Single Cell Studies of Protein Expression Reveal Glioma Cell Phenotypes and Prognostic Tumor Cell Populations

By

Justine Sinnaeve

Dissertation

Submitted to the Faculty of the

Graduate School of Vanderbilt University

in partial fulfillment of the requirements

for the degree of

DOCTOR OF PHILOSOPHY

in

Cancer Biology

June, 30, 2020

Nashville, Tennessee

Approved:

Vito Quaranta

Jin Chen

Ian Macara

Vivian Gama

Jonathan Irish

Rebecca Ihrie

To the incredible network of people I am lucky enough to call my friends and family.

And to the brain tumor patients, whose generous donations in a time of unimaginable strife made this work possible.

ACKNOWLEDGEMENTS

First, thank you to my mentor Dr. Rebecca Ihrie. Thank you for your unwavering belief in my promise as a scientist, even when I often lacked a belief in myself. Thank you for encouraging me to develop my own ideas and for providing the space and resources for me to act on them. I appreciate your support of my professional development on top of my scientific development and am so grateful for the opportunities you allowed me to pursue outside the lab. Finally, thank you for the multitude of opportunities to travel during graduate school. These trips increased my scientific literacy, my network of scientific peers, and my understanding of the place of my work in the context of global scientific efforts. They were also incredibly fun.

Thank you to my committee, Dr. Vito Quaranta, Dr. Jin Chen, Dr. Ian Macara, Dr. Vivian Gama, and Dr. Jonathan Irish, for your support over these five years. Your thoughtful questions and discussion were crucial to my development as a scientist and to improving the work in this dissertation immensely. I'd especially like to thank Vivian for being an understanding mentor over the years with whom I could always share both concerns and laughs. I would also like to thank Jin and Vito for their willingness to help me problem solve when I felt overwhelmed.

Thank you to Carolyn Berry, whose open door I walked through countless times over six years. Your perpetual willingness to listen to my problems, offer advice or chocolate, and make me laugh have made all the difference. Thank you to the UCC, and specifically Dr. Melissa Porter, for providing me with tools to address challenges and change how I approach stress and anxiety. Not only did you transform my graduate school experience, but you have arguably made me a lot better at handling the inevitable challenges that I will face in the rest of my life. Thank you both for your support and patience.

To Rebecca and Jonathan, thank you for assembling a collaborative lab environment that not only expanded my scientific knowledge and competence but also introduced me to people

who have forever changed my life. Thank you Jonathan for your early and continued support of my development as a cytometrist. Without your support, I know that I would not have had the same exposure within the cytometry community and would have missed out on many of the incredible opportunities I had to present my work and travel around the world. Thank you also for forcing me to think critically about every experiment, it has made me an infinitely more thoughtful person.

Thank you to the Ihrie lab for being a safe, fun, and intellectually stimulating place to work for five years. Specifically, thank you Asa, for being a steady force of knowledge and for your patience as you taught me new skills. Thank you also for some of the funniest one-liners I've ever heard. Thank you, Laura, for being a model of a driven, ambitious woman and for sharing my passion for Schitts Creek and Buzzfeed listicles. Thank you Ethan for never failing to make me laugh. Sharing a bay with you for the last year of graduate school kept me (in)sane and I have truly enjoyed talking to you about any number of absurd topics while we dance or walk to get coffee. You are very very smart. Finally, thank you Gabrielle. Thank you for being my very first friend in graduate school, for always leading by impeccable example, and for the passion that you bring to every aspect of life, especially your friendships.

Thank you to the Irish lab for being my extended family at Vanderbilt. I am forever grateful to the original ladies in lab, who made me a better scientist and a better person while bringing so much joy to my graduate school experience. To Sai, for being my unfailing mentor when I was very inexperienced and while you were going through graduate school yourself. I have so much respect for your refusal to compromise your high standards, your dedication, and your integrity. I had so much fun and learned enough to earn an advanced degree under your tutelage and cannot thank you enough for your mentorship and your friendship. To Caroline, for your unwavering support, no questions asked, for years. I have so enjoyed being friends with you during these very pivotal years of your life, and I appreciate your enthusiasm to see me even when I was being a sour, grumpy girl. You made every day brighter and I will miss eating lunch with you. I must also

thank you for “continuing to run the cores” and your dedication to advancing science even in the face of great challenges. To Allie, for being an incredible example of a strong woman in science who acts with confidence, determination, and grace. I have loved learning about how you view the world and am so grateful to have your listening ear when I have a problem, no matter how big or small. I love you all and will forever be honored to call you my friends. When Sai, Allie, and Gabrielle left Vanderbilt, I felt certain I could never be so lucky in my coworkers again. I was proven wrong with the arrival of Sierra and Maddy. The two of you have been a shining light of support and joy over the past two years. Sierra, I could never have imagined to accomplish half of the work in this dissertation without the supercomputer you keep on top of your shoulders. I cannot thank you enough for the hard and thoughtful work you put into the various facets of this project that have not only made it a better body of work but also made it more useful to the scientific community at large. Sharing the trip of a lifetime to Israel together was a dream come true and I have so much respect and love for you dear. Maddy, you are wise, witty, and wonderful. I am constantly impressed by your ability to handle tough situations with humor and confidence. I am not sure I would have completed this degree if I hadn’t felt the (sometimes literal) warm embrace of your support. I appreciate you both so much and I know you have, and will continue to accomplish, amazing things.

Outside of the laboratory, I am blessed by an impossibly fantastic network of friends. To the six women of the BDBF who have been by my side for decades now. You made me into the person I am today (for better or worse) and your continued friendship is invaluable as our lives take unexpected, sometimes challenging, sometimes delightful turns. To Becca, for being the most steadfast, fun, and genuine friend a girl could ask for. To my friends from Duke who always bring honor and have given me a yearly trip to look forward to during graduate school. Your thoughtfulness, wicked intelligence, and willingness to embrace vulnerability are so appreciated. To the posse of IGP boys that put up with me for five years, thank you for so many laughs and the honest and hilarious conversations we shared. To Abbe, Amy, and Lindsey, my fellow Blue

Devils, who have been great role models and even better friends. To Katie Jo Scholl, my roommate for three years of this journey, who brought my sweet baby Duke into our lives. Your pure joy for life and refusal to make things more complicated than they need to be are always a breath of fresh air. It's impossible to have anything but fun when I'm with you. To Andrea Perreault, my next amazing roommate. Thank you for being beyond kind, understanding, and encouraging when I was anything but those things. I've loved working down the hall from you, so that I can hug you at least 15 times a day before going back to our shared home. Your ability to find the light in anything or anyone is astounding and sharing the highs and lows of life with you has been such a treat. And finally, to my sweet angel Rachel Fischer. From day one, when you offered to share queso with me, I knew you were my person. You are so authentic, hilarious, and strong. "Having a day" with you is what made me love Nashville and what got me through a lot of hard times. Thank you, Rachel and Andy, for being the best friends, and probably the only ones, who would stick out these past 5-6 years with me.

Finally, thank you to my family. To my sister Kelly, whose kindness and considerate nature have been an inspiration to me for the past 25 years. Thank you for being my baby girl and being so supportive of me no matter where we both are. To my parents, in short, thank you for everything. Thank you for allowing me to find my passion for science and to follow it here, free from pressure but always with your support. Thank you for building a home that was always filled with love, integrity, and encouragement. Thank you for your commitment to my education, from your help at the kitchen table to Duke to the work in this dissertation. And thank you for being an example of two people who are wildly smart and hardworking and fun as hell. I love you guys.

I am well aware that I have so much to be thankful for and I'm sure there is plenty I have missed here. I am so appreciative of the people in my life, they are what make every day worthwhile. Thank you.

PREFACE

In this dissertation, I describe the generation of a brain tumor mass cytometry dataset and an algorithm for unsupervised identification of cell phenotypes and subsets that are correlated with clinical variables of interest. Furthermore, I describe the analysis of this dataset in the context of tumor location within the brain and the development of additional technical tools for investigating brain tumor biology. Not only were novel tumor cell types uncovered and investigated as described here, but this work also provides both technical and computational tools for studying other malignancies and human diseases. The work described in Chapter II, as well as Appendices A and B have been submitted to peer reviewed scientific journals and has been inserted into the body of the dissertation in an extended version of the form in which it was submitted. None of this work would have been possible without the close collaboration of Dr. Jonathan Irish's lab at Vanderbilt University, especially former graduate student Nalin Leelatian and research assistant Sierra Barone. Chapter I introduces the background of the brain tumors that are explored in the rest of the dissertation, called gliomas. I will provide a broad overview of the literature, as well as review the technological gaps within the field that inspired the research questions and technological innovation addressed in this dissertation. The molecular features of gliomas and methods for studying them are introduced, followed by a discussion of flow cytometry and single cell data analysis tools. Finally, the research objectives of this dissertation are stated. The mass cytometry dataset and algorithm are explained in Chapter II and Appendix B, while the methods used to generate this dataset are outlined in Appendix A. Chapter III introduces a new tool for studying low grade brain tumors using flow cytometry and Chapter IV investigates the relationship between high dimensional protein phenotypes and brain tumor location. In Chapter V, major conclusions from this body of work are discussed, and future directions are explored.

TABLE OF CONTENTS

Page

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

2HG	2-hydroxygluterate
4EBP1	Eukaryotic translation initiation factor 4E (eIF4E)-binding protein 1
α-KG	Alpha-ketoglutarate
ACCENSE	Automatic Classification of Cellular Expression by Nonlinear Stochastic Embedding
ATRX	Alpha thalassemia/mental retardation syndrome X-linked
BCP-ALL	B-Cell Precursor Acute Lymphoblastic Leukemia
BDNF	Brain-derived neurotrophic factor
BMX	Bone marrow tyrosine kinase gene in chromosome X
BSA	Bovine serum albumin
cCASP3	Cleaved caspase 3
CDKN2A	Cyclin-dependent kinase Inhibitor 2A
CI	Confidence interval
Citrus	Cluster identification, characterization, and regression
CNS	Central nervous system
CSF	Cerebrospinal fluid
CXCL12	Chemokine ligand 12
CyTOF®	Cytometry by Time of Flight
DAB	3,3′-Diaminobenzidine
DDPR	Developmentally dependent predictor of relapse
DNA	Deoxyribonucleic acid
EGFR	Epidermal growth factor receptor

EGFRvIII	Epidermal growth factor receptor variant III
EMD	Earth mover’s distance
EOR	Extent of resection
FBS	Fetal bovine serum
FCS	Flow cytometry standard
FFC	Fluorescence flow cytometry
FFPE	Formalin-fixed, paraffin-embedded
FFPE-DISSECT	Formalin-fixed, paraffin-embedded - Disaggregation for intracellular signaling in single epithelial cells from tissue
GBM	Glioblastoma
GFAP	Glial fibrillary acidic protein
GNP	Glioma negative prognostic
GPP	Glioma positive prognostic
GSC	Glioma stem cell
HIER	Heat-induced epitope retrieval
HIF1α	Hypoxia Inducible Factor 1 Subunit Alpha
HR	Hazard ratio
IDH	Isocitrate dehydrogenase
IGF	Insulin growth factor
IHC	Immunohistochemistry
IQR	Interquartile range
ITGA6	Integrin alpha 6
JAK	Janus kinase
L1CAM	L1 cell adhesion molecule protein
MC	Mass cytometry
MDM2	Mouse double minute 2 homolog

MEM	Marker enrichment modeling
MeOH	Methanol
MGMT	O-6-Methylguanine-DNA Methyltransferase
MRI	Magnetic resonance imaging
mTOR	Mechanistic target of rapamycin
NADP+/H	Nicotinamide adenine dinucleotide phosphate
NF1	Neurofibromatosis type 1
NFκB	Nuclear factor kappa-light-chain-enhancer of activated B cells
NSC	Neural stem cell
OLIG2	Oligodendrocyte transcription factor
OS	Overall survival
PBS	Phosphate buffered saline
PCA	Principal component analysis
PCR	Polymerase chain reaction
PDGFRα	Platelet derived growth factor receptor alpha
PDX	Patient-derived xenograft
PFA	Paraformaldehyde
PEDF	Pigment epithelium-derived factor
PFS	Progression-free survival
PI3K	Phosphoinositide 3-kinase
PIGF-2	Placental growth factor 2
PTEN	Phosphatase and tensin homolog
RAPID	Risk Assessment Population IDentification
Rb	Retinoblastoma
RBC	Red blood cell
RNA	Ribonucleic acid

RNA-seq	Ribonucleic acid sequencing
ROI	Region of interest
RTK	Receptor tyrosine kinase
Sap	Saponin
SCAFFOLD	Single-Cell Analysis by Fixed Force- and Landmark-Directed
SCLC	Small cell lung cancer
SEZ	Subependymal zone
SGZ	Subgranular zone
SOX2	SRY (sex determining region Y)-box 2
SPADE	Spanning-tree Progression Analysis of Density-normalized Events
STAT	Signal transducer and activator of transcription
SVZ	Subventricular zone
TCA	Tricarboxylic acid cycle/the citric acid cycle
TCGA	The Cancer Genome Atlas
TMA	Tissue microarray
TMZ	Temozolomide
TPSR	Translational Pathology Shared Resource
t-SNE	t-Distributed Stochastic Neighbor Embedding
UMAP	Uniform Manifold Approximation and Projection
VEGF	Vascular endothelial growth factor
V-SVZ	Ventricular-subventricular zone
VUMC	Vanderbilt University Medical Center
XRT	Radiation

CHAPTER I

BACKGROUND AND RESEARCH DIRECTIONS

Introduction

Glioblastoma is the most common primary malignancy of the adult brain, and such tumors are almost universally fatal, despite considerable efforts over the last few decades to improve patient outcomes. Even with the aggressive standard treatment program of surgery followed by radiation and chemotherapy, patients will ultimately experience tumor recurrence and succumb to disease. Due to the seeming inevitability of tumor recurrence, researchers and clinicians have sought to extend survival through targeted or personalized therapies based on tumor-specific alterations in gene expression and DNA content (reviewed in [1]). Although many promising drug and biologic candidates have entered clinical trials, there have been no improvements to the standard of care therapy since its definition 2005 [2]. As single cell technologies gain wider acceptance and appreciation for individual differences between glioma cells, especially intracellular signaling capacity, grows, the prospect of identifying targetable features of glioma cells within patient tumors becomes more attainable. This chapter will describe gliomas, especially the most aggressive form, glioblastoma (GBM), progress made in understanding the genetic and gene expression changes within tumors and individual tumor cells, as well as the application of single cell protein measurements and high dimensional data analysis to studying human disease. The goal of this dissertation is to harness the potential of single cell protein

measurements to improve overall understanding of GBM and glioma biology broadly, and to identify prognostic cells and potential therapeutic targets (Figure 1.1).

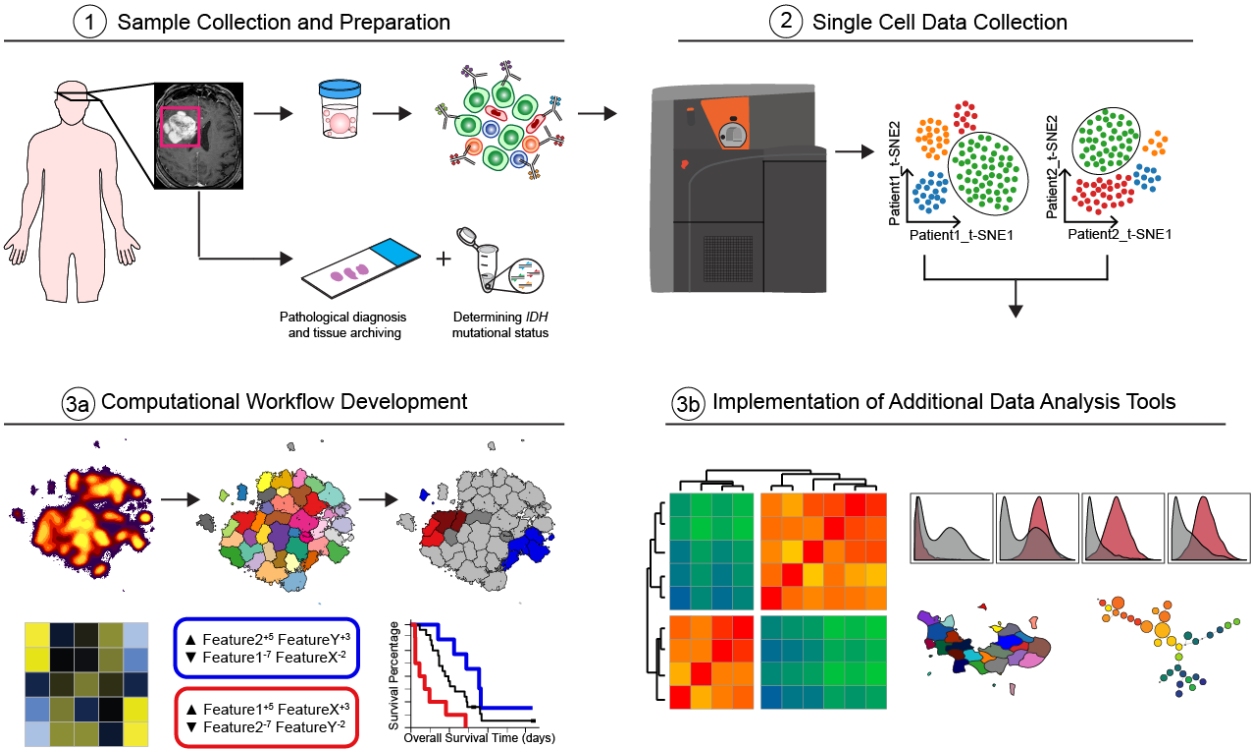


Figure 1.1: **Investigation of glioma biology using single cell techniques.** The goal of this work is to investigate cellular phenotypes in gliomas, as they relate to patient outcomes and tumor location, and to develop tools that could be broadly applied to other single cell studies of human disease. 1) Glioma patient samples were collected from the operating room and immediately dissociated into viable, single cells. Part of the sample was also sent to the pathology lab for standard diagnostic testing and banking of FFPE tissue. Over 100 patient samples were collected in this manner. 2) Single cells were stained for glioblastoma cell identification and acquired by mass cytometry 3a) An unsupervised data analysis pipeline was developed for this high dimensional, single cell, glioblastoma data. 3b) Additional tools were utilized as needed to further interrogate glioma biology, i.e., identify novel cell populations, highlight differences between groups of tumors, or inform hypotheses for future glioma studies. *Adapted from Leelatian and Sinnaeve et al., In press at eLife.*

Glioma pathology

Glioblastoma is one subtype of a category of cancers called gliomas, which originate within the central nervous system (CNS) and, in contrast to many other solid tumors, are not known to metastasize outside the CNS [3]. The majority of gliomas occur in adults, with an average age at diagnosis of 64, and Caucasian men are slightly more likely to develop these tumors than other racial or gender groups [3]. Historically, gliomas were categorized based on histology, their very name is derived from their resemblance to non-malignant glial cells [3]. Based on these histological features as well as clinical characteristics these tumors are graded from grade I benign masses to grade IV highly aggressive, malignant, glioblastomas (GBMs) [3, 4]. Grade I gliomas are generally well-circumscribed, clinically benign, and potentially curable through surgery alone [5]. Grades II and III gliomas comprise a diverse array of neoplasms that are delineated by molecular features discussed below. They are infiltrative but frequently can be managed for years after surgical resection with combinations of clinical monitoring, radiotherapy, or chemotherapy [6]. However, eventually these tumors can recur and progress to higher grade malignancies, including GBM [5]. Glioblastomas are defined by pathological observations of hypervascularization, high mitotic index, infiltration of normal brain, and pseudo-palisading necrosis; these features remain an important part of diagnosis to this day [4]. GBM is rapidly fatal, with an average overall survival of 14.6 months [3] and a time to progression (also called time to recurrence) of 6.9 months from first resection [2]. Upon identification of a suspected glioblastoma via magnetic resonance imaging (MRI), the patient undergoes maximal surgical resection followed by standard of care concurrent radiation and chemotherapy with the DNA alkylating agent temozolomide [2]. Following six weeks of chemo- and radiotherapy, the patient is put on adjuvant temozolomide for six cycles, at one cycle per month, or until disease progression. Virtually all grade IV tumors will recur and patients will succumb to this progression of their disease. Death can occur due to brain stem infiltration, brain herniation, infection, seizure, or hemorrhage within the tumor [7, 8]. This standard of care therapy has not changed in the 14 years

since temozolomide was added to then standard surgery and radiation, with the addition of chemotherapy improving median overall survival by 2.5 months [2]. However, temozolomide is a general DNA alkylating agent that is only effective on actively dividing cells and may not be effective in those GBM tumors with functional DNA repair machinery, discussed below [9, 10]. Generally, tumor recurrence is believed to be due to the survival of infiltrative tumor cells that remain outside the margins of resection and escape subsequent radiation and chemotherapy [11, 12]. Second resections are relatively rare, with only 25% of patients undergoing surgery to remove recurrent tumors. This makes recurrent tumor samples from patients difficult to obtain, limiting the field's understanding of how standard of care therapy is changing tumor biology (in contrast to untreated tumors which are relatively abundant for research purposes) [13].

Molecular features of grade II and III gliomas

In recent decades, the value of molecular features in defining glioma subtypes has emerged [4]. Most notably, mutations in the *isocitrate dehydrogenase* (*IDH1* or *IDH2*) genes have helped to differentiate secondary GBMs, which arise from lower grade, *IDH*-mutant, lesions and progress to grade IV tumors, from primary GBMs, which arise *de novo* as a grade IV malignancy and are *IDH*-wild type [14, 15]. Lower grade gliomas have likewise been categorized by *IDH* mutation status, as well as changes in chromosomal content of chromosomes 1 and 19. Mutations in other tumor suppressors like *p53* and *ATRX* can further delineate lower grade gliomas [4]. There is currently no single mutation or chromosomal aberration that defines glioblastoma. Approximately 70% of low grade gliomas, as well as up to 45% of glioblastomas, exhibit hypermethylation of the promoter region of the DNA repair gene *MGMT* (O^6-methylguanine-DNA methyltransferase), however this DNA modification it is not currently used in classification schemes [10, 16, 17]. MGMT works by removing alkyl groups from guanine, an action that is in direct opposition to the mechanism of action of temozolomide, a DNA alkylating agent. Recent

studies suggest that patients with active, unmethylated, *MGMT* do not benefit from temozolomide treatment and some clinical trials have started stratifying treatment groups based on *MGMT* promoter methylation [16].

Mutations in isocitrate dehydrogenase genes (*IDH1* or *IDH2*) result in better patient outcomes and specific biological alterations in tumor cells (Figure 1.2), and thus are likely to be important in future targeted treatment strategies for these patients [18]. Although there are three isotypes of IDH in humans, only mutations in *IDH1* or *IDH2* have been observed in gliomas, and, less frequently, in other cancers [14, 15, 19]. Up to 80% of grade II and III gliomas carry a mutation in one of the two isoforms, which manifests in an amino acid residue substitution that interferes with IDH binding to its substrate, isocitrate [15, 20, 21]. The most common substitution by far, occurring in 90% of *IDH*-mutant gliomas, is the substitution of a histidine residue for arginine (R132H) in IDH1 [15, 22, 23]. IDH is an enzyme that reversibly converts isocitrate to alpha-ketoglutarate (α-KG) as part of the citric acid cycle in mitochondria and in the cytoplasm to maintain oxidative homeostasis through NADPH production (Figure 1.2) [24]. When mutated, this enzyme loses affinity for its natural substrate, isocitrate, and instead irreversibly converts α-KG to 2-hydroxygluterate (2HG), causing a decrease in α-KG and NADPH and an increase in 2HG (Figure 1.2) [20]. In the cell, this results in wide ranging changes to metabolism, DNA methylation, and responses to hypoxia that are not observed in *IDH*-wild type tumors [25]. Furthermore, *IDH1* and *IDH2* mutant tumors frequently have inactive MGMT because of the hypermethylator phenotype induced by 2HG competition with α-KG-dependent enzymes [16]. Taken together, these molecular alterations have consequences on the impact of chemotherapies that are unique to grade II and III gliomas. Further, they provide insight into potentially useful therapeutics such as targeting mutant IDH protein or exploiting DNA damage repair deficiencies [6]. Understanding the protein expression and signaling consequences of IDH mutations and the resulting epigenetic alterations will be instrumental in developing therapies specifically for these patients.

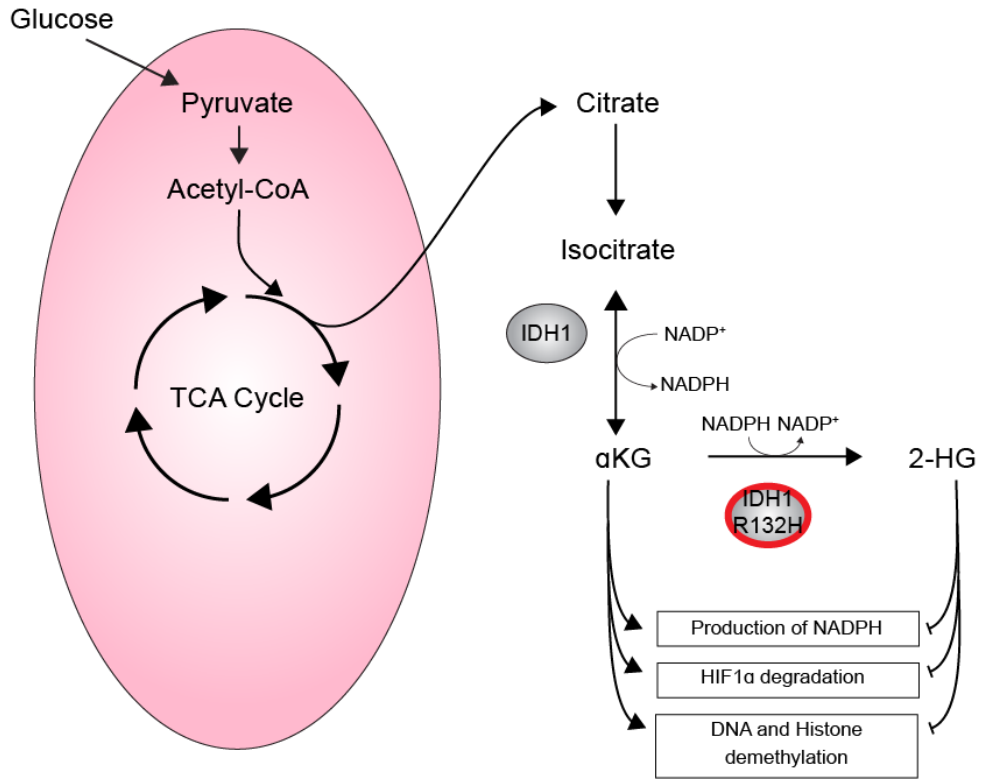


Figure 1.2: **Mutations in *IDH1* alter metabolism in cancer cells resulting in diverse effects**. Citrate, produced during the TCA cycle in the mitochondria (pink oval), is converted to isocitrate, the substrate for IDH1 (gray oval). IDH1 reversibly converts isocitrate to alpha-ketoglutarate (αKG) in the cytoplasm. Mutant IDH1 R132H (red outline) irreversibly converts αKG to 2-HG, widely considered an oncometabolite. The reduction in αKG and increase in 2-HG has a diverse range of effects, included those listed in the boxes (bottom right). *Adapted from [26].*

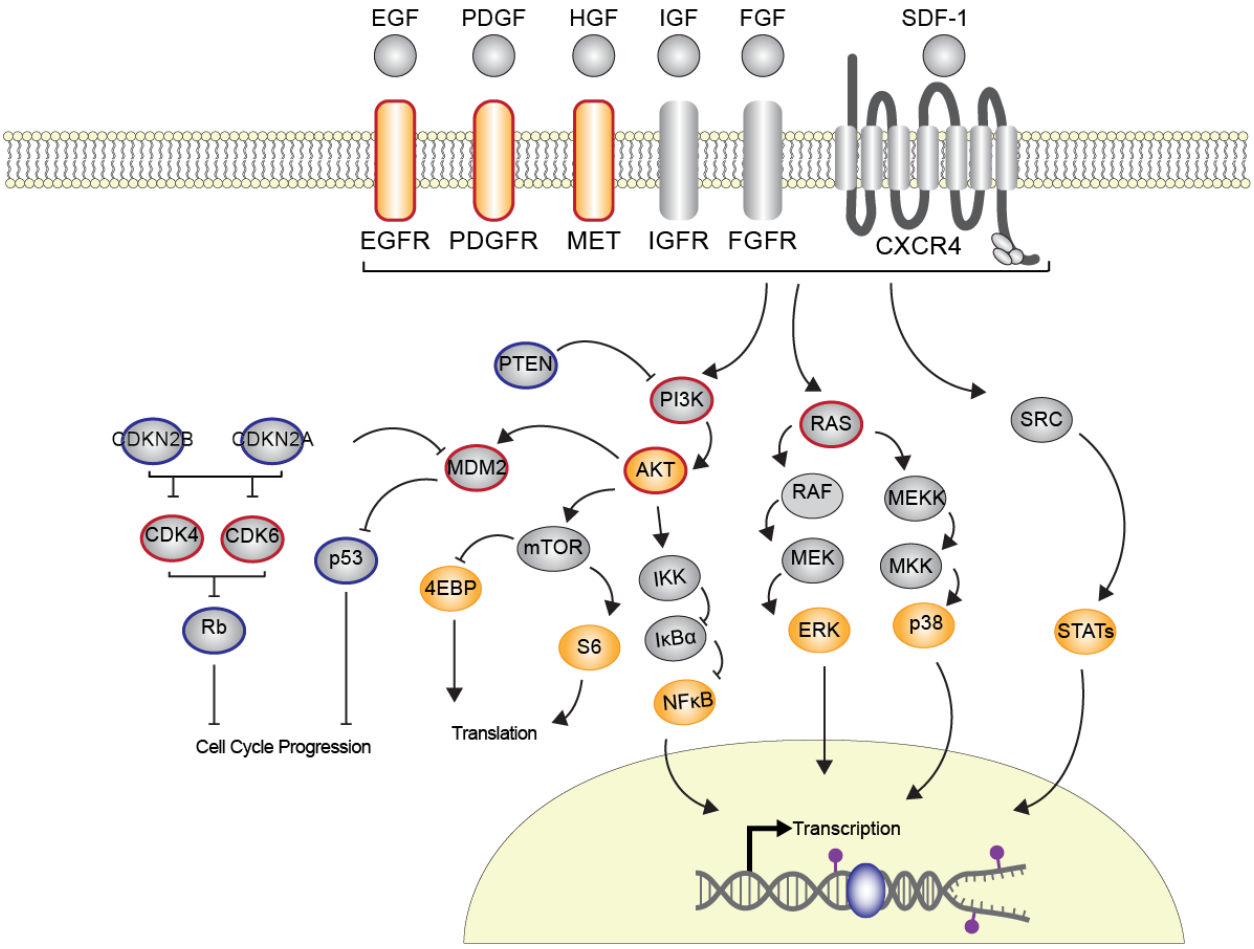


Figure 1.3: **Cell signaling events and molecular alterations in glioblastoma can be measured in single cells.** Extracellular ligands (labeled growth factor circles on top) can bind to transmembrane receptors (rectangular shapes), initiating a signaling cascade driven by intracellular proteins (oval shapes). Each of these signaling pathways may be altered in a tumor environment. Red outline indicates a feature that is considered an oncogene and is commonly altered in glioblastoma. Blue outline indicates a feature that is considered a tumor suppressor and is commonly altered in glioblastoma. Features in yellow were measured in the mass cytometry experiments described in this dissertation.

Molecular features of glioblastoma

With the advent of sequencing technologies, investigation into mutations, amplifications, deletions, and gene expression differences in gliomas, especially GBM, became commonplace. The Cancer Genome Atlas (TCGA) pioneered the deep genetic analysis of GBM beginning in 2005, collecting over 500 GBM samples, an effort that continues to this day [27, 28]. They

performed bulk gene expression analysis using microarrays, copy number analysis, and whole exome sequencing on the samples. Largely from these data, as well as confirmation with independent datasets [29], a molecular picture of GBM began to come into focus. Key alterations were identified, such as amplification of chromosome 7 and deletion of chromosome 10 [28]. Receptor tyrosine kinases (RTKs), p53, and the retinoblastoma (Rb) pathways and downstream signaling were also found to be disrupted in the vast majority of gliomas, with 74% of tumors bearing alterations in all three pathways (Figure 1.3) [27]. This includes amplification of RTK genes such as *EGRF*, *PDGFRα*, and *MET,* or activating mutations in these and other RTK genes, as well as loss of downstream signaling modulators like *PTEN* or *NF1* (Figure 1.3). *EGFR* alterations are especially common, often presenting as extra copies of small extrachromosomal fragments called double minutes, or as a truncated protein called EGFRvIII which, due to a deletion of exons 2-7, lacks much of the extracellular domain and is constitutively active at low levels [30]. Deletion or mutation of *p53* is observed in some tumors, while amplification of the p53 inhibitor *MDM2* is observed in others. Similarly, deletion or mutation of either of the tumor suppressors *CDKN2A* or *Rb1* is observed in over 60% of patients (Figure 1.3) [27]. Collectively, these data provided a wealth of information about common alterations in GBM, including downstream signaling pathways that might be appropriate targets for therapeutics and strategies for designing genetically engineered mouse models.

