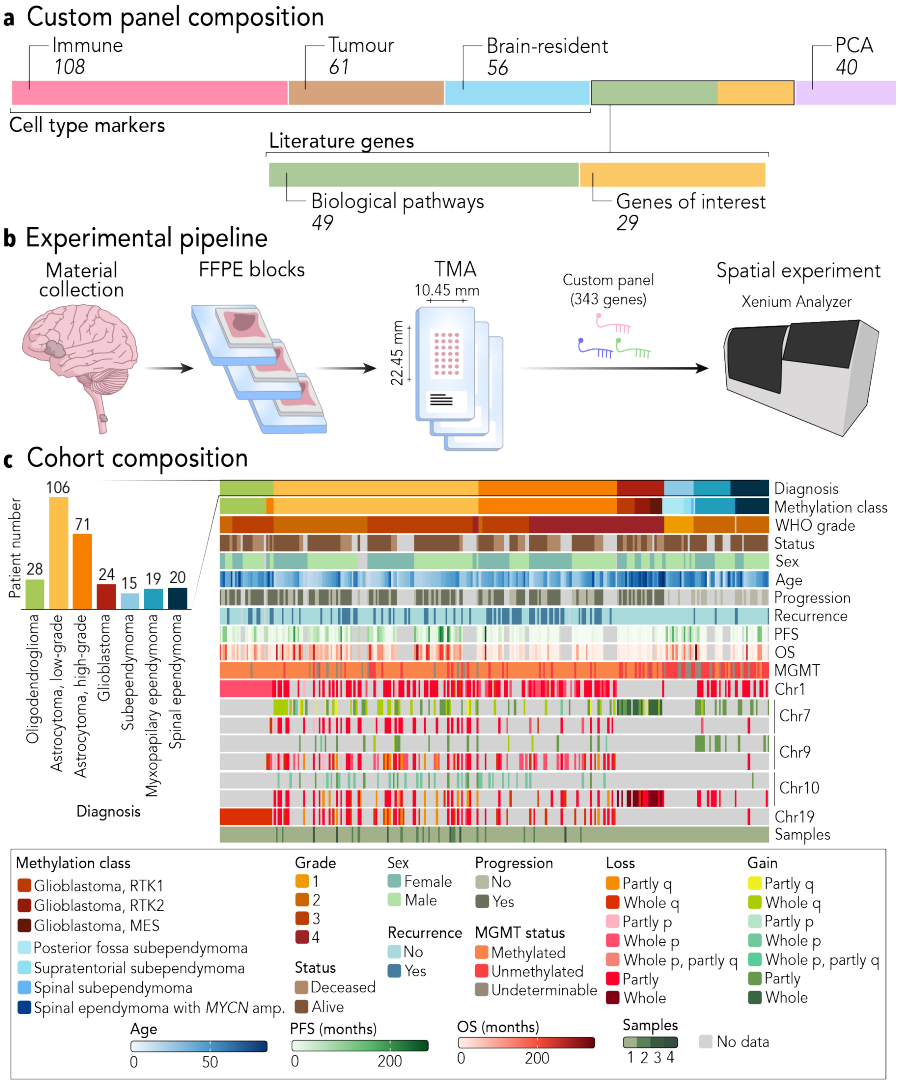

Figure 4.1: *Pan-glioma study design.* Custom panel composition *(a)*. Experimental pipeline *(b)*. Cohort composition and associated clinical data *(c)*. Amp., amplification; Chr, chromosome; OS, overall survival; PFS, progression-free survival; TMA, tissue microarray. Diagnosis colours in *(b)* correspond to the colour legend in *(a)*. Methylation classes matching Diagnosis are omitted.

Subsequent to nuclear segmentation and cell annotation (Section 4.2), the resulting spatial single cell transcriptomics data set consisted of 2,797,406 nuclei with 167,890,066 transcripts (Figure 4.2). Despite the panel's extensive reliance on GB data, no statistical significance was observed in median transcript counts between GB and the other six glioma types (Figure 4.2a, $P > 0.05$, two-sided Dunnett's test). Nevertheless, I identified significant differences in cell count for LGA and EPN-SE samples, as compared with GB (Figure 4.2b, $P < 0.001$ and $P = 0.002$ accordingly). The observed lower cell counts could be attributed to the low grade, as tumours of lower grades are typically associated with lower cell densities [85].

Using a set of cell markers, I performed cell annotation to produce Tumour cells and nine TME cell types, namely Endothelial cells, Pericytes/VSMCs (further "Pericytes"), Oligodendrocytes, Excitatory and Inhibitory neurons, T cells, B cells, Myeloid cells, and Mast cells (Figure 4.2c). It is noteworthy that TME cells from distinct gliomas formed mixed clusters that were characterised by their cellular identity (Figure 4.2d-e). Conversely, Tumour cells from OLIGO, LGA, and HGA were found in a single cluster, while Tumour cells from GB and all three EPN types were separated into distinct clusters. All annotated cell types were shared among patients with all gliomas (Figure 4.3a-b). The most prevalent cell types in gliomas, irrespective of the specific tumour, were Tumour cells, followed by Myeloid cells, Endothelial cells, and Pericytes. Mast cells constituted the rarest cell type (Figure A.8). Despite the presence of a substantial number of Oligodendrocytes and Neurons in certain samples, a greater degree of variation in their numbers was observed within the same glioma types than between different glioma types (Figure A.8). This variation could be indicative of a sampling bias.

Finally, I assessed the capacity of the panel and employed computational workflow to facilitate the recovery of spatial cell distributions (Figure 4.3). A visual examination revealed that, irrespective of a glioma type, cells were successfully recovered throughout all sample areas,

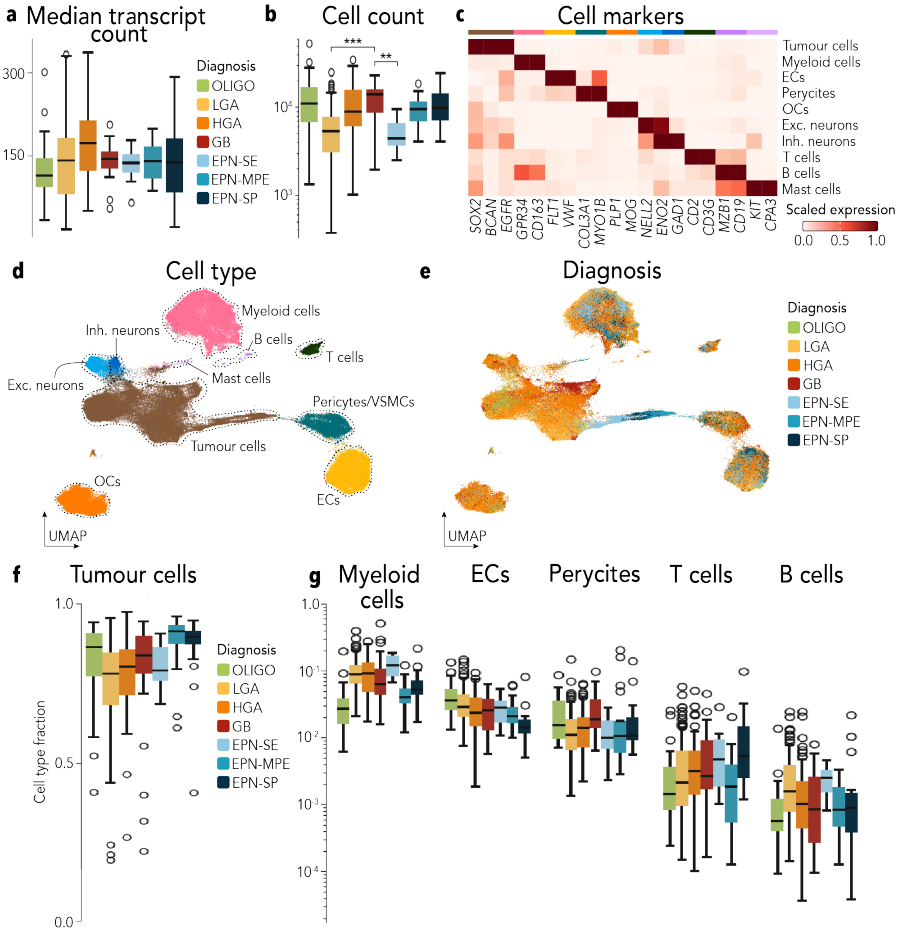

Figure 4.2: *Spatial atlas of gliomas.* Transcript *(a)* and cell *(b)* counts. Cell type markers *(c)*. UMAP of the atlas coloured by cell type *(d)* and glioma type *(e)*. Tumour *(a)* and TME *(b)* cell counts across tumours. EC, endothelial cell; exc., excitatory; inh., inhibitory; OC, oligodendrocyte. **, $P < 0.01$; ***, $P < 0.001$; two-sided Dunnett's test.

with no evidence of missing regions. Within necrotic areas, the presence of Tumour cells was scarce, yet Pericytes and Myeloid cells that infiltrated the areas and presumably engaged in tissue repair [263] could be retrieved. Therefore, despite the potential difficulty in detecting necrotic tumour cells, the workflow once more proved advantageous in elucidating tissue repair mechanisms following cellular death.

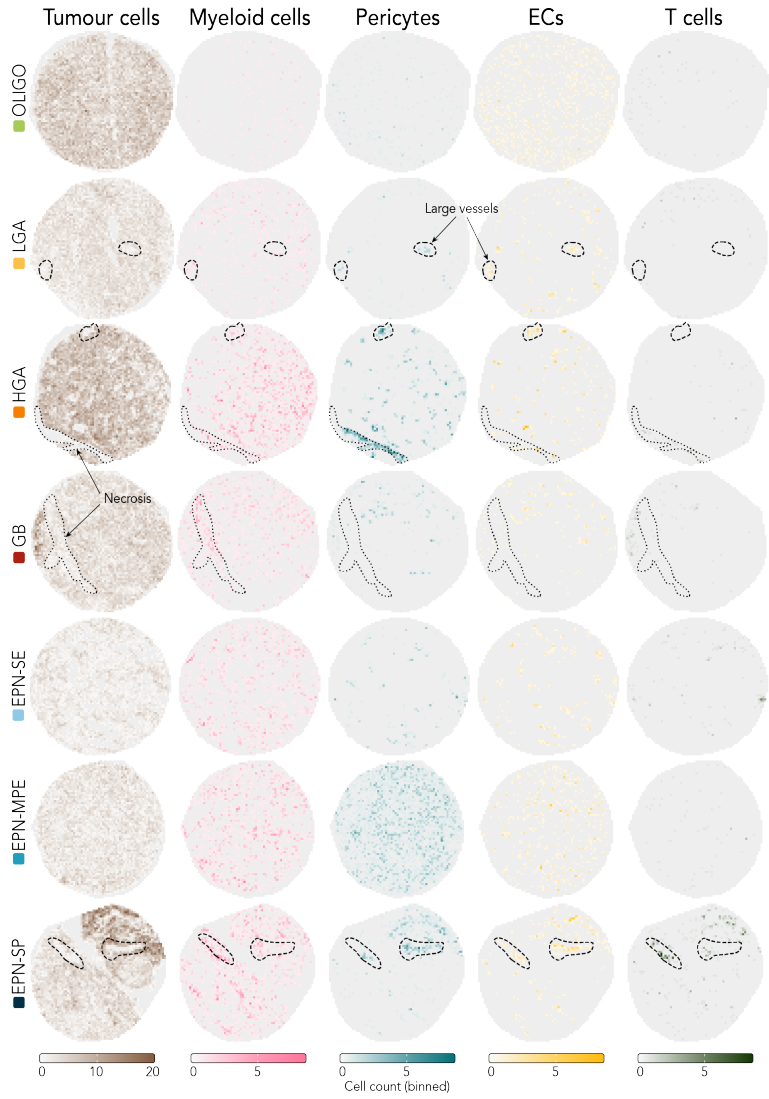

Figure 4.3: *Spatial landscape of gliomas.* Cell type distribution highlighting large vessels and necrotic areas.

In summary, a spatial single cell transcriptomic atlas, comprising 313 samples from 283 patients diagnosed with seven glioma types, was constructed. The custom target

gene panel and the subsequent computational data processing demonstrated efficacy in capturing the cellular and spatial landscape of all the examined gliomas. The generated atlas is a valuable resource, offering an opportunity to explore its content and gain profound insights into glioma biology.

4.3.2 *Glioma expression variability is successfully recovered*

To ascertain the extent to which the panel is capable of capturing expression heterogeneity, I employed consensus NMF to identify nine transcription programs within the Tumour cluster (Figure 4.4, Section 4.2). The program annotation was conducted using the top contributing genes (Figure 4.4a, Figure A.10). Five tumour transcription programs were consistent with known glioma states [92, 115, 247], including AC-, MES-, OPC-, and NPC-like states. I focused on the analysis of seven programs that accounted for more than 80% of total signal in Tumour cells (Figure A.11).