Between 2005 and 2010, several groups attempted to parse these data to break GBM patients into molecular subclasses that might predict patient outcome or guide future therapies. Three to four subgroups were described by various research groups [28, 29, 31], perhaps the most notable being those proposed by Verhaak *et al.* in 2010. Their work described four molecular subclasses, defined by DNA and RNA alterations, and termed Classical, Proneural, Mesenchymal, and Neural [28]. The subclasses were named partly based on the resemblance of each group to non-malignant cell counterparts. Classical tumors are defined by characteristic chromosome 7 amplification and chromosome 10 loss, *EGFR* amplification or mutation, and wild

type *p53*. Proneural tumors overexpress *PDGFRα*, tend to present in younger patients, include secondary GBMs, have mutations or loss of heterozygosity in *p53*, and were named for expression of progenitor genes like *OLIG2* and *SOX2*. Mesenchymal tumors expressed markers like CD44, Vimentin, MET, and YKL40, were especially hypervascularized, and demonstrated increased immune infiltration and necrosis. Finally, neural tumors resembled normal neural cell expression and have not been re-identified by subsequent analyses, suggesting that these samples may in fact be dominated by non-malignant tissue in the highly infiltrative glioma samples that were resected during surgery. Although features resembling the mesenchymal, proneural, and classical subclasses have been described in other work [29, 32], patients in different groups do not demonstrate different outcomes and no targeted therapies have been successfully identified for one group or another [28]. One reason for this is likely due to the fact that these subclasses were derived from analyses of bulk tumor samples, and further work has highlighted the enormity of variation within and between GBM tumors [12, 33-37]. There is well-documented heterogeneity in cell morphology, gene expression, proliferative capacity, tumorigenic capacity, protein expression, and potentiated signaling in individual cells in GBM tumors. In fact, subsequent single cell RNA-seq experiments demonstrated that individual tumors could contain cells from multiple molecular subclasses, and individual cells could even be labeled as high for more than one subclass according to gene expression [33]. Further single cell work continues to reveal the gene and protein expression heterogeneity that can exist within a single tumor and across patients [32, 37, 38]. Such information will be vital to developing personalized therapies and ensuring the elimination of all cancer cells to prevent recurrences that lead to patient mortality.

Glioblastoma stem cells

GBM cellular heterogeneity is thought to arise in part due to cancer stem cells, cells that are slow cycling, self-renewing, tumorigenic, and that can give rise to all cells in the tumor lineage

[39-42]. Broadly, the cancer stem cell theory posits that a fraction of cancer cells are self-renewing progenitors at the apex of a cancer cell hierarchy, capable of generating virtually all cell types found in a tumor [42]. The concept a cancer stem cell was first proposed and demonstrated in leukemias and is thought to be an important part of tumor biology for many other cancers including GBM [43]. Glioma stem cells, or GSCs, were identified in 2003 and 2004 as sphere-forming cells, derived from patient samples, that express CD133 (also called prominin 1) [39-41]. Functionally, GSCs are able to form tumors in immunocompromised mice and to go on to be serially transplanted from these xenografts. Furthermore, GSCs were shown to be able to generate progeny expressing more differentiated markers of a variety of brain cells and losing expression of stem-like markers including CD133. Over the years, it has become clear that CD133 is not the only cancer stem cell marker [44-46], but a universal protein that can be used to isolate these cells remains elusive. Multiple groups have proposed two to four molecular subclasses of cancer stem cells that can transition between states depending on external pressures and internal signaling modules [47-49], while others have identified up to seven putative GSC populations in a single tumor based on gene expression [37]. The relative percentages of these cells may drive tumor aggressiveness and thus patient outcome, and may ultimately be the drivers of the bulk molecular subclasses described above.

Non-malignant stem cell niche and gliomas

This text is adapted from Sinnaeve, J., et al. (2018). "Space Invaders: Brain Tumor Exploitation of the Stem Cell Niche." Am J Pathol 188(1): 29-38. https://doi.org/10.1016/j.ajpath.2017.08.029

Although there is evidence for the existence of cancer stem cells in GBM tumors, the role of non-malignant neural stem cells and the neural stem cell niche in modulating GBM biology is not well understood. In mammals there are two known neurogenic niches, the sub-granular zone

(SGZ) of the dentate gyrus in the hippocampus and the ventricular-subventricular zone (V-SVZ). The cellular constituents, intercellular interactions, and extracellular components of these niches support stem cell maintenance and differentiation which glioma cells may co-opt for a survival advantage (Figure 1.4 and reviewed in [50], [51, 52]). The V-SVZ (sometimes referred to as the subventricular zone (SVZ) or the sub-ependymal zone (SEZ)) is the larger of these two niches and is located immediately adjacent to the lateral ventricles in the cerebrum.

Interest in the V-SVZ heightened with the emergence of the cancer stem cell theory, and was reinforced by similarities in gene expression between non-neoplastic stem cells and cancer cells, as well as by their shared capacity for proliferation [37, 42]. In healthy adult humans, neurogenesis appears to be a rare event [53, 54], although adult V-SVZ generation of mature neurons can occur in the setting of brain injury, as has been shown in adult rats following ischemic stroke (reviewed by [55]). In the setting of brain cancer, it has been proposed that neural stem cells of the V-SVZ are cells of origin for brain cancers, although more recent tumor models implicate additional progenitor and mature cells in tumor development (reviewed in [56, 57]). The development of neoplasia after genetic ablation of tumor suppressors and exogenous upregulation of growth factors in the rodent V-SVZ have further supported these hypotheses (reviewed in [58]). Recent, single cell studies have identified the expression of many stem-like markers in brain tumors and identified a subset of tumor cells that resemble a fetal neural cell type called outer radial glia [37, 59]. These data suggest that either outer radial glia persist into adulthood and contribute to gliomagenesis, or gliomas co-opt early developmental programs resulting in phenotypes of fetal stem cells in adults.

Recently there has been an increased focus on the role of this niche in high-grade (III and IV) gliomas because radiographic tumor contact with this niche is associated with significantly decreased overall survival and progression-free survival, independent of the extent of tumor resection and other prognosticators [60-62]. Strikingly, GBM contact with the SGZ has not been found to influence survival [60]. About 50% of patients present with a V-SVZ contacting tumor and

an even higher percentage have contacting tumors at recurrence [61]. Single cells have been detected in the V-SVZ at autopsy even when the bulk tumor was not in contact with the region, and migration of cells along the subependyma – termed subependymal spread – is a well-described phenomenon [63-65].

A variety of explanations might be appropriate for why there is a consistent difference between outcome in patients with V-SVZ contacting tumors and those without. Although it is tempting to assume that the cell of origin is different for these cases, the cell(s) of origin for GBM have not yet been identified, with some arguing for neural stem cells, neural or oligodendrocyte progenitors, or even mature astrocytes [56-58, 66]. It is possible that tumors in the V-SVZ arise from neural stem cells while those that do not contact the V-SVZ arise from a more differentiated cell in the lineage. However, there is no difference in gene expression between contacting and non-contacting tumors based on bulk sample microarray data nor do tumors of a specific subtype preferentially appear near or far from the V-SVZ [62, 67, 68]. This suggests that there are either multiple different cells of origin for tumors in both locations or that they share the same cell of origin. Intriguingly, individual tumor cells found in the V-SVZ tended to be of the mesenchymal transcriptional subtype regardless of the subtype assigned to the bulk tumor [28, 69]. This finding suggests that the V-SVZ niche either imposes a similar gene expression profile upon the glioma cells or that a certain cell phenotype is particularly capable of V-SVZ infiltration. Taken together, these observations indicate that the V-SVZ, and its unique function as a neural stem cell niche with access to niche-derived factors, major white matter tracts, and cerebrospinal fluid (CSF), likely plays an important role in tumor biology including glioma growth, therapy resistance, dissemination, and immunomodulation (Figure 1.4).

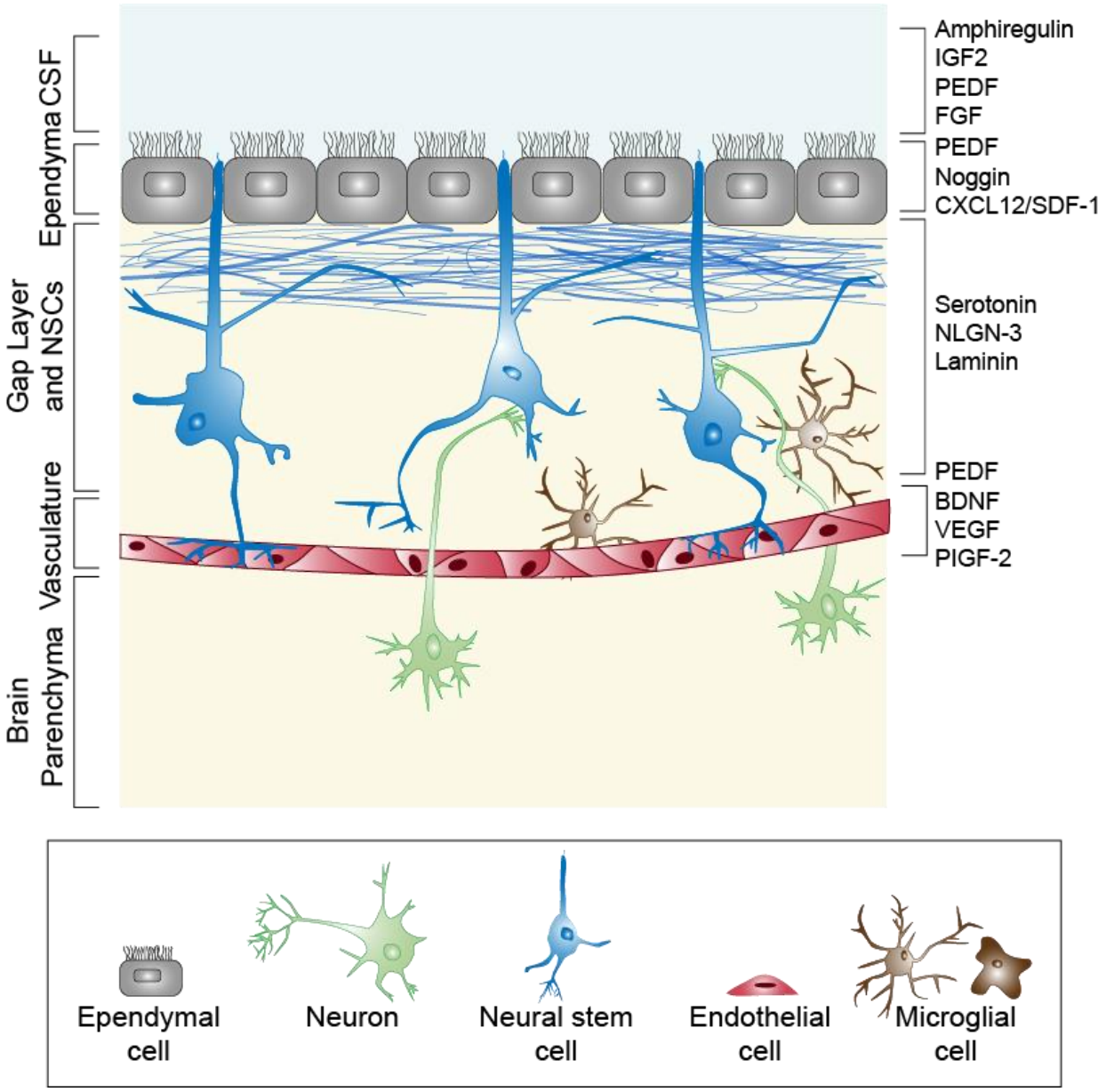


Figure 1.4: **Geography of the ventricular-sub-ventricular zone**. The V-SVZ niche contains ependymal cells (gray) that contact the lateral ventricle and CSF (light blue space). Neural stem cells (blue) have an apical contact with the CSF and a basal contact with the vasculature (red). In the human, astrocytic processes (blue) lie beneath the ependyma. Neurons (green) from the brain parenchyma innervate the niche. Surveying or resting microglia (brown) surveil the niche microenvironment and can become activated in the presence of tumor cells. A subset of factors demonstrated to be involved in these regions and discussed in the text are listed on the right. *Adapted from [70]*

V-SVZ composition

The V-SVZ is comprised of ependymal cells that line the ventricles, neural stem cells with a single primary cilium that protrudes through the ependyma into the CSF, GFAP-positive processes from mature astrocytes, mature neurons, and blood vessels lined by endothelial cells and pericytes (Figure 1.4) [53, 71-77]. Mouse and embryonic or neonatal human V-SVZ also contain rapidly dividing neural progenitor cells but adult human brains are largely quiescent unless perturbed, although this is an active area of research in the field [53, 78]. The region is rich in stimuli and soluble factors that, in healthy brains, serve to maintain the stem cell niche (Figure 1.4). Cerebrospinal fluid, at once a protective cushion for the central nervous system, reservoir of secreted factors, and mechanism for waste removal, contains soluble factors that regulate non-malignant stem cell quiescence and proliferation [51, 79]. These factors include insulin-like growth factor 2 (IGF-2), amphiregulin, and pigment epithelium-derived factor (PEDF), which have been shown to regulate glioma cells [80-82]. Moving into the V-SVZ proper, the cellular milieu itself has been shown to express various soluble and membrane-bound factors that can modulate glioma cell biology. These can include PEDF, noggin, and CXCL12 which suppress glioma stem cell differentiation and promote gliomasphere formation [81], maintain glioma cell tumor-initiating capacity [83, 84], and induce glioma cell homing to the niche [85, 86], respectively. An underappreciated but developing area of research is the impact of neuronal stimulation from non-malignant neurons onto glioma cells via traditional synapses as well as gap junctions [87-90]. Finally, the V-SVZ is highly vascularized but the pericytes that normally tightly ensheath endothelial cells are sparse, making vessels more permissive to the exchange of factors like growth signals, nutrients, and oxygen [91]. This more promiscuous architecture is exaggerated in the tumor setting of highly tortuous and leaky blood vessels, a hallmark of glioblastoma. Soluble growth factors like PEDF, BNDF, PIGF-2, and VEGF, hormones, nutrients, and oxygen from the blood can access the V-SVZ and the CSF via choroid plexus blood vessels [91-97]. These unique features of the V-SVZ may influence GBM cell protein expression, signaling, and behavior via

proliferation, migration, or de-differentiation and the topic merits further research. Improvements in tumor therapy could include targeting niche factors and disrupting niche-tumor cell interactions. Recent advances in both the preparation of single cell suspensions and the collection of high-dimensional data will enhance our ability to map specific populations of cancer and niche cells, providing a better understanding of the impact of V-SVZ niche diversity on tumor behavior [98, 99].

Single cell interrogation of gliomas

Single cell technologies have been instrumental in revealing new glioma biology over the past decade [12, 32, 33, 37, 38, 47, 99-102]. Single cell RNA-seq, in particular, has been elegantly applied to uncovering gene expression features of cells making up tumors. Findings from these studies include the vast heterogeneity of gene expression patterns within and among tumors [33, 38, 101], reconstructed phylogenies of individual tumors from proposed glioma stem cells to clonal progeny [37, 102], and the influence of non-malignant tumor microenvironment cells [32, 100]. These reports have illuminated the limitations of the previously identified three or four tumor molecular subclasses and emphasized the importance of considering diverse cell types in designing future therapies or selecting patients for clinical trials. This is especially apparent in the findings that multiple receptor tyrosine kinases are altered in individual cells, which suggests that treatments based on blocking a single receptor will be unsuccessful [33, 38]. Tumor cell types are also increasingly being related back to non-malignant populations of cells found in developing and adult human brains, with implications for tumor cell of origin and strategies for elimination of such cells [37, 47, 102]. RNA-seq has also been used to uncover different cell states between which individual cells may transition. Such transitions may be a mechanism by which tumor heterogeneity is established and maintained, lending GBM cells a plasticity that may be beneficial for escaping therapeutic elimination and contributing to eventual tumor progression [47]. On top

of these studies, functional examination of single glioma cells or derivative clones have been instructive in understanding how cells might respond to putative therapies [12, 37, 99]. Meyer *et al.* (PNAS, 2015) found that clones derived from single GBM cells had differential responses to both the standard of care alkylating agent temozolomide and a National Cancer Institute library of small molecule drugs. Some clones demonstrated resistance to therapy even before any application of drug started, a potential explanation for the fact that current standard of care therapy will fail to eliminate all cells that remain after surgery. Subsequent studies demonstrate that even new, targeted therapies are likely to be ineffective when used in isolation [99]. Single GBM cells were treated with kinase inhibitors and 12 different proteins and phosphoproteins were measured. Baseline measurements demonstrated unpotentiated GBM cell signaling and relationships between activated pathways but upon treatment with an mTOR inhibitor, compensatory signaling mechanisms were uncovered. By combining molecules that targeted mTOR signaling and the ensuing alternative signaling cascades, GBM cells were much more effectively inhibited. These studies demonstrate the immense potential of single cell protein and functional measurements in informing future GBM treatments that may eliminate more total cells, cells that specifically contribute to tumor recurrence, or specific cell populations. Furthermore, single cell technologies have advanced the field's understanding of the vast diversity and potential of glioma stem cells from their first description as CD133-positive cells in the early 2000s [39, 40] to over 21 cell types expressing glioma stem cell markers in a recent publication [37].

Single cell measurement of protein expression and modification

It is well-established that GBM tumors are comprised of many different cell types with different genetic aberrations and different gene expression [33, 36], and that tumors can be found in specialized niches with distinct non-malignant cell types, as described above. However, information on how the observed genetic and epigenetic changes or access to growth factors are

able to modulate individual cell protein expression and intracellular signaling is sorely lacking. Some notable progress has been made using single cell RNA-seq but this technique is limited to measuring transcript levels, and current studies are limited to a few hundred to 33,000 total cells for analysis [33, 37]. These approaches have not yet improved clinical practice or outcome for patients with glioblastoma. Therefore, it would be beneficial to be able to measure many different proteins of interest at once to be able to parse lineage markers, signaling events, and stemness features (Figure 1.3).

Mass cytometry is uniquely suited to address this challenge. Mass cytometry is an extension of the technology of fluorescence flow cytometry, wherein features are measured in single cells using antibodies or oligonucleotides against antigens of interest (Figure 1.5) [103-107]. In traditional fluorescence flow cytometry, single cells are stained with fluorophore-conjugated antibodies and are sequentially passed through lasers within a cytometer to excite the fluorophores. Emissions from these fluorophores are recorded, and a per-cell measurement of fluorescence intensity for each captured signal is reported. In mass cytometry, antibodies are instead covalently tagged with non-physiological heavy metals (atomic mass 89-209) and single cells are detected by a time of flight mass spectrometer (Figure 1.5). This permits 35+ features to be measured on thousands or millions of single cells, without the concerns of spectral overlap that currently limit flow cytometry; however, advances in fluorescence flow cytometry have increased the number of parameters that can be multiplexed, approaching numbers currently considered standard for mass cytometry [105, 108-110].

Suspension mass cytometry is a potentially valuable platform for solid tumor analysis, as it is relatively low cost, well-powered to detect rare and novel cell types, and able to sensitively measure many mechanistic determinants of cancer cell identity such as phosphorylated transcription factors, which are inaccessible to sequencing modalities [105, 111, 112]. The technology has been successfully used to identify cells associated with disease progression or patient outcomes in blood cancers [113-115]. However, mass cytometry requires samples that

are in single cell suspensions, presenting a technical challenge for analysis of solid tumors. Thus, the application of mass cytometry to study solid tumors, especially brain tumors, has lagged behind studies in fluid systems, such as the blood [111]. Excluding studies of tumor infiltrating immune cells, there are only a handful of studies applying mass cytometry to examine solid tissues [98, 116, 117]. The first challenge in employing the technology is to prepare samples in a way that is suitable for single cell staining and acquisition (Figure 1.5). Some groups have circumvented this challenge by employing mass-based labeling of intact tissue, either for mass-based imaging [118, 119] or for subsequent dissociation after staining [120]. The next challenge is to develop a panel of validated antibodies to detect phenotypes of interest in tumors that have primarily been described using gene expression or genome sequencing and derived from tissue of which there are few healthy samples for controls. Methods to address both of these challenges will be discussed in this dissertation (Chapter III and Appendix A). Once tumor samples have been stained with the antibody panel and data acquired on the mass cytometer, there is an abundance of high dimensional data that can illuminate biological findings of interest, such as what proteins are enriched in specific cell types, how protein expression or cell abundance varies between patient groups, and what cells are associated with patient outcomes (Figure 1.1). However, despite the potential of high-dimensional data, parsing the vast amount of information to gain meaningful insight is a significant challenge.

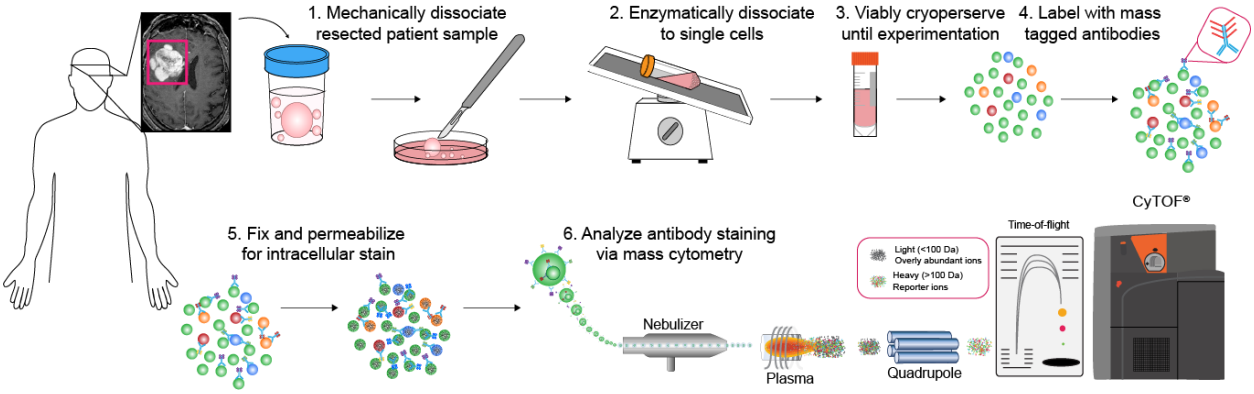


Figure 1.5: **Intra-operative patient samples can be dissociated into viable single cell suspension for analysis by flow cytometry**. Samples are collected from patients undergoing standard tumor resections. The samples are transported back to the laboratory where they are manually and enzymatically dissociated and viably cryopreserved. Upon thaw, the cells are stained with monoisotopic metal-tagged antibodies (atomic mass 89-209) against both extracellular and intracellular antigens. Single cells then pass through a nebulizer and are atomize and ionized by argon plasma, creating a cloud of ions corresponding to each cell, which is then analyzed by time of flight mass spectrometry. *Adapted from [112, 121].*

Table 1.1: **Algorithms for the analysis of single cell data**

Algorithm*	Applications	Description	Considerations
PCA [122]	1. Dimensionality Reduction	Transforms data to new axes in which the first component captures most of the variance in the dataset – Assumes data are parametric	1. Linear
DBSCAN [123]	1. Clustering	Density based clustering in which points packed closely together (also thought of as points with many neighbors) are assigned to the same cluster – Points in low density regions are outliers	1. Unsupervised
SPADE [124]	1. Dimensionality Reduction 2. Visualization 3. Clustering	First, high dimensional data is density downsampled, followed by agglomerative clustering, and the arrangement of upsampled clusters on a minimum spanning tree. – User must choose the desired number of clusters and the features on which to cluster – Useful for viewing the global data structure	1. Very limited or no prior knowledge required 2. Unsupervised 3. Non-deterministic
viSNE [125]	1. Dimensionality Reduction 2. Visualization	Based on the t-distributed stochastic neighbor embedding algorithm, providing a single cell view of high dimensional relationships in 2-dimensional space. – Emphasizes local structure while giving a view of global structure – Limited by the number of cells that is computationally feasible to analyze – Can be valuable to detect technical artifacts	1. Unsupervised 2. Non-deterministic (optional seed available)

Automatic Classification of Cellular Expression by Nonlinear Stochastic Embedding (ACCENSE) [126]	1. Dimensionality Reduction 2. Clustering	Combines non-linear dimensionality reduction (t-SNE) with density-based clustering.	1. Unsupervised
Citrus [127]	1. Clustering 2. Phenotype Defining 3. Modeling	Hierarchical clustering of cells based on user-defined features. Clusters and features of interest are identified that are associated with known endpoints of interest. Then Citrus trains a model based on identified features. – Random subsampling of cells (user-defined) – Cells can be assigned to >1 cluster – Requires at least 8 samples per group	1. Supervised
Wanderlust [128]	1. Dimensionality Reduction 2. Developmental Trajectory	Generates k-nearest neighbor graphs of all single cells and finds the average position of each cell, based on randomly chosen “waypoint” cells, on iterative trajectory calculations on the KNN graphs. Infers developmental trajectory from the average position of the cells in the graph – User must define initiating cell – Assumes all intermediate cells, including rare cells, are in the sample	1. Assumes linear cellular progression (no branched development) 2. Scales well to large datasets
FlowSOM	1. Clustering 2. Visualization	First, a self-organizing map is trained on the single cell data such cells are arranged in nodes based on similarity in high dimensional space. Nodes are then connected via a minimum spanning tree and similar nodes are combined into the final clusters. – Performs well with respect to speed and cluster stability and accuracy [130]	1. Unsupervised 2. Scales well to large datasets

[129]		– Can be sensitive to random starts resulting in outlier runs which perform worse than others	
Phenograph [113]	1. Clustering 2. Phenotype Defining	Models high-dimensional data using a nearest-neighbor graph, identifying highly interconnected nodes. – Makes no assumptions about size, number or form of subpopulations – Only one user-defined parameter – Can scale up to large amounts of input data	1. Unsupervised
Single-Cell Analysis by Fixed Force- and Landmark-Directed maps (Scaffold) [131]	1. Reference map	Builds models using force-directed graphs that incorporate prior knowledge. Uses landmark populations to direct the graphs. – Users are able to overlay new data onto map – Enables systems-level comparisons and evaluation of the global data structure – Can handle millions of cells	1. Requires prior knowledge
Cytosplore [132]	1. Workflow 2. Phenotype Description	First, uses SPADE to partition data, followed by A-tSNE to analyze high-level partitions in detail. Uses Gaussian Mean Shift (GMS [133]) to determine clusters based on the A-tSNE results. – Designed to deal with multiple scales within data from lineage defining features (large differences between subsets) and subtle differences within cell populations (features on a smaller scale) – Emphasis on visualization and interactive features for data exploration	2. Supervised 3. Assumes prior knowledge of the system and expected cell types
Hierarchical Stochastic Neighbor Embedding (HSNE)	1. Dimensionality Reduction	Builds small neighborhoods using k-nearest neighbor graphs and then identifies landmarks within the neighborhoods. Each landmark defines an area of influence.	1. Non-linear

[134]		– Shows structure at different scales by creating a hierarchical representation of data – Can be used to avoid downsampling cells	
Xshift [135]	1. Clustering	Uses K-nearest neighbor density estimation to identify clusters and arranges the clusters based on the number of nearest neighbors. Then clusters are merged based on mahalanobis distance. – Can process large datasets without subsampling	1. Unsupervised
A-tSNE [136]	1. Dimensionality Reduction 2. Visualization	Uses approximated k-nearest neighbors to approximate t-SNE embeddings in an adaptation of the t-SNE algorithm designed to minimize computation time. – Allows user to interact with and modify the analysis as it is ongoing	
CellCNN [137]	1. Clustering 2. Phenotype Description	Employs a convolutional neural network to identify cell populations from single cell measurements and associated phenotypes. – Designed to detect rare populations	1. Supervised.
DDPR [114]	1. Risk stratification	Manually gated, healthy populations were analyzed using Principal Component Analysis and then malignant cells were projected onto healthy populations using mahalanobis distance. – Elastic net machine learning used to identify features at the time of diagnosis that could predict relapse	1. Supervised 2. Requires prior knowledge
Uniform manifold approximation and projection (UMAP) [138]	1. Dimensionality Reduction	– Preserves both local and global structure – Faster than t-SNE – Retains information about developmental continua	1. Nonlinear 2. Unsupervised

Fourier transform (FFT)-accelerated interpolation-based t-SNE (FIt-SNE) [139]	1. Dimensionality Reduction 2. Visualization	An accelerated variation of t-SNE. Instead of comparing each point with all other points, FIt-SNE defines a small number of interpolation nodes and compares each cell to those nodes. The nodes are also compared to each other. The algorithm then interpolates interactions between all points from this data. – Eliminates the need for downsampling – Heatmap visualization based on one-dimensional t-SNE	1. Unsupervised
Opt-SNE [140]	1. Dimensionality Reduction 2. Visualization	An automated tool for choosing the optimal t-SNE parameters to facilitate embedding of datasets of millions of cells – Reduces computation time and improves visualization	1. Unsupervised

*In addition to the publications describing each algorithm, this table was informed by [105, 130, 141-144].

High dimensional, single cell data analysis

Analysis of single cell cytometry data has recently moved from an era of human-driven identification of cell types using known markers (expert gating) to embrace machine learning tools that can automatically reveal and characterize novel or abnormal cells (Figure 1.1) [144-147]. As this shift occurs, many tools for understanding high dimensional flow cytometry data have emerged, which can be used alone or in combination (Table 1.1). Several tools, such as t-distributed Stochastic Neighbor Embedding (t-SNE) and subsequent variations (like hierarchical-SNE or opt-SNE) and Uniform Manifold Approximation and Projection (UMAP) are used to reduced data dimensionality, taking 35+ dimensions and projecting them onto two new axes [125, 138, 140, 148]. They work by iteratively projecting the high dimensional data onto a two-dimensional plot, attempting to improve the 2D embedding with each iteration, based on reducing

the deviation of cell-cell relationships from the high dimensional cloud. These tools help explore the structure of multidimensional data and reveal subpopulations that can be overlooked in expert manual analysis as well as gain insight into patterns of marker expression or spot artifacts in datasets [125]. However, analysis does not end at dimensionality reduction. The goal of many mass cytometry experiments is to identify cell clusters for purposes ranging from describing a new cell type, to tracking populations over time or treatment, to comparing subset abundances between patients or conditions. To this end, a plethora of clustering algorithms have been developed within the past decade whose variety and abundance offer many options for scientists to best address their particular question or dataset (Table 1.1). A key feature of these algorithms is whether they are unsupervised (identifying groups of cells based on similar measured features) or supervised (identifying groups of cells based on external biological or clinical variables) [130]. Spanning-tree Progression Analysis of Density-normalized Events (SPADE), particularly useful for finding rare populations, requires users to choose a number of clusters before density-downsampling and building a minimum spanning tree [124]. FlowSOM also requires a user-defined cluster number and builds a minimum spanning tree after the construction of a self-organizing map [129]. Phenograph, meanwhile, automatically determines the number of clusters based on a user-defined nearest neighbors parameter [113]. None of these tools requires any classification or sorting of the samples prior to analysis. On the other hand, there are tools like Citrus, an automated cell subset discovery tool that uses prior knowledge of categorical labels, such as “diseased” or “healthy”, to identify cell clusters differentially associated with those labels [127]. CellCNN is another supervised analysis tool that requires prospective assignment of samples to categories and uses convolutional neural networks to learn a filter that predicts whether new cells match one of the groups [137]. Other cell subset discovery approaches do not supervise the analysis with knowledge of clinical outcomes but do use prior biological knowledge to identify cell subpopulations and then test whether differential outcomes are associated these cell subsets, such as developmentally dependent predictor of relapse (DDPR) [114]. At the end

of any of these analyses, the user has a number of cell populations that can be probed for biologically-relevant phenotypes, associations with clinically-relevant events, or changes between conditions.

Malignant cells in human tumors are remarkably diverse in their functional cell identities, and such intra-tumor cellular heterogeneity has been linked to patient outcomes [149, 150]. In blood cancers, single cell profiles of signaling networks have revealed cancer cells present at diagnosis whose abundance is linked to clinical outcomes, including patient survival [113-115, 151, 152]. Due to the relative ease of collecting and processing blood and bone marrow, many of the above tools were developed for, and bench-marked by, analysis of such samples, where known healthy controls are abundant [111]. Unlike these tissues, normal developing brain samples are difficult to obtain. Therefore, constructing normal brain cell developmental lineages is difficult and thus precludes the use of analytical tools reliant on such a dataset. Furthermore, GBMs are not easily divided into meaningful categories, as all patients progress, overall survival is very short, and standard of care is universal for every GBM. Therefore, patients are difficult to split into good or bad survival or into different therapeutic groups. To address these challenges in understanding data obtained from highly aggressive diseases, new experimental and computational tools are needed.

Conclusions and research objectives

Primary brain tumors, especially glioblastomas, are striking for their rapid and extremely poor prognosis as well as for their cellular heterogeneity at many levels of biological processes. This heterogeneity has been examined in the context of genomic and transcriptomic alterations, and to some extent, functional measurements of stem cell properties. Such studies have led to the definition of molecular subclasses of tumors and been the impetus for clinical trials. Although both inter- and intra-tumoral dysregulation of signaling in glioblastoma, particularly the disruption

of receptor tyrosine kinase (RTK) homeostasis, is hypothesized to drive disease aggressiveness, very little is known about the activation states of signaling effector proteins in glioblastoma and how these signaling changes may be associated with cell subpopulations and patient clinical outcomes [12, 35]. Disappointingly, the standard of care therapy has not changed since 2005 [2], and personalized therapeutic regimens are not available for patients. More recent single cell studies suggest that combinatorial therapy approaches may be beneficial to target multiple, divergent cell populations or to prevent resistance to a single drug. Novel, molecularly-driven criteria may give valuable insights into the biology of tumor progression and identify patients more likely to benefit from targeted therapeutics in development for this devastating malignancy.

A dearth of resources including appropriately prepared samples and validated technologies has resulted in few studies of single cell protein measurements in these tumors. This dissertation will describe the development of a mass cytometry panel for investigation of single glioblastoma cells from primary patient samples. Using this panel and CyTOF® technology (Fluidigm), diverse populations of cells were identified in patient samples. The creation of a new automated and unsupervised data analysis pipeline was used to further parse the dataset generated by mass cytometry. Risk Assessment Population IDentification (RAPID), as the algorithm is titled, was able to identify negative and positive prognostic cells within patient tumors. These negative and positive prognostic populations had unique phenotypic features, including specific intracellular signaling events and co-expression of lineage markers and stem cell proteins. Furthermore, the phenotypes were recaptured using a low-dimensional gating approach as well as in a new dataset of immunohistochemistry on tissue samples. RAPID was designed to be modular, and multiple approaches were tested at each step of the analysis, to ensure that the outcomes and workflow did not depend on one specific data analysis technique. Additionally, it was tested on an independent dataset from a different cancer and institution.

A complement to this work was a technical advance, described in Chapter II, for the study of lower grade gliomas. An antibody specific to the IDH1 R132H mutant protein, found in a large

number of grades II and III gliomas, was optimized for use in flow cytometry [153, 154]. The antibody had previously only been used for immunohistochemistry, but a series of epitope retrieval trials revealed a strategy for detecting it in glioma cells. It was successfully used on cell lines as well as on FFPE tissue using the previously described FFPE-DISSECT protocol [155]. The technical challenges of using this antibody on dissociated glioma samples were also explored.

In addition to interrogating the prognostic importance of individual cell types, glioblastoma mass cytometry data were also analyzed to determine if there were any signaling or protein expression differences between GBM tumors that contact the V-SVZ and those that do not. It has been described that patients bearing V-SVZ contacting tumors have worse outcomes than patients whose tumors do not contact this niche, but available data has not explained why this discrepancy exists. Although it is tempting to assume these tumors must contain different DNA alterations, gene expression profiles, or protein levels, this dissertation will report no detectable differences. Ideas for addressing the biological differences between the two categories of tumors are explored in Chapter V.

CHAPTER II

UNSUPERVISED MACHINE LEARNING REVEALS RISK STRATIFYING GLIOBLASTOMA TUMOR CELLS

Authors: Nalin Leelatian*, Justine Sinnaeve*, Akshitkumar M. Mistry, Sierra M. Barone, Asa A. Brockman, Kirsten E. Diggins, Allison R. Greenplate, Kyle D. Weaver, Reid C. Thompson, Lola B. Chambless, Bret C. Mobley, Rebecca A. Ihrie, and Jonathan M. Irish

*Equal contribution

This work is an accepted manuscript at *eLife* with expected publication in July 2020.

Preface

Glioblastoma tumors epitomize an urgent, unmet medical need due to their rapid, negative prognosis, documented heterogeneity, and dearth of efficacious treatment options. In close collaboration with Dr. Jonathan Irish's lab, and specifically then-graduate student Nalin Leelatian, I sought to use mass cytometry to better understand the contribution of protein expression to cellular phenotypes and intra- and inter-tumor heterogeneity. This led to the development of a mass cytometry antibody panel, specifically for antigens of interest in glioblastoma, and its application to 28 glioblastoma patient samples collected at Vanderbilt University Medical Center.

It became clear that many of the tools frequently applied to mass cytometry data would not be able to answer our questions about our glioblastoma dataset. Based on my work with Nalin Leelatian and described in her dissertation [156], I worked with Sierra Barone in the Irish lab to develop a new computational tool that could identify tumor cell clusters and indicate if they correlated with continuous clinical variables in an unsupervised manner. This work, the generation of a mass cytometry dataset on glioblastoma and the analysis of this data using the new algorithm are described in this chapter. The tumor processing and mass cytometry methods described in this chapter are detailed in Appendix A. Extended data from analysis of the mass cytometry experiments are shown in Appendix B.

Abstract

A central goal of cancer research is to reveal tumor cell subsets linked to clinical outcomes to generate new drug development and biomarker hypotheses. A key gap in this area is the need for tools based on continuous survival outcomes that can identify putative risk stratifying cells in pilot cohorts. We introduce a machine learning algorithm, Risk Assessment Population IDentification (RAPID), that is unsupervised and automated, identifies phenotypically distinct cell populations, and determines whether these populations stratify patient survival. With an initial mass cytometry dataset of 2 million cells from 28 glioblastomas, RAPID identified tumor cells whose abundance independently and continuously stratified patient risk of death. Statistical validation of these populations included repeated runs of stochastic workflow steps, repeated subsampling of cells, and testing of different algorithms and settings within the workflow. Biological validation used an orthogonal platform common within clinical trials, immunohistochemistry, and a larger cohort of 73 glioblastoma patients to confirm the findings from the pilot cohort. RAPID was also validated to find known risk stratifying cells and features

using published data from blood cancer. Cell subsets revealed by RAPID were continuously associated with patient outcomes and independent from known prognostic features. Thus, RAPID provides an automated, unsupervised approach for finding statistically and biologically significant cells using robust cytometry data from patient cohorts.

Introduction

A modern goal of quantitative analysis of single cell data in human cancers is to move beyond human-driven identification of cell types using known markers (expert gating) to machine learning tools that can reveal and characterize novel and abnormal cells [144-147]. Citrus, an automated analysis tool based on assignment of samples to binary categories (e.g. ("healthy" and "disease") before testing whether cell populations are associated with these categories, was designed with this purpose in mind (Table 2.1) [127]. However, many important clinical features of patient tissue samples are reported as continuous variables, such as time to progression, overall survival, or percentage of immune infiltrate, which can be challenging to convert to arbitrary binary categories and may not be driven by a single unified cellular phenotype [113, 114, 117]. Similarly, known, healthy cell populations from different stages of development or differentiation may be required for some approaches, such as developmentally dependent predictor of relapse (DDPR [114]) or Phenograph ([113], and are not always available or fully represented for all datasets. This is especially acute for some tissues, such as brain, which may be quiescent in adults and not routinely sampled in clinical care or research. In phenotypically variable diseases such as glioblastoma, it would also be powerful to test whether cell subpopulations identified via unsupervised means are associated with differential risk of death on a continuous scale [157].

Tools are needed, therefore, that can take into account continuous clinical variables that may be censored, such as overall survival or progression-free survival, and which operate in an unsupervised manner. In building an automated cytometry workflow, algorithm developers must decide whether users will supervise the discovery of cell subsets using clinical knowledge. Tools for automated, unsupervised cell discovery and characterization include SPADE [112], t-SNE [125], UMAP [138], FlowSOM [129], and Marker Enrichment Modeling (MEM [158]) (additional examples are found in Table 1.1). These tools help explore the structure of multidimensional data, review enriched features of groups, and reveal subpopulations that can be overlooked in expert manual analysis [144, 145, 147, 158, 159]. Ultimately, tools that work with high dimensional data should help users to translate findings from an algorithmic machine learning tool to common practice by identifying lower-dimensional correlates that can be used to validate signatures using a complementary, clinically tractable approach. This utility was a focus of the tool design and validation strategy used here. A computational workflow should also be validated via repeated subsampling of data to ensure the phenotypes identified are robust, by testing of different dimensionality reduction tools, by testing across multiple datasets, and by validation of prognostic signatures using complementary approaches. Finally, a practical challenge of modern single cell discovery projects is that they may often be at a project point where they are working with a smaller initial cohort (around 25 patients). This study size is powered to closely correlate cell subsets with patient outcomes using signaling cytometry data, as this study and others have shown for blood cancers [113, 114, 117, 151, 152, 160], but necessitates extensive statistical and biological validation, as discussed below.

RAPID (Risk Assessment Population IDentification) is a newly created algorithm that was designed using single cell cytometry data and which addresses the key challenges of clinical research using discovery cohorts of patients (https://github.com/cytolab/RAPID). This open-access tool can couple single cell experiments to clinical outcome and other variables in an unsupervised manner and provide information that can be translated into simplified tests on other

platforms. For this study, the algorithm was assessed for 1) cluster stability [161] for both cells and phenotypes; 2) modularity [144, 145], which would allow the algorithm to function with a range of dimensionality reduction approaches, such as no dimensionality reduction, t-distributed stochastic neighbor embedding (t-SNE [125]), or uniform manifold approximation and projection (UMAP [138]), clustering tools, such as FlowSOM [129] or dbscan [162], and enrichment analysis tools, such as marker enrichment modeling (MEM [158]); 3) transparency, evaluated as the ability to derive simple models of data structure [157], such as decision trees or flow cytometry gating hierarchies, so that new datasets could be easily assessed; 4) independence - whether risk stratifying cell populations are independent of known predictors (age, others); and 5) reproducibility and translational potential, tested by gathering additional data using traditional, one-dimensional immunohistochemistry (IHC) that is widely used in clinical testing.

Here, the utility and validity of the RAPID algorithm were tested using two datasets with varying levels of prior knowledge, numbers of patients and cells, and outcome trajectories. The first was a new dataset of 28 glioblastoma patient samples and is described in detail below. Central findings from this first dataset were then validated using 73 additional samples analyzed using a different technology. The second was a previously published dataset of 54 bone marrow samples from B cell precursor acute lymphoblastic leukemia [114]. This study was chosen as an example of a dataset in which prognostic features had already been independently identified, and so validation was assessed by whether known features were revealed by RAPID.

When applied to single cell cytometry data from human tumors, as shown here, the aim of RAPID was to reveal and characterize populations of risk stratifying cells. For this goal, glioblastoma, the cancer type in the first dataset, represents an excellent challenge, since glioblastoma is a highly heterogeneous solid tumor that is amenable to single cell approaches [98, 116, 117, 146] and where there is a great opportunity for molecular prognostic features to have an impact on new treatments and clinical care. Glioblastoma is the most common primary brain tumor in adults, is highly aggressive, and the median survival of glioblastoma patients after

diagnosis has remained approximately 12-15 months for over a decade [2, 163]. Furthermore, is known to contain cells with diverse genomic and transcriptomic features reflecting abnormal neural lineages [3, 33, 37, 98, 99]. Previous studies in glioblastomas have either measured signaling states in bulk primary tumors [28, 31, 164] or characterized genomic and transcriptomic profiles in a limited number of single cells (<33,000) [33, 34, 37, 47, 99, 165]. While differing subclasses of glioblastomas were proposed a decade ago [28], these categories do not correspond to large differences in prognosis and are not always reflected by individual cells [33]. Mosaic amplifications of receptor tyrosine kinase (RTK) genes are commonly observed in subsets of cells within a single glioblastoma tumor [35], suggesting that single cell analysis of glioblastoma should include signaling measurements. In other cancer types, phospho-protein signaling has repeatedly revealed cancer cell subsets that are closely linked to patient clinical outcomes [113, 114, 117, 151, 152, 160]. These results suggest that a protein-level approach in a small pilot cohort may reveal phenotypically distinct cancer cell subsets whose abundance provides new ways to stratify glioblastoma outcomes. While it is known that upstream regulators of pro-growth and pro-survival signaling are altered in brain tumors, little is known about the activation states of signaling effector proteins in single glioblastoma cells, as these features are inaccessible to sequencing modalities [12, 35, 105, 111].

Another challenge that the RAPID algorithm was designed to address was the need to work with heterogeneous cell phenotypes and populations that might be rare and variable across patients. Cytometry data are a good match for this type of algorithm, as a large number of cells are collected from each tumor sample, the data have an excellent signal-to-noise ratio and support quantitative comparisons, and cytometry enables direct measurement of signaling pathway activation [111, 115, 152, 160]. Here, two new technologies were created in parallel: 1) a tailored set of 34 antibodies for single cell mass cytometry of glioblastoma focused on phospho-protein signaling effectors, stem cell and cell identity proteins, and transcription factors critical to neural development, and 2) an unsupervised cell discovery workflow termed RAPID (Risk

Assessment Population IDentification). RAPID implements t-SNE, FlowSOM, and MEM analysis of single cell mass cytometry data to reveal risk stratifying cell populations [157]. The aim of combining these technologies was to reveal and characterize populations of risk stratifying glioblastoma cells, in the process illuminating new therapeutic targets or combinations of targets for patients. Additionally, the RAPID workflow can be applied to any high dimensional, single cell data evaluated in the context of a continuous variable. When glioblastoma mass cytometry data were analyzed by RAPID, both negative- and positive-prognostic phenotypes were identified, with protein-level phenotypes not described by prior studies. Statistical description of prognostic phenotypes within the RAPID algorithm then enabled the design of a simple workflow using traditional IHC, which stratified outcome in a separate set of 73 glioblastoma patient tissues.

Methods

Patient samples

Surgical resection specimens of 28 *IDH*-wild type glioblastomas collected at Vanderbilt University Medical Center between 2014 and 2016 were processed into single cell suspensions following an established protocol (Appendix A and [121]). Only samples that were confirmed to be *IDH*-wild type glioblastomas by standard pathological diagnosis were used. All samples were collected with patient informed consent in compliance with the Vanderbilt Institutional Review Board (IRBs #030372, #131870, #181970), and in accordance with the declaration of Helsinki.

Patient characteristics and collection of clinical data

Additional patient characteristics are included in Table 2.3 for all samples in this study. All patients were adults (≥ 18 years of age) at the time of their maximal safe surgical resection of

their cerebral (supratentorial) glioblastomas. Extent of surgical resection was independently classified as either gross total or subtotal resection by a neurosurgeon and a neuroradiologist. Gross total resection was defined as agreement by both viewers of no significant residual tumor enhancement on patients' gadolinium-enhanced magnetic resonance imaging (MRI) of the brain obtained within 24 hours after surgery. All patients were considered for treatment with postoperative chemotherapy (temozolomide) and radiation according to the standard of care [2], after determination of *MGMT* promoter methylation status by pyrosequencing (Cancer Genetics, Inc., Los Angeles, CA, USA). Multiplex polymerase chain reaction (PCR) was used to determine *IDH1/2* mutational status. Patients' postoperative course was followed until February 2019, noting time to first, definitive radiographic progression or recurrence of glioblastoma as agreed upon by the treating neuro-oncologist and neuroradiologist, and the time to patients' death. All deaths were deemed to be due to the natural course of patients' glioblastoma. Median overall survival of the analyzed 28 patients with *IDH*-wild type glioblastoma was 388.5 days (13 months) and median PFS was 187.5 days (6.3 months), which is typical for the disease [2, 3].

Mass cytometry analysis

Cells derived from patient samples were prepared as previously described (Appendix A and [121]). A multi-step staining protocol was used, which included 1) live surface stain, 2) 0.02% saponin permeabilization intracellular stain, and 3) intracellular stain after permeabilization with ice-cold methanol. All antibodies used, including clone information, and the steps when used are given in Table 2.4. After staining, cells were resuspended in deionized water containing standard normalization beads (Fluidigm) [166], and collected on a CyTOF 1.0 instrument located in the Cancer and Immunology Core facility at Vanderbilt University. Mass cytometry standardization beads were used to remove batch effects and to set the variance stabilizing arcsinh scale transformation for each channel following field-standard protocols [121, 146, 167]. Rhodium

viability stain and cleaved caspase-3 antibody were included in staining to exclude non-viable and apoptotic cells, respectively. Detection of total histone H3 was used to identify intact, nucleated cells [98]. A 34-dimensional mass cytometry antibody panel was used to analyze over 2 million viable cells from 28 tumors (ranging from 4,860 to 336,284 cells per tumor). Data were normalized with MATLAB-based normalization software [166], and were arcsinh transformed (cofactor 5), prior to analysis using the Cytobank platform [168]. Positively identified cells were defined by having signal above 10 on any channel on which an antibody was used to detect antigen. A patient-specific t-SNE view was generated, using 26 of the measured markers for all tumor and stromal cells from each patient's tumor [125] (Table 2.4). Immune ($CD45^+$) and endothelial cells ($CD31^+$) were computationally excluded from each individual patient prior to subsequent downstream analysis. Remaining $CD45^-CD31^-$ cells were included in a common t-SNE analysis, generated using 24 of 34 measured markers (Table 2.4). Distribution of each of the 28 patients' cells on the common t-SNE axes and mass intensity for each marker are shown in Appendix B. This common t-SNE analysis was used for automated analysis of risk stratifying cell subsets in RAPID (below).

Implementation of RAPID in R

FCS files for each patient sample (28) containing only cells of interest (non-immune, non-endothelial cells) were input in R (4,710 cells from each patient, 131,880 cells total). Cell subset identification was performed using the previously published FlowSOM R package [129]. Original features (24 measured markers), t-SNE values (t-SNE1_glioblastoma and t-SNE2_glioblastoma) from t-SNE, or UMAP values (UMAP1_gliolbastoma and UMAP2_gliolbastoma) from UMAP analysis of $CD45^-CD31^-$ glioblastoma cells from 28 patients were used as parameters for cell subset clustering. Within the RAPID workflow, the optimal number of clusters was determined by first identifying, for each feature, the smallest number of clusters that minimizes the intra-cluster

signal variance for that feature. Then, the optimal cluster number of the dataset was determined by taking the median of the optimal numbers for each individual feature. Once the cluster number was determined, the abundance of cell subsets and their clinical significance was assessed using outcome-guided analysis. Patients were divided into Low and High groups, based on the distribution (interquartile variance, IQR) of the abundance of a given cell subset across the cohort. A univariate Cox regression analysis was then used to estimate the effect size (hazard ratio, HR, of death) on survival and quantify its statistical significance with a p-value. The RAPID program output included: 1) two t-SNE (or UMAP) plots (.png), one color coded by each FlowSOM cluster and one color coded by prognostic status and p-value; 2) Kaplan-Meier survival curves for cell subsets; 3) .txt files of MEM and Median values for each feature, enrichment scores, and IQR values; 4) a new FCS file with File ID, cluster ID, and prognostic status for each cell; and 5) an .rds file with survival statistics for each cluster. In this study, abundance of Glioblastoma Negative Prognostic (GNP) and Glioblastoma Positive Prognostic (GPP) cells in each tumor was quantified as percentages per total glioblastoma cells (i.e. immune and endothelial cells were already excluded). Total GNP and GPP cell abundance were determined for each patient by adding the events in all GNP (or GPP subsets, respectively) together. GNP high patients were identified as containing more GNP cells than the IQR of total GNP abundance. GPP high patients were defined in the same manner. MEM analysis was performed in R, using the previously published R package [158]. In short, MEM captured and quantified cell subset-specific feature enrichment by scaling the magnitude (median) differences between clusters, depending on the spread (IQR) of the data. These values were then computed in comparison to the remaining cells in a given dataset. MEM values were interpreted as either being positively enriched (▲, UP positive values) or negatively enriched (▼, DN negative values). The variation of a given cellular feature across GNP or GPP cell subsets was quantified as ± standard deviations (SD). For the primary dataset used in this study (131,880 cells), RAPID ran in 15 minutes from start to finish after dimensionality reduction.

Cluster stability testing

Ten independent t-SNE or UMAP analyses were performed on equal numbers of randomly sampled cells from each patient (4,710 cells per patient, 131,880 total cells). RAPID was used to analyze each of these ten t-SNE/UMAP runs. For each sub-sampling of cells and the respective t-SNE/UMAP, an additional 99 FlowSOM clusterings were performed without setting a seed for reproducible results. After each analysis, an F-measure was calculated per cluster, measuring both the precision and recall of cell assignment. After 100 total FlowSOM runs, each of the original clusters had an average F-measure, interpreted here as a measure of cluster stability.

Survival and statistical analysis

Time from surgical resection to death (overall survival, OS) and time from surgical resection to the initial radiographic recurrence or death before radiographic assessment (progression-free survival, PFS) were depicted using right-censored Kaplan-Meier curves and analyzed in R. Survival time points were censored if, at last follow up, the patient was known to be alive or had not had radiographic progression. Differences in the survival curves of groups were compared using the Cox univariate regression model, reporting a hazard ratio (HR) with 95% confidence intervals between the survival curves.

A Cox proportional-hazards regression model was created to assess the influence of GNP and GPP cells on OS and PFS as continuous variables while accounting for other factors known to affect survival, including age at diagnosis, *MGMT* promoter methylation status, extent of surgical resection (EOR), treatment with temozolomide (TMZ), and radiation (XRT). The hazard model can be written as:

$$HR = \frac{h(t)}{h_0(t)} = e^{(b_{GNP}GNP + b_{age}Age + b_{MGMT}MGMT + b_{EOR}EOR + b_{XRT}XRT + b_{TMZ}TMZ)}$$

where $\frac{h(t)}{h_0(t)}$ represents the ratio of hazard comparing the risk of death at time *t* to the baseline hazard (obtained when all variables are equal to zero) and e^{b_x} represents the hazard ratio of variable x. The data were fit using R software, version 3.5 (R foundation for Statistical Computing, Vienna, Austria). The proportional-hazards assumption was tested in all multivariate models and supported by a non-significant relationship between Schoenfeld residuals and time for each covariate included in the model ($p > 0.38$; degree of freedom = 1) and the overall model ($p = 0.96$; degrees of freedom = 6 and 7). Statistical significance α was set at 0.05 for all statistical analyses, one- or two-tailed noted in figure legends.

An F-measure was used to quantify the level of agreement between classifications of patients or cells between alternative analysis strategies as wells as multiple RAPID iterations. The F-measure is the harmonic mean of the precision and recall given by the equation F = 2 * (Precision * Recall) / (Precision + Recall) where Precision = True Positive / (True Positive + False Positive) and Recall = True Positive / (True Positive + False Negative). An F-measure of 1 indicates perfect agreement between two different strategies or iterations as opposed to an F-measure of 0 which would mean no agreement between classifications of patients or cells from two strategies or iterations. Patients could be classified as GNP high, GNP and GPP low, or GPP high, while cells were classified as GNP, GPP, or neither. None of the patients in this study were classified as both GNP high and GPP high. To calculate the F-measure of patient categorization, the classification of the 28 patients into the three prognostic groups from the t-SNE implementation of RAPID was used as the reference point from which to compare patient classification resulting from the UMAP implementation of RAPID. Similarly, the stability of the RAPID workflow in assigning cells to GNP, GPP, or non-significant clusters was tested by using the t-SNE or UMAP implementation of RAPID (FlowSOM seed 38) as the reference from which to compare 100 iterations of RAPID (random FlowSOM seed per iteration). Calculation of the F-measure was implemented using R software, version 3.5.

Computer specifications

R was downloaded from https://cran.r-project.org/bin/ and implemented using the R Studio GUI https://www.rstudio.com/products/rstudio/download/#download. PC users also needed to download R Tools https://cran.r-project.org/bin/windows/Rtools/ and MAC users needed to download X11 Quartz https://www.xquartz.org/. RAPID was implemented, using these tools, on several personal computers. It was developed on a Dell Precision 7820 with a solid-state hard drive and 64GB RAM.

TMA sample selection

Formalin-fixed paraffin-embedded (FFPE) glioblastoma specimens were identified using the Vanderbilt Surgical Pathology database. The absence of *IDH* mutation was determined by multiplex PCR coupled with base extension assay (SNaPshot reaction mixture, Life Technologies, Carlsad, CA, USA), followed by capillary electrophoresis on an ABI Genetic Analyzer 3130XL and GeneMapper v.4.1. Following confirmation of the previously rendered histologic diagnosis, hematoxylin and eosin stained slides were scanned on the Panoramic P250 (3DHistech) whole slide scanner. Areas containing viable tumor were identified and circled by two pathologists (BM, NL).

TMA construction and staining

Blocks were delivered to the Vanderbilt University Medical Center TPSR (Translational Pathology Shared Resource), where cores were extracted from the encircled areas. Donor blocks and recipient blocks were loaded into the Tissue Microarray Grandmaster (3DHistech). The virtual slide images were aligned and overlaid on the tissue block and cores were removed from the donor block based on the pathologist annotation. Three 1mm core samples were collected

from each tumor and placed in the recipient block. IHC of serial sections of two TMA blocks (<10 μm thick) were stained with primary antibodies conjugated to HRP and 3,3′-Diaminobenzidine (DAB) detection for EGFR and S100B, and counter stained with Hematoxylin by the Translational Pathology Shared Resource (TPSR) at Vanderbilt University. Digital images were obtained with an Ariol SL-50 automated scanning microscope and the Leica SCN400 Slide Scanner from VUMC Digital Histology Shared Resource.

Marker	Clone	Company
S100B	polyclonal	Dako
EGFR	A-10	Santa Cruz Biotechnology

TMA imaging and analysis

Whole slide imaging was performed in the Digital Histology Shared Resource at Vanderbilt University Medical Center (www.mc.vanderbilt.edu/dhsr). For each marker, a QuPath project was created and all slide images were uploaded to be processed in batch. In QuPath, regions of interest (ROI's) were designated by circling each tumor core. Each ROI was computationally linked to the patient by a unique identifier, allowing cores from the same patient to be grouped. For each marker, the "Estimate Stain Vectors" function in QuPath was used to find the appropriate deconvolution parameters to isolate the signal intensity contribution from Hematoxylin and DAB respectively. The deconvolution parameters are listed below:

Marker	Hematoxylin			DAB			Background		
S100B	0.60484	0.67532	0.42204	0.20996	0.50234	0.83879	224	223	221
EGFR	0.72353	0.63737	0.26508	0.24952	0.52384	0.81445	221	219	220

For each ROI, nuclear segmentation on the Hematoxylin Optical Density (OD) was optimized using the "Watershed cell detection" function in QuPath, and the cytoplasm around each nucleus was estimated by performing a 3 μm expansion from the nuclear outline. All measurements from all detections were exported for analysis in R. In R, specific parameters (Name, Cell..DAB.OD.mean, Cytoplasm..DAB.OD.mean, and Nucleus..DAB.OD.mean) were extracted for every detection (cell) from every patient. These parameters identify the ROI/core from which the cell was segmented, its corresponding patient ID, the mean optical density of the deconvoluted DAB signal in each entire segmented cell, the DAB signal in only the cytoplasm, and the signal exclusively in the nucleus respectively. The full TMA map linking QuPath IDs, Patient_IDs, Block, and Core_IDs was also imported. In addition, for each marker, the median DAB intensity was calculated for each patient (averaged over three cores). The thresholds and measurements on which these thresholds were applied are summarized below:

Marker	Measurement	Threshold - Block A	Threshold - Block B
S100B	Cell_DAB	0.4	0.4
EGFR	Cell_DAB	0.2	0.2

Patients were categorized as GNP-like if their TMA cores had S100B staining intensity above the first quartile of S100B intensities (>0.6728) and had EGFR staining below the 50th percentile (<0.4199). Patients were categorized as GPP-like if their TMA cores scored in the top tertile of EGFR intensity (>0.6929).

Data availability

Annotated flow data files are available at the following link https://flowrepository.org/id/FR-FCM-Z24K. Patient specific views of population abundance and channel mass signals for all analyzed patients are shown in Appendix B.

Code availability

RAPID code is currently available on GitHub, together with a published dataset for analysis, at: https://github.com/cytolab/RAPID.

Results

RAPID identifies stratifying cell subsets in an automatic and unsupervised manner

The RAPID algorithm workflow is depicted in Figure 2.1 using results from Dataset 1. Following patient-specific identification of major cell types (Figure 2.1a), the algorithm (Figure 2.1b) randomly sampled an equal number of glioblastoma cells from each patient's tumor and analyzed the cells on a single, common t-SNE. This even sampling was conducted to generate a t-SNE analysis where each patient contributed equally. Subsequent statistical testing (Figure 2.1c) included repeated subsampling to ensure that sampled cells were representative of the original tumors. After multiple statistical tests, the most robust and reproducible cell types identified by RAPID were validated biologically, including using a new data type and a larger cohort (Figure 2.1d).

The RAPID algorithm was unsupervised and included two key statistical decisions. The first decision was the automation of the number of target clusters sought at the clustering step (Figure 2.1b, middle). This was achieved through repeated analysis with the chosen clustering tool, in this case FlowSOM [129], followed by statistical analysis. RAPID iteratively tested a range (cluster number 5-50) of unsupervised self-organizing maps from FlowSOM to identify an appropriate number of stable clusters containing phenotypically homogenous cells. The minimum number of clusters that minimized intra-cluster variance for each feature was calculated after all iterations were completed and set as the optimized target cluster number (see Methods). Clustering with other tools or clustering on untransformed axes, was both slower and less

accurate in identifying stable, phenotypically distinct clusters, consistent with published observations (data not shown and [130]). The second decision was in assessing cluster abundance in patients (Figure 2.1b, right). RAPID assigned patients to high or low abundance for each automatically identified cluster based on a statistical cut point, set as the interquartile range of the population abundance across the samples (see Methods). These two decisions resulted in automation of steps that are typically manual in cytometry analysis.

After finding clusters in an unsupervised manner and determining which patients' tumors contained a high level of each cluster, the last step in a run of RAPID was to test whether each cluster stratified risk of death. For this last test, RAPID applied a univariate Cox survival analysis to determine the correlation between the abundance of tumor cells in each cluster and patient survival outcome (Table 2.2). Clusters were identified as prognostic by assessing the hazard ratio (HR) of death in patients who had either high or low abundance of the cell cluster. Negative and positive prognostic clusters were colored red or blue, respectively, if they were significantly associated ($p<0.05$) with an HR that was >1 (negative, red) or <1 (positive, blue). The RAPID algorithm used statistical analysis of the common t-SNE, feature variance, and population abundance to automatically set all computational analysis parameters, independent of clinical outcomes.

The output of RAPID includes a PDF containing a color-coded, 2D t-SNE plot depicting all FlowSOM clusters, a 2D t-SNE plot colored by clusters which were significantly associated with patient outcome, and Kaplan-Meier survival plots of patients for each subset (additional files described in Methods) (Figure 2.1b). To compactly report and depict the phenotype of algorithmically identified cell subsets, RAPID used Marker Enrichment Modeling (MEM) labels [158]. Thus, feature enrichment was reported on a +10 to -10 scale, where +10 indicated that the feature was especially enriched in those cells and -10 indicated that the feature was specifically excluded from those cells, relative to all other cells in other clusters. The MEM label here was thus an objective description of what made each population distinct from the other clusters

identified by RAPID. In summary, RAPID provided an unsupervised, automated, statistical approach to revealing and characterizing clinically significant cells.

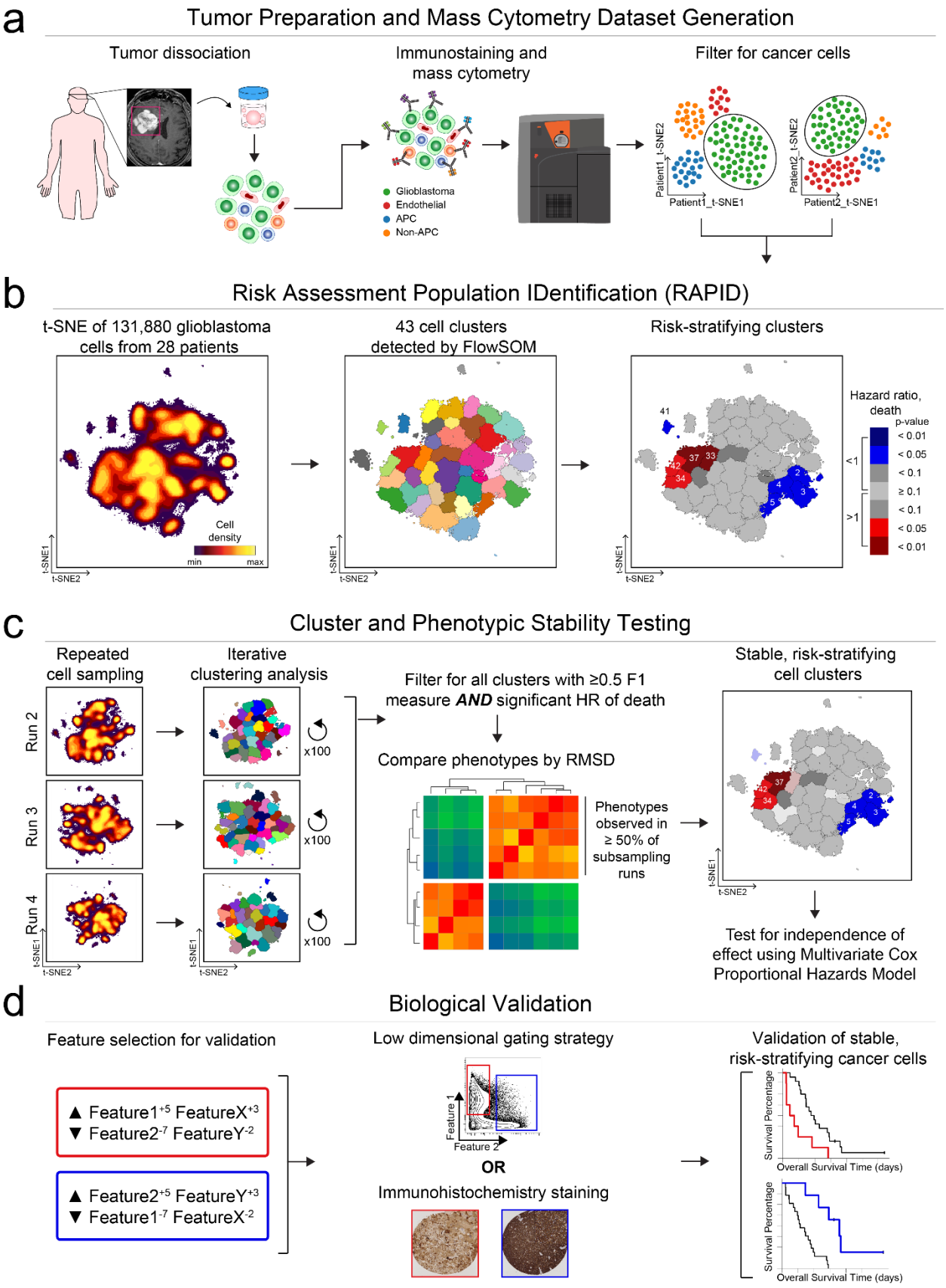

a
Tumor Preparation and Mass Cytometry Dataset Generation
Tumor dissociation
Immunostaining and mass cytometry
Filter for cancer cells
Glioblastoma
Endothelial
APC
Non-APC
Patient1_t-SNE1
Patient1_t-SNE2
Patient2_t-SNE1
Patient2_t-SNE2
b
Risk Assessment Population IDentification (RAPID)
t-SNE of 131,880 glioblastoma cells from 28 patients
43 cell clusters detected by FlowSOM
Risk-stratifying clusters
Cell density
min
max
Hazard ratio, death
p-value
< 0.01
< 0.05
< 0.1
≥ 0.1
< 0.1
< 0.05
< 0.01
<1
>1
41
37
33
42
34
2
4
3
5
t-SNE1
t-SNE2
c
Cluster and Phenotypic Stability Testing
Repeated cell sampling
Iterative clustering analysis
Run 2
Run 3
Run 4
x100
Filter for all clusters with ≥0.5 F1 measure AND significant HR of death
Compare phenotypes by RMSD
Phenotypes observed in ≥ 50% of subsampling runs
Stable, risk-stratifying cell clusters
Test for independence of effect using Multivariate Cox Proportional Hazards Model
d
Biological Validation
Feature selection for validation
▲ Feature1+5 FeatureX+3
▼ Feature2-7 FeatureY-2
▲ Feature2+5 FeatureY+3
▼ Feature1-7 FeatureX-2
Low dimensional gating strategy
Feature 1
Feature 2
OR
Immunohistochemistry staining
Validation of stable, risk-stratifying cancer cells
Survival Percentage
Overall Survival Time (days)

Figure 2.1: **RAPID identifies single cell phenotypes associated with continuous clinical variables that are stable and validated via complementary approaches.** (a) Graphic of tumor processing and data collection. After data collection and standard pre-processing, non-immune, non-endothelial glioblastoma cells were computationally isolated for analysis by RAPID. (b) RAPID workflow on glioblastoma cells identified from 28 patients and computationally pooled for t-SNE analysis. Cell subsets were automatically identified by FlowSOM and were systematically assessed for association with patient overall or progression-free survival. 43 glioblastoma cell subsets were identified and were color-coded based on hazard ratio of death and p-values (HR>1, red; HR<1, blue). Cell density, FlowSOM clusters, and cluster significance are depicted on t-SNE plots. (c) RAPID results were tested for stability. Each tumor was randomly subsampled for 4,710 cells multiple times. Each of these cell subsampling runs was subject to 100 iterative FlowSOM analyses and an F-measure was calculated for each cluster. Only clusters with an F-measure of greater than 0.5 were considered stable. Then, the phenotype of stable clusters associated with patient outcome were assessed via RMSD and used to determine stable phenotypes. (d) Validation of the findings from the mass cytometry data was done using lower dimensional gating strategies and an orthogonal technology to confirm the biological findings.