Similarly to the previous study discussed in this dissertation (see Chapter 3), I identified two transcription programs associated with the mesenchymal state: a Hypoxic MES2-like program and Gliosis MES1-like program. Interestingly, I did not identify statistically significant difference in Hypoxic MES2-like program proportions between LGAs and HGAs samples (Figure 4.4b, $P > 0.05$, two-sided Mann-Whitney U test with Benjamini-Hochberg correction). However, the difference in Hypoxic MES2-like tumour cell proportion between EPN types was found significant ($P = 3.7120 \times 10^{-5}$, two-sided Kruskal test with Benjamini-Hochberg correction). The Hypoxic MES2-like cells were most prevalent in GB in comparison to other gliomas (Figure 4.4b), aligning with necrosis being a known distinctive morphologic feature of GB [239]. Of the three EPN types, EPN-MPE exhibited the highest proportion of the Hypoxic MES2-like tumour cells. The Gliosis MES1-like program was characterised by a combination of mesenchymal, neuronal, and astrocytic markers. Cells express-

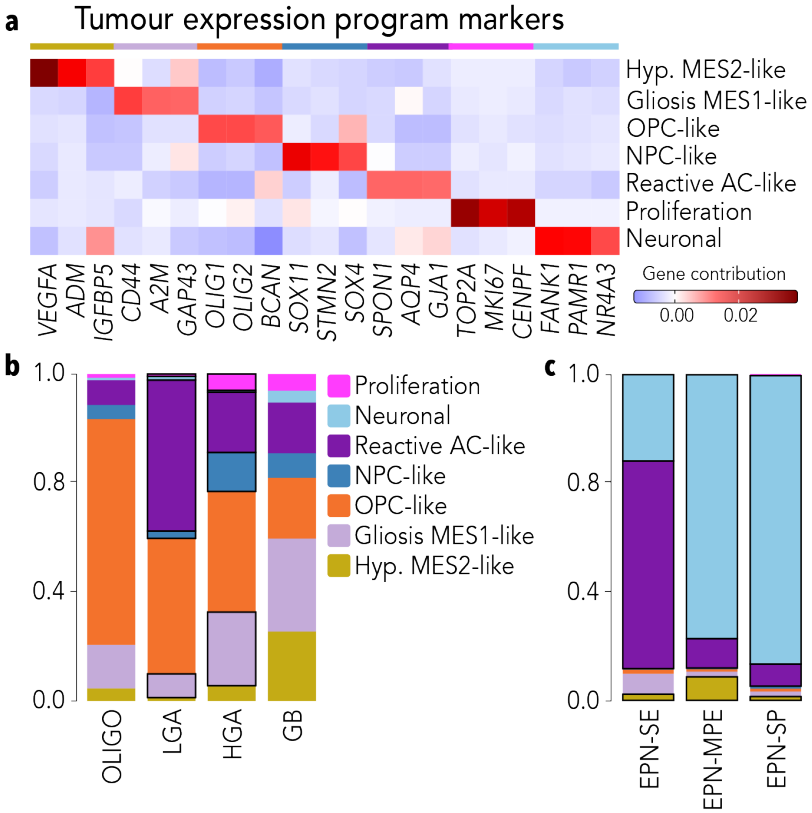

Figure 4.4: *Tumour transcription programs.* Program marker genes *(a)*. Program distribution across adult-type diffuse gliomas *(b)* and EPNs *(c)*. Programs with significantly different proportions are highlighted with bold frames. Spatial distribution of tumour programs *(c)*. Reac., reactive.

ing this program were most prominent in HGA and GB (Figure 4.4b). I did not identify statistically significant differences in Gliosis MES1-like cell proportions between neither astrocytomas nor EPNs ($P > 0.05$, two-sided Mann-Whitney U test and Kruskal test with Benjamini-Hochberg correction, correspondingly).

Two transcription programs represented progenitor-like Tumour cells. The OPC-like program was characterised by OPC (*OLIG1*, *OLIG2*) and stem cell (*BCAN*) markers, whereas the NPC-like program was associated with both NPC1 and NPC2 markers (Figure 4.4a). A fifth program, annotated as a Proliferation program, was enriched in genes associated with cell cycle. Notably, the two progenitor-like

programs as well as the Proliferation program were largely restricted to adult-type diffuse gliomas, while in EPNs Tumour cells expressing these transcription programs were rare (Figure 4.4c). Furthermore, in OLIGO samples, the OPC-like program accounted for the largest proportion of Tumour cells (Figure 4.4b). Additionally, I observed statistically significant differences in the NPC-like and Proliferation program proportions between HGA and LGA samples ($P = 8.5634 \times 10^{-6}$ and $P = 6.1808 \times 10^{-11}$, two-sided Mann-Whitney U test with Benjamini-Hochberg correction, correspondingly).

Finally, I identified a Reactive AC-like program, that likely represented brain-resident astrocytes, and a Neuronal program, that was associated with neuronal markers (*NR4A3*, *KCNMA1*) and genes assigned to the Ependymal program in previous studies (*FANK1*, *PAMR1*) [102]. Interestingly, the Reactive AC-like cells were particularly abundant in low grade tumours, namely EPN-SEs and LGAs. In EPN-SE samples, this program accounted for the majority of Tumour cells, whereas in the two other EPN types most Tumour cells expressed the Neuronal program.

To summarise, the findings presented herein indicate the efficacy of the designed panel and the optimised computational workflow in recovering expression within the tumour clusters. The data set contained sufficient information to enable the identification of glioma programs previously reported in the literature and to characterise their distribution across seven glioma types.

4.3.3 *Spatial neighbourhoods introduce additional complexity*

The primary objective of a spatial study is to discern novel spatial biomarkers when examining complex cell communities, thus surpassing the information attainable at the level of individual cell types. To this end, I constructed a neighbourhood graph and clustered it to produce five spatial neighbourhoods (Figure 4.5, Section 4.2). Given that one of the neighbourhoods was predominantly associated with EPNs, I focused on the remaining four, that were bet-

ter sampled across all glioma types. Furthermore, since adult-type diffuse gliomas were better represented in the cohort compared to EPNs, I utilised them to compare spatial neighbourhood distribution in different gliomas.

The Hypoxic MES2-like tumour cells formed a discrete Hypoxic neighbourhood with an admixture of Myeloid cells (Figure 4.5a). The prevalence of the neighbourhood was most pronounced in GB, followed by OLIGO, HGA, and EPN-MPE samples (Figure 4.5b). The LGAs and two other EPNs exhibited a minimal presence of the Hypoxic neighbourhood, a finding that is consistent with the established grading criteria, wherein low grades correspond to the absence of necrosis [72].

The second neighbourhood, termed Wound-healing, comprised Gliosis MES1-like tumour cells, a smaller fraction of Myeloid cells, as well as OPC-like and Reactive AC-like cells (Figure 4.5a). The largest proportion of the Wound-healing neighbourhood was identified in GB samples, at approximately 30%, followed by HGA, OLIGO, and LGA (Figure 4.5b). It is noteworthy that, in EPN-MPE, which contained the largest fraction of the Hypoxic neighbourhood, the Wound-healing neighbourhood constituted a minor proportion. In terms of spatial relationships, the presence of Hypoxic and Wound-healing neighbourhoods in close proximity to each other possibly indicates an area of tissue damage and its active repair (Figure 4.5c).

The Proliferatory neighbourhood primarily consisted of OPC-like cells and constituted the most substantial proportion observed in OLIGOs. Finally, the Perivascular neighbourhood combined vascular and immune cells, together with AC-like cells, which are known to be integral to the integrity of the blood-brain barrier [264]. This neighbourhood was found to be abundant across all adult-type diffuse gliomas and EPN-SE (Figure 4.5b). In terms of spatial location, the Perivascular neighbourhood was characterised by its localisation in non-necrotic areas (Figure 4.5c).

To summarise, the established workflow employed for the design of a custom panel, in conjunction with the analysis of the resulting data, has facilitated an investigation

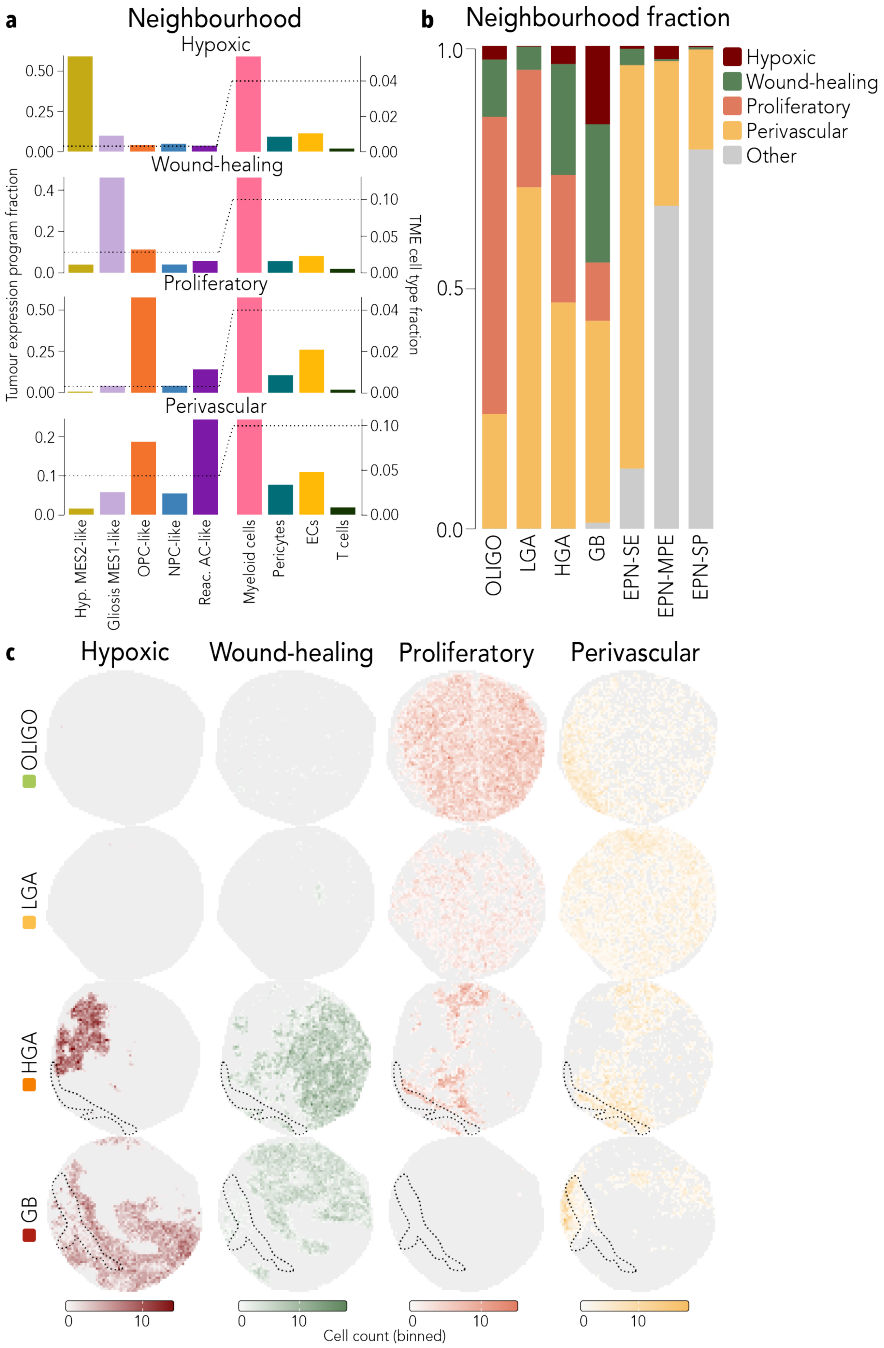

Figure 4.5: *Spatial neighbourhoods of gliomas.* Cell type composition of spatial neighbourhoods *(a)*. Neighbourhood distribution across gliomas *(b)* and their spatial location *(c)*.

into the cellular and spatial landscape of seven gliomas at the single cell resolution.

4.4 DISCUSSION

This Chapter presented a distinctive study utilising a novel spatially-resolved single cell transcriptomics methodology. Within the scope of the study, I conducted a comprehensive evaluation of a multitude of computational methods and formulated a set of considerations and best practices for panel design and spatial data analysis. The results presented here demonstrated that a panel tailored for spatial cancer profiling can be developed even in the absence or restricted availability of tumour-specific scRNA-seq data. As I outlined above, the panel demonstrated its efficacy in the recovery of cells and transcript data that was consistent across all seven glioma types. In view of the restricted nature of patient data, the development of a gene panel in the absence of prior information about single cell expression in the target cancer type can assist in the investigation of spatial architecture in rare tumour entities.