Identification of risk stratifying glioblastoma cells in Dataset 1

RAPID was designed for datasets like Dataset 1, a pilot glioblastoma mass cytometry dataset including cells collected from 28 patients with *isocitrate dehydrogenase (IDH)* wild type glioblastoma at the time of primary surgical resection (Table 2.3). This dataset is currently available online (https://flowrepository.org/id/FR-FCM-Z24K). The median progression-free survival (PFS) and overall survival (OS) after diagnosis were 6.3 and 13 months, respectively, typical of the trajectory of this disease [2]. Resected tissues were immediately dissociated into single cell suspensions as previously reported (Appendix A and [121]) and the resulting cells were stained with a customized antibody panel, which was designed to capture the expression of known cell surface proteins, intracellular proteins, and phospho-signaling events (Table 2.4). Collectively, the antigens included in this panel positively identified >99% of viable single cells within any given tumor sample (see Methods). To identify glioblastoma cells prior to RAPID, as in Figure 2.1a, a patient-specific t-SNE was created using 26 of the measured markers for the tumor and stromal cells from each patient's tumor [125] (Figure 2.2 and Table 2.4). Patient specific t-SNE maps revealed non-glioblastoma populations of immune ($CD45^+$) and endothelial ($CD45^-$ $CD31^+$) cells, consistent with prior mass cytometry and sequencing studies of gliomas [33, 47, 98,

146, 158]. Immune and endothelial cells from each individual patient were computationally excluded prior to subsequent downstream analysis (Figure 2.1, Figure 2.2), and $CD45^{-}CD31^{-}$ cells were labeled as glioblastoma cells.

Plots of cell density on the t-SNE axes revealed phenotypically distinct subpopulations of glioblastoma cells within a single patient's tumor (example patient LC26, Figure 2.2 and visualization of all patients in Appendix B). Intra-tumoral subsets were distinguished by differences in expression of core neural identity proteins and by aberrant co-expression of neural lineage and stem cell proteins. In the example case of tumor LC26, abnormal phenotypes in glioblastoma cells included co-expression of astrocytic S100B and stem-like CD133 or co-expression of markers associated with different molecular subtypes of glioblastoma, such as mesenchymal (CD44) and classical (EGFR) (Figure 2.2) [28]. These results with protein confirmed the existence of non-canonical cell types that had previously been observed in single-cell RNA-seq [33]. The abnormal co-expression of identity proteins seen here, as well as previously reported single cell studies relying on inferred DNA alterations [47], indicate that the large majority of the $CD45^{-}CD31^{-}$ cells were likely cancer lineage cells.

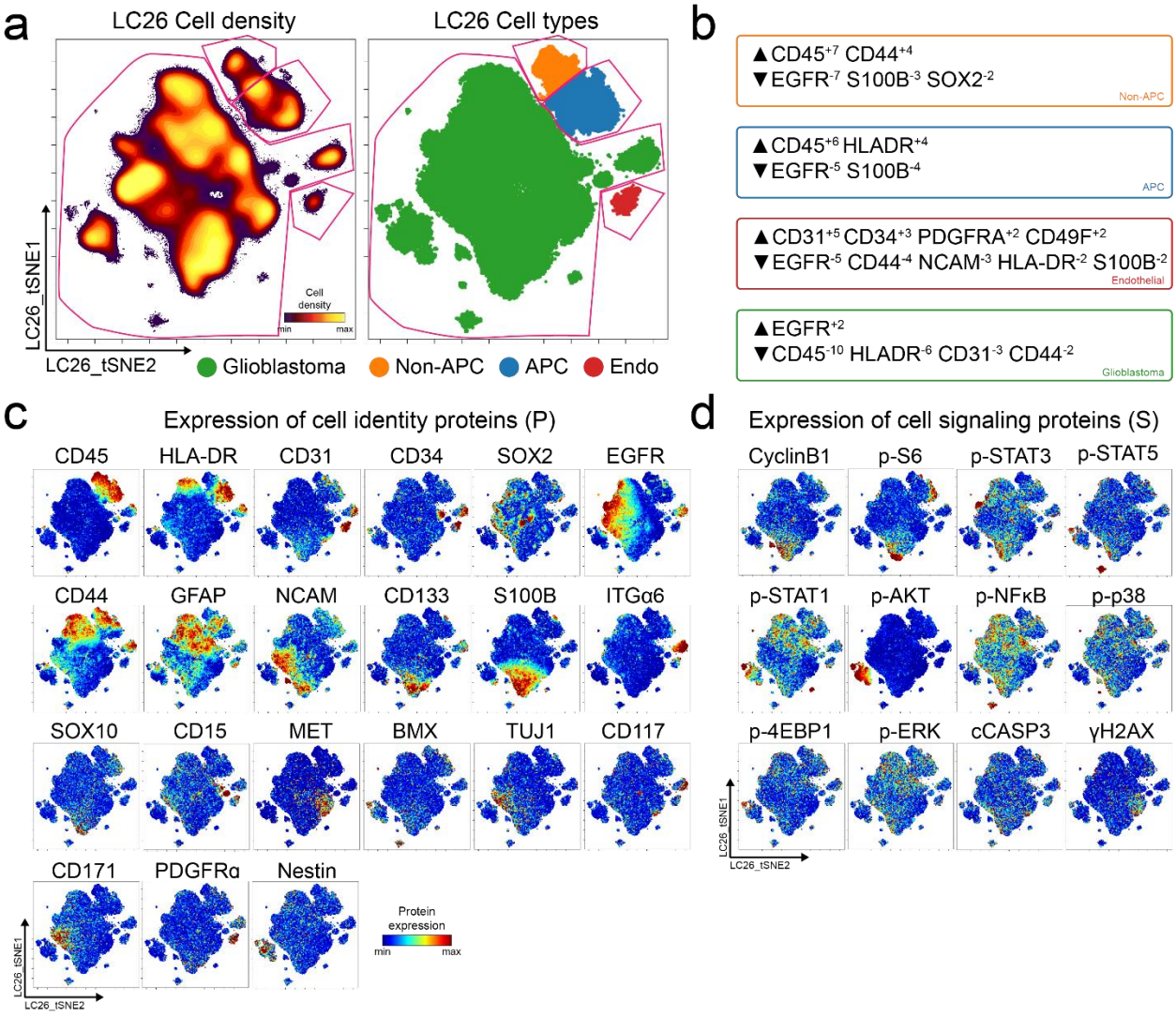


Figure 2.2: **Single cell quantification of identity proteins and phospho-protein signaling in glioblastoma**. (a) t-SNE plots of cell density (left) and major cell types in a patient tumor (LC26) colored by expert gating (right) for antigen presenting cells (APC, blue), other immune cells (non-APC, orange), endothelial cells (Endo, red), and glioblastoma cells (green) using CD45, CD31, and HLA-DR to identify cells. Pink lines indicate where expert gates were drawn. (b) MEM protein enrichment scores for populations indicated by color in (a), using the other three populations as reference. (c) Per-cell expression levels of 21 identity proteins, (d) 9 phosphorylated signaling effectors, proliferation marker cyclin B1, apoptotic signaling factor cleaved caspase 3 (cCASP3), and DNA damage marker γH2AX in LC26 are depicted. Heat indicates protein or phospho-protein expression per cell; scale is specific to each measured feature.

Using an equal number of subsampled glioblastoma cells from each patient (see Methods), a single, common t-SNE map was created to represent glioblastoma cell protein phenotypes across all patients (N = 131,880 cells; 4,710 cells x 28 patients, using 24 measured features). The RAPID algorithm, using the pooled data from all patients, identified 43 phenotypically distinct cell clusters, and then determined for each tumor whether a patient was high or low for a particular cluster using the interquartile range of abundance for that cluster. For example, for glioblastoma cluster 24, the interquartile range was 0.67% to 3.36%, resulting in a cut point of 2.69%. Those patients with ≤ 2.69% were designated 'low' for cluster 24 while those with > 2.69% were assigned to the 'high' group. Additional cut points, based on splitting populations into quartiles or tertiles, were tested and resulted in consistent prognostic phenotypes (the average F-measure of patients being consistently assigned to the high, low, or neither categories identified below was 0.86).

The RAPID algorithm identified four Glioblastoma Negative Prognostic (GNP) clusters (red; clusters 33, 34, 37, and 42) and five Glioblastoma Positive Prognostic (GPP) clusters (blue; clusters 2, 3, 4, 5, and 41) whose abundance was associated with overall survival (Figure 2.1b). MEM labels were used to identify the enriched features of risk stratifying glioblastoma cells (Figure 2.3 and 2.4). MEM labels were calculated for both total proteins (P), such as S100B and EGFR, and signaling effectors (S), such as p-STAT5. GNP cells aberrantly co-expressed neural-lineage proteins (astrocytic S100B and stem-like SOX2). Additionally, GNP cells displayed phosphorylation of RTK signaling effectors known to promote cell survival, growth, and proliferation (e.g. p-STAT5^{Y694}, p-S6$^{S235/S236}$, p-STAT3^{Y705}) (Figure 2.3 and 2.4). The MEM protein enrichment values (average and standard deviation) for GNP cells included neural lineage determinants (▲S100B$^{+5\pm1.6}$, SOX2$^{+5\pm1}$) and phospho-proteins (▲p-STAT3$^{+3\pm2.1}$, p-STAT5$^{+2\pm1.8}$, p-S6$^{+3\pm1.4}$) and identified proteins that were specifically lacking in GNP cells relative to other glioblastoma cell clusters (▼EGFR$^{-2\pm0.1}$, GFAP$^{-4\pm0.7}$, CD44$^{-4\pm0}$) (Figure 2.5). In contrast, GPP cells were positively enriched for EGFR (▲EGFR$^{+5\pm0.8}$) and consistently lacked pro-survival phospho-

proteins (▼p-S6$^{-4\pm3.7}$, p-STAT5$^{-2\pm0.8}$, p-STAT3$^{-2\pm1.6}$) and one of the proliferation markers measured (▼cyclin B1$^{-3\pm3.3}$) (Figure 2.5).

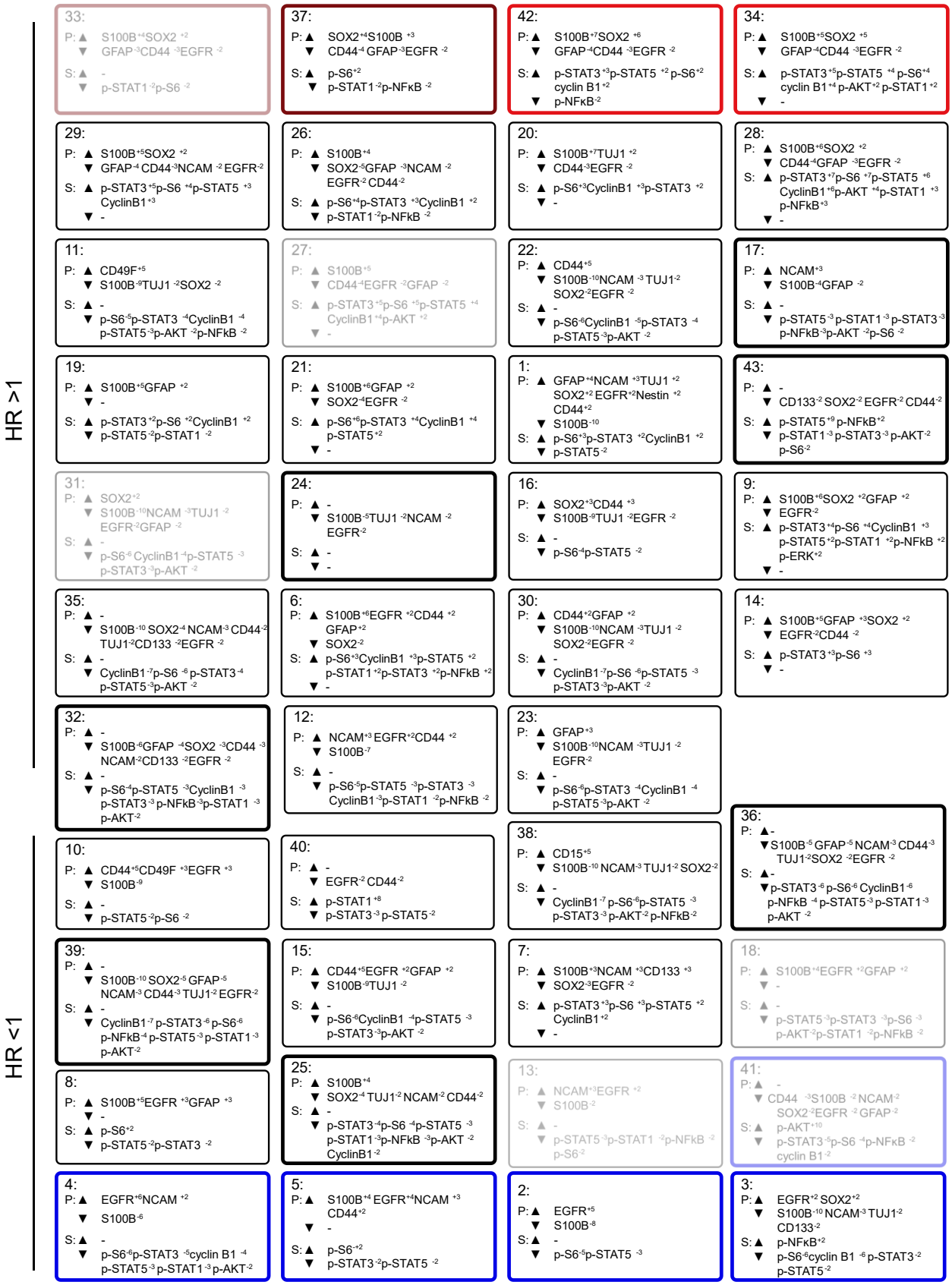

HR >1
HR <1
33:
P: ▲ S100B+4SOX2 +2
▼ GFAP-3CD44 -3EGFR -2
S: ▲ -
▼ p-STAT1 -2p-S6 -2
37:
P: ▲ SOX2+4S100B +3
▼ CD44-4 GFAP-3EGFR -2
S: ▲ p-S6+2
▼ p-STAT1 -2p-NFκB -2
42:
P: ▲ S100B+7SOX2 +6
▼ GFAP-4CD44 -3EGFR -2
S: ▲ p-STAT3 +3p-STAT5 +2 p-S6+2
cyclin B1+2
▼ p-NFκB-2
34:
P: ▲ S100B+5SOX2 +5
▼ GFAP-4CD44 -3EGFR -2
S: ▲ p-STAT3 +5p-STAT5 +4 p-S6+4
cyclin B1+4 p-AKT+2 p-STAT1 +2
▼ -
29:
P: ▲ S100B+5SOX2 +2
▼ GFAP-4 CD44-3NCAM -2 EGFR-2
S: ▲ p-STAT3 +5p-S6 +4p-STAT5 +3
CyclinB1+3
▼ -
26:
P: ▲ S100B+4
▼ SOX2-5GFAP -3NCAM -2
EGFR-2 CD44-2
S: ▲ p-S6+4p-STAT3 +3CyclinB1 +2
▼ p-STAT1 -2p-NFκB -2
20:
P: ▲ S100B+7TUJ1 +2
▼ CD44-3EGFR -2
S: ▲ p-S6+3CyclinB1 +3p-STAT3 +2
▼ -
28:
P: ▲ S100B+6SOX2 +2
▼ CD44-4GFAP -3EGFR -2
S: ▲ p-STAT3 +7p-S6 +7p-STAT5 +6
CyclinB1+6p-AKT +4p-STAT1 +3
p-NFkB+3
▼ -
11:
P: ▲ CD49F+5
▼ S100B-9TUJ1 -2SOX2 -2
S: ▲ -
▼ p-S6-5p-STAT3 -4CyclinB1 -4
p-STAT5 -3p-AKT -2p-NFkB -2
27:
P: ▲ S100B+5
▼ CD44-4EGFR -2GFAP -2
S: ▲ p-STAT3 +5p-S6 +5p-STAT5 +4
CyclinB1 +4p-AKT +2
▼ -
22:
P: ▲ CD44+5
▼ S100B-10NCAM -3 TUJ1-2
SOX2-2EGFR -2
S: ▲ -
▼ p-S6-6CyclinB1 -5p-STAT3 -4
p-STAT5 -3p-AKT -2
17:
P: ▲ NCAM+3
▼ S100B-4GFAP -2
S: ▲ -
▼ p-STAT5 -3 p-STAT1 -3 p-STAT3 -3
p-NFkB -3p-AKT -2p-S6 -2
19:
P: ▲ S100B+5GFAP +2
▼ -
S: ▲ p-STAT3 +2p-S6 +2CyclinB1 +2
▼ p-STAT5 -2p-STAT1 -2
21:
P: ▲ S100B+6GFAP +2
▼ SOX2-4EGFR -2
S: ▲ p-S6+6p-STAT3 +4CyclinB1 +4
p-STAT5+2
▼ -
1:
P: ▲ GFAP+4NCAM +3TUJ1 +2
SOX2+2 EGFR+2Nestin +2
CD44+2
▼ S100B-10
S: ▲ p-S6+3p-STAT3 +2CyclinB1 +2
▼ p-STAT5 -2
43:
P: ▲ -
▼ CD133-2 SOX2-2 EGFR-2 CD44-2
S: ▲ p-STAT5 +9 p-NFkB+2
▼ p-STAT1 -3 p-STAT3 -3 p-AKT-2
p-S6-2
31:
P: ▲ SOX2+2
▼ S100B-10NCAM -3TUJ1 -2
EGFR-2GFAP -2
S: ▲ -
▼ p-S6-6 CyclinB1-4p-STAT5 -3
p-STAT3 -3p-AKT -2
24:
P: ▲ -
▼ S100B-5TUJ1 -2NCAM -2
EGFR-2
S: ▲ -
▼ -
16:
P: ▲ SOX2+3CD44 +3
▼ S100B-9TUJ1 -2EGFR -2
S: ▲ -
▼ p-S6-4p-STAT5 -2
9:
P: ▲ S100B+6SOX2 +2GFAP +2
▼ EGFR-2
S: ▲ p-STAT3 +4p-S6 +4CyclinB1 +3
p-STAT5 +2p-STAT1 +2p-NFkB +2
p-ERK+2
▼ -
35:
P: ▲ -
▼ S100B-10 SOX2-4 NCAM-3 CD44-2
TUJ1-2CD133 -2EGFR -2
S: ▲ -
▼ CyclinB1-7p-S6 -6 p-STAT3-4
p-STAT5 -3p-AKT -2
6:
P: ▲ S100B+6EGFR +2CD44 +2
GFAP+2
▼ SOX2-2
S: ▲ p-S6+3CyclinB1 +3p-STAT5 +2
p-STAT1 +2p-STAT3 +2p-NFkB +2
▼ -
30:
P: ▲ CD44+2GFAP +2
▼ S100B-10NCAM -3TUJ1 -2
SOX2-2EGFR -2
S: ▲ -
▼ CyclinB1-7p-S6 -6p-STAT5 -3
p-STAT3 -3p-AKT -2
14:
P: ▲ S100B+5GFAP +3SOX2 +2
▼ EGFR-2CD44 -2
S: ▲ p-STAT3 +3p-S6 +3
▼ -
32:
P: ▲ -
▼ S100B-6GFAP -4SOX2 -3CD44 -3
NCAM-2CD133 -2EGFR -2
S: ▲ -
▼ p-S6-4p-STAT5 -3CyclinB1 -3
p-STAT3 -3 p-NFkB -3p-STAT1 -3
p-AKT-2
12:
P: ▲ NCAM+3 EGFR+2CD44 +2
▼ S100B-7
S: ▲ -
▼ p-S6-5p-STAT5 -3p-STAT3 -3
CyclinB1-3p-STAT1 -2p-NFkB -2
23:
P: ▲ GFAP+3
▼ S100B-10NCAM -3TUJ1 -2
EGFR-2
S: ▲ -
▼ p-S6-6p-STAT3 -4CyclinB1 -4
p-STAT5 -3p-AKT -2
36:
P: ▲-
▼S100B-5 GFAP-5 NCAM-3 CD44-3
TUJ1-2SOX2 -2EGFR -2
S: ▲-
▼p-STAT3 -6 p-S6-6 CyclinB1-6
p-NFkB -4 p-STAT5 -3 p-STAT1 -3
p-AKT -2
10:
P: ▲ CD44+5CD49F +3EGFR +3
▼ S100B-9
S: ▲ -
▼ p-STAT5 -2p-S6 -2
40:
P: ▲ -
▼ EGFR-2 CD44-2
S: ▲ p-STAT1 +8
▼ p-STAT3 -3 p-STAT5-2
38:
P: ▲ CD15+5
▼ S100B-10 NCAM-3 TUJ1-2 SOX2-2
S: ▲ -
▼ CyclinB1-7 p-S6-6p-STAT5 -3
p-STAT3 -3 p-AKT-2 p-NFkB-2
39:
P: ▲ -
▼ S100B-10 SOX2-5 GFAP-5
NCAM-3 CD44-3 TUJ1-2 EGFR-2
S: ▲ -
▼ CyclinB1-7 p-STAT3-6 p-S6-6
p-NFkB-4 p-STAT5-3 p-STAT1-3
p-AKT-2
15:
P: ▲ CD44+5EGFR +2GFAP +2
▼ S100B-9TUJ1 -2
S: ▲ -
▼ p-S6-6CyclinB1 -4p-STAT5 -3
p-STAT3 -3p-AKT -2
7:
P: ▲ S100B+3NCAM +3CD133 +3
▼ SOX2-3EGFR -2
S: ▲ p-STAT3 +3p-S6 +3p-STAT5 +2
CyclinB1+2
▼ -
18:
P: ▲ S100B+4EGFR +2GFAP +2
▼ -
S: ▲ -
▼ p-STAT5 -3p-STAT3 -3p-S6 -3
p-AKT-2p-STAT1 -2p-NFkB -2
8:
P: ▲ S100B+5EGFR +3GFAP +3
▼ -
S: ▲ p-S6+2
▼ p-STAT5 -2p-STAT3 -2
25:
P: ▲ S100B+4
▼ SOX2-4 TUJ1-2 NCAM-2 CD44-2
S: ▲ -
▼ p-STAT3 -4p-S6 -4p-STAT5 -3
p-STAT1 -3p-NFkB -3p-AKT -2
CyclinB1-2
13:
P: ▲ NCAM+3EGFR +2
▼ S100B-2
S: ▲ -
▼ p-STAT5 -3p-STAT1 -2p-NFkB -2
p-S6-2
41:
P: ▲ -
▼ CD44 -3S100B -2 NCAM-2
SOX2-2EGFR -2 GFAP-2
S: ▲ p-AKT+10
▼ p-STAT3 -6p-S6 -4p-NFκB -2
cyclin B1-2
4:
P: ▲ EGFR+6NCAM +2
▼ S100B-6
S: ▲ -
▼ p-S6-6p-STAT3 -6cyclin B1 -4
p-STAT5 -3 p-STAT1 -3 p-AKT-2
5:
P: ▲ S100B+4 EGFR+4NCAM +3
CD44+2
▼ -
S: ▲ p-S6-+2
▼ p-STAT3 -2p-STAT5 -2
2:
P: ▲ EGFR+5
▼ S100B-8
S: ▲ -
▼ p-S6-5p-STAT5 -3
3:
P: ▲ EGFR+2 SOX2+2
▼ S100B-10 NCAM-3 TUJ1-2
CD133-2
S: ▲ p-NFκB+2
▼ p-S6-6cyclin B1 -6 p-STAT3-2
p-STAT5-2

Figure 2.3: **Quantitative MEM labels of the enriched identity proteins and signaling features of all glioblastoma cell subsets identified by RAPID.** Enrichment of identity proteins (P) and phosphorylated signaling effectors (S) of glioblastoma cell subsets identified by RAPID was quantified using MEM. GNP and GPP cells are labeled in red and blue, respectively. Populations detected in every patient sample (abundances ranging from 0.02% to 28.05) are outlined in bold. Populations deemed unstable (either by F-measure <0.5 or representing phenotypes displayed in less than 50% of cell subsampling runs) are faded.

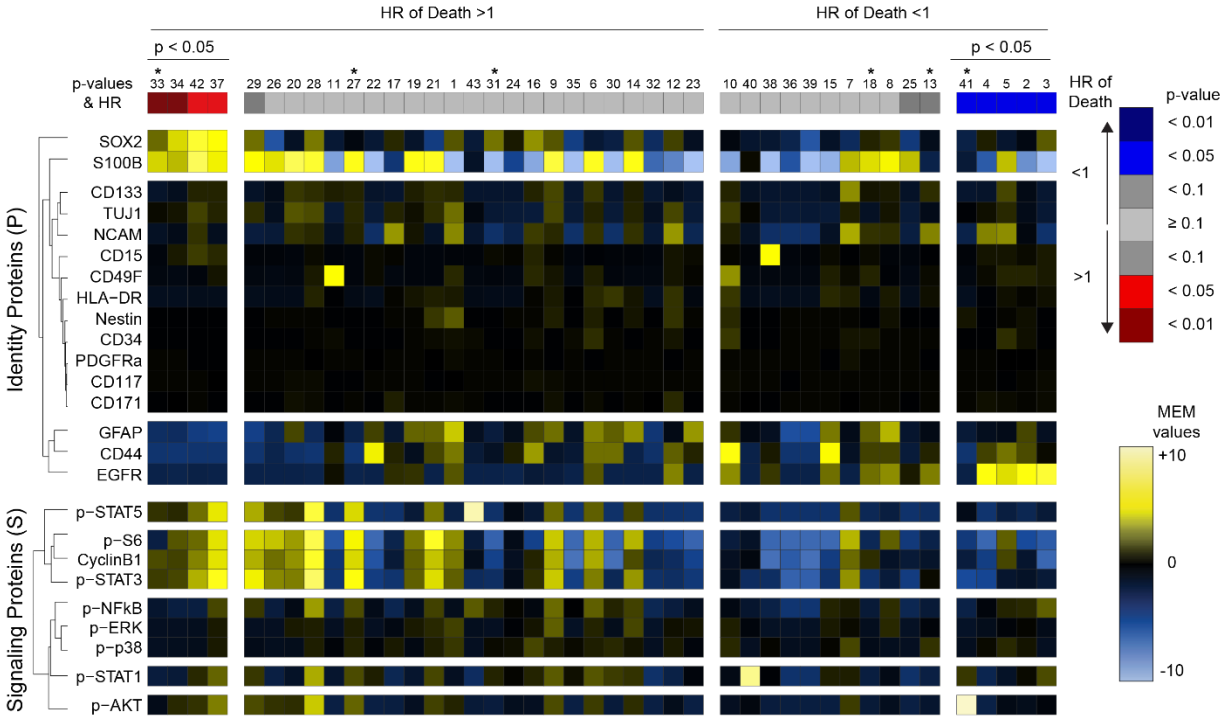


Figure 2.4: **Glioblastoma cell subsets showed differential enrichment of identity proteins and phosphorylated signaling effectors.** Forty-three glioblastoma cell subsets automatically identified by FlowSOM are arranged according to their associations with overall survival (HR>1, left; HR<1, right) and statistical significance of that association (p-values). The heatmap represents the MEM values of glioblastoma cell subsets (columns). GNP cells are labeled in red, while GPP cells are labeled in blue. Hierarchical clustering was performed based on MEM values and is depicted on the left of the heatmap for measured features. HR = hazard ratio of death. Asterisks (*) above indicate that clusters are not stable (F-measure of <0.5 or phenotypes identified in less than 50% of cell subsampling runs).

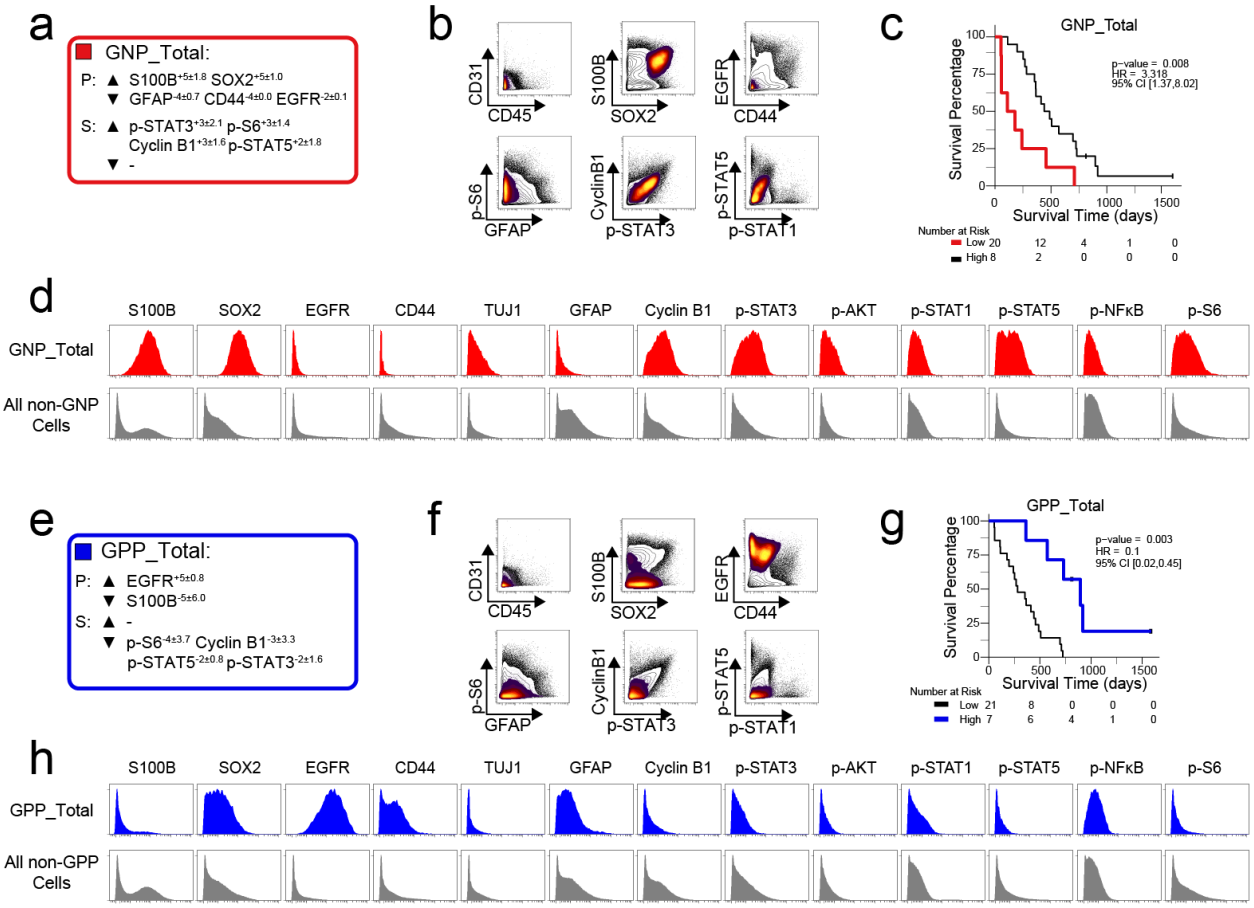


Figure 2.5: **Divergent phenotypes are associated with patient outcomes.** (a) Enrichment (upwards arrowhead) or lack (downwards arrowhead) of identity proteins (P) and phosphorylated signaling effectors (S) on Glioblastoma Negative Prognostic cell subsets was quantified using MEM. Average MEM scores are shown for three GNP subsets ± the standard deviation. (b) Combined GNP cell subsets (density contours) were mapped over biaxial plots of all other tumor cells (black contours). (c) Overall survival of patients for high (> 2.96%) total GNP content compared to patients with low (< 2.96%) GNP content. (d) Histogram plots of GNP cells (red) and all other glioblastoma cells (gray) illustrate the expression of identity proteins and phosphorylated signaling effectors. (e) Enrichment (upwards arrowhead) or lack (downwards arrowhead) of identity proteins (P) and phosphorylated signaling effectors (S) on Glioblastoma Positive Prognostic cell subsets was quantified using MEM. Average MEM scores are shown for three GNP subsets ± the standard deviation. (f) Combined GPP cell subsets (density contours) were mapped over biaxial plots of all other tumor cells (black contours). (g) Overall survival of patients for high (> 8.65%) total GPP content compared to patients with low (< 8.65%) GPP content. (h) Histogram plots of each GPP cell subset (blue) and all other glioblastoma cells (gray) illustrate the expression of proteins and phosphorylated signaling effectors.

Non-malignant cells, including immune and endothelial cells, were excluded from initial RAPID analyses and subsequent biaxial gating confirmed that the GNP and GPP subsets were not unexpected residual $CD45^+$ or $CD31^+$ cells (Figure 2.5). However, infiltrating immune cells can comprise a large proportion of non-cancer cells in glioblastomas and have highly variable overall abundance across patients [169]. Notably, GPP-high (n=7) patients' tumors all contained more than 9% $CD45^+$ cells (median % = 25.3±13.8), whereas all GNP-high (n=8) patients' tumors contained less than 9% $CD45^+$ cells (median %= 3.3±2.4, $p < 0.001$, Figure 2.6).

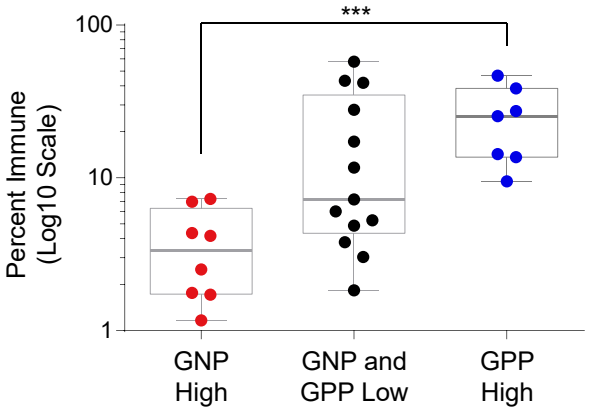


Figure 2.6: **Abundance of immune cells correlated with the abundance of prognostic cell subsets.** Box and whisker plot of immune abundance (%, log10 scale) on the y-axis and patients divided into three groups: GNP high (red, >2.96% GNP cells), GPP high (blue, >8.65% GPP), or GNP and GPP low (gray). Box encompasses the 25th to 75th percentile, gray horizontal line indicates the median, and whiskers extend to the minimum and maximum values. *** p=0.0008, two-tailed t-test.

Identification of risk stratifying B-cell leukemia cells in Dataset 2

FCS files from a previously published mass cytometry study of B-cell precursor acute lymphoblastic leukemia (BCP-ALL) by an independent lab were input into the RAPID workflow to test whether the RAPID algorithm could re-discover prognostic cell subsets in other disease

settings [114]. Dataset 2 is available online (originally: https://github.com/kara-davis-lab/DDPR/releases, in this study: https://github.com/cytolab/RAPID). This dataset contained almost twice the number of patients (n=54) but less than half the number of total cells compared to Dataset 1 (48,600) because of a single patient with only 900 live, lineage-negative blast cells [114]. A total of 47 clusters were identified by RAPID, 3 of which were negative prognostic cell subsets that were associated with time to relapse (Figure 2.7). Importantly, features identified in the original publication as part of the signature associated with relapse (black text, Figure 2.7) were re-identified using RAPID. In the protein feature MEM values, enrichment of CD38 and CD34 was consistent with previously reported trends in pre-pro B cell-like phenotypes in BCP-ALL. Most notably, the signaling features p-S6, p-SYK, and p-4EBP1, which were important features positively associated with relapse in the DDPR model, were enriched in the negative prognostic populations identified by RAPID. Thus, RAPID was able to identify cells and features associated with time to relapse in another disease setting, generating a signature of negative-prognostic cells consistent with the original findings by another research group.

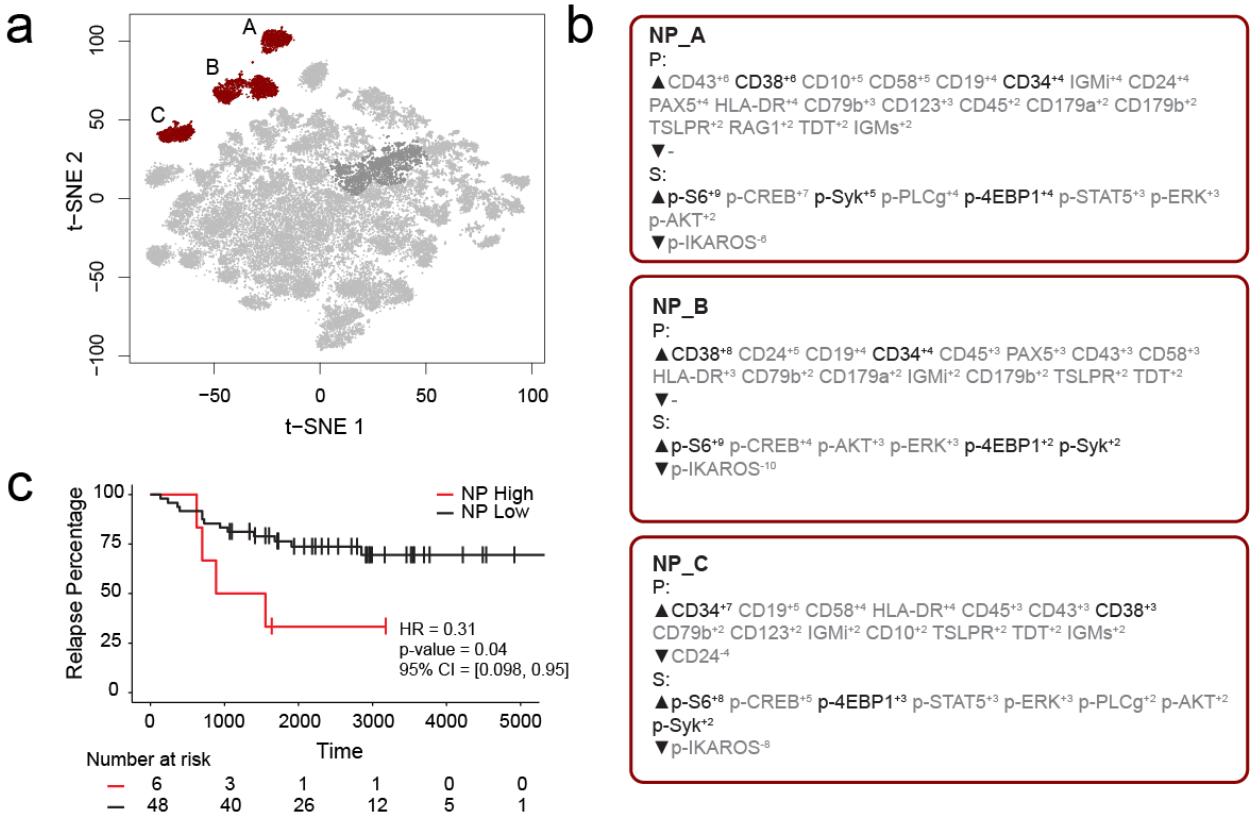


Figure 2.7: **RAPID analysis of a published B-cell leukemia dataset to identify negative prognostic cell subsets.** (a) t-SNE plot of 54 B-cell leukemia patient samples with negative prognostic populations (A, B, C) colored in red. (b) MEM labels for three negative prognostic cell subsets (NP_A, NP_B, NP_C). Features important in the original discovery of predictors of relapse are colored in black. (c) Kaplan-Meier Curve comparing time to relapse in patients with high abundance of negative prognostic cells (identified by RAPID) to patients with low abundance of negative prognostic cells.

Statistical validation 1: Clusters identified by RAPID were statistically robust

To determine the stability of the clusters identified by RAPID, 99 additional runs of FlowSOM were performed within the RAPID workflow (Figure 2.1c). Due to the stochastic nature of FlowSOM, the clusters identified in each subsequent run could contain different cells. For each of the clusters, an F-measure was calculated, based on the accuracy of cell assignment within a cluster in subsequent iterations of FlowSOM (see Methods, Table 2.2). Of the original 43 clusters, five had an average F-measure of less than 0.5 (average F-measure of all clusters = 0.75). These

five clusters, including cluster 33, previously identified as a GNP cluster, were considered unstable and were not included in subsequent analyses (indicated by shading in Figures 2.1 and 2.3, and asterisks in Figure 2.4 and Table 2.2).

Statistical validation 2: Clusters identified by RAPID were not dependent on individual patients or sub-samplings

A key design decision in RAPID was the use of an equal number of cell events from each patient to avoid tumors disproportionately impacting the analysis based on the number of cells collected. However, this decision limits a given RAPID analysis run to a number of cells equal to the smallest collected from any one patient. For the tumors studied here, the number of live glioblastoma cells ranged from 4,710 to 330,000 cells per patient. To test whether the cells subsampled for RAPID were representative of the total tumor sample and eliminate the possibility that randomly subsampled cells from larger samples are not representative, 9 additional t-SNE analyses were generated, each with a different sample of 4,710 cells selected at random, with replacement, from each patient (Figure 2.8). Each of these 9 t-SNE projections were then used in a new RAPID analysis, creating 10 total analyses (the original and 9 new tests). Of these, a total of 55 clusters from the 10 runs were considered stable (F-measure >0.5) and prognostic (see Methods, Figures 2.8 and 2.9). An F-measure could not be calculated on a cell-by-cell basis because the cells varied between analyses, but the average F-measure based on patient categorization (GNP-high, GPP-high, and GNP and GPP low) was 0.79 between t-SNE runs.

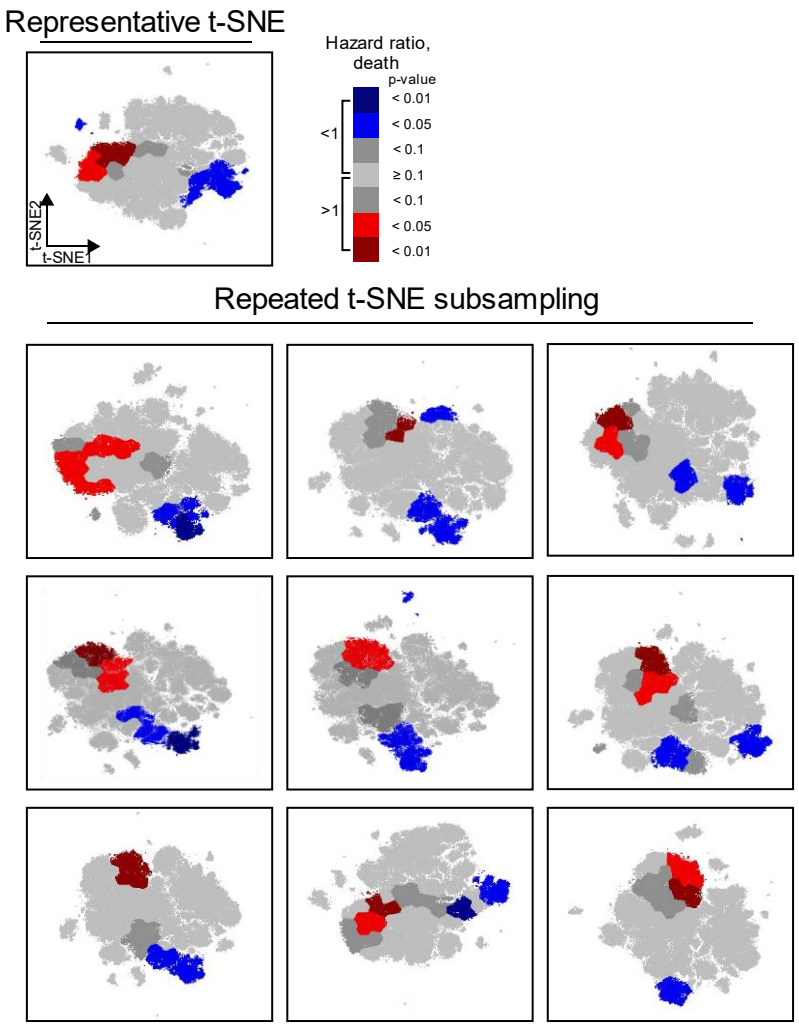


Figure 2.8: **GNP and GPP subsets are identified in repeated cell subsampling and t-SNE analyses.** The representative t-SNE is shown on top (contributed to Figure 2.1, 2.3, 2.4, 2.5, and 2.13). Nine additional t-SNE plots are generated by subsampling the tumors, with replacement, for 4,710 cells (contributed to Figure 2.9). GNP clusters are shown in red colors and GPP clusters are shown in blue, as noted in the heat scale.

To quantify the degree of similarity between the 47 newly identified prognostic clusters and the 8 representative GNP (34, 37, 42) and GPP (2, 3, 4, 5, 41) clusters, the root-mean-square deviation (RMSD) in the MEM enrichment values was calculated [158, 170]. GNP subsets from subsequent runs were highly similar to the GNP subsets identified by the initial analysis described

above, and the same was observed for GPP subsets (Figure 2.9; GNP v GNP average RMSD = 92.8, GPP v GPP average RMSD = 88.9, and GNP v GPP average RMSD = 80.9). However, some phenotypes were only observed in a small number of t-SNE runs. For example, the phenotype representing cluster 41 was only seen in one other t-SNE. Because this cell type was not observed in at least 50% of the cell sub-samplings, it was considered phenotypically unstable and removed from subsequent analyses (indicated by shading in Figures 2.1 and 2.3, and asterisks in Figure 2.4 and Table 2.2).

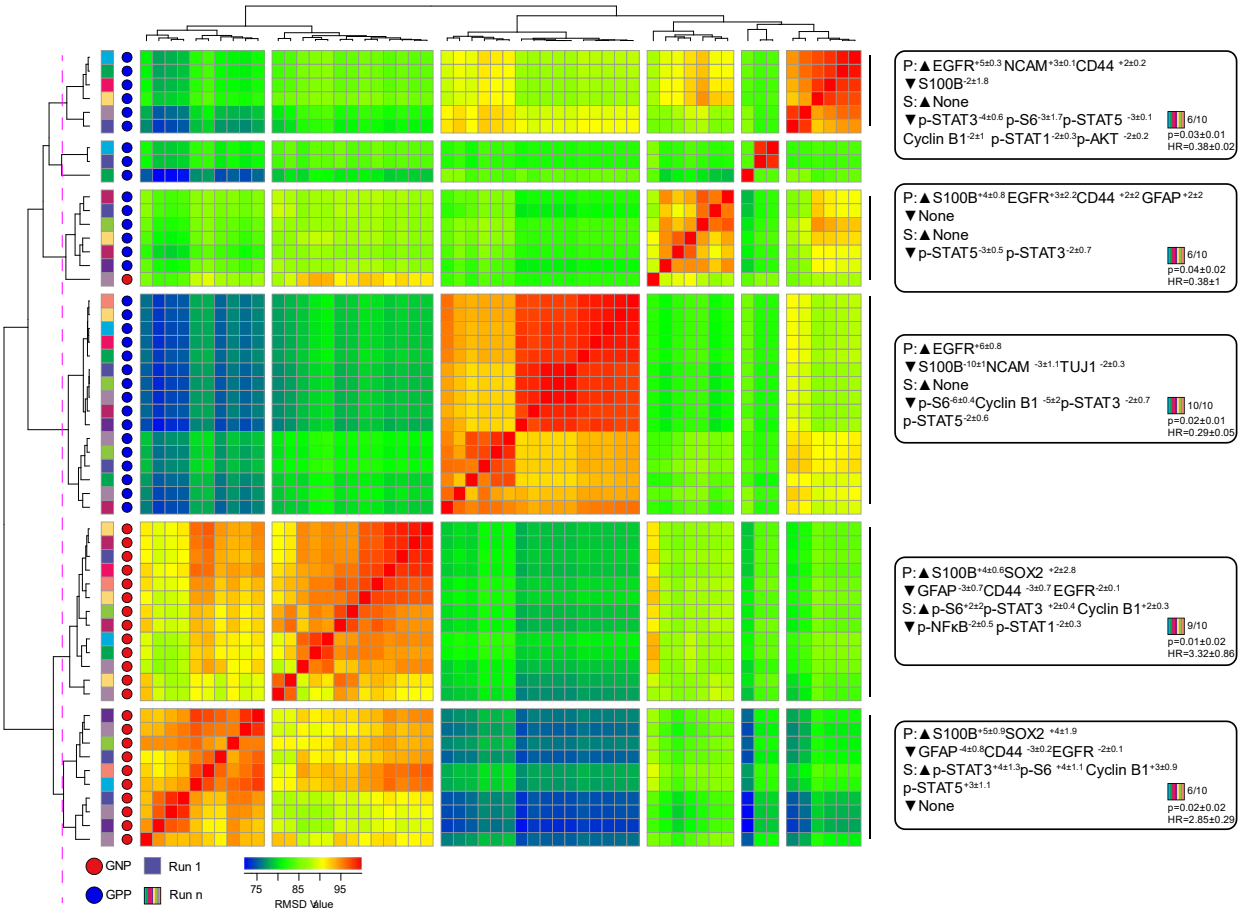


Figure 2.9**. Subsampling of glioblastoma cells repeatedly resulted in GNP and GPP subsets with similar phenotypes.** RMSD map comparing MEM scores for stable GNP and GPP subsets identified in the main figures and from nine additional t-SNE runs. GNP subsets are noted by red

circles and GPP subsets are noted by blue circles. Colored boxes to the left of the red or blue circles indicate the t-SNE run from which the subset is derived. Median MEM labels (± standard deviation) are shown for five major populations to the right. The number of t-SNE analyses represented in each group, as well as median p-value and hazard ratio (HR) are noted in the bottom right corner of each MEM label.

Statistical validation 3: Comparable clusters were identified by RAPID using UMAP instead of t-SNE

To test the modularity of RAPID, the algorithm was implemented using different dimensionality reduction values as input parameters, replacing t-SNE with UMAP, a tool that emphasizes both local and global data structure [138]. RAPID identified 31 populations using UMAP input; 4 of these were prognostic and significantly associated with OS (1 GNP_{UMAP} and 3 GPP_{UMAP}) (Figure 2.10). GNP_{UMAP} MEM scores reflected the characteristic S100B and SOX2 co-expression observed in the GNP populations along with an active pro-survival basal signaling status. GPP_{UMAP} subsets were similarly defined by co-expression of EGFR and CD44 and a general lack of the measured phosphorylated signaling effectors (Figure 2.10). When the cells identified using t-SNE were overlaid on the UMAP axes, they occupied similar phenotypic space as UMAP-identified clusters, and vice versa (F-measure for cell assignment to GNP, GPP, or neither = 0.87, Figure 2.10). Thus, when UMAP was used in the RAPID algorithm, GNP and GPP populations were identified that had comparable phenotypes to those identified previously in t-SNE analyses, confirming that RAPID is not dependent upon a specific dimensionality reduction tool (Figure 2.10).

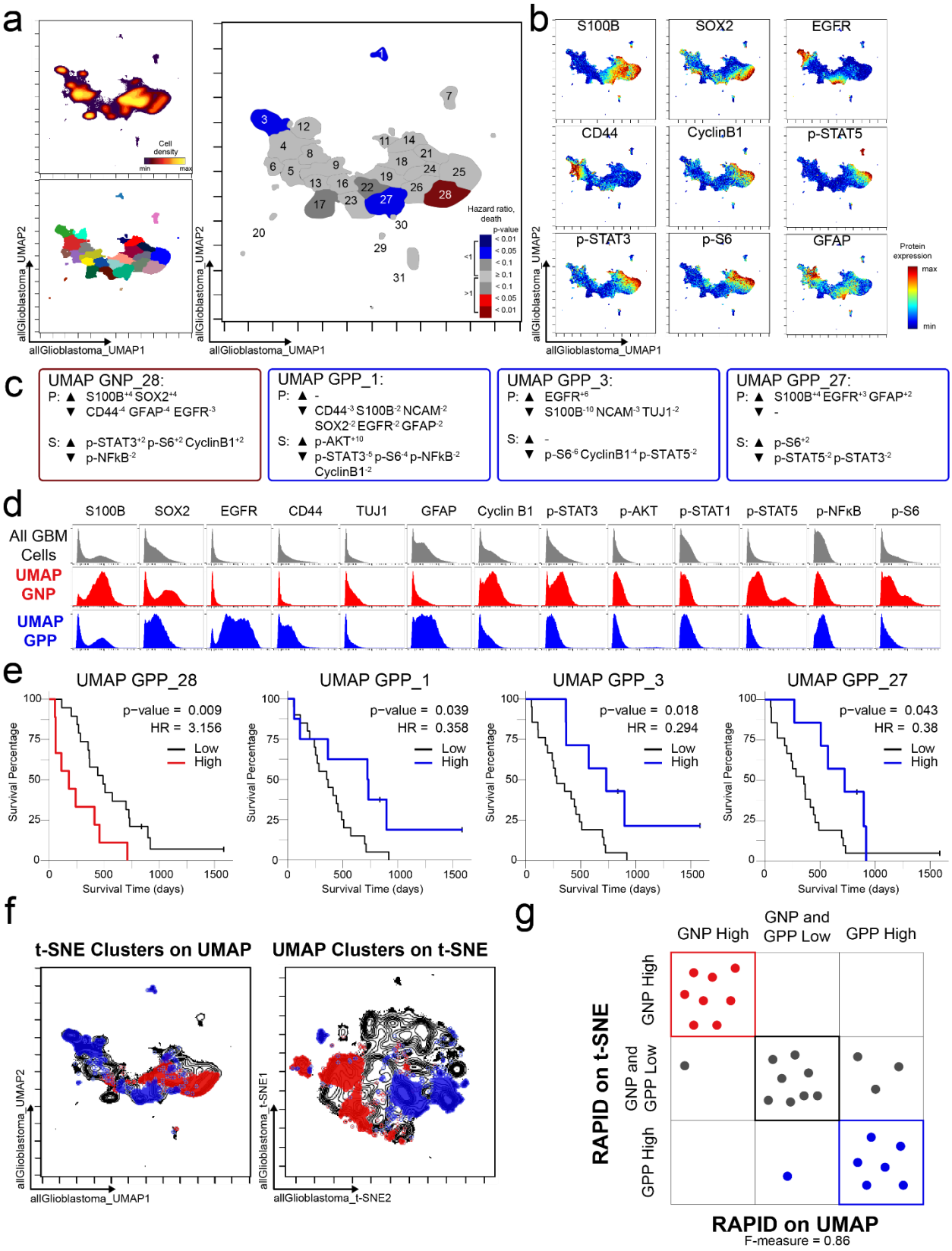

a
Cell density
min
max
allGlioblastoma_UMAP2
allGlioblastoma_UMAP1
Hazard ratio, death
p-value
< 0.01
< 0.05
< 0.1
≥ 0.1
< 0.1
< 0.05
< 0.01
<1
>1
b
S100B
SOX2
EGFR
CD44
CyclinB1
p-STAT5
p-STAT3
p-S6
GFAP
Protein expression
max
min
c
UMAP GNP_28:
P: ▲ S100B+4 SOX2+4
▼ CD44-4 GFAP-4 EGFR-3
S: ▲ p-STAT3+2 p-S6+2 CyclinB1+2
▼ p-NFkB-2
UMAP GPP_1:
P: ▲ -
▼ CD44-3 S100B-2 NCAM-2 SOX2-2 EGFR-2 GFAP-2
S: ▲ p-AKT+10
▼ p-STAT3-5 p-S6-4 p-NFkB-2 CyclinB1-2
UMAP GPP_3:
P: ▲ EGFR+6
▼ S100B-10 NCAM-3 TUJ1-2
S: ▲ -
▼ p-S6-6 CyclinB1-4 p-STAT5-2
UMAP GPP_27:
P: ▲ S100B+4 EGFR+3 GFAP+2
▼ -
S: ▲ p-S6+2
▼ p-STAT5-2 p-STAT3-2
d
S100B
SOX2
EGFR
CD44
TUJ1
GFAP
Cyclin B1
p-STAT3
p-AKT
p-STAT1
p-STAT5
p-NFκB
p-S6
All GBM Cells
UMAP GNP
UMAP GPP
e
UMAP GPP_28
p-value = 0.009
HR = 3.156
UMAP GPP_1
p-value = 0.039
HR = 0.358
UMAP GPP_3
p-value = 0.018
HR = 0.294
UMAP GPP_27
p-value = 0.043
HR = 0.38
Low
High
Survival Percentage
Survival Time (days)
f
t-SNE Clusters on UMAP
UMAP Clusters on t-SNE
allGlioblastoma_UMAP2
allGlioblastoma_UMAP1
allGlioblastoma_t-SNE1
allGlioblastoma_t-SNE2
g
GNP High
GNP and GPP Low
GPP High
RAPID on t-SNE
RAPID on UMAP
F-measure = 0.86

Figure 2.10**: GNP and GPP cells were also identified using dimensionality reduction tool UMAP in the RAPID algorithm.** (a) UMAP analysis of 131,880 cells from 28 patients. Upper left plot - heat on cell density; lower left plot – colored by FlowSOM cluster; right plot – colored by GNP (red)/GPP (blue) designation and p-value. (b) Per-cell expression levels of 5 identity proteins, 3 phosphorylated signaling effectors, and proliferation marker cyclin B1 are depicted. (c) Enrichment of identity proteins (P) and phosphorylated signaling effectors (S) of glioblastoma cell subsets was quantified using MEM. GNP and GPP cells are labeled in red and blue, respectively. (d) Histogram analysis depicts the expression of key identity proteins and phosphorylation signaling effectors of GNP (red) and GPP (blue) compared to all glioblastoma (GBM) cells (gray, top row). (e) Overall survival curves for four UMAP-identified populations associated with survival. Cox-proportional hazard model was used to determine a hazard ratio (HR) of death. Censored patients are indicated by vertical ticks. (f) GNP (red) and GPP (blue) cells identified via t-SNE ("t-SNE GNP" or "t-SNE GPP") and UMAP ("UMAP GNP" or "UMAP GPP") are overlaid on either UMAP or t-SNE axes. (g) Categorization of each patient (dots) based on GNP high (red), GPP high (blue), or neither (gray) according to abundance based on RAPID using t-SNE or RAPID using UMAP (F-measure = 0.86).

To determine the stability of the clusters identified by RAPID on UMAP, 99 additional runs of FlowSOM were performed within the RAPID workflow as described in Statistical Validation 1. For each of the clusters an F-measure was calculated (see Methods), and of the original 31 clusters, two had an average F-measure of less than 0.5 (average F-measure of all clusters = 0.77). All four of the GNP or GPP clusters identified were stable as determined by an F-measure of greater than 0.5.

Like the t-SNE analyses, UMAP analyses were performed using an equal number of cell events. As in Statistical Validation 2, 9 additional UMAP analyses were generated, each with a different sample of 4,710 cells selected at random, with replacement, from each patient. Each of these 9 UMAP projections were then used in a new RAPID analysis, creating 10 total analyses (the original and 9 new tests). Of these, a total of 16 clusters from the 10 runs were considered stable (F-measure >0.5) and prognostic (see Methods, Figure 2.11). To quantify the degree of similarity between the 12 newly identified prognostic clusters and the 4 representative GNP (28) and GPP (1, 3, 27) clusters, the root-mean-square deviation (RMSD) in the MEM enrichment values was calculated [158, 170]. GNP subsets from subsequent runs were highly similar to the GNP subsets identified by the initial analysis described above, and the same was observed for

GPP subsets (Figure 2.11: GNP v GNP average RMSD = 95.6, GPP v GPP average RMSD = 88.8, and GNP v GPP average RMSD = 80.6). However, like the results observed using t-SNE, some phenotypes were only observed in a small number of UMAP runs.

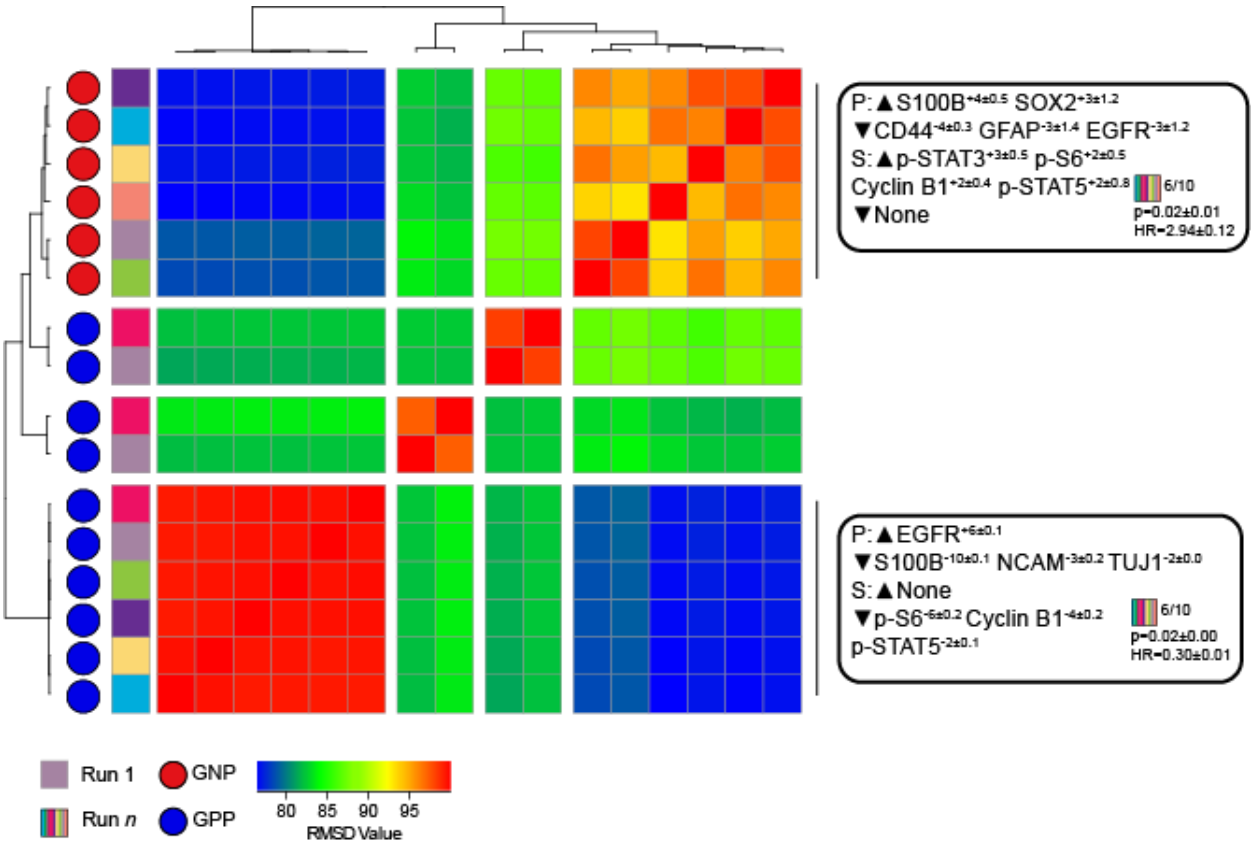


Figure 2.11: **Repeated cell subsampling of UMAP analyses resulted in GNP and GPP subsets with similar phenotypes.** RMSD map comparing MEM scores for stable GNP and GPP subsets identified in Figure 2.9 and from nine additional UMAP runs. GNP subsets are noted by red circles and GPP subsets are noted by blue circles. Colored boxes to the left of the red or blue circles indicate the UMAP run from which the subset is derived. Median MEM labels (± standard deviation) are shown for two major populations to the right. The number of UMAP analyses represented in each group, as well as median p-value and hazard ratio (HR) are noted in the bottom right corner of each MEM label.

Although in our experience, mass cytometry data analysis benefits from a dimensionality reduction step, this is not required for the implementation of RAPID. In the absence of a such a step, the input parameters for FlowSOM are the features measured (in this case, the 24 markers used to generate the t-SNE or UMAP plots described above). When RAPID was performed without initial dimensionality reduction, 32 FlowSOM clusters were identified. Of these, two were significantly associated with patient OS, one GNP and one GPP. These clusters had the characteristic phenotypes similar to the GNP and GPP clusters identified using t-SNE or UMAP (Figure 2.12). The GNP cluster (GNP_Original Markers) was enriched for $SOX2^{+5}$ and $S100B^{+4}$ as well as the signaling features p-$STAT3^{+3}$, p-$STAT5^{+2}$, p-$S6^{+2}$ and the proliferation marker Cyclin $B1^{+2}$ (Figure 2.12). This cluster was also specifically lacking $GFAP^{-4}$, $CD44^{-4}$, and $EGFR^{-2}$. The GPP cluster (GPP_Original Markers) had enriched $EGFR^{+6}$ while many signaling features were absent (Figure 2.12).