Furthermore, while there have been numerous studies that examined the transcriptomic and spatial diversity of distinct gliomas [101, 115, 247, 265], this is the first study to provide a comprehensive pan-glioma overview. This is of particular importance in the context of ependymal tumours, which have been the focus of fewer scRNA-seq studies over the recent years. According to the findings presented in this Chapter and in line with previous studies [101, 102], Reactive AC-like, Hypoxic MES2-like, and Gliosis MES1-like transcription states were observed across all seven glioma types. Interestingly, although all EPN-SP and EPN-MPE tumours were classified as grade 2, EPN-MPEs samples exhibited the greatest proportion of Hypoxic MES2-like cells, likely reflecting the prevailing notion that the histopathological grading of ependymal tumours is of limited value [87].

Several tumour programs were unique to certain gliomas. For instance, the Proliferating, OPC-like, and NPC-like tu-

mour cells, were predominantly observed in adult-type diffuse gliomas, with their presence in EPNs being minimal. Furthermore, one program, the Neuronal program, was largely limited to ependymal tumours and aligned with a program defined in one previous study of supratentorial EPNs [102]. Moreover, all three EPN types under consideration were mostly dominated by one differentiated tumour program, with the AC-like state dominating grade 1 and the Neuronal state dominating grade 2 tumours. Interestingly, in previous studies [101], undifferentiated transcription states were found in supratentorial and posterior fossa EPNs. Since all supratentorial and posterior fossa EPNs discussed in this study were diagnosed as EPN-SEs, the observed discrepancies in tumour transcription states may provide further evidence in support of such separation.

Furthermore, in contrast to the current four-state model of GB and a three-state model of IDH-mutant gliomas, I demonstrated here that the four adult-type diffuse gliomas exhibit shared programs and differ solely in their relative proportions. Moreover, the MES-like state corresponded to two distinct programs: one associated with a hypoxic response and the other linked to a wound-healing signature. Finally, I identified significant differences in the tumour state composition between LGA and HGA samples, a finding that lends support to the grouping of astrocytomas into low- and high-grade groups. Nevertheless, the clinical importance of these disparities is yet to be elucidated.

Lastly, the present study provided insights into the spatial architecture of gliomas. Except for the Perivascular community, each recurrent neighbourhood corresponded predominantly to a single tumour program. Among adult-type diffuse gliomas, GB exhibited the largest area of Hypoxic neighbourhoods, while the LGAs demonstrated only a minor presence of this spatial community. Although the characterisation of ependymal tumours was constrained by the limited number of samples available for analysis, it was evident that grade 1 EPN-SEs had the largest proportion of Perivascular neighbourhoods. Finally, both low-

grade glioma types, LGA and EPN-SE, displayed the greatest fraction of Perivascular neighbourhoods.

Notwithstanding the contributions of this study, it is important to acknowledge its limitations. Firstly, the limited panel size and the paucity of available EPN scRNA-seq data made it challenging to incorporate markers that would encompass all transcriptional variations in gliomas. This limitation is particularly evident in the case of EPNs, for which ependymal-like or stem-like tumour programs that were previously reported [101, 102] could not be identified. The advent of enhanced technologies, which have the potential to profile a greater quantity of genes, may provide a more comprehensive solution in the coming years. Secondly, especially with regard to ependymal tumours, the available data set was constrained in terms of the number of samples, and the presence of multiple methylation classes within each EPN type rendered the comparison difficult.

In summary, the findings presented in this Chapter delineate the distinguishing features of seven glioma types in terms of tumour transcription and spatial architecture. The atlas produced in the present study is intended to function as a valuable resource and a basis for subsequent studies. The utilisation of TMAs enabled a high-throughput spatial profiling of several hundreds of samples, thereby empowering prospective detection of clinically relevant biomarkers.

5

DISCUSSION AND OUTLOOK

Notwithstanding the considerable variability in the cellular composition of tumour bulk, in certain instances, the neoplastic contribution can be as modest as one-tenth of the total cell count [266]. It is therefore paramount to emphasise the necessity of a nuanced examination of the TME. Moreover, a growing body of research is revealing the ways in which spatial interactions between TME and cancer, as well as the intricate neighbourhoods that emerge from these interactions, hold the potential to guide cancer treatment strategies [60, 66, 68, 111, 113]. In the present dissertation I set out the rationale for the significance of investigating the spatial aspect of cancer evolution, and employ the methodology of spatial single cell transcriptomics to illuminate the intricacies of cancer ecosystems observed in gliomas.

The majority of preceding studies investigating spatial structures of tumours have utilised the Visium technology, a method that relies on the profiling of spots, despite the capacity to generate a whole-transcriptome readout [40, 60, 110, 113, 115, 213, 228, 267, 268]. Despite the advent of novel spatial profiling methodologies that offer single-cell resolution, the analysis of the associated data remains non-standardised and unvalidated. Within the scope of this dissertation, I conducted an extensive comparison of numerous analytical methods, and presented the optimal approaches along with newly developed strategies in Chapter 2. At the inception of my PhD, the availability of resources to inform researchers engaged in spatial single cell transcriptomics studies was severely limited. In the context of our research group, I was the first member to establish the foundation for future spatial analysis initiatives within the group. Despite the potential for technological advancements to alleviate the analytical demands placed upon researchers in future research projects, the hope is

that the present contribution will facilitate the design of studies and the generation of valuable insights.

The present dissertation has furthermore made a contribution to research on glioma. Chapter 3 detailed the pioneering spatial investigation into the complex spatial structure of necrotic tissue in GB patients. The study presented a spatially resolved single-cell atlas consisting of over a million cells from nine patients diagnosed with GB and RN. Notably, I demonstrated that, in contrast to progressive GB, tumour cells in RN samples exhibited low levels of *EGFR* expression despite the genetic amplification, and did not proliferate or express a progenitor signature. This is noteworthy, as it suggests a potential reason why the large number of tumour cells identified by transcriptomics are not apparent during histopathological examination. Moreover, the finding that in RN there was an abundance of BAMs in the brain parenchyma offers an avenue for future research, particularly in the development of a non-invasive diagnostic test. The potential for blocking BAM cytokine release to alleviate clinical symptoms in GB patients merits further investigation. Chapter 4 presents the development of a large-scale pan-glioma study. In this study, I demonstrated how a targeted panel of genes can be designed even in the absence of exhaustive scRNA-seq atlas. The study provides a comprehensive analysis of tumour expression and spatial patterns in seven glioma types in adults. This analysis served as a foundation for further investigations in our group aimed at elucidating the associations between various clinical parameters and the identified tumour characteristics. The designed panel and the compiled data set were utilised to generate a second, upgraded panel that comprised 500 genes and covered a more extensive range of CNS tumour types.

Despite a firm conviction concerning the significance of the work delineated in this study, I believe it is imperative to address certain challenges to identify prospects for future efforts. For instance, effective segmentation, that would accommodate the wide range of tissue types and technological variations, remains one of the most persist-

ing challenges in the field. The employment of nuclear transcripts for the definition of cellular identities, in conjunction with cytoplasmic transcripts for the identification of more subtle variations in expression, may prove advantageous. The combination of spatial transcriptomics with antibody labelling is another option that may assist in identifying cell membranes, thereby facilitating accurate signal assignment to cells. Nonetheless, in certain instances, it can be challenging to find a suitable membrane marker. Another problem, particularly within the context of the brain, is the presence of cells with many dendrites that are difficult to segment. In such cases, the development of new computational models with the capacity to separate spatial signals could prove to be advantageous [197].

It is also evident that certain hurdles demand enhancements of a technological nature. A notable constraint of *in situ* techniques is the rate at which they permit the acquisition of images. As current imaging methods require a relatively high level of magnification, spatial studies are limited to profiling either a small number of whole-slide samples (as evidenced by the study in Chapter 3) or a greater number of samples of a reduced size (as evidenced by the study in Chapter 4). The choice therefore lies in determining whether to opt for sufficiently large fields of view for robust spatial characterisation or to ensure the availability of extensive data to empower statistically substantiated conclusions. Furthermore, both studies reported in this dissertation focused on the profiling of only a few hundred genes, which hindered the ability to conduct a more comprehensive investigation into the spatial architecture of gliomas. Moreover, a meticulous gene selection process was imperative, which required reference data sets and resulted in an increased timelines. The advent of novel approaches with higher multiplexity and enhanced image acquisition times has the potential to facilitate a thorough characterisation of cancer ecosystems.

Notwithstanding the aforementioned challenges, a plethora of opportunities for spatial cancer research are readily apparent. Recent studies investigating the spatial architectures of cancer have begun to elucidate the mech-

anisms through which various types of TME and cancer cells establish spatial communities and engage in complex dynamic interactions [47, 65, 66]. Advances in technology are set to delineate the influence of these spatial structures on the treatment of cancer [60, 66, 68, 111, 113]. This could be of particular benefit in the case of ependymal tumours, as only a small number of independent prognostic factors have been identified thus far, and the incorporation of molecular biomarkers could present a promising new avenue [85]. In Chapter 4, I paved the way by describing recurrent spatial neighbourhoods that are shared across seven gliomas or are unique to certain types. In Chapter 3, I utilised single cell transcriptomics to identify spatial and functional heterogeneity among myeloid cell populations between samples with two distinct histological diagnoses. The future studies within the domain of glioma research could further delve into the influence of glioma proximity to neurons and its interconnectivity with astrocytes on cancer evolution and therapeutic resistance [269].

Another emergent field is multimodal spatial profiling, that could enhance our understanding of the molecular mechanisms underlying cancer evolution [270]. The potential for simultaneous spatial characterisation of the same sample at the transcriptomic, proteomic and genetic levels could assist in addressing the current challenge of aligning molecular phenotypes found in consecutive tissue sections. Spatial genomics represents a particularly exciting modality, with the potential to improve prediction of the clinical course of a disease and underpin the diverse clinical outcomes of genetically identical clones, attributable to the varying characteristics of the TME [32, 140]. Moreover, the evolution of advanced technologies in the field of artificial intelligence promises to transform the diagnostics by integrating the analysis of histopathological features and the assessment of spatially resolved molecular signals.

Currently, there are numerous ongoing initiatives aimed at generating useful resources that delineate the molecular characteristics of various cancers, with examples including The Cancer Genome Atlas [271]. In a similar vein, the deployment of spatial profiling methodologies holds promise

in the generation of comprehensive and informative censuses of spatial cancer communities across diverse human tissues. Such collections of reference data could function as sources of knowledge, serving to inform prospective studies aimed at addressing nuanced biological questions.

BIBLIOGRAPHY

1. Seferbekova, Z. *et al.* Spatial biology of cancer evolution. *Nature reviews. Genetics* (Dec. 2022).
2. McGranahan, N. *et al.* Clonal status of actionable driver events and the timing of mutational processes in cancer evolution. *Science translational medicine* **7,** 283ra54 (Apr. 2015).
3. Dentro, S. C. *et al.* Characterizing genetic intra-tumor heterogeneity across 2,658 human cancer genomes. *Cell* **184,** 2239–2254.e39 (Apr. 2021).
4. Nik-Zainal, S. *et al.* The life history of 21 breast cancers. *Cell* **149,** 994–1007 (May 2012).
5. Andor, N. *et al.* Pan-cancer analysis of the extent and consequences of intratumor heterogeneity. *Nature medicine* **22,** 105–113 (Jan. 2016).
6. Moore, L. *et al.* The mutational landscape of human somatic and germline cells. *Nature* **597,** 381–386 (Sept. 2021).
7. Lynch, M. D. *et al.* Spatial constraints govern competition of mutant clones in human epidermis. *Nature communications* **8,** 1119 (Oct. 2017).
8. Li, R. *et al.* A body map of somatic mutagenesis in morphologically normal human tissues. *Nature* **597,** 398–403 (Sept. 2021).
9. Fabre, M. A. *et al.* The longitudinal dynamics and natural history of clonal haematopoiesis. *Nature* **606,** 335–342 (June 2022).
10. Jaiswal, S. *et al.* Age-related clonal hematopoiesis associated with adverse outcomes. *The New England journal of medicine* **371,** 2488–2498 (Dec. 2014).

11. Genovese, G. *et al.* Clonal hematopoiesis and blood-cancer risk inferred from blood DNA sequence. *The New England journal of medicine* **371,** 2477–2487 (Dec. 2014).

12. Bernstein, N. *et al.* Analysis of somatic mutations in whole blood from 200,618 individuals identifies pervasive positive selection and novel drivers of clonal hematopoiesis. *Nature genetics* **56,** 1147–1155 (June 2024).

13. West, J. *et al.* Normal tissue architecture determines the evolutionary course of cancer. *Nature communications* **12,** 2060 (Apr. 2021).

14. Martincorena, I. *et al.* Tumor evolution. High burden and pervasive positive selection of somatic mutations in normal human skin. *Science (New York, N.Y.)* **348,** 880–886 (May 2015).