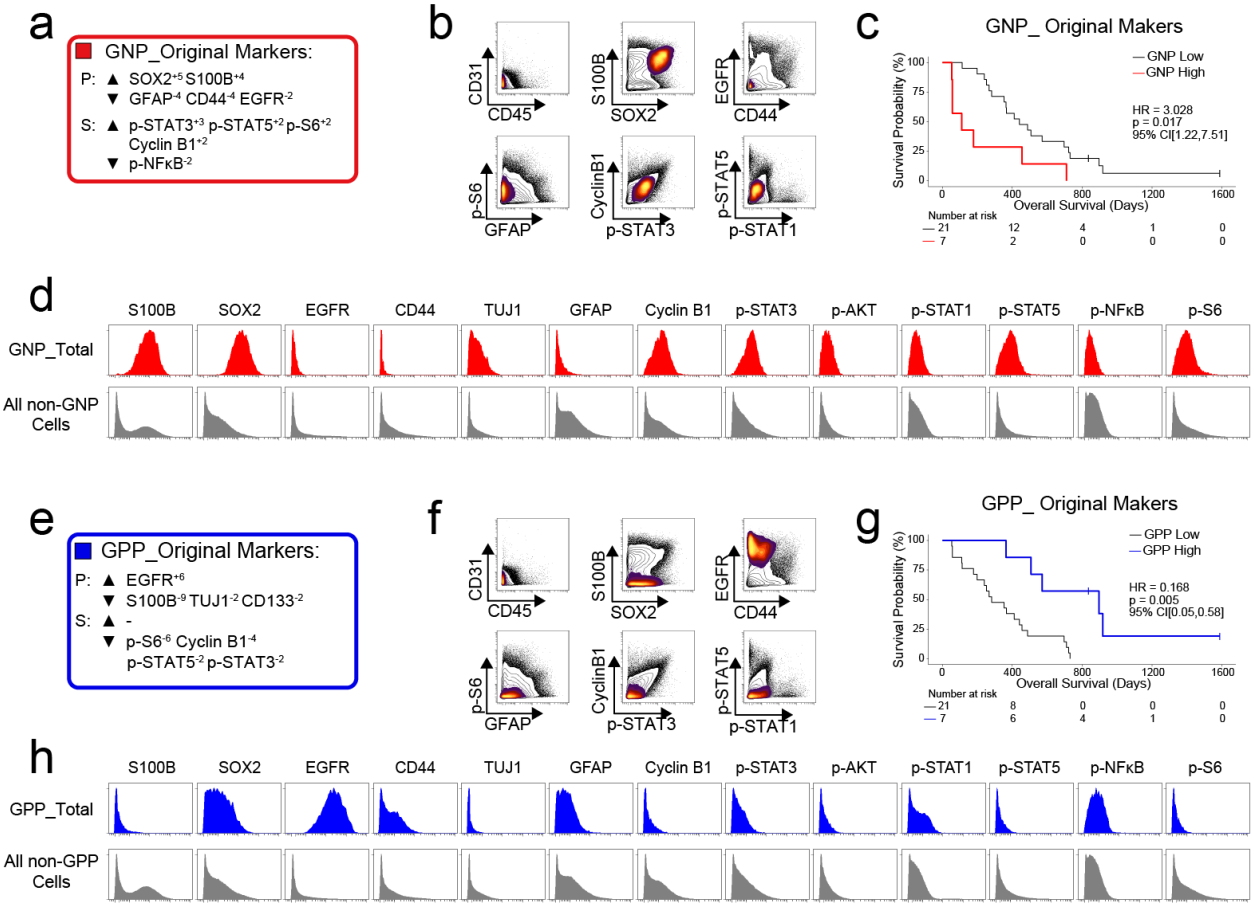


Figure 2.12: **GNP and GPP cells identified using RAPID without a dimensionality reduction step stratifies patients**. (a) Enrichment (upwards arrowhead) or lack (downwards arrowhead) of identity proteins (P) and phosphorylated signaling effectors (S) on the Glioblastoma Negative Prognostic cell subset was quantified using MEM. MEM scores are shown for the GNP subset. (b) The GNP cell subset (density contours) was mapped over biaxial plots of all other tumor cells (black contours). (c) Overall survival of patients for high GNP content compared to patients with low GNP content. (d) Histogram plots of GNP cells (red) and all other glioblastoma cells (gray) illustrate the expression of identity proteins and phosphorylated signaling effectors. (e) Enrichment (upwards arrowhead) or lack (downwards arrowhead) of identity proteins (P) and phosphorylated signaling effectors (S) on the Glioblastoma Positive Prognostic cell subset was quantified using MEM. MEM scores are shown for the GNP subset. (f) The GPP cell subset (density contours) was mapped over biaxial plots of all other tumor cells (black contours). (g) Overall survival of patients for high GPP content compared to patients with low GPP content. (h) Histogram plots of the GPP cell subset (blue) and all other glioblastoma cells (gray) illustrate the expression of proteins and phosphorylated signaling effectors.

Statistical validation 4: Risk stratifying cells were continuously associated with outcomes and independent of other glioblastoma stratifying features

At the conclusion of the RAPID analysis, to ensure that results were not an artifact of the high-low cut point choice and to determine if the effect of cell subset abundance was continuous and independent of other features known to stratify glioblastoma survival, a multivariate Cox proportional-hazards model analysis was performed incorporating known predictive features and GNP or GPP cell abundance. The included known predictors were age [171, 172], O^6-methylguanine DNA methyltransferase (*MGMT)* promoter methylation status [10, 173], and treatment variables including the extent of surgical resection [174, 175], therapy with temozolomide [2], and radiation [176, 177]. Multivariate survival analysis of GNP cell abundance on a continuous scale, keeping the other predictors constant, indicated that each 1% increase in GNP cells was associated with an approximately 7% increase in mortality compared to baseline (OS HR=1.07 [95% CI 1.02-1.12], p=0.003). Similarly, a 1% increase in GPP cells was associated with an approximately 7% decrease in mortality rate (OS HR=0.93 [0.87-1.0], p=0.05) and an approximately 4% increase in time to tumor progression, as compared to baseline (PFS HR=0.96 [0.93-0.998], p=0.04). When GNP and GPP were assessed simultaneously, abundance of GNP cells was the primary predictor of mortality (OS HR=1.05 [1.00-1.10], p=0.04), while abundance of GPP cells was the primary predictor of time to tumor progression (PFS HR =0.96 [0.92-1.00]; p=0.03). Thus, the abundances of GNP and GPP cell subsets were associated with distinct and contrasting patient outcomes (Figure 2.5), and their predictive value was independent of each other and known prognostic factors of patient survival.

Since assessing progression-free survival (PFS) can be especially useful in the clinic for cancers with longer median survival, RAPID was also used for the identification of glioblastoma cell clusters with differential PFS, as opposed to OS. Of the 43 subsets identified by RAPID, 4 subsets were significantly associated with PFS (subsets 20, 33, and 43 with unfavorable PFS (GNP_{PFS}) and subset 3 was associated with favorable PFS (GPP_{PFS}), Figure 2.13).

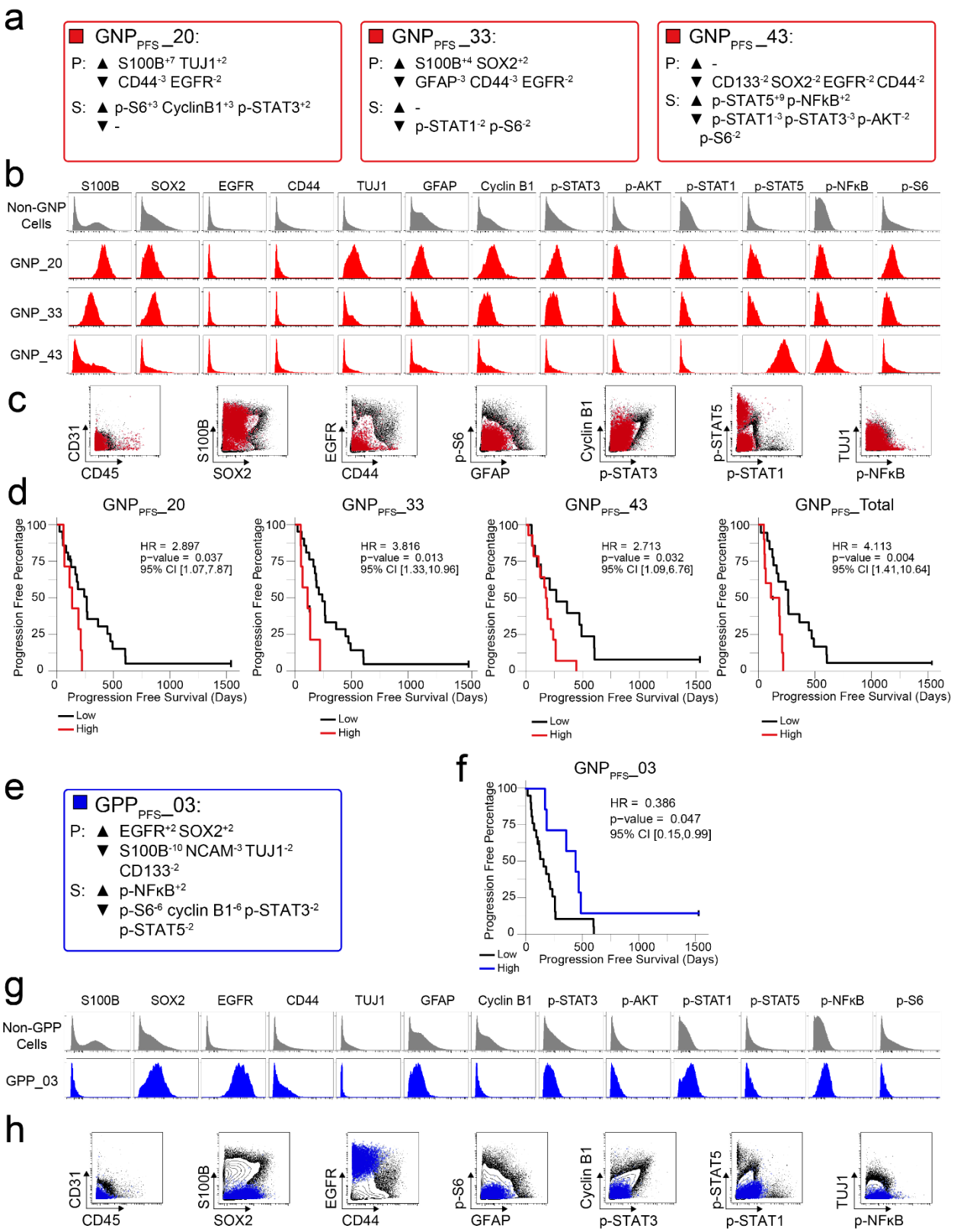

a
GNP_PFS_20:
P: ▲ S100B+7 TUJ1+2
▼ CD44-3 EGFR-2
S: ▲ p-S6+3 CyclinB1+3 p-STAT3+2
▼ -
GNP_PFS_33:
P: ▲ S100B+4 SOX2+2
▼ GFAP-3 CD44-3 EGFR-2
S: ▲ -
▼ p-STAT1-2 p-S6-2
GNP_PFS_43:
P: ▲ -
▼ CD133-2 SOX2-2 EGFR-2 CD44-2
S: ▲ p-STAT5+9 p-NFκB+2
▼ p-STAT1-3 p-STAT3-3 p-AKT-2 p-S6-2
b
S100B SOX2 EGFR CD44 TUJ1 GFAP Cyclin B1 p-STAT3 p-AKT p-STAT1 p-STAT5 p-NFκB p-S6
Non-GNP Cells
GNP_20
GNP_33
GNP_43
c
CD31 CD45
S100B SOX2
EGFR CD44
p-S6 GFAP
Cyclin B1 p-STAT3
p-STAT5 p-STAT1
TUJ1 p-NFκB
d
GNP_PFS_20
HR = 2.897
p-value = 0.037
95% CI [1.07,7.87]
GNP_PFS_33
HR = 3.816
p-value = 0.013
95% CI [1.33,10.96]
GNP_PFS_43
HR = 2.713
p-value = 0.032
95% CI [1.09,6.76]
GNP_PFS_Total
HR = 4.113
p-value = 0.004
95% CI [1.41,10.64]
Progression Free Percentage
Progression Free Survival (Days)
Low
High
e
GPP_PFS_03:
P: ▲ EGFR+2 SOX2+2
▼ S100B-10 NCAM-3 TUJ1-2 CD133-2
S: ▲ p-NFκB+2
▼ p-S6-6 cyclin B1-6 p-STAT3-2 p-STAT5-2
f
GNP_PFS_03
HR = 0.386
p-value = 0.047
95% CI [0.15,0.99]
g
S100B SOX2 EGFR CD44 TUJ1 GFAP Cyclin B1 p-STAT3 p-AKT p-STAT1 p-STAT5 p-NFκB p-S6
Non-GPP Cells
GPP_03
h
CD31 CD45
S100B SOX2
EGFR CD44
p-S6 GFAP
Cyclin B1 p-STAT3
p-STAT5 p-STAT1
TUJ1 p-NFκB

Figure 2.13: **RAPID identified four populations associated with time to disease progression**. (a) Enrichment of identity proteins (P) and phosphorylated signaling effectors (S) of GNP cell subsets revealed by analysis of disease progression (GNP_{PFS}) was quantified using MEM. (b) Histogram plots of each GNP_{PFS} cell subset (red) and all other glioblastoma cells (gray) illustrate the expression of proteins and phosphorylated signaling effectors. (c) Combined GNP_{PFS} cell subsets (red circles) were mapped over biaxial plots of all other tumor cells (black contours). (d) For each subset, PFS was compared between patients with high vs low cell abundance (see Methods). (e) Enrichment of identity proteins (P) and phosphorylated signaling effectors (S) of the GPP_{PFS} cell subset was quantified using MEM. (f) PFS was compared between patients with high vs low GPP_{PFS} cell abundance (g) Histogram plots of the GPP_{PFS} cell subset (blue) and all other glioblastoma cells (gray) illustrate the expression of proteins and phosphorylated signaling effectors. (h) The GPP_{PFS} cell subset (blue circles) was mapped over biaxial plots of all other tumor cells (black contours).

Tumors are mosaics of multiple subsets, but number of subsets does not correlate with outcome

In the representative t-SNE run (Figure 2.1), RAPID identified 43 phenotypically distinct glioblastoma cell subsets within the tumors analyzed by mass cytometry in this study (Figure 2.1 and 2.5). The abundance of the 43 clusters varied extensively across patients (Table 2.2). Tumors contained a median of 14 clusters at >1% with a range from 5 cell clusters in LC06 to a maximum of 27 cell clusters represented in LC25 (Table 2.2, additional data in Appendix B). Although intra-tumor diversity has been hypothesized to contribute to poor response to treatment and survival, here, the number of glioblastoma cell clusters present within a tumor at >1% abundance (a surrogate for intra-tumor diversity) was not observed to be associated with differential survival (ρ=0.047, p=0.812). In contrast, the abundance of each of the 7 stable and prognostic glioblastoma cell clusters was closely correlated with overall survival (Figure 2.5).

Biological validation 1: A transparent algorithm enables creation of a simple cell identification strategy that captures the cells identified in Dataset 1

After patterns are recognized by a machine learning approach, it is useful to learn from key features and create a straightforward test using alternative technologies or simpler models.

One such model is a decision tree using one- or two-dimensional cytometry gating [157], consistent with traditional strategies in immunology and hematopathology. The GNP and GPP MEM labels generated using mass cytometry data were therefore used to design a two-dimensional prognostic strategy based on S100B and EGFR expression (Figure 2.14). Using only these two proteins, patients could be grouped as GNP-like, GPP-like or GNP and GPP Low, and these groups again exhibited stratified clinical outcomes (HR=6.56, GNP-like median OS = 111.5 days, GPP-like median OS = 896 days, Figure 2.14). Thus, a simple gating model based on the two most divergent features identified by RAPID was able to meaningfully separate patients into clinically distinct groups.

Biological validation 2: A larger cohort of glioblastoma samples was stratified using IHC based on phenotypes discovered by RAPID

Unlike fluorescence or mass flow cytometry, IHC is routinely used in surgical pathology. To confirm the ability of S100B and EGFR in separating clinically distinct patient populations using an orthogonal approach, a tissue microarray (TMA) of 73 glioblastoma patient samples was developed. Serial TMA sections were stained with antibodies against S100B and EGFR and the overall signal intensity was determined using QuPath software for each feature (see Methods). By comparing S100B and EGFR staining intensity, patients were scored as GNP-like, GPP-like, or GNP and GPP Low (Figure 2.14). A Kaplan-Meier analysis comparing overall survival between patients enriched with GNP-like cells to those with GPP-like cells confirmed that GNP-like cell enrichment is associated with a shorter overall survival (HR=2.3, GNP-like median OS = 298 days, GPP-like median OS = 560 days, Figure 2.14). These results validated the suspension mass cytometry findings and demonstrated that once revealed by RAPID, GNP-like and GPP-like cells could be identified in new samples by complementary approaches used in laboratory and clinical settings.

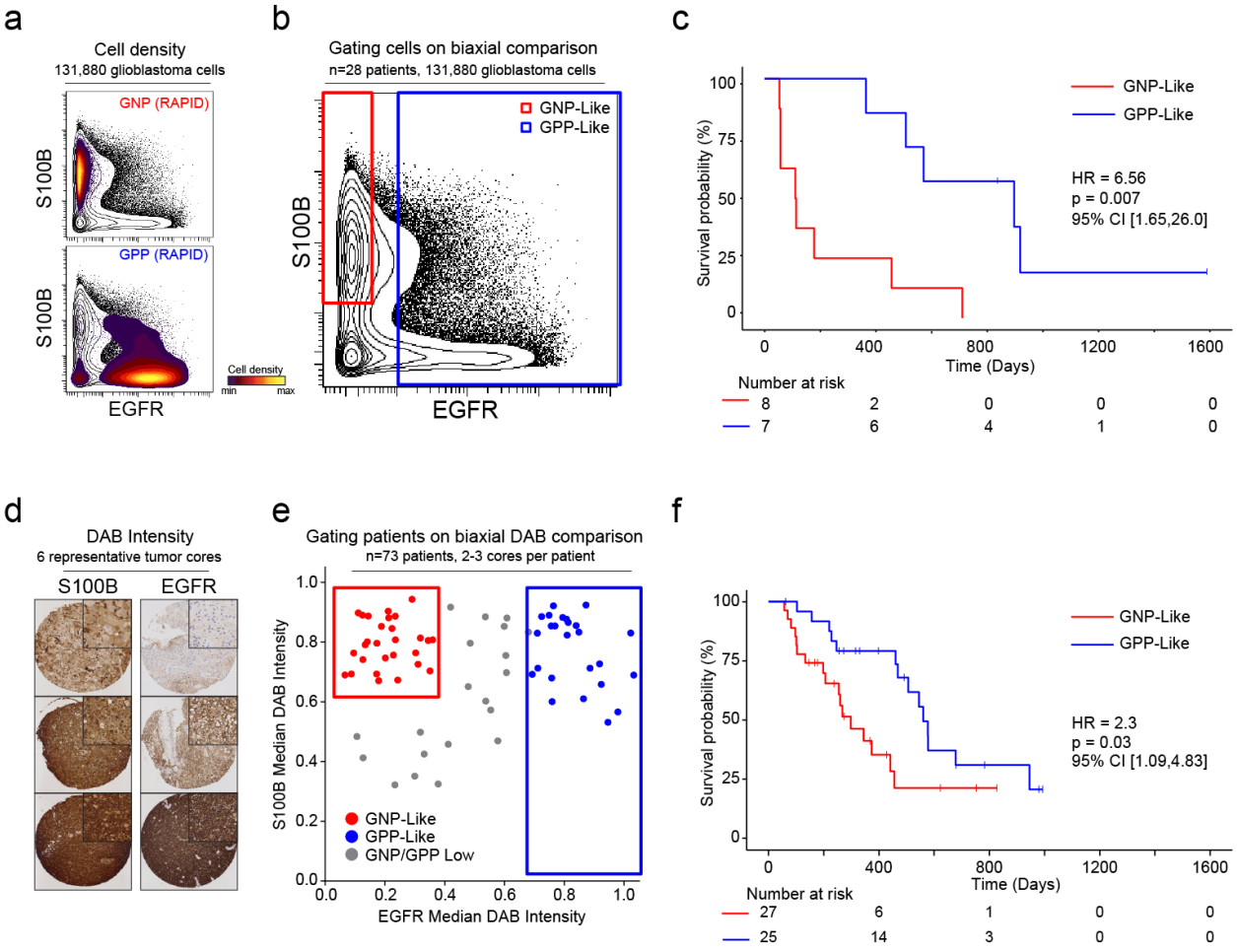


Figure 2.14: **A simple gating strategy based on S100B and EGFR can stratify patients using mass cytometry or immunohistochemistry data.** a) Biaxial plot of S100B (y-axis) and EGFR (x-axis). Gray contours depict all 131,880 cells from all patients. Density contour overlays depict GNP (top) or GPP (bottom) cells identified by the RAPID algorithm. b) Biaxial plot of S100B (y-axis) and EGFR (x-axis). Gray contours depict all 131,880 cells from all patients as in a). Red box indicates gate for $S100B^{+}/EGFR^{-}$ cells, called GNP-like. Blue box indicates gate for $EGFR^{+}$ cells, called GPP-like. c) Kaplan Meier curve comparing overall survival (in days) of patients with high percentages of GNP-like cells in red (red gate in a, > 65.7% =high) and patients with high percentages of GPP-like cells in blue (blue gate in a, >31.2% =high). The hazard ratio of death, calculated using a cox proportional hazards model, is 6.56 (p=0.0007). d) Example TMA cores stained for S100B (left) or EGFR (right). Brown signal is from 3,3′-Diaminobenzidine (DAB). e) Graph depicting DAB signal intensity for S100B (y-axis) or EGFR (x-axis) from tissue microarray immunohistochemistry on 73 glioblastoma patient samples. The red box outlines patients described as GNP-like ($S100B^{high}/EGFR^{low}$) and the blue box outlines patients designated GPP-like ($EGFR^{high}$). All other patients are shown in gray. f) A Kaplan-Meier curve showing overall survival (in days) of patients in the GNP-like (red) or GPP-like (blue) groups. The hazard ratio of death, calculated using a Cox proportional hazards model, is 2.3 (p-value = 0.03).

Discussion

The focus of this study was the creation of an unsupervised approach that could work with pilot datasets to suggest prognostic cell types for validation. The RAPID workflow automatically assigned single tumor cells into computational clusters based on phenotypic similarity, generated a quantitative phenotypic descriptor of each population, and determined the correlation between the abundance of populations and clinical outcomes. Ultimately, the algorithm was tested using numerous statistical approaches, validated with two datasets, and validated as revealing biologically robust cells detectable on other platforms in a larger follow up cohort with formalin-fixed, paraffin-embedded tissue. Prior workflows and algorithms were developed to identify cell populations of interest in cancer samples and emphasized supervised modeling, as with Citrus [127] and Cytofast [178], or comparison to known subsets, as with DDPR [114] and Phenograph [113]. These approaches could not be used with Dataset 1, either because they required a level of prior knowledge about non-malignant adult human brain cells which was not available, or because they required supervision using categorical outcomes, which are not always clearly delineated for continuous variables. Another advantage of RAPID is that it does not require a target cluster number, which is important when it is not known how many phenotypically distinct subsets will be observed in a given cancer type. Cell subsets in tumors can be challenging to manually annotate as they may reasonably be assigned to multiple known cell types, as was apparent here and in prior studies [33, 47]. RAPID is unsupervised, provides a quantitative label of features enriched in each cluster, and is modular, such that a variety of dimensionality reduction and clustering tools can be used. Currently, a user inputs raw data files (e.g., FCS files from cytometry platforms or equivalent data types from other platforms) and annotated patient survival data. RAPID outputs quantitatively described cell clusters and their significance with respect to patient outcome. While the focus of this study was cytometry data, the design is suitable to other single cell data types where clinical outcomes or similar continuous variables have been scored

for pilot cohorts, typically at least 25 individuals. Published datasets were not available for single-cell RNA-seq that matched the criteria for RAPID, including having thousands of cells per sample, more than 25 individuals with annotated clinical outcomes, and multiple features scored consistently for every cell. As single cell RNA-seq and imaging cytometry technologies advance, we anticipate RAPID will be useful for such datasets, especially given how widespread t-SNE, UMAP, and related approaches are within these fields.

The utility of RAPID includes its ability to identify stable, robust clusters that are independent of known prognosticators, provide users with opportunities to customize the workflow with a variety of tools, and inform subsequent studies on validation datasets or using different technologies. Here, RAPID was extensively probed for its performance in each of these areas. By repeated subsampling of each tumor and iterative FlowSOM analyses, clusters with consistent cell content and phenotypes observed in the majority of subsamplings were identified (Figure 2.9). Furthermore, these clusters were independently associated with continuous clinical variables - patient overall survival and progression-free survival. A subsequent, low dimensional decision tree applied to both mass cytometry data and a new set of patient samples stained via IHC was also able to stratify patients, suggesting that the biology learned from the high dimensional approach could be used to inform complementary approaches (Figure 2.14). Critically, RAPID was also used to analyze a dataset from different tissue in a different disease collected at a different institution, Dataset 2 in Figure 2.7 [114]. In this application of RAPID, features previously identified by the original authors to be associated with time to relapse were re-captured, identifying cellular phenotypes concordant with prior results without requiring the normal developmental trajectory reference used in the original analysis.

In this study, RAPID analysis of glioblastoma patient samples demonstrated a link between altered signaling and possible abnormal lineage programs in glioblastoma [179]. Within Dataset 1 analyzing 28 *IDH*-wild type pre-therapy glioblastoma patient samples, the RAPID workflow automatically uncovered two prognostic phenotypic signatures which were independent

of other known predictors of outcome. Glioblastoma Negative Prognostic (GNP) cells, characterized by enrichment for expression of S100B, SOX2, p-STAT3, and p-STAT5, were associated with decreased overall survival, while Glioblastoma Positive Prognostic (GPP) cells, characterized by co-enrichment of EGFR and CD44 proteins, were associated with longer overall survival. Once revealed in high-dimensional data, a simple gating scheme using S100B and EGFR could be used to stratify outcome in a separate, expanded set of samples using traditional pathological approaches. High-dimensional cytometry and RAPID were critical to revealing novel prognostic cells in glioblastoma data in two ways. First, assessment of a large number of cells per tumor – over 2 million viable single cells, with at least 4,710 glioblastoma cells from each patient – enabled the use of an unsupervised approach in the identification of rare, novel cell subsets across patients. Second, per-cell quantification of phosphorylated signaling effector proteins revealed potential mechanisms of tumor cell regulation that are not readily apparent in bulk tumor data, genomic analyses, or lower dimensional approaches such as one- to four-color imaging. Supervised analysis of single cell data has previously uncovered signaling events tied to patient survival in hematologic malignancies [113-115, 151], and a similar pattern was observed here. Critically, therapeutically targetable signaling events were identified as a signature of prognostic cell populations, suggesting potential novel therapeutic strategies for patients with these characteristics.

The GNP signature was defined by abnormal neural development features such as co-expression of stem cell transcription factor SOX2 and astrocyte lineage marker S100B [180, 181] and simultaneous high basal phosphorylation of multiple signaling effectors downstream of receptor tyrosine kinases reported to be important in tumor biology [182-187] (Figure 2.5). STAT3 and S6 phosphorylation, identified here in GNP cells, agreed with prior studies indicating the importance of p-STAT3 in T cell suppression [182] and mTOR-dependent signaling in tumor formation [186, 187]. STAT3 specifically has been described to drive a proneural to mesenchymal transition in glioma cells, with mesenchymal phenotypes generally associated with worse patient

outcomes [184, 185]. RAPID also uncovered a connection between p-STAT5 and glioblastoma outcome previously unidentified in primary patient samples. STAT5 signaling is required in development of many tissues to block apoptosis and drive cell cycle entry [150]; for example, p-STAT5 is an essential feature of negative prognostic acute myeloid leukemia signaling profiles [113, 151]. The signaling events of the negative and positive prognostic cells can now be studied in glioblastoma research models, such as patient xenografts and glioblastoma organoids [37, 188-191], using new combinations of targeted therapies, such as JAK inhibitors that target molecules upstream of STAT5 and STAT3, in combination with PI3K/mTOR pathway inhibitors, which will target molecules upstream of AKT and S6 signaling. In this way, new combinations of existing therapies may prove useful in targeting the signaling that defines the negative prognostic cells seen here.

It is also of note that while SOX2, a key regulator of early developmental state and adult neural stem cells, was enriched in some GNP cells, multiple other markers of stem and progenitor-like cells (Nestin, CD133, CD15, BMX, ITGA6/CD49f, CD117, SOX10, L1CAM) were measured in this study and were not specifically associated with poor outcome, distinct from studies proposing a “stemness” transcriptional signature driving aggressive tumors. Recent work using single cell gene expression has described the existence of multiple cellular states in glioblastoma tumors and the ability of cells to transition between states [47]. Similar to most transcript-based studies, RAPID analyses were performed on cells collected at a single timepoint, precluding a direct investigation of the ability of GNP or GPP cells to transition to other phenotypes; however, it is possible that phosphorylated, active STAT3, STAT5, and S6 may enable transition between progenitor-like states as they do in earlier development, and thus influence patient outcome [192, 193]. Isolation of GNP and GPP cells from primary samples, or the establishment of gliomasphere lines from GNP-high and GPP-high tumors, will be invaluable in probing the stability of these phenotypic states, the associated gene expression modules, and functional importance of each cell type (i.e. sphere formation, tumorigenic potential, and response to therapies). Another key

research question for the future will be whether the signature features of the risk stratifying cells seen here will also be seen in other types of intractable human malignancies. Intriguingly, p-STAT5 and p-STAT3 signaling profiles reminiscent of the negative prognostic cells from glioblastoma have been seen in leukemia [113, 151, 160] and ovarian cancer [117].

The GPP signature, in contrast, was defined by EGFR and CD44 co-enrichment, diminished evidence of proliferation, and specific lack of STAT5 phosphorylation. GPP cells were further associated with higher proportions of tumor-infiltrating immune cells. This result suggests an understanding of prognostic cell content or biomarkers may be relevant for immunotherapy research in glioblastoma. Previous DNA and RNA-driven molecular subtyping predicts EGFR expression in the classical subset of glioblastoma tumors and CD44 expression in mesenchymal tumors [28]. As these categories were primarily based on bulk tumor data, cells co-expressing EGFR and CD44 (classified as GPP cells in this study) may have previously been missed, although single glioma cells have been shown to simultaneously amplify sequence or co-express transcripts for important signaling regulators [33, 35]. EGFR has been extensively studied as a driver of gliomas in the past (reviewed in [194]), and the association of this gene and transcript with outcome has been a matter of debate [194-196]. Genetically, glioblastomas commonly have amplified EGFR [28, 164]; however, we noted examples of tumors with robust EGFR copy number amplification that contained both high and low percentages of GPP cells (data not shown), highlighting the importance of measuring protein expression in addition to genomic content. Although EGFR signaling has been linked with increased p-S6, through mTOR, and p-STAT3/5 in tumor models, these associations were not observed in the GNP or GPP subsets [197, 198]. Instead, some GPP cells showed enrichment of p-NFκB (Figure 2.5), a transcription factor that activates pro-apoptotic programs and DNA damage repair programs [184, 199, 200].

This study finds that expression of EGFR protein is associated with better overall survival. One reason for the difference between this study and other reports may be that EGFR protein levels were measured in individual cells rather than copy number analysis or transcript levels in

bulk tumor samples; our own analyses and others' have indicated that copy number or transcript level are not necessarily predictive of protein expression [31, 201, 202]. Although antibody-based methods for protein detection, like those used here, depend on the specificity of each selected clone, it is important to note that two different, rigorously validated antibodies (mass cytometry, clone AY13; TMA, clone A-10) gave the same results (Figure 2.14). S100B has been explored as a serum biomarker [203], and S100B is known for its impact on macrophages, including microglia [204]. These features of negative and positive prognostic cells extend the single cell phospho-specific flow cytometry approach to a new solid tumor that is in urgent need of new biological insights and targets.

Recent studies have revealed significant variation in immune cell abundance and relative proportions of immune cell subsets across glioblastomas [32, 100]. Here, unfavorable GNP cells were associated with diminished tumor-infiltrating immune cells and GPP cells were associated with higher proportions of immune cells in the tumor microenvironment. These results invite the question of whether an altered immune microenvironment precedes development of an aggressive glioblastoma or whether more aggressive tumors suppress anti-tumor immunity. These findings argue that immunotherapy is likely to be more efficacious in tumors containing GPP cells, but additional research is needed to understand whether GNP cells directly suppress microglia or immigrant leukocytes.

When applied to a new glioblastoma dataset as well as a previously published study of blood cancer, RAPID reliably identified cells whose abundance was predictive of good or poor outcome. Cellular identification was robust, stable, and reproducible, and independent of the specific dimensionality reduction tools used. Critically, the discoveries from RAPID were able to inform a scoring system for detection of GNP-like and GPP-like phenotypes in IHC data that stratified patient outcome in 73 patient samples. RAPID also led to the development of a lower-dimensional cytometry pipeline which could be optimized for clinical stratification. There is now the exciting potential to extend the hypotheses suggested by RAPID into clinical research studies

using either traditional flow cytometry or IHC on widely available formalin-fixed, paraffin-embedded samples, as in the biological validation here (Figure 2.14). Thus, techniques accessible to clinical research, such as IHC, could be informed by the results from RAPID and envisioned as a way to assign glioblastoma patients to treatment groups in early phase clinical trials. Overall, the combination of single cell analyses and the automated RAPID algorithm can be applied to the discovery of critical onco-signaling events in other types of intractable human malignancies, providing a needed complement to genomic classification.

Table 2.1: Comparison of Citrus and RAPID

	Citrus	RAPID
Finding cell clusters	Unsupervised (hierarchical clustering, cells may be in > 1 cluster)	Unsupervised (various: FlowSOM*, dbSCAN, KNN)
Determining number of clusters to seek	Unsupervised (must be >5% of sample)	Unsupervised (automatically chosen based on prioritizing low intra-cluster variance)
Modeling cluster features	Supervised, multivariate (lasso regularized logistic regression, nearest shrunken centroid)	Unsupervised, univariate (median or MEM statistical description of cluster)
Splitting patients into groups	Supervised, prior to clustering (expert assigns patients to groups)	Unsupervised, post-clustering (cluster abundance determines cut points, tested with a Cox model of hazards)

*used in the work described above

Table 2.2 RAPID Subset and Patient Information

Tumor / RAPID Subset	2	3	4	5	41*	13*	25	8	18*	7	15	39	36	38	40
LC06	2.68	87.39	1.06	0.02	0.76	0.11	0.34	0.02	0.00	0.00	0.32	1.97	1.02	0.00	0.00
W11	3.59	1.61	25.16	28.34	0.57	2.00	4.08	1.55	1.53	1.40	1.38	8.49	5.07	0.36	0.57
LC26	5.50	2.38	10.83	2.91	7.28	1.95	6.31	1.10	0.93	3.50	13.38	2.72	3.18	1.19	0.70
RT14	12.93	2.80	1.68	1.85	0.02	0.57	13.10	2.23	1.40	0.06	2.80	0.21	0.40	0.06	1.34
K01	3.63	2.55	8.28	0.36	1.04	0.30	0.38	0.08	0.13	0.00	31.42	14.03	8.49	1.08	1.38
LC04	1.34	0.38	11.63	0.89	0.13	1.13	3.40	53.63	1.10	0.00	13.61	0.15	0.30	0.00	2.72
LC25	6.62	0.23	1.59	1.68	0.66	1.74	1.59	1.27	0.51	10.28	2.19	6.31	6.92	2.97	2.68
W02	0.72	0.59	6.50	0.53	0.21	2.08	9.38	43.14	2.63	0.00	1.34	0.57	1.40	0.00	2.55
W05	3.14	0.19	0.79	0.06	0.19	0.11	1.25	0.17	0.04	0.13	3.33	20.02	5.37	0.19	1.93
LC22	2.46	0.23	1.42	0.13	0.02	0.55	0.49	0.19	0.04	0.42	12.70	1.42	1.15	0.08	1.40
W12	2.51	0.70	1.23	0.08	1.08	0.36	3.01	3.74	0.59	0.28	1.08	0.32	0.49	0.13	2.34
W14	0.34	0.17	0.36	0.66	0.55	0.06	0.19	0.00	0.06	0.08	5.92	6.52	3.91	1.63	0.21
LC27	1.08	0.34	0.72	0.00	0.08	0.30	0.96	0.08	0.25	0.00	3.12	0.70	1.30	1.89	0.36
RT01	0.83	0.02	0.11	0.02	0.08	0.04	0.04	0.00	0.02	0.11	0.17	3.72	3.14	0.00	2.72
RT10	0.66	0.00	0.06	0.02	0.02	0.25	0.47	0.13	0.02	1.61	1.85	0.11	0.21	0.00	0.28
LC21	1.25	0.00	0.72	0.57	0.02	1.17	21.80	3.72	8.11	0.13	0.23	0.36	0.98	0.00	1.00
W15	0.21	0.06	0.08	0.06	0.08	0.70	0.62	0.04	0.15	12.14	4.48	0.59	1.19	0.36	0.32
RT15	0.30	0.02	0.13	0.51	0.11	5.12	0.98	0.00	0.30	61.53	0.34	1.46	2.08	0.32	0.74
LC08	0.00	0.00	0.00	0.00	0.32	0.21	0.38	0.02	0.04	0.74	1.30	0.57	1.04	0.00	1.46
LC11	0.02	0.00	0.25	0.02	0.47	0.38	5.03	0.21	0.74	0.38	0.02	0.74	1.63	0.00	0.02
LC18	0.06	0.02	0.17	0.00	0.28	0.00	1.91	0.11	0.08	0.00	3.99	0.66	0.93	5.07	1.10
W04	0.06	0.00	0.04	0.02	0.34	0.08	3.18	0.38	0.87	0.87	0.17	0.08	0.32	0.00	2.82
LC03	0.04	0.00	0.00	0.02	0.64	0.00	1.63	0.89	0.06	0.04	0.11	0.11	0.15	0.00	3.82
LC09	0.02	0.02	0.00	0.00	5.86	0.02	2.21	0.11	0.87	0.04	0.00	0.47	1.44	0.00	5.75
LC10	0.00	0.00	0.00	0.00	0.04	0.00	0.51	0.02	0.11	0.30	0.00	0.06	0.49	0.00	0.00
RT07	0.02	0.00	0.04	0.00	0.00	0.00	0.11	0.02	0.00	0.02	0.02	0.04	0.42	0.00	0.64
LC12	0.02	0.00	0.11	0.02	0.08	0.00	0.06	0.02	0.02	0.49	0.00	0.08	0.06	0.00	0.00
LC02	0.00	0.00	0.02	0.00	0.11	0.02	0.38	0.02	0.15	0.02	0.00	0.06	0.25	0.00	0.02
IQR	2.51	0.44	1.41	0.53	0.50	0.77	2.86	1.13	0.83	0.75	3.34	2.02	1.93	0.36	2.08
P Value	**0.02**	**0.02**	**0.04**	**0.04**	**0.04**	0.06	0.06	0.10	0.16	0.23	0.30	0.34	0.45	0.64	0.77
HR	0.30	0.29	0.40	0.39	0.39	0.40	0.45	0.47	0.53	0.57	0.62	0.63	0.71	0.79	0.88
F-measure	0.772	0.92	0.816	0.854	0.9808	0.4925	0.6756	0.85	0.4379	0.95	0.8406	0.9127	0.8959	1	0.9769

Table 2.2 RAPID Subset and Patient Information

Tumor / RAPID Subset	10	23	12	32	14	30	6	35	9	16	24	31*	43	1	21	19
LC06	0.04	0.53	0.49	0.02	0.00	0.02	0.02	0.40	0.00	0.40	0.89	0.30	0.74	0.00	0.00	0.00
W11	0.72	1.19	0.81	0.70	0.00	0.00	1.15	1.04	0.04	0.36	0.70	0.06	1.38	0.25	0.02	3.46
LC26	2.93	5.71	2.44	1.02	0.06	3.10	0.53	4.46	0.28	1.91	1.83	0.36	1.21	0.11	0.06	3.76
RT14	0.59	0.70	0.85	1.61	0.06	0.15	16.14	0.40	0.76	0.23	0.76	0.06	0.64	0.00	2.29	21.76
K01	1.80	2.57	1.40	0.47	0.00	0.47	0.00	2.89	0.00	4.76	1.13	0.32	2.25	0.17	0.00	0.04
LC04	0.23	1.21	0.45	0.17	0.28	0.02	0.76	0.30	0.42	0.23	0.47	0.04	1.83	0.08	0.04	1.61
LC25	1.78	4.42	1.68	2.12	2.23	0.38	0.40	1.19	5.90	3.95	3.80	0.49	6.41	0.76	0.17	3.57
W02	0.38	2.29	0.62	0.89	1.95	0.13	1.72	1.40	1.42	0.11	1.08	0.02	1.91	0.04	0.13	8.58
W05	0.91	14.31	20.13	1.34	0.04	1.13	1.23	3.04	0.02	2.21	1.70	0.42	3.38	2.21	0.06	2.89
LC22	8.22	7.79	2.55	0.32	0.21	2.46	8.68	1.57	0.36	4.16	3.63	0.59	3.21	2.42	0.51	3.69
W12	0.15	9.32	0.32	0.81	13.67	0.32	2.91	0.89	22.36	0.64	3.57	0.11	3.04	0.08	1.85	9.58
W14	0.15	10.83	0.42	0.40	0.00	3.14	0.08	3.14	0.00	25.29	9.75	3.25	1.87	0.02	0.02	0.19
LC27	0.17	5.10	0.42	0.42	0.00	22.19	0.13	39.94	0.02	3.50	3.29	5.99	0.34	0.02	0.02	1.00
RT01	1.04	4.84	1.74	0.15	0.00	0.08	0.08	1.34	0.00	1.10	4.01	0.53	4.76	63.52	0.21	0.04
RT10	4.80	3.99	0.25	0.34	0.21	0.17	1.51	0.30	0.55	0.68	2.19	0.15	2.80	1.61	42.51	1.68
LC21	0.17	2.48	3.93	2.68	0.45	0.47	14.06	1.34	0.68	0.06	1.27	0.19	0.89	0.11	0.62	21.76
W15	1.51	9.30	0.25	0.66	0.21	0.79	0.13	0.87	0.40	37.58	2.59	0.57	0.93	0.15	0.89	1.66
RT15	0.11	0.89	0.40	1.10	0.25	0.02	1.19	0.25	0.96	2.17	0.40	0.13	1.38	0.13	1.27	2.97
LC08	0.11	1.44	0.70	0.53	1.34	3.82	0.25	2.95	17.62	2.40	28.03	3.29	1.53	0.30	0.87	1.76
LC11	0.00	0.28	3.50	3.38	0.72	0.06	0.42	0.40	4.12	0.06	1.21	0.15	2.00	0.11	4.16	12.27
LC18	0.13	5.71	0.15	1.85	0.83	7.60	0.15	9.94	0.47	7.26	12.82	8.77	0.66	0.02	0.21	3.59
W04	0.02	0.53	0.85	0.87	4.78	0.00	0.36	0.02	19.62	0.04	0.59	0.00	3.27	0.06	1.63	22.40
LC03	0.08	1.17	0.79	0.53	28.03	0.11	0.51	0.38	32.89	0.11	1.57	0.13	1.55	0.32	0.47	9.55
LC09	0.00	0.08	0.19	2.23	1.42	0.00	0.00	0.17	5.52	0.02	0.25	0.08	2.36	0.00	0.32	14.06
LC10	0.00	0.25	0.00	1.08	0.53	0.00	0.00	0.13	0.76	0.00	0.42	0.02	0.81	0.00	1.32	1.21
RT07	0.02	0.02	0.00	0.40	0.17	0.00	0.00	0.00	0.06	0.00	0.72	0.11	0.40	0.00	0.49	0.23
LC12	0.00	0.00	0.02	0.25	0.00	0.00	0.00	0.00	0.13	0.00	0.19	0.04	1.04	0.00	0.02	0.04
LC02	0.00	0.00	0.02	0.21	0.30	0.00	0.00	0.08	0.51	0.02	0.21	0.04	1.08	0.00	0.04	0.85
IQR	0.91	4.72	1.22	0.77	0.92	0.85	1.13	1.61	2.04	2.61	2.69	0.44	1.54	0.25	0.94	7.86
P Value	0.90	0.99	0.97	0.83	0.84	0.82	0.82	0.79	0.71	0.70	0.64	0.54	0.53	0.52	0.46	0.42
HR	0.95	1.00	1.02	1.09	1.10	1.11	1.10	1.13	1.18	1.19	1.22	1.31	1.28	1.32	1.40	1.43
F-measure	0.9631	0.718	0.6094	0.5315	0.6333	0.8688	0.84	0.8817	0.84	0.8004	0.7088	0.4444	0.9997	0.98	0.8818	0.5211

Table 2.2 RAPID Subset and Patient Information

Tumor / RAPID Subset	17	22	27*	11	28	20	26	29	33*	34	37	42	# above 5%	# above 50%
LC06	0.02	0.00	0.00	0.42	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	1	1
W11	1.13	0.19	0.02	0.34	0.00	0.04	0.32	0.02	0.13	0.00	0.19	0.02	4	0
LC26	2.44	1.27	0.04	0.96	0.00	0.53	0.59	0.00	0.30	0.02	0.19	0.02	6	0
RT14	0.74	0.21	0.81	0.32	0.00	0.19	7.28	0.34	1.00	0.00	0.62	0.00	5	0
K01	1.68	5.80	0.00	1.06	0.00	0.00	0.00	0.00	0.00	0.00	0.04	0.00	5	0
LC04	0.55	0.34	0.04	0.00	0.02	0.06	0.13	0.02	0.04	0.02	0.21	0.00	3	1
LC25	7.03	0.13	0.32	0.34	0.02	0.62	1.13	0.25	1.42	0.13	2.04	0.08	7	0
W02	1.40	0.87	0.13	0.23	0.02	0.40	0.34	0.06	0.42	0.02	1.68	0.08	4	0
W05	4.88	0.13	0.02	1.78	0.00	0.02	0.79	0.04	0.21	0.00	0.19	0.00	4	0
LC22	18.81	3.72	0.51	0.62	0.06	1.00	1.63	0.15	0.13	0.02	0.19	0.06	5	0
W12	3.65	0.11	0.49	0.89	0.23	2.89	0.62	0.17	0.98	0.08	2.08	0.28	4	0
W14	2.02	17.47	0.00	1.13	0.00	0.00	0.00	0.00	0.06	0.02	0.06	0.00	6	0
LC27	0.49	4.01	0.02	1.13	0.00	0.02	0.13	0.02	0.25	0.00	0.19	0.00	4	0
RT01	3.12	0.08	0.55	1.23	0.21	0.00	0.21	0.08	0.02	0.00	0.00	0.00	1	1
RT10	0.34	0.81	5.73	1.06	0.40	0.55	20.89	0.45	0.13	0.06	0.13	0.02	3	0
LC21	0.87	0.11	0.11	1.83	0.00	1.10	1.97	0.06	1.83	0.00	0.89	0.02	4	0
W15	2.12	10.23	2.65	1.21	1.55	0.32	1.32	0.21	0.36	0.08	0.30	0.04	4	0
RT15	4.54	0.13	2.12	0.23	0.25	0.36	2.93	0.45	0.59	0.15	0.55	0.13	2	1
LC08	7.86	4.90	1.70	0.53	0.81	5.27	2.21	1.25	0.98	0.30	0.93	0.23	4	0
LC11	14.27	0.00	14.46	0.93	3.80	2.93	11.63	3.84	3.27	0.32	1.53	0.23	5	0
LC18	1.25	13.18	0.21	2.53	0.00	1.25	0.51	0.25	3.10	0.04	2.76	0.36	8	0
W04	6.71	0.13	2.57	0.28	0.66	15.97	2.34	0.49	3.38	0.17	2.51	0.51	4	0
LC03	1.59	0.00	0.57	0.00	0.11	3.10	0.79	0.40	2.10	0.45	4.54	0.66	3	0
LC09	15.73	0.00	1.00	1.00	0.40	23.57	0.87	0.57	8.77	0.15	3.65	0.76	7	0
LC10	2.68	0.00	24.93	0.06	30.34	0.30	4.93	14.14	6.79	3.35	4.03	0.38	4	0
RT07	0.06	0.00	11.23	0.02	17.64	0.02	1.25	20.70	7.03	24.20	13.10	0.76	6	0
LC12	0.04	0.00	3.31	0.08	32.78	0.00	0.30	4.14	1.21	34.93	11.59	8.90	4	0
LC02	2.42	0.02	1.91	0.17	3.29	0.47	0.38	2.68	4.10	12.51	36.14	31.49	3	0
IQR	3.79	1.87	2.19	0.84	0.70	1.12	1.72	0.47	2.22	0.20	2.38	0.37		
P Value	0.39	0.35	0.31	0.28	0.25	0.23	0.14	0.07	**0.00**	**0.05**	**0.01**	**0.02**		
HR	1.47	1.56	1.60	1.57	1.70	1.73	2.00	2.29	4.23	2.50	3.32	3.03		
F-measure	0.6974	0.7595	0.4938	0.966	0.6259	0.6577	0.5846	0.5481	0.4996	0.6108	0.6724	0.5381		

Table 2.2 RAPID Subset and Patient Information

Tumor / RAPID Subset	Stable GNP Total	Stable GPP Total	GPP - GNP	Biaxial GNP-Like	Biaxial GPP-Like	Immune Total
LC06	0.0	91.1	91.1	0.0	93.8	27.4
W11	0.2	58.7	58.5	3.3	70.6	38.5
LC26	0.2	21.6	21.4	7.8	47.2	13.7
RT14	0.6	19.3	18.6	18.5	43.8	25.3
K01	0.0	14.8	14.8	0.5	45.9	9.5
LC04	0.2	14.2	14.0	1.2	83.9	14.3
LC25	2.3	10.1	7.9	23.2	20.7	46.6
W02	1.8	8.3	6.6	12.0	47.2	5.3
W05	0.2	4.2	4.0	5.6	22.3	41.9
LC22	0.3	4.2	4.0	12.5	23.2	43.3
W12	2.4	4.5	2.1	45.8	12.3	6.0
W14	0.1	1.5	1.4	0.4	4.4	4.9
LC27	0.2	2.1	2.0	2.3	5.0	17.3
RT01	0.0	1.0	1.0	1.1	28.4	57.7
RT10	0.2	0.7	0.5	70.5	6.3	7.2
LC21	0.9	2.5	1.6	31.4	24.1	3.0
W15	0.4	0.4	0.0	20.1	2.1	27.9
RT15	0.8	1.0	0.1	60.0	3.0	11.7
LC08	1.5	0.0	-1.5	38.0	0.1	3.8
LC11	2.1	0.3	-1.8	57.6	6.5	1.8
LC18	3.2	0.3	-2.9	17.3	1.2	7.3
W04	3.2	0.1	-3.1	81.0	1.4	4.4
LC03	5.6	0.1	-5.6	86.2	0.7	4.2
LC09	4.6	0.0	-4.5	71.6	0.4	7.0
LC10	7.8	0.0	-7.8	89.2	0.1	1.8
RT07	38.1	0.1	-38.0	81.8	0.1	1.2
LC12	55.4	0.1	-55.3	89.7	0.8	1.7
LC02	80.1	0.0	-80.1	88.5	0.4	2.5

IQR
P Value
HR
F-measure

Table 2.3 Glioblastoma Patient Characteristics

Sample ID	Gender	Age	TMZ	RT	EOR	MGMT meth	Stable GNP Total (%)	Stable GPP Total (%)	PFS (days)	PFS Status	OS (days)	OS Status
LC06	Male	41	Yes	Yes	GTR	No	0.0	91.1	1539	Censored	1588	Alive
W11	Male	69	Yes	Yes	GTR	No	0.2	58.7	491	Progressed	896	Dead
LC26	Male	71	Yes	Yes	STR	Yes	0.2	21.6	363	Progressed	836	Alive
RT14	Female	55	Yes	Yes	GTR	Yes	0.6	19.3	472	Progressed	571	Dead
K01	Female	59	Yes	Yes	STR	No	0.0	14.8	176	Progressed	364	Dead
LC04	Male	65	Yes	Yes	GTR	No	0.2	14.2	263	Progressed	918	Dead
LC25	Female	60	Yes	Yes	GTR	No	2.3	10.1	185	Progressed	731	Dead
W02	Male	75	Yes	Yes	STR	Yes	1.8	8.3	446	Progressed	507	Dead
W05	Male	60	Yes	Yes	STR	No	0.2	4.2	241	Progressed	282	Dead
LC22	Male	64	Yes	Yes	STR	No	0.3	4.2	263	Progressed	366	Dead
W12	Male	50	Yes	Yes	STR	Yes	2.4	4.5	190	Progressed	723	Dead
W14	Male	47	Yes	Yes	STR	Yes	0.1	1.5	125	Progressed	488	Dead
LC27	Male	62	Yes	Yes	STR	No	0.2	2.1	79	Progressed	252	Dead
RT01	Female	69	Yes	Yes	STR	No	0.0	1.0	162	Progressed	198	Dead
RT10	Male	56	No	No	STR	Yes	0.2	0.7	22	Progressed	113	Dead
LC21	Female	69	Yes	Yes	GTR	Yes	0.9	2.5	267	Progressed	267	Dead
W15	Male	50	Yes	Yes	STR	Yes	0.4	0.4	603	Progressed	697	Dead
RT15	Female	80	No	Yes	STR	No	0.8	1.0	98	Progressed	353	Dead
LC08	Female	55	Yes	Yes	STR	No	1.5	0.0	210	Progressed	441	Dead
LC11	Female	76	Yes	Yes	STR	Yes	2.1	0.3	111	Progressed	411	Dead
LC18	Male	69	Yes	Yes	STR	No	3.2	0.3	133	Progressed	240	Dead
W04	Male	60	Yes	Yes	STR	No	3.2	0.1	220	Progressed	456	Dead
LC03	Male	60	No	No	STR	No	5.6	0.1	57	Progressed	57	Dead
LC09	Male	68	Yes	Yes	STR	Yes	4.6	0.0	66	Progressed	110	Dead
LC10	Male	78	No	Yes	STR	No	7.8	0.0	117	Censored	178	Dead
RT07	Female	70	Yes	Yes	STR	No	38.1	0.1	49	Progressed	57	Dead
LC12	Male	40	Yes	Yes	GTR	Yes	55.4	0.1	606	Progressed	710	Dead
LC02	Male	67	No	No	STR	No	80.1	0.0	53	Progressed	53	Dead

TMZ = temozolomide
RT = radiation
EOR = extent of resection
STR = subtotal resection
GTR = gross total resection
MGMT meth = MGMT promoter methylation

Table 2.4 – Mass cytometry antibody panels

Target	Mass	Clone	Signaling & proteins		Stain		
			Panel	t-SNE	Live	Sap	MeOH
Rhodium	103	-	●		✓		
Cyclin B1	139	GNS-1	●				✓
TUJ1	141	TUJ1	●	■			✓
cCasp3	142	5A1E	●				✓
CD117	143	104D2	●	■	✓		
S100B	144	19/S100B	●	■			✓
CD31	145	WM59	●	■*	✓		
γH2AX	147	JBW301	●				✓
CD34	148	581	●	■	✓		
p-4E-BP1 (T37/T46)	149	236B4	●				✓
p-STAT5 (Y694)	150	47	●	■			✓
BMX	151	40/BMX	●				✓
p-AKT (S473)	152	D9E	●	■			✓
p-STAT1 (Y701)	153	58D6	●	■			✓
CD45	154	HI30	●	■*	✓		
NCAM/CD56	155	HCD56	●	■	✓		
p-p38 (T180/Y182)	156	D3F9	●	■			✓
p-STAT3 (Y705)	158	4/P-STAT3	●	■			✓
ITGα6/CD49F	159	GoH3	●	■	✓		
CD133	160	AC133	●	■	✓		
PDGFRα	161	16A1	●	■	✓		
SOX2	163	O30-678	●	■		✓	
SSEA-1/CD15	164	W6D3	●	■	✓		
EGFR	165	AY13	●	■	✓		
p-NFκB p65 (S529)	166	K10-895.12.50	●	■			✓
L1CAM	167	5G3	●	■	✓		
Nestin	168	10C2	●	■			✓
CD44	169	BJ18	●	■	✓		
GFAP	170	1B4	●	■			✓
p-ERK1/2 (T202/Y204)	171	D13.14.4E	●	■			✓
p-S6 (S235/S236)	172	N7-548	●	■			✓
SOX10	173	A-2	●				✓
HLA-DR	174	L243	●	■	✓		
p-HH3	175	HTA28	●				✓
Histone H3	176	D1H2	●				✓

● = included in the panel
■ = included for generation of t-SNE map
* Excluded from t-SNE analyses of only glioblastoma cells
Live = live surface stain
Sap = 0.02% saponin stain
MeOH = stain after ice-cold methanol permeabilization

CHAPTER III

DETECTION OF IDH1 R132H VIA FLOW CYTOMETRY IN DISSOCIATED GLIOMA CELLS AND TISSUE SAMPLES

Authors: Justine Sinnaeve*, Bret C. Mobley*, Cherie Scurrah, Alan J. Simmons, Nalin Leelatian, Donna C. Ferguson, Akshitkumar M. Mistry, Kyle D. Weaver, Reid C. Thompson, Lola B. Chambless, Jonathan M. Irish, Ken S. Lau, Rebecca A. Ihrie.

* Equal contribution

This work is being prepared for submission to Cytometry B: Clinical Cytometry (5/11/2020)

Preface

Although low grade gliomas (grade II/III) have better prognosis than glioblastomas, they are also devastating malignancies that can be fatal in their own right or can progress to glioblastoma to the same effect. As discussed in the introduction, there are a few key molecular features that are unique to lower grade gliomas, specifically mutations in *isocitrate dehydrogenase* (*IDH1*). However, there are limited tools for detecting this protein in single cell studies of gliomas. RNA-seq is limited in its ability to detect specific mutations and cell yield measures in the hundreds to tens of thousands. I worked closely with Dr. Bret Mobley to address

this issue, developing a protocol to detect mutant IDH1 in glioma cell lines and formalin-fixed, paraffin-embedded (FFPE) tissue for analysis using flow cytometry. This chapter details that work, identifying two epitope retrieval methods that are sufficient for epitope detection. Furthermore, this chapter serves as an example of the thorough and comprehensive approach to antibody validation that is prized in the cytometry field and in the Ihrie and Irish labs. It can serve as a framework for other scientists looking to validate difficult antibodies and target antigens in other contexts.

Abstract

Isocitrate dehydrogenase (*IDH*) mutations are common occurrences in gliomas, and are also observed, to a lesser degree, in other tumor types like acute myeloid leukemias. Mutant IDH protein expression provides a valuable opportunity to positively identify cancer-lineage cells in flow cytometry studies of these tumors. Multiple antigen retrieval and permeabilization methods were tested on the glioma cell lines BT142 and U87 to expose the IDH1 R132H epitope. A previously described antibody, clone H09 generated against the R132H epitope, was used to detect the mutant protein in a novel setting using both fluorescence and mass cytometry. FFPE-DISSECT, a method for flow cytometry analysis of FFPE tissue, was also used on archival glioma samples to analyze dissociated samples. A heat-induced epitope retrieval process using citrate buffer successfully exposed the epitope of interest for flow cytometry on dissociated gliomaspheres and FFPE tissue. Methanol permeabilization also revealed the epitope on dissociated cells. The H09 antibody was conjugated to a fluorophore and a heavy metal for use in fluorescence and mass cytometry. These tools were successfully used to detect mutant IDH expression in glioma cell lines. The IDH1 R132H protein was also detected in glioma cells lines and in archived patient samples by flow cytometry. However, in acutely dissociated patient

samples, the mutant protein was detected in equivalent levels in *IDH1*-wild type tumors and tumors with confirmed *IDH1* mutations. Flow cytometric detection of IDH1 R132H in acutely dissociated primary glioma patient samples therefore presents a technical challenge, and requires further troubleshooting before experiments using the D09 clone can be reliably performed on these samples.

Introduction

Mutations in the genes encoding isocitrate dehydrogenase (IDH), *IDH1* or *IDH2*, are present in up to 80% of low-grade and intermediate-grade diffuse gliomas, a subset of brain tumors which includes World Health Organization (WHO) grade II and grade III astrocytomas and oligodendrogliomas [14, 15]. *IDH* mutations are also found in secondary glioblastomas (GBMs) that are hypothesized to develop from diffuse lower-grade precursors [14, 15, 205]. The presence of an *IDH* mutation in a diffuse glioma is clinically significant because patients exhibit substantially longer survival times [14, 23, 205]. Importantly, *IDH1* is mutated only in cancer-lineage cells and has not been observed to be altered in normal brain tissue [153, 154]. Based on these observations, *IDH* status is now an important component of the WHO classification system of gliomas [4, 23].

IDH mutations disrupt core metabolic and epigenetic processes in cells, suggesting that these mutations contribute to a mechanistically distinct subclass of cancers from *IDH*-wild type tumors. Although this is conceivable, the technical limitations of detecting mutations using single cell approaches have precluded consistent identification of mutant cells within the tumor mass and limited the further study of this idea [100]. Single-cell RNA sequencing is susceptible to incomplete or limited coverage of the transcriptome or potentially low transcript abundance, meaning that while some cells can be identified as positive for this mutation, there is the potential

for a high number of false-negative events [206, 207]. Therefore, detection of the protein product of mutated *IDH1* presents an opportunity to positively identify cancer lineage cells in heterogeneous samples.

IDH1 encodes isocitrate dehydrogenase 1, a cytoplasmic and peroxisomal protein that catalyzes the oxidative carboxylation of isocitrate to α-ketoglutarate (α-KG) [24]. Mutant IDH enzymes gain the ability to catalyze the NADPH-dependent reduction of α-KG to the oncometabolite 2-hydroxyglutarate (2HG) [20]. 2HG competitively inhibits a multitude of α-KG-dependent enzymes important for DNA damage repair, epigenetic modifications, and basement membrane generation among other processes [208]. The most frequent mutations in *IDH1* involve the R132 residue located at the substrate binding site, which interferes with IDH1-isocitrate binding [20, 21]. An arginine-for-histidine (R132H) is the most common substitution, present in greater than 90% of mutant gliomas [15, 22, 23]. The *IDH1* R132H mutation is currently detected via immunohistochemistry (IHC) in tissue sections from formalin-fixed, paraffin-embedded (FFPE) tumor specimens using an antibody demonstrated to be specific for the R132H mutation [23, 153, 154]. Neuropathologists routinely use this technique, as well as PCR, in clinical practice to detect the mutation. While these techniques can determine the mutational status of a patient sample, PCR does not distinguish which cells have the *IDH* mutation, nor if the mutant protein is expressed, and diagnostic IHC typically measures one antigen of interest in any given cell on any given slide.

Recent work detailing a high degree of genetic heterogeneity in gliomas, as well as the importance of measuring multiple features in individual cells to gain an understanding of cellular state and function, suggest that flow cytometry can be a useful, complementary technique in discovery research studying *IDH*-mutant gliomas [47, 99, 100, 102]. Flow cytometry has already been instrumental in understanding phenotypic heterogeneity in cancers, including identifying and describing cell subsets of interest [113, 114, 117, 151, 158]. However, the identification of brain cancer cells with this technique is challenging without a method for detection of a cancer-lineage

specific antigen. Thus, detection of mutant IDH1 as a protein marker of cancer cell lineage, presents a unique opportunity to positively identify such cells in research studies of these diffusely infiltrative tumors.

In this work, we introduce an approach to detect the *IDH1* R132H mutation by flow cytometry by adapting a heat-induced epitope retrieval protocol [209, 210]. Using this technique, we demonstrate that IDH1 R132H is detected in glioma cell lines, as well as in cells from archived formalin-fixed paraffin-embedded (FFPE) material, via flow cytometry.

Methods

Cell Culture

All cells were grown under standard conditions (at 37˚C and 5% CO_2). BT142 (BT142 mut/-, ATCC ACS-1018), U87, Ramos, and Jurkat cells were obtained from commercial sources (BT142) and collaborators (all other lines). Ramos and Jurkat cells were grown non-adherently in tissue culture treated T75 flasks (Grenier Bio-One) in RPMI-1640 (Corning #10-040-CV) with 10% FBS and 1x gentamicin according to the distributor's suggested protocol. Ramos and Jurkat cells were grown as single cell suspensions and were passaged approximately every 3 days. BT142 and U87 cells were grown as non-adherent spheres in sterile, untreated petri dishes in DMEM/F12 + Glutamax (Gibco/Life Technologies, MA, #10565018) supplemented with 50µg/mL gentamicin, 1X hormone mix (0.06% Glucose, 0.01125% NaHCO3, 0.5mM HEPES,40mg Apo-Transferrin, 3.86mg Putrescine, 4nM progesterone, 6nM sodium selenite, 0.2mg/mL insulin), 5mM HEPES, 4mg/mL BSA, 20ng/mL EGF, 20ng/mL FGF, and 4µg/mL heparin [211]. Spheres were dissociated with Accutase (Stem Cell Technologies) and passaged when they reached a diameter of 200 µm.

Immunohistochemistry

Spheres were harvested, fixed for 10 minutes in 1.6% paraformaldehyde (PFA), resuspended in HistoGel (Thermo Scientific), and transferred to a histology cassette for processing and paraffin embedding. Sections of paraffin-embedded spheres were dewaxed and rehydrated using a xylene/graded ethanol series, and then underwent heat-induced epitope retrieval in 10mM citrate buffer (pH 6.0) in a pressure cooker for 12 minutes. Following retrieval, slides were incubated for 30 minutes with anti-human IDH1 R132H mouse monoclonal antibody (Dianova catalog# DIA H09) at a dilution of 1:400, and developed with the HiDef Detection HRP Polymer System (Cell Marque) using a 5 minute 3,3'-Diaminobenzidine (DAB) incubation.