15. Li, R. *et al.* Macroscopic somatic clonal expansion in morphologically normal human urothelium. *Science (New York, N.Y.)* **370,** 82–89 (Oct. 2020).

16. Lawson, A. R. J. *et al.* Extensive heterogeneity in somatic mutation and selection in the human bladder. *Science (New York, N.Y.)* **370,** 75–82 (Oct. 2020).

17. Martincorena, I. *et al.* Somatic mutant clones colonize the human esophagus with age. *Science (New York, N.Y.)* **362,** 911–917 (Nov. 2018).

18. Yokoyama, A. *et al.* Age-related remodelling of oesophageal epithelia by mutated cancer drivers. *Nature* **565,** 312–317 (Jan. 2019).

19. Brunner, S. F. *et al.* Somatic mutations and clonal dynamics in healthy and cirrhotic human liver. *Nature* **574,** 538–542 (Oct. 2019).

20. Bizzotto, S. & Walsh, C. A. Genetic mosaicism in the human brain: from lineage tracing to neuropsychiatric disorders. *Nature reviews. Neuroscience* **23,** 275–286 (May 2022).

21. Buhigas, C. *et al.* The architecture of clonal expansions in morphologically normal tissue from cancerous and non-cancerous prostates. *Molecular cancer* **21,** 183 (Sept. 2022).

22. Grossmann, S. *et al.* Development, maturation, and maintenance of human prostate inferred from somatic mutations. *Cell stem cell* **28,** 1262–1274.e5 (July 2021).

23. Humphries, A. & Wright, N. A. Colonic crypt organization and tumorigenesis. *Nature reviews. Cancer* **8,** 415–424 (June 2008).

24. Lee-Six, H. *et al.* The landscape of somatic mutation in normal colorectal epithelial cells. *Nature* **574,** 532–537 (Oct. 2019).

25. Coorens, T. H. H. *et al.* The somatic mutation landscape of normal gastric epithelium. *Nature* **640,** 418–426 (Apr. 2025).

26. Yamaguchi, M. *et al.* Spatiotemporal dynamics of clonal selection and diversification in normal endometrial epithelium. *Nature communications* **13,** 943 (Feb. 2022).

27. Moore, L. *et al.* The mutational landscape of normal human endometrial epithelium. *Nature* **580,** 640–646 (Apr. 2020).

28. Anglesio, M. S. *et al.* Cancer-associated mutations in endometriosis without cancer. *The New England journal of medicine* **376,** 1835–1848 (May 2017).

29. Suda, K. *et al.* Clonal expansion and diversification of cancer-associated mutations in endometriosis and normal endometrium. *Cell reports* **24,** 1777–1789 (Aug. 2018).

30. Kakiuchi, N. *et al.* Frequent mutations that converge on the NFKBIZ pathway in ulcerative colitis. *Nature* **577,** 260–265 (Jan. 2020).

31. Risom, T. *et al.* Transition to invasive breast cancer is associated with progressive changes in the structure and composition of tumor stroma. *Cell* **185,** 299–310.e18 (Jan. 2022).

32. Lomakin, A. *et al.* Spatial genomics maps the structure, nature and evolution of cancer clones. *Nature,* 1–9 (Nov. 2022).

33. Louis, D. N. *et al.* The 2021 WHO classification of tumors of the Central Nervous System: A summary. *Neuro-oncology* **23,** 1231–1251 (Aug. 2021).

34. Matoso, A. & Epstein, J. I. Grading of prostate cancer: Past, present, and future. *Current urology reports* **17,** 25 (Mar. 2016).

35. Bloom, H. J. & Richardson, W. W. Histological grading and prognosis in breast cancer; a study of 1409 cases of which 359 have been followed for 15 years. *British journal of cancer* **11,** 359–377 (Sept. 1957).

36. Brierley, J. D. *et al. TNM classification of malignant tumours* (John wiley & sons, 2025).

37. Mo, C.-K. *et al.* Tumour evolution and microenvironment interactions in 2D and 3D space. *Nature* **634,** 1178–1186 (Oct. 2024).

38. Nishimura, T. *et al.* Evolutionary histories of breast cancer and related clones. *Nature* **620,** 607–614 (Aug. 2023).

39. Heiser, C. N. *et al.* Molecular cartography uncovers evolutionary and microenvironmental dynamics in sporadic colorectal tumors. *Cell* **186,** 5620–5637.e16 (Dec. 2023).

40. Erickson, A. *et al.* Spatially resolved clonal copy number alterations in benign and malignant tissue. *Nature* **608,** 360–367 (Aug. 2022).

41. Woodcock, D. J. *et al.* Prostate cancer evolution from multilineage primary to single lineage metastases with implications for liquid biopsy. *Nature communications* **11,** 5070 (Oct. 2020).

42. Rasche, L. *et al.* The spatio-temporal evolution of multiple myeloma from baseline to relapse-refractory states. *Nature communications* **13,** 4517 (Aug. 2022).

43. Noble, R. *et al.* Spatial structure governs the mode of tumour evolution. *Nature ecology & evolution* **6,** 207–217 (Feb. 2022).

44. Waclaw, B. *et al.* A spatial model predicts that dispersal and cell turnover limit intratumour heterogeneity. *Nature* **525,** 261–264 (Sept. 2015).

45. Cooper, L. A. *et al.* PanCancer insights from The Cancer Genome Atlas: the pathologist's perspective. *The Journal of pathology* **244,** 512–524 (Apr. 2018).

46. Lambert, A. W., Pattabiraman, D. R. & Weinberg, R. A. Emerging biological principles of metastasis. *Cell* **168,** 670–691 (Feb. 2017).

47. Karras, P. *et al.* A cellular hierarchy in melanoma uncouples growth and metastasis. *Nature* (Sept. 2022).

48. Gundem, G. *et al.* The evolutionary history of lethal metastatic prostate cancer. *Nature* **520,** 353–357 (Apr. 2015).

49. Hu, Z. *et al.* Multi-cancer analysis of clonality and the timing of systemic spread in paired primary tumors and metastases. *Nature genetics* **52,** 701–708 (July 2020).

50. Nguyen, B. *et al.* Genomic characterization of metastatic patterns from prospective clinical sequencing of 25,000 patients. *Cell* **185,** 563–575.e11 (Feb. 2022).

51. Priestley, P. *et al.* Pan-cancer whole-genome analyses of metastatic solid tumours. *Nature* **575,** 210–216 (Nov. 2019).

52. Martínez-Jiménez, F. *et al.* Pan-cancer whole-genome comparison of primary and metastatic solid tumours. *Nature* **618,** 333–341 (June 2023).

53. Brastianos, P. K. *et al.* Genomic characterization of brain metastases reveals branched evolution and potential therapeutic targets. *Cancer discovery* **5,** 1164–1177 (Nov. 2015).

54. De Mattos-Arruda, L. *et al.* The genomic and immune landscapes of lethal metastatic breast cancer. *Cell reports* **27,** 2690–2708.e10 (May 2019).

55. Noorani, A. *et al.* Genomic evidence supports a clonal diaspora model for metastases of esophageal adenocarcinoma. *Nature genetics* **52,** 74–83 (Jan. 2020).

56. Brown, D. *et al.* Phylogenetic analysis of metastatic progression in breast cancer using somatic mutations and copy number aberrations. *Nature Communications* **8,** 1–13 (Apr. 2017).

57. Marusyk, A. *et al.* Non-cell-autonomous driving of tumour growth supports sub-clonal heterogeneity. *Nature* **514,** 54–58 (Oct. 2014).

58. Colom, B. *et al.* Mutant clones in normal epithelium outcompete and eliminate emerging tumours. *Nature,* 1–5 (Oct. 2021).

59. Gaynor, L. *et al.* Crypt density and recruited enhancers underlie intestinal tumour initiation. *Nature* **640,** 231–239 (Apr. 2025).

60. Danenberg, E. *et al.* Breast tumor microenvironment structures are associated with genomic features and clinical outcome. *Nature genetics* **54,** 660–669 (May 2022).

61. Jackson, H. W. *et al.* The single-cell pathology landscape of breast cancer. *Nature* **578,** 615–620 (Feb. 2020).

62. He, S. *et al.* Starfysh integrates spatial transcriptomic and histologic data to reveal heterogeneous tumor-immune hubs. *Nature biotechnology* **43,** 223–235 (Feb. 2025).

63. Wang, X. Q. *et al.* Spatial predictors of immunotherapy response in triple-negative breast cancer. *Nature* **621,** 868–876 (Sept. 2023).

64. Hammerl, D. *et al.* Spatial immunophenotypes predict response to anti-PD1 treatment and capture distinct paths of T cell evasion in triple negative breast cancer. *Nature communications* **12,** 5668 (Sept. 2021).

65. Nirmal, A. J. *et al.* The spatial landscape of progression and immunoediting in primary melanoma at single-cell resolution. *Cancer discovery* **12,** 1518–1541 (June 2022).

66. Ji, A. L. *et al.* Multimodal Analysis of Composition and Spatial Architecture in Human Squamous Cell Carcinoma. *Cell* **182,** 497–514.e22 (July 2020).

67. Grünwald, B. T. *et al.* Spatially confined sub-tumor microenvironments in pancreatic cancer. *Cell* **184,** 5577–5592.e18 (Oct. 2021).

68. Keren, L. *et al.* A Structured Tumor-Immune Microenvironment in Triple Negative Breast Cancer Revealed by Multiplexed Ion Beam Imaging. *Cell* **174,** 1373–1387.e19 (Sept. 2018).

69. Capper, D. *et al.* DNA methylation-based classification of central nervous system tumours. *Nature* **555,** 469–474 (Mar. 2018).

70. Ostrom, Q. T. *et al.* CBTRUS statistical report: Primary brain and other central nervous system tumors diagnosed in the United States in 2016-2020. *Neuro-oncology* **25,** iv1–iv99 (Oct. 2023).

71. Louis, D. N. *et al.* The 2016 World Health Organization Classification of Tumors of the Central Nervous System: a summary. *Acta neuropathologica* **131,** 803–820 (June 2016).

72. Louis, D. N. *et al.* cIMPACT-NOW update 6: new entity and diagnostic principle recommendations of the cIMPACT-Utrecht meeting on future CNS tumor classification and grading. *Brain pathology (Zurich, Switzerland)* **30,** 844–856 (July 2020).

73. Minami, J. K. *et al.* CDKN2A deletion remodels lipid metabolism to prime glioblastoma for ferroptosis. *Cancer cell* (May 2023).

74. Verhaak, R. G. W. *et al.* Integrated genomic analysis identifies clinically relevant subtypes of glioblastoma characterized by abnormalities in PDGFRA, IDH1, EGFR, and NF1. *Cancer cell* **17,** 98–110 (Jan. 2010).

75. Nair, N. U. *et al.* Chromosome 7 gain compensates for chromosome 10 loss in glioma. *Cancer research* **84,** 3464–3477 (Oct. 2024).

76. Drexler, R. *et al.* DNA methylation subclasses predict the benefit from gross total tumor resection in IDH-wildtype glioblastoma patients. *Neuro-oncology* **25,** 315–325 (Feb. 2023).

77. Kessler, T. *et al.* Prognostic markers of DNA methylation and next-generation sequencing in progressive glioblastoma from the EORTC-26101 trial. *Clinical cancer research: an official journal of the American Association for Cancer Research* **29,** 3892–3900 (Oct. 2023).

78. Weller, M. *et al.* EANO guidelines on the diagnosis and treatment of diffuse gliomas of adulthood. *Nature reviews. Clinical oncology* **18,** 170–186 (Mar. 2021).

79. Capper, D. *et al.* EANO guideline on rational molecular testing of gliomas, glioneuronal, and neuronal tumors in adults for targeted therapy selection. *Neuro-oncology* **25,** 813–826 (May 2023).

80. Hegi, M. E. *et al.* MGMT gene silencing and benefit from temozolomide in glioblastoma. *The New England journal of medicine* **352,** 997–1003 (Mar. 2005).

81. Mellinghoff, I. K. *et al.* Vorasidenib in IDH1- or IDH2-mutant low-grade glioma. *The New England journal of medicine* **389,** 589–601 (Aug. 2023).

82. Platten, M. *et al.* A vaccine targeting mutant IDH1 in newly diagnosed glioma. *Nature* **592,** 463–468 (Apr. 2021).

83. Schumacher, T. *et al.* A vaccine targeting mutant IDH1 induces antitumour immunity. *Nature* **512,** 324–327 (Aug. 2014).

84. Omuro, A. *et al.* Radiotherapy combined with nivolumab or temozolomide for newly diagnosed glioblastoma with unmethylated MGMT promoter: An international randomized phase III trial. *Neuro-oncology* **25,** 123–134 (Jan. 2023).

85. Rudà, R. *et al.* EANO guidelines for the diagnosis and treatment of ependymal tumors. *Neuro-oncology* **20,** 445–456 (Mar. 2018).

86. Pajtler, K. W. *et al.* Molecular classification of ependymal tumors across all CNS compartments, histopathological grades, and age groups. *Cancer cell* **27,** 728–743 (May 2015).

87. Ellison, D. W. *et al.* Histopathological grading of pediatric ependymoma: reproducibility and clinical relevance in European trial cohorts. *Journal of negative results in biomedicine* **10,** 7 (May 2011).

88. Pombo Antunes, A. R. *et al.* Understanding the glioblastoma immune microenvironment as basis for the development of new immunotherapeutic strategies. *eLife* **9** (Feb. 2020).

89. Vargas López, A. J. Glioblastoma in adults: a Society for Neuro-Oncology (SNO) and European Society of Neuro-Oncology (EANO) consensus review on current management and future directions. *Neuro-oncology* **23,** 502–503 (Mar. 2021).

90. Holland, E. C. Glioblastoma multiforme: the terminator. *Proceedings of the National Academy of Sciences of the United States of America* **97,** 6242–6244 (June 2000).

91. Wang, Q. *et al.* Tumor Evolution of Glioma-Intrinsic Gene Expression Subtypes Associates with Immunological Changes in the Microenvironment. *Cancer cell* **32,** 42–56.e6 (July 2017).

92. Neftel, C. *et al.* An Integrative Model of Cellular States, Plasticity, and Genetics for Glioblastoma. *Cell* **178,** 835–849.e21 (Aug. 2019).

93. Couturier, C. P. *et al.* Single-cell RNA-seq reveals that glioblastoma recapitulates a normal neurodevelopmental hierarchy. *Nature communications* **11,** 3406 (July 2020).

94. Wang, L. *et al.* The phenotypes of proliferating glioblastoma cells reside on a single axis of variation. *Cancer discovery* **9,** 1708–1719 (Dec. 2019).

95. Mossi Albiach, A. *et al.* Futile wound healing drives mesenchymal-like cell phenotypes in human glioblastoma. *bioRxiv,* 2023.09.01.555882 (Sept. 2023).

96. Suvà, M. L. & Tirosh, I. The glioma stem cell model in the era of single-cell genomics. *Cancer cell* **37,** 630–636 (May 2020).

97. Venteicher, A. S. *et al.* Decoupling genetics, lineages, and microenvironment in IDH-mutant gliomas by single-cell RNA-seq. *Science (New York, N.Y.)* **355** (Mar. 2017).

98. Tirosh, I. *et al.* Single-cell RNA-seq supports a developmental hierarchy in human oligodendroglioma. *Nature* **539,** 309–313 (Nov. 2016).

99. Blanco-Carmona, E. *et al.* Tumor heterogeneity and tumor-microglia interactions in primary and recurrent IDH1-mutant gliomas. *Cell reports. Medicine* **4,** 101249 (Nov. 2023).

100. Chaligne, R. *et al.* Epigenetic encoding, heritability and plasticity of glioma transcriptional cell states. *Nature genetics* **53,** 1469–1479 (Oct. 2021).

101. Gojo, J. *et al.* Single-cell RNA-seq reveals cellular hierarchies and impaired developmental trajectories in pediatric ependymoma. *Cancer cell* **38,** 44–59.e9 (July 2020).

102. Jeong, D. *et al.* Single-cell multidimensional profiling of tumor cell heterogeneity in supratentorial ependymomas. *bioRxiv,* 2024.08.07.607066 (Aug. 2024).

103. Zong, H., Parada, L. F. & Baker, S. J. Cell of origin for malignant gliomas and its implication in therapeutic development. *Cold Spring Harbor perspectives in biology* **7,** a020610 (Jan. 2015).

104. Zong, H., Verhaak, R. G. W. & Canoll, P. The cellular origin for malignant glioma and prospects for clinical advancements. *Expert review of molecular diagnostics* **12,** 383–394 (May 2012).

105. Patel, A. P. *et al.* Single-cell RNA-seq highlights intratumoral heterogeneity in primary glioblastoma. *Science* **344,** 1396–1401 (June 2014).

106. Yabo, Y. A., Niclou, S. P. & Golebiewska, A. Cancer cell heterogeneity and plasticity: A paradigm shift in glioblastoma. *Neuro-oncology* **24,** 669–682 (May 2022).

107. Lee, J.-K. *et al.* Spatiotemporal genomic architecture informs precision oncology in glioblastoma. *Nature genetics* **49,** 594–599 (Apr. 2017).

108. Puchalski, R. B. *et al.* An anatomic transcriptional atlas of human glioblastoma. *Science (New York, N.Y.)* **360,** 660–663 (May 2018).

109. Hambardzumyan, D. & Bergers, G. Glioblastoma: Defining tumor niches. *Trends in cancer* **1,** 252–265 (Dec. 2015).

110. Miller, T. E. *et al.* Programs, origins and immunomodulatory functions of myeloid cells in glioma. *Nature,* 1–11 (Feb. 2025).

111. Karimi, E. *et al.* Single-cell spatial immune landscapes of primary and metastatic brain tumours. *Nature* (Feb. 2023).

112. Wang, L. *et al.* A single-cell atlas of glioblastoma evolution under therapy reveals cell-intrinsic and cell-extrinsic therapeutic targets. *Nature cancer* **3,** 1534–1552 (Dec. 2022).

113. Zheng, Y. *et al.* Spatial cellular architecture predicts prognosis in glioblastoma. *Nature communications* **14,** 4122 (July 2023).

114. Brat, D. J. *et al.* Pseudopalisades in glioblastoma are hypoxic, express extracellular matrix proteases, and are formed by an actively migrating cell population. *Cancer research* **64,** 920–927 (Feb. 2004).

115. Greenwald, A. C. *et al.* Integrative spatial analysis reveals a multi-layered organization of glioblastoma. *Cell* **187,** 2485–2501.e26 (May 2024).

116. Varn, F. S. *et al.* Glioma progression is shaped by genetic evolution and microenvironment interactions. *Cell* **185,** 2184–2199.e16 (June 2022).

117. Venkataramani, V. *et al.* Glioblastoma hijacks neuronal mechanisms for brain invasion. *Cell* **185,** 2899–2917.e31 (Aug. 2022).

118. Wälchli, T. *et al.* Shaping the brain vasculature in development and disease in the single-cell era. *Nature reviews. Neuroscience* **24,** 271–298 (May 2023).

119. Hausmann, D. *et al.* Autonomous rhythmic activity in glioma networks drives brain tumour growth. *Nature* **613,** 179–186 (Jan. 2023).

120. Venkatesh, H. S. *et al.* Electrical and synaptic integration of glioma into neural circuits. *Nature* **573,** 539–545 (Sept. 2019).

121. Zong, C. *et al.* Genome-wide detection of single-nucleotide and copy-number variations of a single human cell. *Science (New York, N.Y.)* **338,** 1622–1626 (Dec. 2012).

122. Yuan, J. & Sims, P. A. An automated microwell platform for large-scale single cell RNA-Seq. *Scientific reports* **6,** 33883 (Sept. 2016).

123. Islam, S. *et al.* Characterization of the single-cell transcriptional landscape by highly multiplex RNA-seq. *Genome research* **21,** 1160–1167 (July 2011).

124. Picelli, S. *et al.* Full-length RNA-seq from single cells using Smart-seq2. *Nature protocols* **9,** 171–181 (Jan. 2014).

125. Gerlinger, M. *et al.* Intratumor heterogeneity and branched evolution revealed by multiregion sequencing. *The New England journal of medicine* **366,** 883–892 (Mar. 2012).

126. Yates, L. R. *et al.* Subclonal diversification of primary breast cancer revealed by multiregion sequencing. *Nature medicine* **21,** 751–759 (July 2015).

127. Casasent, A. K. *et al.* Multiclonal invasion in breast tumors identified by Topographic Single Cell Sequencing. *Cell* **172,** 205–217.e12 (Jan. 2018).

128. Ellis, P. *et al.* Reliable detection of somatic mutations in solid tissues by laser-capture microdissection and low-input DNA sequencing. *Nature protocols* **16,** 841–871 (Feb. 2021).

129. Chen, J. *et al.* Spatial transcriptomic analysis of cryosectioned tissue samples with Geo-seq. *Nature protocols* **12,** 566–580 (Mar. 2017).

130. Nichterwitz, S. *et al.* Laser capture microscopy coupled with Smart-seq2 for precise spatial transcriptomic profiling. *Nature communications* **7,** 12139 (July 2016).

131. Emmert-Buck, M. R. *et al.* Laser capture microdissection. *Science (New York, N.Y.)* **274,** 998–1001 (Nov. 1996).

132. Espina, V. *et al.* Laser-capture microdissection. *Nature protocols* **1,** 586–603 (2006).

133. Morton, M. L. *et al.* Identification of mRNAs and lincRNAs associated with lung cancer progression using next-generation RNA sequencing from laser micro-dissected archival FFPE tissue specimens. *Lung cancer (Amsterdam, Netherlands)* **85,** 31–39 (July 2014).

134. Combs, P. A. & Eisen, M. B. Sequencing mRNA from cryo-sliced Drosophila embryos to determine genome-wide spatial patterns of gene expression. *PloS one* **8,** e71820 (Aug. 2013).

135. Junker, J. P. *et al.* Genome-wide RNA Tomography in the zebrafish embryo. *Cell* **159,** 662–675 (Oct. 2014).

136. Ståhl, P. L. *et al.* Visualization and analysis of gene expression in tissue sections by spatial transcriptomics. *Science (New York, N.Y.)* **353,** 78–82 (July 2016).

137. Stickels, R. R. *et al.* Highly sensitive spatial transcriptomics at near-cellular resolution with Slide-seqV2. *Nature biotechnology* **39,** 313–319 (Mar. 2021).

138. Rodriques, S. G. *et al.* Slide-seq: A scalable technology for measuring genome-wide expression at high spatial resolution. *Science (New York, N.Y.)* **363,** 1463–1467 (Mar. 2019).

139. Liu, Y. *et al.* High-spatial-resolution multi-omics sequencing via deterministic barcoding in tissue. *Cell* **183,** 1665–1681.e18 (Dec. 2020).

140. Zhao, T. *et al.* Spatial genomics enables multi-modal study of clonal heterogeneity in tissues. *Nature* **601,** 85–91 (Jan. 2022).

141. Deng, Y. *et al.* Spatial-CUT&Tag: Spatially resolved chromatin modification profiling at the cellular level. *Science (New York, N.Y.)* **375,** 681–686 (Feb. 2022).

142. Deng, Y. *et al.* Spatial profiling of chromatin accessibility in mouse and human tissues. *Nature* **609,** 375–383 (Sept. 2022).

143. Vickovic, S. *et al.* High-definition spatial transcriptomics for in situ tissue profiling. *Nature methods* **16,** 987–990 (Oct. 2019).

144. Cho, C.-S. *et al.* Microscopic examination of spatial transcriptome using Seq-Scope. *Cell* **184,** 3559–3572.e22 (June 2021).

145. Chen, A. *et al.* Spatiotemporal transcriptomic atlas of mouse organogenesis using DNA nanoball-patterned arrays. *Cell* **185,** 1777–1792.e21 (May 2022).

146. Oliveira, M. F. d. *et al.* High-definition spatial transcriptomic profiling of immune cell populations in colorectal cancer. *Nature genetics* **57,** 1512–1523 (June 2025).

147. Rudkin, G. T. & Stollar, B. D. High resolution detection of DNA-RNA hybrids in situ by indirect immunofluorescence. *Nature* **265,** 472–473 (Feb. 1977).

148. Schröck, E. *et al.* Multicolor spectral karyotyping of human chromosomes. *Science (New York, N.Y.)* **273,** 494–497 (July 1996).

149. Raj, A. *et al.* Imaging individual mRNA molecules using multiple singly labeled probes. *Nature methods* **5,** 877–879 (Oct. 2008).

150. Lubeck, E. & Cai, L. Single-cell systems biology by super-resolution imaging and combinatorial labeling. *Nature methods* **9,** 743–748 (June 2012).

151. Lyubimova, A. *et al.* Single-molecule mRNA detection and counting in mammalian tissue. *Nature protocols* **8,** 1743–1758 (Sept. 2013).

152. Codeluppi, S. *et al.* Spatial organization of the somatosensory cortex revealed by osmFISH. *Nature methods* **15,** 932–935 (Nov. 2018).

153. Eng, C.-H. L. *et al.* Transcriptome-scale super-resolved imaging in tissues by RNA seqFISH. *Nature* **568,** 235–239 (Apr. 2019).

154. Lubeck, E. *et al.* Single-cell in situ RNA profiling by sequential hybridization. *Nature methods* **11,** 360–361 (Apr. 2014).

155. Saka, S. K. *et al.* Immuno-SABER enables highly multiplexed and amplified protein imaging in tissues. *Nature biotechnology* **37,** 1080–1090 (Sept. 2019).

156. Gerdes, M. J. *et al.* Highly multiplexed single-cell analysis of formalin-fixed, paraffin-embedded cancer tissue. *Proceedings of the National Academy of Sciences of the United States of America* **110,** 11982–11987 (July 2013).

157. Gut, G., Herrmann, M. D. & Pelkmans, L. Multiplexed protein maps link subcellular organization to cellular states. *Science (New York, N.Y.)* **361,** eaar7042 (Aug. 2018).

158. Lin, J.-R. *et al.* Highly multiplexed immunofluorescence imaging of human tissues and tumors using t-CyCIF and conventional optical microscopes. *eLife* **7,** e31657 (July 2018).

159. Wang, F. *et al.* RNAscope: a novel in situ RNA analysis platform for formalin-fixed, paraffin-embedded tissues. *The Journal of molecular diagnostics: JMD* **14,** 22–29 (Jan. 2012).

160. Ke, R. *et al.* In situ sequencing for RNA analysis in preserved tissue and cells. *Nature methods* **10,** 857–860 (Sept. 2013).

161. Janesick, A. *et al.* High resolution mapping of the tumor microenvironment using integrated single-cell, spatial and in situ analysis. *Nature communications* **14,** 8353 (Dec. 2023).

162. Moffitt, J. R. *et al.* Molecular, spatial, and functional single-cell profiling of the hypothalamic preoptic region. *Science (New York, N.Y.)* **362,** eaau5324 (Nov. 2018).

163. Gyllborg, D. *et al.* Hybridization-based in situ sequencing (HybISS) for spatially resolved transcriptomics in human and mouse brain tissue. *Nucleic acids research* **48,** e112 (Nov. 2020).

164. Goltsev, Y. *et al.* Deep profiling of mouse splenic architecture with CODEX multiplexed imaging. *Cell* **174,** 968–981.e15 (Aug. 2018).

165. Long, M. *et al.* Comparing Xenium 5K and Visium HD data from identical tissue slide at a pathological perspective. *Journal of experimental & clinical cancer research: CR* **44,** 219 (July 2025).

166. Giesen, C. *et al.* Highly multiplexed imaging of tumor tissues with subcellular resolution by mass cytometry. *Nature methods* **11,** 417–422 (Apr. 2014).

167. Angelo, M. *et al.* Multiplexed ion beam imaging of human breast tumors. *Nature medicine* **20,** 436–442 (Apr. 2014).

168. Keren, L. *et al.* MIBI-TOF: A multiplexed imaging platform relates cellular phenotypes and tissue structure. *Science advances* **5,** eaax5851 (Oct. 2019).

169. Ptacek, J. *et al.* Advances in multiplexed ion beam imaging (MIBI) for immune profiling of the tumor microenvironment. *Journal for immunotherapy of cancer* **8,** A51–A51 (Nov. 2020).

170. Mund, A. *et al.* Deep Visual Proteomics defines single-cell identity and heterogeneity. *Nature biotechnology* **40,** 1231–1240 (Aug. 2022).

171. Rovira-Clavé, X. *et al.* Spatial epitope barcoding reveals clonal tumor patch behaviors. *Cancer cell* **40,** 1423–1439.e11 (Nov. 2022).

172. Luecken, M. D. & Theis, F. J. Current best practices in single-cell RNA-seq analysis: a tutorial. *Molecular systems biology* **15,** e8746 (June 2019).

173. Karlsson, M. *et al.* A single-cell type transcriptomics map of human tissues. *Science advances* **7,** eabh2169 (July 2021).

174. Rozenblatt-Rosen, O. *et al.* Building a high-quality Human Cell Atlas. *Nature biotechnology* **39,** 149–153 (Feb. 2021).

175. Franzén, O., Gan, L.-M. & Björkegren, J. L. M. PanglaoDB: a web server for exploration of mouse and human single-cell RNA sequencing data. *Database: the journal of biological databases and curation* **2019,** baz046 (Jan. 2019).

176. Zhang, X. *et al.* CellMarker: a manually curated resource of cell markers in human and mouse. *Nucleic acids research* **47,** D721–D728 (Jan. 2019).

177. Stegle, O., Teichmann, S. A. & Marioni, J. C. Computational and analytical challenges in single-cell transcriptomics. *Nature reviews. Genetics* **16,** 133–145 (Mar. 2015).

178. Dai, M., Pei, X. & Wang, X.-J. Accurate and fast cell marker gene identification with COSG. *Briefings in bioinformatics* **23** (Mar. 2022).

179. Baran, Y. & Doğan, B. scMAGS: Marker gene selection from scRNA-seq data for spatial transcriptomics studies. *Computers in biology and medicine* **155,** 106634 (Feb. 2023).

180. Dumitrascu, B. *et al.* Optimal marker gene selection for cell type discrimination in single cell analyses. *Nature communications* **12,** 1186 (Feb. 2021).

181. Amitay, Y. *et al.* CellSighter: a neural network to classify cells in highly multiplexed images. *Nature communications* **14,** 4302 (July 2023).

182. Chen, X., Chen, S. & Thomson, M. Minimal gene set discovery in single-cell mRNA-seq datasets with ActiveSVM. *Nature Computational Science* **2,** 387–398 (June 2022).

183. Kim, C. *et al.* MarcoPolo: a method to discover differentially expressed genes in single-cell RNA-seq data without depending on prior clustering. *Nucleic acids research* **50,** e71 (July 2022).

184. Liang, S. *et al.* Single-cell manifold-preserving feature selection for detecting rare cell populations. *Nature computational science* **1,** 374–384 (May 2021).

185. Missarova, A. *et al.* geneBasis: an iterative approach for unsupervised selection of targeted gene panels from scRNA-seq. *Genome biology* **22,** 333 (Dec. 2021).

186. Vlot, A. H. C., Maghsudi, S. & Ohler, U. Cluster-independent marker feature identification from single-cell omics data using SEMITONES. *Nucleic acids research* **50,** e107 (Oct. 2022).

187. Liberzon, A. *et al.* Molecular signatures database (MSigDB) 3.0. *Bioinformatics (Oxford, England)* **27,** 1739–1740 (June 2011).

188. Osorio, D. *et al.* Single-cell expression variability implies cell function. *Cells (Basel, Switzerland)* **9,** 14 (Dec. 2019).

189. Maćkiewicz, A. & Ratajczak, W. Principal Components Analysis (PCA). *Computers & Geosciences* **19,** 303–342 (Mar. 1993).

190. Kuemmerle, L. B. *et al.* Probe set selection for targeted spatial transcriptomics. *bioRxiv,* 2022.08.16.504115 (Aug. 2022).

191. Zou, H., Hastie, T. & Tibshirani, R. Sparse principal component analysis. *Journal of computational and graphical statistics: a joint publication of American Statistical Association, Institute of Mathematical Statistics, Interface Foundation of North America* **15,** 265–286 (2006).

192. Fischer, S. & Gillis, J. How many markers are needed to robustly determine a cell's type? *iScience* **24,** 103292 (Nov. 2021).

193. Ohshima, K. *et al.* Integrated analysis of gene expression and copy number identified potential cancer driver genes with amplification-dependent overexpression in 1,454 solid tumors. *Scientific reports* **7,** 641 (Apr. 2017).

194. Song, M. *et al.* Benchmarking copy number aberrations inference tools using single-cell multi-omics datasets. *Briefings in bioinformatics* **26,** bbaf076 (Mar. 2025).

195. Petukhov, V. *et al.* Cell segmentation in imaging-based spatial transcriptomics. *Nature biotechnology* **40,** 345–354 (Mar. 2022).

196. Greenwald, N. F. *et al.* Whole-cell segmentation of tissue images with human-level performance using large-scale data annotation and deep learning. *Nature biotechnology* **40,** 555–565 (Apr. 2022).

197. Heidari, E. *et al.* Segger: Fast and accurate cell segmentation of imaging-based spatial transcriptomics data. *bioRxiv,* 2025.03. 14.643160 (Mar. 2025).

198. Chen, H., Li, D. & Bar-Joseph, Z. SCS: cell segmentation for high-resolution spatial transcriptomics. *Nature methods* (July 2023).

199. Genomics, 1. *Understanding Xenium Outputs: Cell and nucleus segmentation files.* https://www.10xgenomics.com/support/software/xenium-onboard-analysis/1.9/analysis/xoa-output-understanding-outputs. Accessed: 15 Apr. 2025.

200. Marco Salas, S. *et al.* Optimizing Xenium In Situ data utility by quality assessment and best-practice analysis workflows. *Nature methods* **22,** 813–823 (Apr. 2025).

201. Fu, X. *et al.* BIDCell: Biologically-informed self-supervised learning for segmentation of subcellular spatial transcriptomics data. *Nature communications* **15,** 509 (Jan. 2024).

202. Park, J. *et al.* Cell segmentation-free inference of cell types from in situ transcriptomics data. *Nature communications* **12,** 3545 (June 2021).

203. He, Y. *et al.* ClusterMap for multi-scale clustering analysis of spatial gene expression. *Nature communications* **12,** 5909 (Oct. 2021).

204. Prabhakaran, S. Sparcle: assigning transcripts to cells in multiplexed images. *Bioinformatics advances* **2** (June 2022).

205. Ahlmann-Eltze, C. & Huber, W. Comparison of transformations for single-cell RNA-seq data. *Nature methods* **20,** 665–672 (May 2023).

206. Bhuva, D. D. *et al.* Library size confounds biology in spatial transcriptomics data. *bioRxiv,* 2023.03.15.532733 (Mar. 2023).

207. Gayoso, A. *et al.* A Python library for probabilistic analysis of single-cell omics data. *Nature Biotechnology.* ISSN: 1546-1696. https://doi.org/10.1038/s41587-021-01206-w (Feb. 2022).

208. Kharchenko, P. V. The triumphs and limitations of computational methods for scRNA-seq. *Nature methods* **18,** 723–732 (July 2021).

209. Luecken, M. D. *et al.* Benchmarking atlas-level data integration in single-cell genomics. *Nature methods,* 1–10 (Dec. 2021).

210. Kiselev, V. Y., Andrews, T. S. & Hemberg, M. Challenges in unsupervised clustering of single-cell RNA-seq data. *Nature reviews. Genetics* **20,** 273–282 (May 2019).

211. Evan, G. I. & Vousden, K. H. Proliferation, cell cycle and apoptosis in cancer. *Nature* **411,** 342–348 (May 2001).

212. Stein-O'Brien, G. L. *et al.* Enter the matrix: Factorization uncovers knowledge from omics. *Trends in genetics: TIG* **34,** 790–805 (Oct. 2018).

213. Gavish, A. *et al.* Hallmarks of transcriptional intratumour heterogeneity across a thousand tumours. *Nature* (May 2023).

214. Hoogstrate, Y. *et al.* Transcriptome analysis reveals tumor microenvironment changes in glioblastoma. *Cancer cell* **0** (Mar. 2023).

215. Barkley, D. *et al.* Cancer cell states recur across tumor types and form specific interactions with the tumor microenvironment. *Nature genetics* **54,** 1192–1201 (Aug. 2022).

216. Hafshejani, S. F. & Moaberfard, Z. Initialization for nonnegative matrix factorization: A comprehensive review. *arXiv [math.OC]* (Sept. 2021).

217. Sankowski, R. *et al.* Multiomic spatial landscape of innate immune cells at human central nervous system borders. *Nature medicine* (Dec. 2023).

218. Kim, H. *et al.* CellNeighborEX: deciphering neighbor-dependent gene expression from spatial transcriptomics data. *Molecular systems biology* **19,** e11670 (Nov. 2023).

219. Musin, O. R. *Properties of the Delaunay triangulation* in *Proceedings of the thirteenth annual symposium on Computational geometry - SCG '97* (ACM Press, New York, New York, USA, 1997).

220. Behanova, A., Klemm, A. & Wählby, C. Spatial statistics for understanding tissue organization. *Frontiers in physiology* **13,** 832417 (Jan. 2022).

221. Ali, M. *et al.* GraphCompass: spatial metrics for differential analyses of cell organization across conditions. *Bioinformatics (Oxford, England)* **40,** i548–i557 (June 2024).

222. Palla, G. *et al.* Squidpy: a scalable framework for spatial omics analysis. *Nature methods* (Jan. 2022).

223. Schürch, C. M. *et al.* Coordinated cellular neighborhoods orchestrate antitumoral immunity at the colorectal cancer invasive front. *Cell* **182,** 1341–1359.e19 (Sept. 2020).

224. Jackson, K. C. *et al.* Identification of spatial homogeneous regions in tissues with concordex. *bioRxiv,* 2023.06.28.546949 (July 2024).

225. Birk, S. *et al.* Large-scale characterization of cell niches in spatial atlases using bio-inspired graph learning. *bioRxiv,* 2024.02.21.581428 (Feb. 2024).

226. Singhal, V. *et al.* BANKSY unifies cell typing and tissue domain segmentation for scalable spatial omics data analysis. *Nature genetics* (Feb. 2024).

227. Vannan, A. *et al.* Spatial transcriptomics identifies molecular niche dysregulation associated with distal lung remodeling in pulmonary fibrosis. *Nature genetics* **57,** 647–658 (Mar. 2025).

228. Moreno, C. R. *et al.* Harmonized single-cell landscape, intercellular crosstalk and tumor architecture of glioblastoma. *bioRxiv,* 2022.08.27.505439 (Aug. 2022).

229. Hausmann, D. *et al.* Autonomous rhythmic activity in glioma networks drives brain tumour growth. *Nature* (Dec. 2022).

230. Lopez, R. *et al.* Deep generative modeling for single-cell transcriptomics. *Nature methods* **15,** 1053–1058 (Dec. 2018).

231. Xu, C. *et al.* Probabilistic harmonization and annotation of single-cell transcriptomics data with deep generative models. *Molecular systems biology* **17,** e9620 (Jan. 2021).

232. Kotliar, D. *et al.* Identifying gene expression programs of cell-type identity and cellular activity with single-cell RNA-Seq. *eLife* **8** (July 2019).

233. Pedregosa, F. *et al.* Scikit-learn: Machine Learning in Python. *arXiv [cs.LG],* 2825–2830 (Jan. 2012).

234. Vellayappan, B. *et al.* Diagnosis and management of radiation necrosis in patients with brain metastases. *Frontiers in oncology* **8,** 395 (Sept. 2018).

235. Ali, F. S. *et al.* Cerebral radiation necrosis: Incidence, pathogenesis, diagnostic challenges, and future opportunities. *Current oncology reports* **21,** 66 (June 2019).

236. Parvez, K., Parvez, A. & Zadeh, G. The diagnosis and treatment of pseudoprogression, radiation necrosis and brain tumor recurrence. *International journal of molecular sciences* **15,** 11832–11846 (July 2014).

237. Kossmann, M. R. P. *et al.* Histopathologically confirmed radiation-induced damage of the brain - an in-depth analysis of radiation parameters and spatio-temporal occurrence. *Radiation oncology (London, England)* **18,** 198 (Dec. 2023).

238. Soffietti, R. *et al.* Delayed adverse effects after irradiation of gliomas: clinicopathological analysis. *Journal of neuro-oncology* **3,** 187–192 (June 1985).

239. Rong, Y. *et al.* 'Pseudopalisading' necrosis in glioblastoma: a familiar morphologic feature that links vascular pathology, hypoxia, and angiogenesis. *Journal of neuropathology and experimental neurology* **65,** 529–539 (June 2006).

240. Seferbekova, Z. *et al.* Spatial Transcriptomics Characterisation of Radionecrotic Changes in Glioblastoma Patients. *medRxiv,* 2025.09.25.25336313 (Oct. 2025).

241. Wen, P. Y. *et al.* RANO 2.0: Update to the Response Assessment in Neuro-Oncology criteria for high- and low-grade gliomas in adults. *Journal of clinical oncology: official journal of the American Society of Clinical Oncology* **41,** 5187–5199 (Nov. 2023).

242. Wolf, F. A., Angerer, P. & Theis, F. J. SCANPY: large-scale single-cell gene expression data analysis. *Genome biology* **19,** 15 (Feb. 2018).

243. Andreatta, M., Garnica, J. & Carmona, S. J. Identification of malignant cells in single-cell transcriptomics data. *Communications biology* **8,** 1264 (Aug. 2025).

244. Labussière, M. *et al.* TERT promoter mutations in gliomas, genetic associations and clinico-pathological correlations. *British journal of cancer* **111,** 2024–2032 (Nov. 2014).

245. Sankowski, R. *et al.* Mapping microglia states in the human brain through the integration of high-dimensional techniques. *Nature neuroscience* **22,** 2098–2110 (Dec. 2019).

246. Schultheiss, T. E. *et al.* Radiation response of the central nervous system. *International journal of radiation oncology, biology, physics* **31,** 1093–1112 (Mar. 1995).

247. De Jong, G. *et al.* A spatiotemporal cancer cell trajectory underlies glioblastoma heterogeneity. *bioRxiv,* 2025.05.13.653495 (May 2025).

248. Weller, M. *et al.* European Association for Neuro-Oncology (EANO) guideline on the diagnosis and treatment of adult astrocytic and oligodendroglial gliomas. *The lancet oncology* **18,** e315–e329 (June 2017).

249. Ivkovic, S., Canoll, P. & Goldman, J. E. Constitutive EGFR signaling in oligodendrocyte progenitors leads to diffuse hyperplasia in postnatal white matter. *The Journal of neuroscience: the official journal of the Society for Neuroscience* **28,** 914–922 (Jan. 2008).

250. Hubenak, J. R. *et al.* Mechanisms of injury to normal tissue after radiotherapy: a review: A review. *Plastic and reconstructive surgery* **133,** 49e–56e (Jan. 2014).

251. Huang, L.-H. *et al.* The effects of storage temperature and duration of blood samples on DNA and RNA qualities. *PloS one* **12,** e0184692 (Sept. 2017).

252. Kim, W.-K. *et al.* CD163 identifies perivascular macrophages in normal and viral encephalitic brains and potential precursors to perivascular macrophages in blood. *The American journal of pathology* **168,** 822–834 (Mar. 2006).

253. Pey, P. *et al.* Phenotypic profile of alternative activation marker CD163 is different in Alzheimer's and Parkinson's disease. *Acta neuropathologica communications* **2,** 21 (Feb. 2014).

254. Zhang, Z. *et al.* Parenchymal accumulation of CD163+ macrophages/microglia in multiple sclerosis brains. *Journal of neuroimmunology* **237,** 73–79 (Aug. 2011).

255. Becher, B., Spath, S. & Goverman, J. Cytokine networks in neuroinflammation. *Nature reviews. Immunology* **17,** 49–59 (Jan. 2017).

256. Amlerova, Z. *et al.* Reactive gliosis in traumatic brain injury: a comprehensive review. *Frontiers in cellular neuroscience* **18,** 1335849 (Feb. 2024).

257. Watson, S. S. *et al.* Fibrotic response to anti-CSF-1R therapy potentiates glioblastoma recurrence. *Cancer cell* **42,** 1507–1527.e11 (Sept. 2024).

258. Weller, M. *et al.* EANO guideline for the diagnosis and treatment of anaplastic gliomas and glioblastoma. *The lancet oncology* **15,** e395–403 (Aug. 2014).

259. Roy-O'Reilly, M. *et al.* Soluble CD163 in intracerebral hemorrhage: biomarker for perihematomal edema. *Annals of clinical and translational neurology* **4,** 793–800 (Nov. 2017).

260. Yabo, Y. A., Niclou, S. P. & Golebiewska, A. Cancer cell heterogeneity and plasticity: A paradigm shift in glioblastoma. *Neuro-oncology* (Dec. 2021).

261. Xie, Y. *et al.* Key molecular alterations in endothelial cells in human glioblastoma uncovered through single-cell RNA sequencing. *JCI insight* **6** (Aug. 2021).

262. Wälchli, T. *et al.* Single-cell atlas of the human brain vasculature across development, adulthood and disease. *Nature* **632,** 603–613 (Aug. 2024).

263. Nakamura, K. & Ago, T. Pericyte-mediated molecular mechanisms underlying tissue repair and functional recovery after ischemic stroke. *Journal of atherosclerosis and thrombosis* **30,** 1085–1094 (Sept. 2023).

264. Andersen, B. M. *et al.* Glial and myeloid heterogeneity in the brain tumour microenvironment. *Nature reviews. Cancer* **21,** 786–802 (Dec. 2021).

265. Hack, K. *et al.* P01.15.A SINGLE NUCLEUS TRANSCRIPTOMICS OF SPINAL EPENDYMOMA TYPES AND SUBTYPES RECOGNIZES INTRATUMORAL HETEROGENEITY. *Neuro-oncology,* noaf210 (Oct. 2025).

266. Aran, D., Sirota, M. & Butte, A. J. Systematic pan-cancer analysis of tumour purity. *Nature communications* **6,** 8971 (Dec. 2015).

267. Andersson, A. *et al.* Spatial deconvolution of HER2-positive breast cancer delineates tumor-associated cell type interactions. *Nature communications* **12,** 6012 (Oct. 2021).

268. Wu, S. Z. *et al.* A single-cell and spatially resolved atlas of human breast cancers. *Nature genetics* **53,** 1334–1347 (Sept. 2021).

269. Winkler, F. *et al.* Cancer neuroscience: State of the field, emerging directions. *Cell* **186,** 1689–1707 (Apr. 2023).

270. Velten, B. *et al.* Identifying temporal and spatial patterns of variation from multimodal data using MEFISTO. *Nature methods* **19,** 179–186 (Feb. 2022).

271. Cancer Genome Atlas Research Network *et al.* The Cancer Genome Atlas Pan-Cancer analysis project. *Nature genetics* **45,** 1113–1120 (Oct. 2013).

APPENDIX

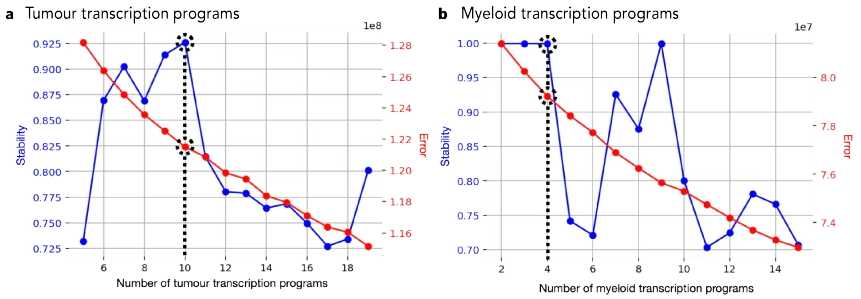

Figure A.1: *Identification of transcription programs.* Error and stability depending on the number of factors for tumour *(a)* and myeloid *(b)* transcription programs. Adapted with permission from [240].

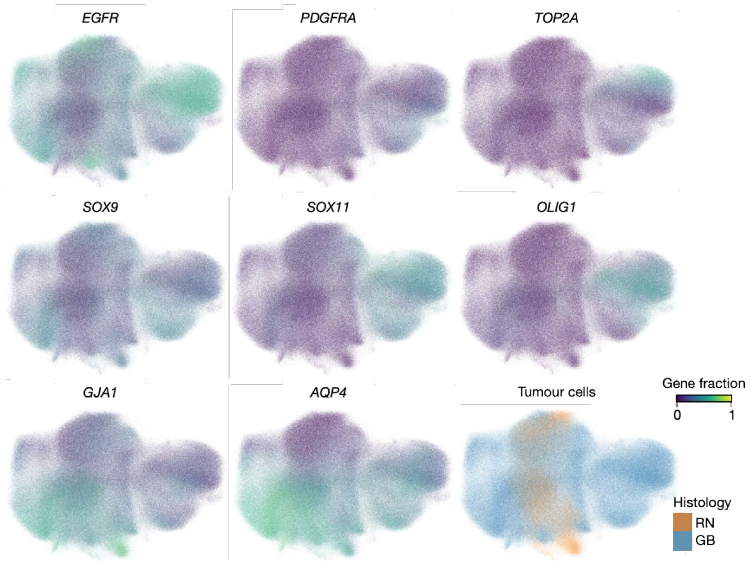

Figure A.2: *Tumour cell cluster expression.* Expression of astrocytic and tumour markers in the Tumour cluster and the cell origin. Adapted with permission from [240].

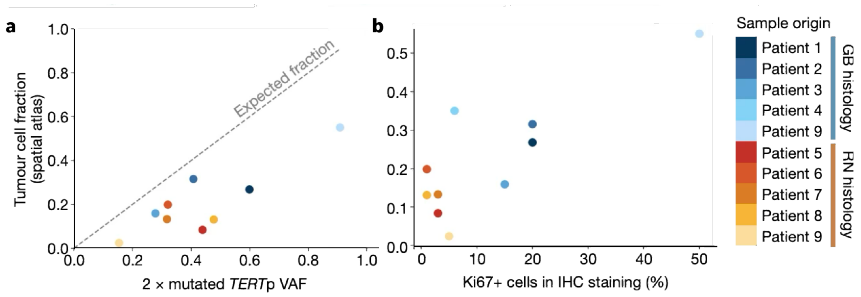

Figure A.3: *Assessment of Tumour cell fractions.* Observed Tumour fractions based on the spatial single cell atlas and the expected fractions based on mutated cell signal *(a)* and IHC staining *(b)*. Adapted with permission from [240].

Patient	Diagnosis	Chr7	*EGFR*	Chr7/*EGFR*
Patient 9	RN	0.051	0.271	5.314
Patient 9	GB	0.277	1.501	5.419
Patient 5	RN	0.152	1.544	10.158
Patient 2	GB	0.4685	1.533	3.272

Table A.1: A validation of *EGFR* amplification. Re-printed with permission from [240].

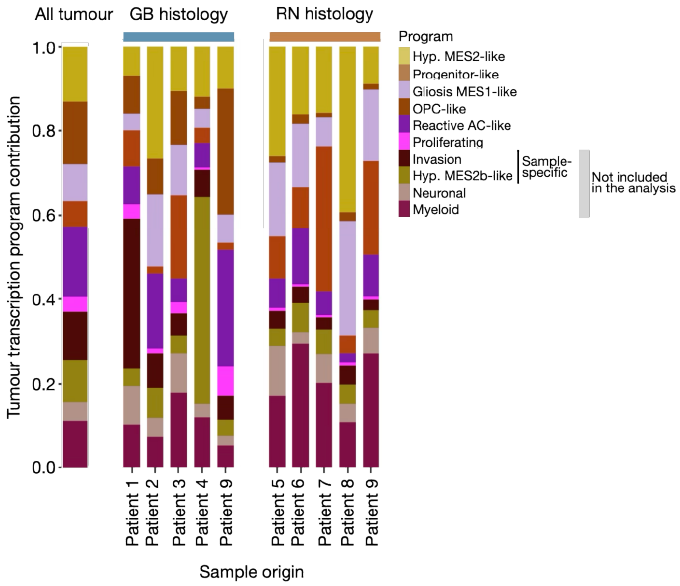

Figure A.4: *Tumour transcription program distribution.* Two sample-specific and two minor programs were excluded from the analysis. Adapted with permission from [240].

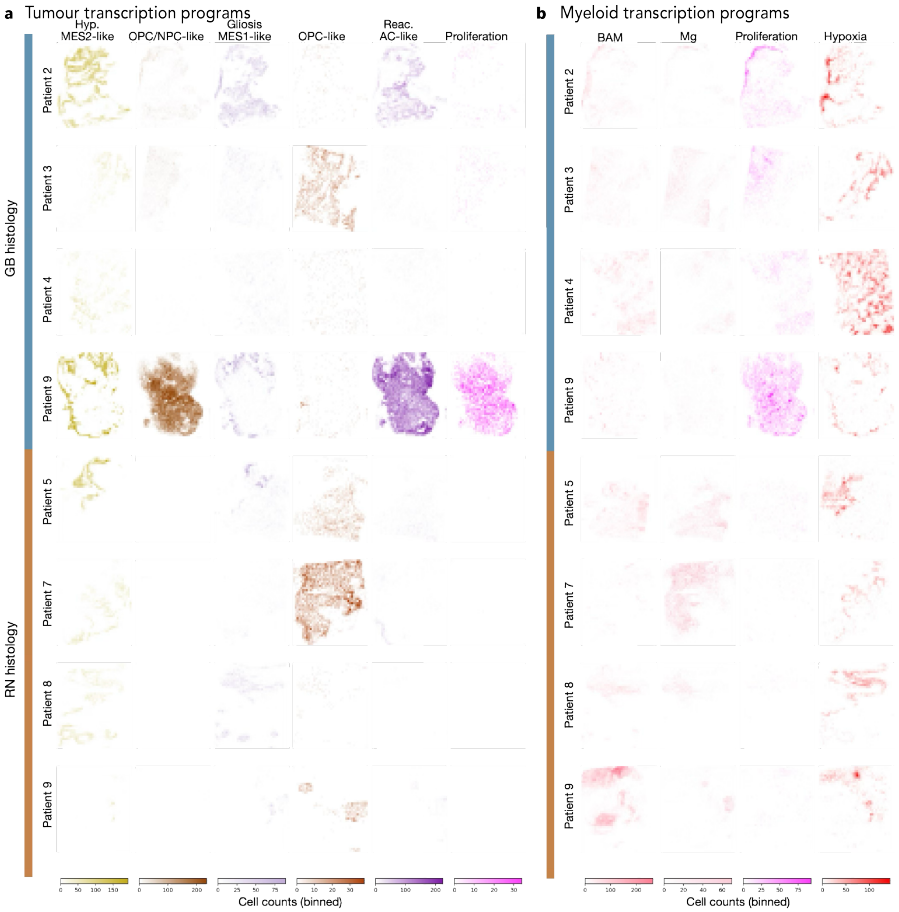

Figure A.5: *The spatial distribution of transcription programs.* Tumour *(a)* and Myeloid *(b)* programs distribution for the rest of the samples. Adapted with permission from [240].

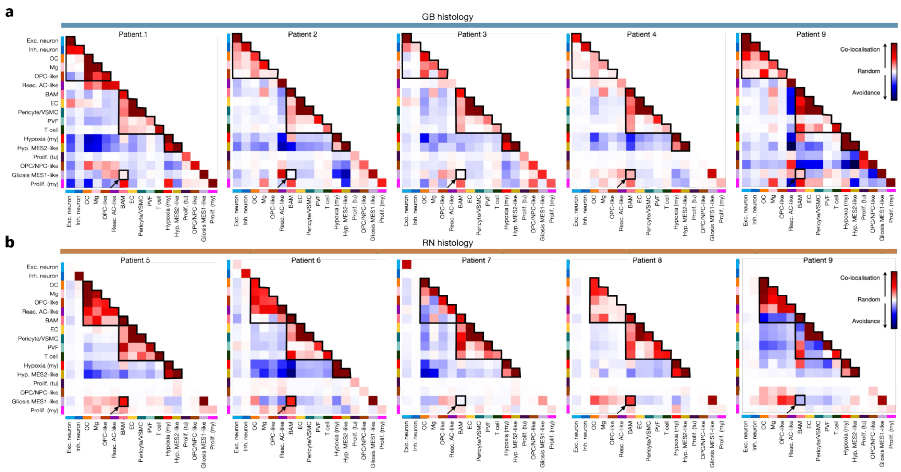

Figure A.6: *Spatial communities.* Cell type and tumour state colocalisation in samples with GB *(a)* and RN *(b)* histologies. Adapted with permission from [240].

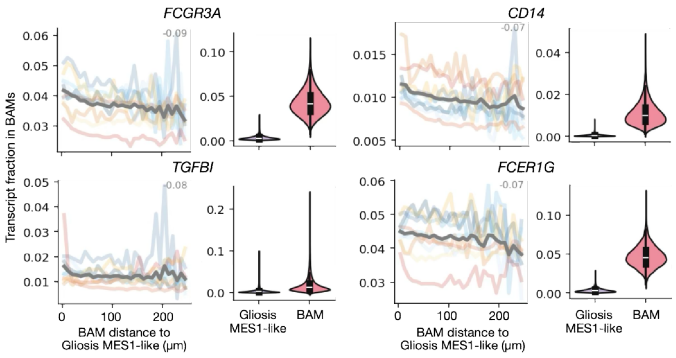

Figure A.7: *Genes enriched in BAMss neighbouring Gliosis MES1-like cells.* Numbers indicate the Pearson correlation coefficient between distance and gene expression in BAMs. Adapted with permission from [240].

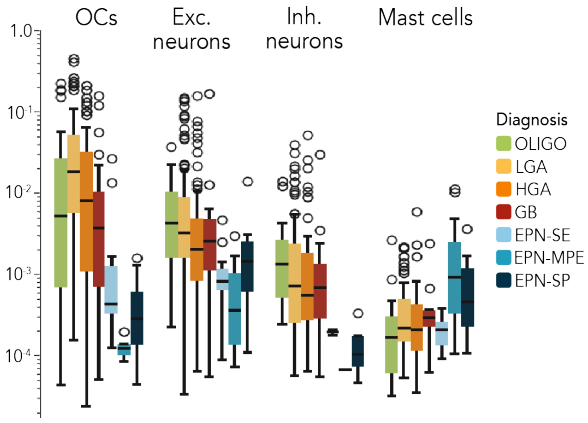

Figure A.8: *Minor cell type distribution across gliomas..*

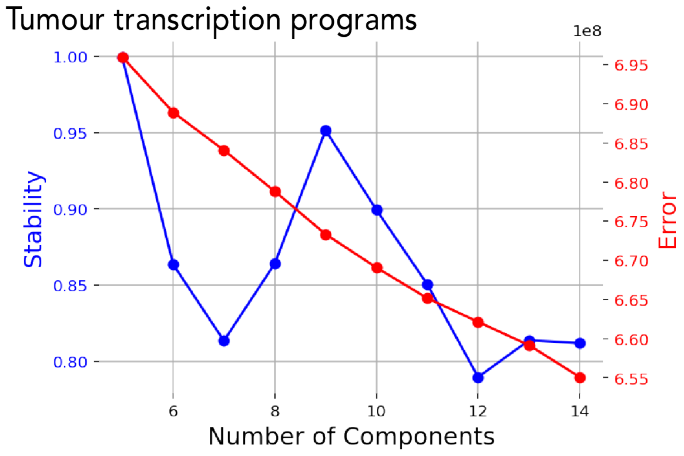

Figure A.9: *Identification of tumour transcription programs.* Error and stability depending on the number of factors for tumour transcription programs.

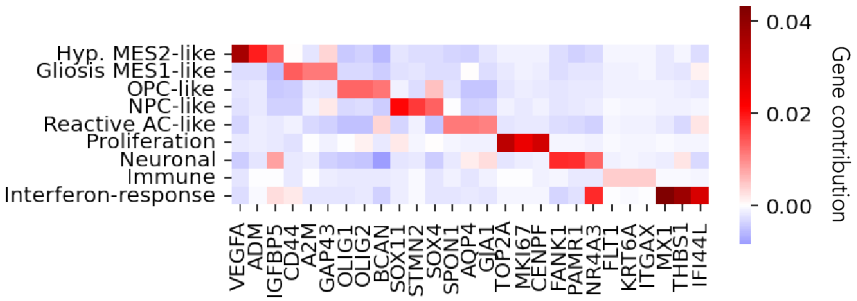

Figure A.10: *Tumour transcription program markers.*

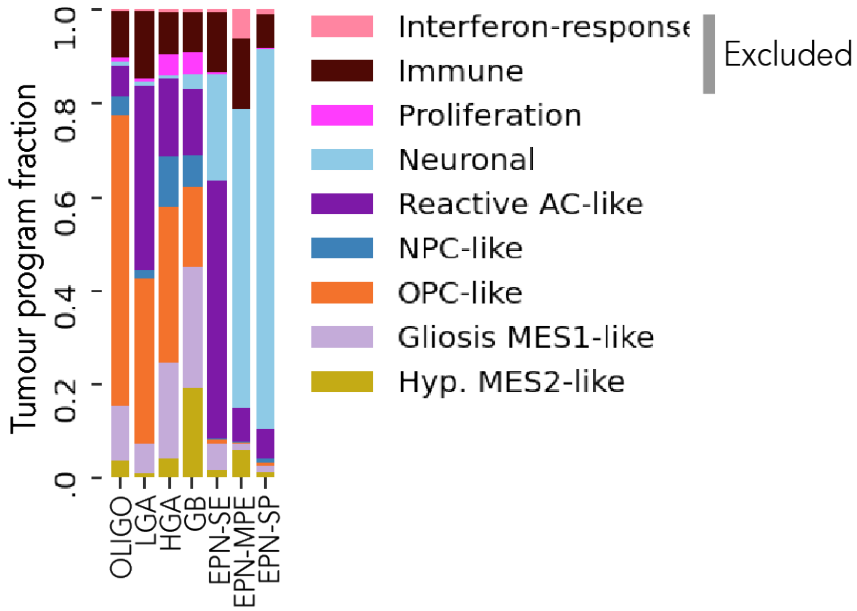

Figure A.11: *Tumour transcription program distribution.* Two minor programs were excluded from the analysis.

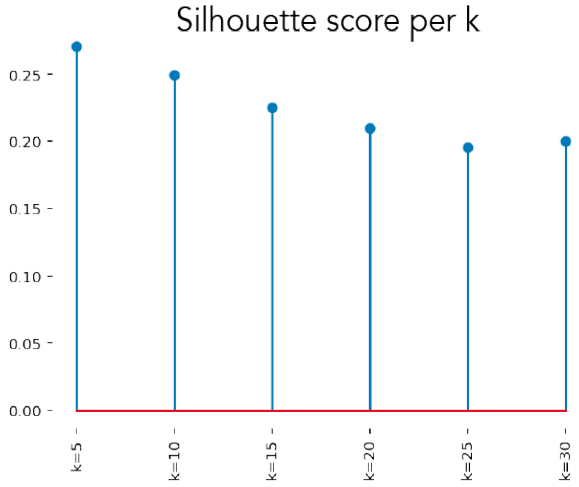

Figure A.12: *Choosing the number of neighbourhoods.* Five clusters corresponded to the highest score.