Patient samples

Surgical resection specimens of *IDH1*-mutant and *IDH1*-wild type diffuse gliomas were collected between 2014 and 2017 with informed patient consent, in compliance with Vanderbilt institutional review board (IRB) #131870 approval, in accordance with the Declaration of Helsinki, and were de-identified. Paraffin samples were gathered from the surgical pathology archive under IRB exemption #180238. Clinical data associated with collected samples was obtained in compliance with IRB #181970. Diagnoses were confirmed by standard neuropathological analyses of the corresponding archived formalin-fixed paraffin-embedded material, including assay for *IDH1* and *IDH2* mutations by SNaPshot PCR multiplex base extension assay (Life Technologies), and for chromosomes 1p and 19q loss of heterozygosity using PCR and primers specific to microsatellite markers [212]. To the best of our knowledge, none of the patients with glioma had a blood malignancy or whole blood cell count concerning for it. Human gliomas were processed into viable single cell suspensions following an established protocol [121]. Briefly, glioma samples were transported at room temperature in sterile saline without delay to the laboratory and processing began within 30 minutes of resection from patients. Glioma samples

were minced with a scalpel in experimental medium (DMEM/F12 + glutamax (Gibco/Life Technologies, MA, #10565018) with a defined hormone and salt mix (see above, [211]) and 50 µg/mL gentamicin) to obtain pieces of approximately $1mm^3$. Samples were resuspended in experimental medium and incubated with 1mg/mL collagenase II (Sigma Aldrich) and 0.25mg/mL of DNase I (Sigma Aldrich) for 1 hour in a 37°C incubator with 5% CO_2 on a nutating platform mixer at 18 rpm. Cell suspensions were then strained with 70 and 40 µm cell strainers. The cells were incubated with ACK lysis buffer (Lonza) for 60 seconds and then resuspended in experimental medium. The cells were spun down to remove lysis buffer and medium. Cells were resuspended to desired volume in experimental medium plus 4mg/mL BSA, 4ug/mL Heparin Solution, 20ng/mL EGF (Peprotech) and 20ng/mL bFGF (Sigma Aldrich) (Complete medium). Single-cell suspensions were viably cryopreserved in complete medium +10% DMSO.

Flow Cytometry

BT142 and U87 spheres and Jurkat and Ramos cells were collected and spun down at 200 x *g* for 10 minutes. Spheres were resuspended in Accutase (2mL per petri dish) and incubated for 10 minutes at 37˚C; this step was not included for suspension cells (Jurkat and Ramos). Cell suspension was triturated and brought to 10mL with control media and spun down at 200 x *g* for 10 minutes to obtain single cell suspensions. Glioma patients' samples were rapidly thawed from cryopreservation in a bead or water bath warmed to 37˚C and were diluted 1:10 in experimental media. Cells were spun down at 200 x *g* for 5 minutes. Cells were resuspended in 1mL of experimental media in a round bottom FACS tube and 2µL of 500x Pacific Blue-succinimidyl ester or Alexa 700-succinimidyl ester (50ng/mL) was added. Cells were incubated for 15 minutes at 37˚C to permit dye uptake in permeable (dead) cells. Pacific Blue or Alexa 700 was washed out via addition of 1 mL phosphate-buffered saline (PBS) and cells were spun down at 200 x *g* for 10 minutes. Cells were fixed in 1.6% paraformaldehyde in PBS for 10 minutes at room temperature.

PFA was washed out with 1mL PBS and cells were spun down at 200 x *g* for 10 minutes. Cells were washed in 1mL PBS with 1% bovine serum albumin (BSA) and resuspended in 10mL of 10mM citrate buffer (pH 6.0) in a 50mL conical tube for heat-induced epitope retrieval. A 600mL beaker containing 300mL ddH2O was microwaved in an Emerson microwave, model MW 8117 W, for 2.5 minutes. Each conical tube was swirled in the heated water so that the total volume (10mL) was submerged for 2 minutes. After 2 minutes, the cells were brought to room temperature by incubation on ice (about 10 minutes). The water was heated for an additional 45 seconds prior to each subsequent tube. This continued until all tubes had been heated. Cells were spun down at 800 x *g* for 10 minutes, transferred to a 5 mL round bottom tube, and washed in 1mL PBS + 1% BSA. Cells were incubated for 30 minutes at room temperature in 50µL staining volume with the following antibodies: anti-IDH1 R132H (mouse monoclonal unconjugated Dianova catalog# DIA H09) at 1:50 (0.2µg per 50µL test), anti-CD45 (rabbit monoclonal D9M8I Cell Signaling Alexa Fluor 700 conjugate catalog #32189) at 1:100, anti-CD45 (mouse monoclonal HI30 BD Biosciences BV786 conjugate catalog #563716) at 1:40, anti-CD31 (mouse monoclonal WM59 BD Biosciences BUV395 or Ax647 conjugate catalog #565290/#561654) at 1:50, and where applicable anti-IDH1 R132H (mouse monoclonal custom conjugate to Pacific Orange [details given in Antibody Conjugation] Dianova catalog# DIA H09) at 1:100 (approximately 0.5µg per test). Cells were washed with 1mL PBS+ 1% BSA. For experiments involving U87 and BT142 cells, secondary antibody was applied (donkey anti-mouse 488 Invitrogen catalog# A21202) at a concentration of 1:1000 in PBS+ 1% BSA. Secondary antibody was rinsed out with 1mL PBS + 1% BSA and cells rinsed with 1mL PBS prior to flow cytometry. Trials of saponin permeabilization (as an alternative to retrieval with heat and citrate) involved one wash of 1mL 0.02% saponin in PBS prior to antibody incubation in 0.02% saponin in PBS. Trials of permeabilization with methanol involved post-fixation exposure to ice-cold methanol overnight at -20˚C, followed by antibody staining on the following day. Flow cytometry was performed on a BD Fortessa or a 5 laser BD LSRII. Data were analyzed using Cytobank software [168].

Mass Cytometry

Cell lines were collected for flow cytometry as described above and in Appendix A and [121]. A multi-step staining protocol was used, which included 1) live surface stain, 2) when necessary, a stain after 0.02% Saponin permeabilization, and 3) intracellular stain after permeabilization with ice-cold methanol. Once a single cell suspension was prepared, cells were resuspended in 1mL of experimental media in a FACs tube and 2µL of 500x of Rhodium (final concentration of 50µM) was added. Cells were incubated for 1 minute at room temperature to permit exclusion of dead cells. Rhodium was washed out with 1mL PBS +1% BSA. At this point, if live staining was to be performed cells were incubated with antibodies noted in Table 3.1 in a total volume of 50 µL. Following live staining, cells were fixed in 1.6% paraformaldehyde in PBS for 10 minutes at room temperature. PFA was washed out with 1mL PBS. Cells were washed in 1mL PBS with 1% bovine serum albumin (BSA) and proceeded to either 0.02% saponin staining, methanol permeabilization, or heat-induced epitope retrieval with citrate buffer as described above. Following the appropriate permeabilization approach, cells were incubated for 30 minutes at room temperature in 50µL staining volume with antibodies indicated in Table 3.1. After staining, cells were resuspended in deionized water containing standard normalization beads (Fluidigm)[166], and collected on a Helios instrument located in the Mass Cytometry Center of Excellence at Vanderbilt University. Mass cytometry standardization beads were used to remove batch effects and to set the variance stabilizing arcsinh scale transformation for each channel following field-standard protocols [98, 121, 146]. Rhodium viability stain and cleaved caspase-3 antibody were included in staining to exclude non-viable and apoptotic cells, respectively. Detection of total histone H3 was used to identify intact, nucleated cells [98]. A 26-dimensional mass cytometry antibody panel was used to analyze 4 million viable cells from 8 tumors (500,000 cells per tumor). Data were normalized with Fluidigm CyTOF software normalization function, and were arcsinh transformed (cofactor 5), prior to analysis using the Cytobank platform [168]. An

equal number of live, single events were included in a common t-SNE analysis, generated using 19 of 26 measured markers (Table 3.1).

FFPE-DISSECT and Cytometry of Disaggregated FFPE Material

50 µm sections were freshly cut from each block of cell lines or glioma tissue and placed in 1.5mL Eppendorf tubes (Fisher). Samples were processed as previously reported [213]. Briefly, samples were heated to 65°C for 25 minutes to melt wax and then washed three times with 1 mL of Histo-Clear (National Diagnostics) for 8 minutes each. Tissues were then rehydrated in two washes each of 100%, 70%, and 50% ethanol and then three washes of PBS. Samples were washed for 10 minutes in PBS with 0.3% Triton X-100 and then washed for a final time in PBS before incubation in the HIER buffer (DAKO). Samples were incubated in the buffer under high heat and pressure (pressure cooker) for 20 minutes (actively heating for the first 4 minutes), followed by 20 minutes cooling on the bench. Samples were then washed three additional times in PBS and stored at 4°C until staining. Tissues were blocked at room temperature for 15 minutes in 2.5% normal donkey serum (Jackson ImmunoResearch) in PBS and stained overnight at room temperature with IDH1 R132H antibody (Dianova catalog# DIA H09) diluted to a concentration of 1:200 in the same buffer. Following three washes in PBS, secondary antibody (donkey anti-mouse Alexa Fluor 647 conjugate, Life Technologies) was applied at a dilution of 1:500 in 2.5% normal donkey serum in PBS for 1 hour at room temperature. After 3 washes in PBS, samples were incubated for 30 minutes in 4% PFA to cross-link antibodies to their targets. Samples were washed 3 times in PBS and then incubated for 20 minutes at 37°C in 200 µl of PBS with 1 mg/mL each of collagenase (Calbiotech) and dispase (Life Technologies). 800µl of PBS with .003% Triton X-100 was then added and the tissue was passaged 5 to 10 times through a 27.5-gauge needle to mechanically dissociate it into single cells. Cells were filtered through a 35µM filter to remove

debris and washed once in 0.003% Triton X-100. Cells were resuspended in Hoechst dye for 20 minutes and then washed in 0.003% Triton X-100 before being analyzed on the flow cytometer.

Patient Samples for FFPE-DISSECT

IDH1-wild type and -mutant infiltrating glioma samples were identified from the Vanderbilt Surgical Pathology archive under an approved IRB protocol (#141428). Three wild type and three mutant cases with sufficient tumor burden, collected between 2013 and 2014, were chosen for analysis. The diagnoses were verified by a neuropathologist (B.C.M.).

Antibody Conjugation to Pacific Orange

Purified IDH1 R132H antibody was obtained from Dianova in PBS + 0.05% sodium azide (NaN_3). NaN_3 was diluted out via buffer exchange to PBS using 50kDa filters (Amicon). Briefly, 120µg of antibody was loaded into a filter (75µL). PBS was added to a total of 500µL (425µL PBS) and the sample was spun at 12,000 x *g* for 8 minutes. The flow-through was discarded and the remaining volume (~20µL) containing the antibody was collected by addition of PBS (to total 90µL) and spinning the inverted filter into a collection tube at 1000 x *g* for 1 minute. 100µg of buffer exchanged antibody, in 90µL PBS, was combined with 1M NaH_2CO_3 and added to lyophilized Pacific Orange supplied by Thermo Scientific (Cat # P30014). Antibody and Pacific Orange dye were incubated at room temperature for 1 hour with regular inversion to mix. The conjugated antibody was stored in the dark at 4˚C.

Antibody Conjugation to 171

Purified IDH1 R132H antibody was obtained from Dianova in PBS + 0.05% sodium azide (NaN_3). NaN_3 was diluted out via buffer exchange to PBS using 50kDa filters (Amicon). Metal conjugation was performed using the Fluidigm Maxpar® X8 Antibody Labeling Kit. Briefly, commercially available metals of interest are incubated with a chelating polymer for 1 hour. The purified antibody is partially reduced using tris(2-carboxyethyl)phosphine (TCEP) for 30 minutes at 37°C. The metal-polymer is then added to the reduced antibody.

Data Analysis and Statistics

FCS files were uploaded to cytobank.org [168]. For fluorescence flow cytometry experiments, compensation was calculated using the automatic compensation calculation tool and bead compensation tubes. Intact cells were gated using FSC-A and SSC-A. Singlets were gated using FSC-A and FSC-W. Viable cells were gated based on exclusion of Pacific Blue or Alexa 700 succinimidyl ester. Single cells were then analyzed for intensity of antibody conjugates. Histograms were generated in Cytobank. The arcsinh transformed values of median fluorescence intensity (MFI) were compared between tubes. MFI values were input into Prism (Graphpad) and a t-test was performed on IDH1 R132H signal from U87, BT142, and glioma patient samples. For mass-based experiments, compensation beads were excluded, and intact cells were gated on gaussian parameters (fluidigm.com) and Iridium. Viable cells were gated based on exclusion of Rhodium and cleaved caspase 3. Single cells were analyzed for mass intensity. Histograms were generated in Cytobank. A common t-SNE analysis was performed in Cytobank on equal numbers of live patient cells from each sample (n=8) using 20 of the measured markers.

Data Availability

Flow cytometry data files (.fcs) will be available on flow repository pending acceptance of this work for publication.

Results

IDH1 R132H protein is detected via flow cytometry on BT142 cells

The *IDH1* RI32H mutation is retained in BT142 cells, which were generated from a WHO grade III *IDH1*-mutant glioma with oligoastrocytoma histology [214]. The GBM cell line U87 does not have a mutation in *IDH1* and is reported to be wild type for the protein [215]. The *IDH1* status of BT142 and U87 cells was first confirmed via immunoperoxidase staining of gliomaspheres (Figure 3.1a). These cell lines were used as positive and negative controls, respectively, for all subsequent flow cytometry experiments. Antibody titration is an important consideration when utilizing an antibody across multiple platforms and the optimal concentration may differ between technologies. Therefore, a series of antibody concentrations for flow cytometry were tested, using a mix of BT142 and U87 cells. The dilution of 1:50 (0.2µg per 50uL test) resulted in the largest difference between median fluorescence intensity of the positive and negative populations (Figure 3.1b).

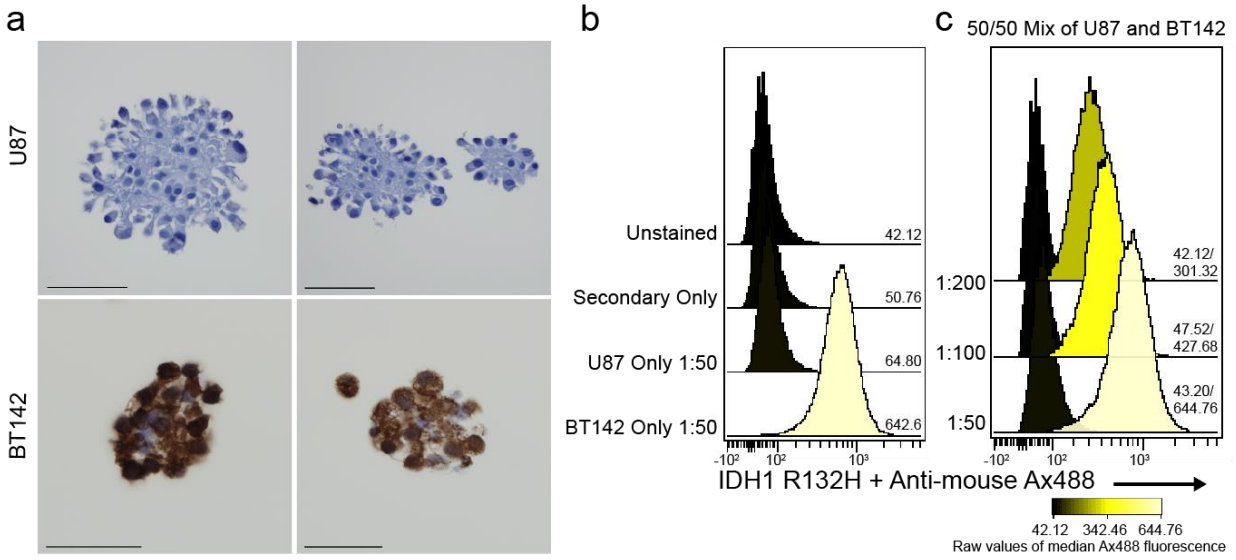


Figure 3.1: **IDH1 R32H is detected via flow cytometry** (a) Immunoperoxidase staining for IDH1 R132H in paraffin-embedded, sectioned BT142 (positive) and U87 (negative) spheres. Scale bars = 50μm. (b) Histogram of relative cell count for each condition (y axis) and signal from IDH1 R132H antibody staining (x axis). First row is unstained cells, second row is anti-mouse secondary antibody only, and bottom rows are individual cell lines stained with primary and secondary antibodies. (c) IDH1 R132H antibody titration for flow cytometry using a mixture (U87 and BT142) of sphere-derived cells. 1:200, 1:100, and 1:50 dilutions of primary IDH1 R132H antibody are depicted in the rows. U87 and BT142 populations are separated computationally and represented as separate histograms. Numbers and heat signify median fluorescence intensity of Ax488 signal.

Two methods can be used to retrieve the IDH1 R132H epitope for detection

Due to the cytoplasmic localization of IDH1 R132H, cells must be permeabilized prior to detection of the mutant protein. Commonly used, previously published permeabilization techniques that have been used on glioma samples [121] were tested on BT142 and U87 cells, followed by cell staining for IDH1 R132H (Figure 3.2a). PBS (no retrieval), 0.02% saponin, or citrate buffer alone did not result in specific antibody staining using these control cell lines (Figure 3.2b). However, 10mM pH 6 citrate buffer combined with gentle heating (see methods) resulted in distinct positive (BT142) and negative (U87) populations via flow cytometry. This permeabilization method is routinely used for detection of mutant *IDH1* in tissue sections by

neuropathologists [216, 217]. Similarly, permeabilization via ice cold 100% methanol, which is widely used in Western blotting and intracellular staining for flow cytometry [121, 218, 219], resulted in robust detection of positive and negative cell populations. All subsequent flow cytometry experiments were performed using either citrate and heating or ice-cold methanol, noted in each section below.

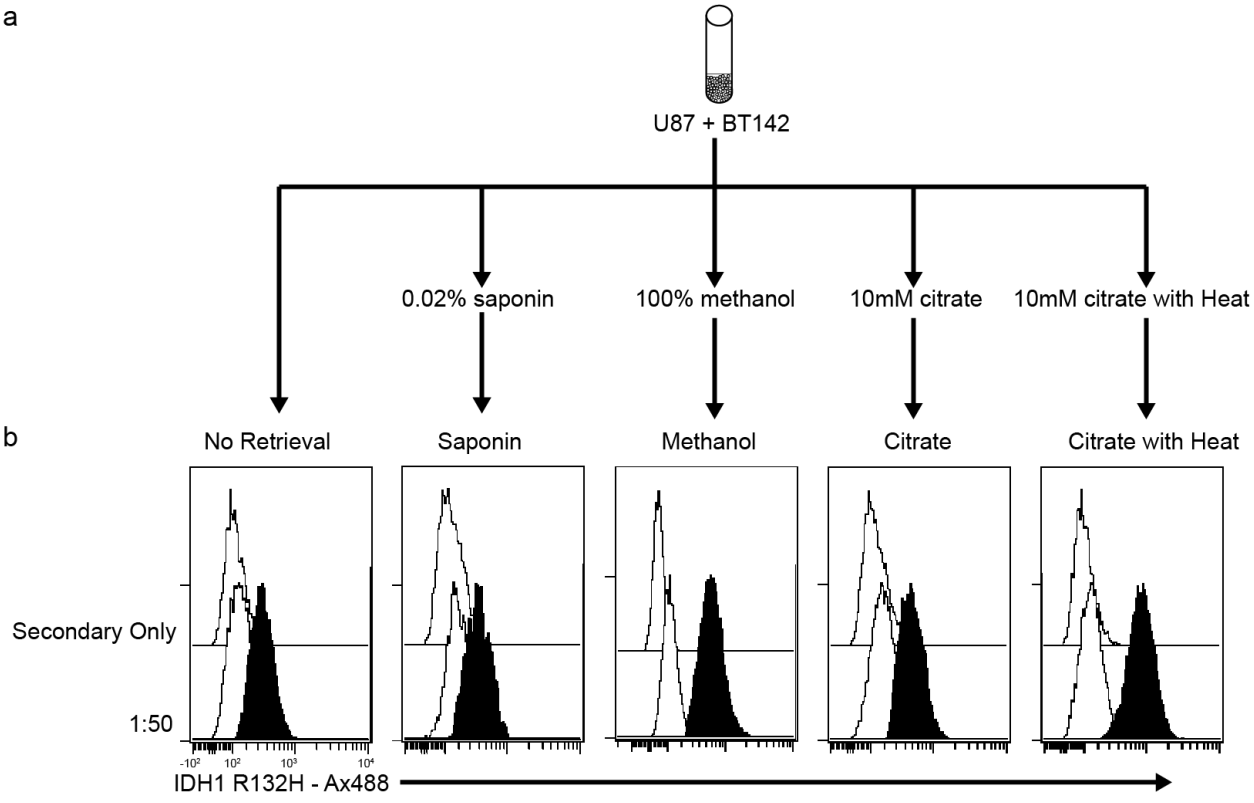


Figure 3.2: **Methanol and citrate buffer with heat are used to expose the IDH1 R132H epitope** (a) Flow cytometry staining strategy for mixed U87 and BT142 single cells. Cells were fixed and then stained directly (No Retrieval), permeabilized with 0.02% saponin for staining (Saponin), permeabilized with ice cold 100% methanol overnight (Methanol), or incubated in citrate buffer with (Citrate with Heat) or without heating (Citrate) prior to staining. (b) IDH1 R132H antibody signal on mixture of U87 and BT142 cells for each retrieval condition. Histograms depict relative cell count for each condition (Y axis) and signal from IDH1 R132H antibody staining (x axis). First row of the histogram is secondary antibody staining only. Second row depicts histograms for U87 (white fill) and BT142 (black fill) populations after staining with 1:50 dilution of IDH1 R132H primary antibody and anti-mouse secondary. Methanol condition was collected on a different day

from remaining conditions and thus median fluorescence on the arcsinh scale could not be calculated on a single scale for this data as in other figures.

Heat-induced epitope retrieval can be combined with staining for other antigens

One major advantage of flow cytometry is the ability to stain for multiple antigens in single cells. To test if the heat-induced epitope retrieval method could be combined with staining for other antigens, Jurkat and Ramos cells were subjected to this method and then stained with CD31 or CD45 respectively. Ramos cells displayed similar levels of CD45 staining intensity regardless of whether cells were subject to heat-induced epitope retrieval (Figure 3.3). Jurkat cells stained for CD31 in three different conditions, 1) live staining, 2) staining prior to heat-induced epitope retrieval, and 3) staining after heat-induced epitope retrieval, demonstrated signal for CD31 in all conditions (Figure 3.3b). The highest CD31 signal was observed in cells which had been stained live without retrieval but staining pre- or post-epitope retrieval was also sufficient to generate signal above controls (Figure 3.3b). These results show that heat-induced epitope retrieval did not significantly impair the detection of CD45 or CD31 and can be combined with staining for other antigens when appropriate controls are included.

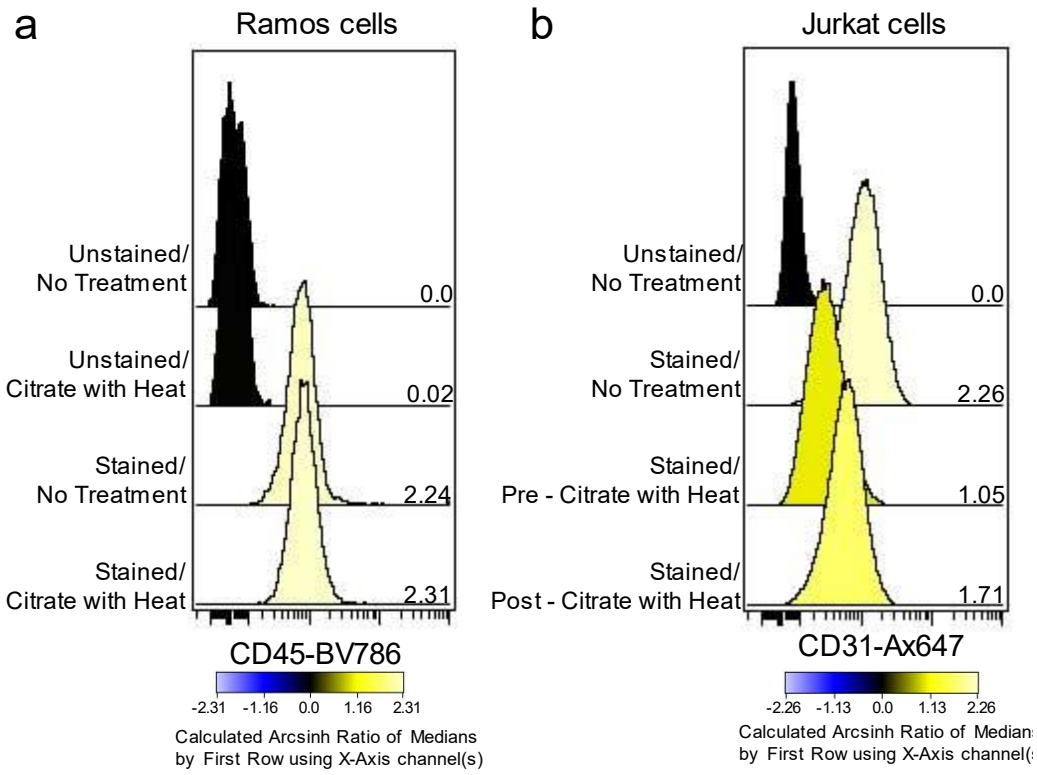


Figure 3.3: **Heat-induced epitope retrieval does not prevent staining for additional antigens of interest**. (a) Histogram of relative Ramos cell count for each condition (y axis) and signal from CD45-BV786 antibody staining (x axis). First row is unstained cells, second row are unstained cells treated with heat and citrate buffer. Third and fourth rows contain cells stained with antibody either live or after citrate with heat treatment. (b) Histogram of relative Jurkat cell count for each condition (y axis) and signal from CD31-Ax647 antibody staining (x axis). First row is unstained cells, second row are live stained cells. Third and fourth rows contain cells that been stained with antibody either prior to or after citrate with heat treatment. Heat signifies median fluorescence intensity of fluorophore indicated below each histogram.

IDH1 R132H was custom conjugated for fluorescence and mass cytometry

The work described above was completed using a primary antibody against IDH1 R132H and a secondary antibody conjugated to a fluorophore for detection. An IDH1 R132H antibody directly conjugated to a fluorophore or metal would allow combination with antibodies generated in the same species (mouse) and simplify the staining protocol. Anti-IDH1 R132H was therefore conjugated to a fluorophore (Pacific Orange) for traditional flow cytometry and a heavy metal for

mass cytometry (171Yb) (see Methods). Staining U87 and BT142 cells with the Pacific Orange conjugated antibody resulted in expected staining patterns, with BT142 cells demonstrating IDH1 R132H signal and U87 cells indistinguishable from unstained controls (Figure 3.4). The primary antibody was also conjugated to a heavy metal for use in mass cytometry. Staining of healthy peripheral blood mononuclear cells (PBMC) and BT142 cells shows staining on both populations, with higher intensity on BT142 cells (Figure 3.5). Healthy PBMC samples are expected to be *IDH1*-wild type, so any observed signal was interpreted as non-specific staining. By using CD45 staining, PBMC ($CD45^+$) and BT142 ($CD45^-$) can be separated. When the two populations are compared, a dilution of 1:50 reduced observed signal on PBMC while retaining signal on BT142 cells (Figure 3.5). To test if methanol permeabilization was equivalent to heat-induced epitope retrieval on mass cytometry, the metal-conjugated antibody was used to stain U87 and BT142 cells post-methanol. Based on the separation in median staining intensity between the BT142 cells and the U87 cells at a dilution of 1:100 (10ug/mL per tube), this condition was chosen for subsequent mass cytometry experiments (Figure 3.6).

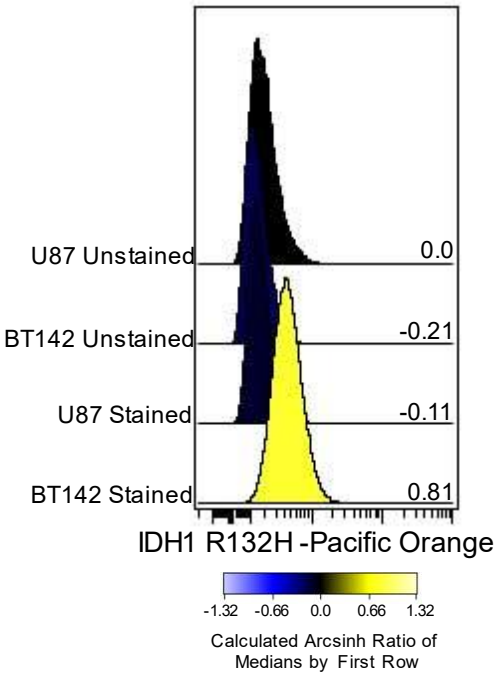


Figure 3.4: **Anti-IDH1 R132H conjugated to Pacific Orange specifically stains BT142 cells**. Histogram of relative cell count for each condition (y axis) and signal from IDH1 R132H-Pacific Orange antibody staining (x axis). First and second rows are unstained cells. Third and fourth rows depict stained cells. Number and color signify arcsinh-transformed median fluorescence intensity of fluorophore, relative to first row, for each histogram.

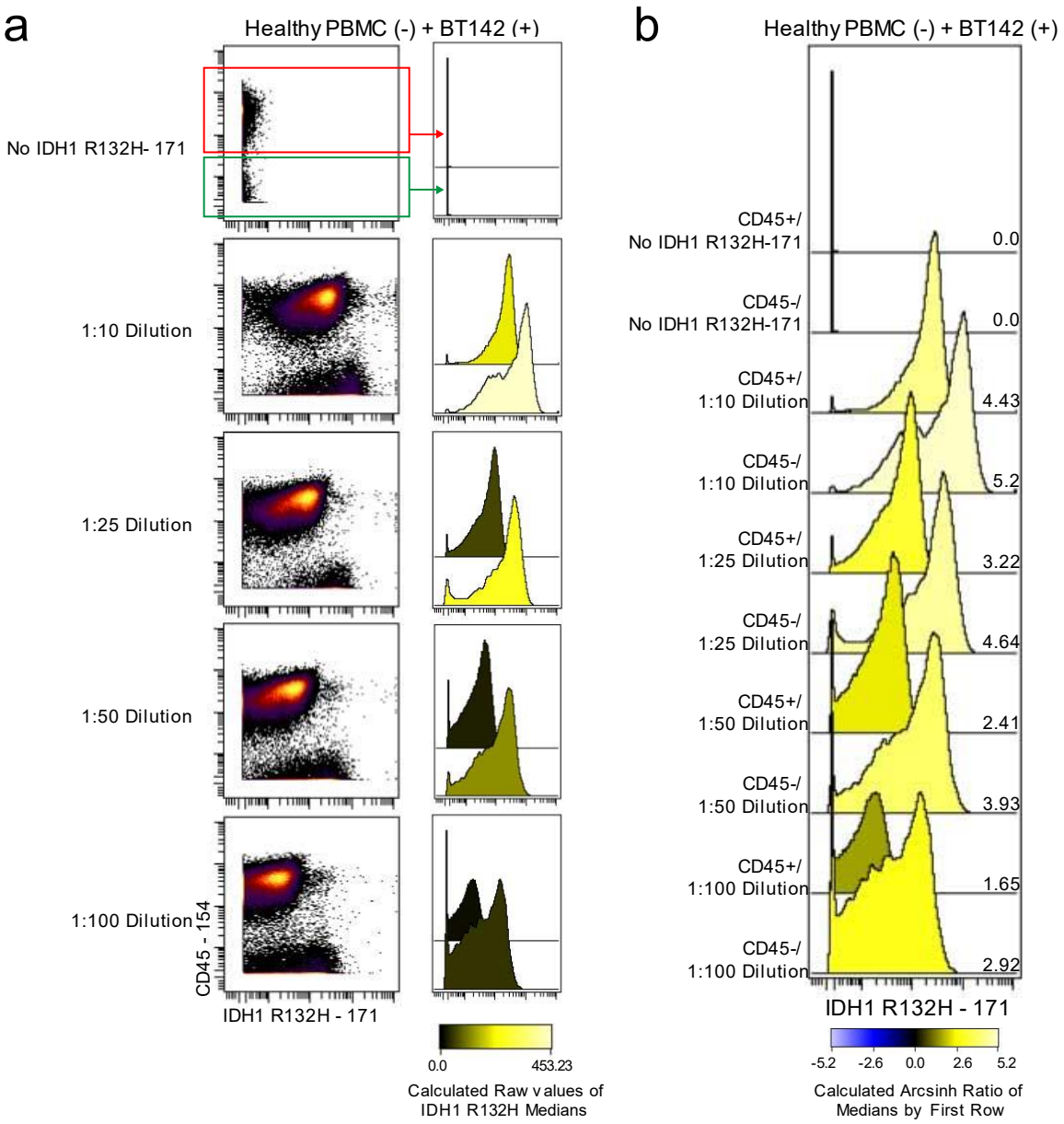


Figure 3.5: **After heat-induced epitope retrieval, PBMC cells have non-specific IDH1 R132H staining but can be separated from BT142 cells by CD45 staining**. (a) Each graph depicts a mixture of BT142 and PBMC cells stained with CD45-154 (y axis) and IDH1 R132H -171 (x axis). Dot plots (right) show CD45-154 versus decreasing dilutions of IDH1 R132H-171. Cells from these plots were gated based on CD45 expression (PBMC = $CD45^+$/Red Box, BT142 = $CD45^-$/ Green Box). These populations are depicted in the histograms to the right. (b) Overlaid histogram from data in (a). Manually gated CD45 -positive and -negative populations stained for decreasing dilutions of IDH1 R132H-171. Heat indicates median mass intensity for IDH1 R132H-171, relative to first row.

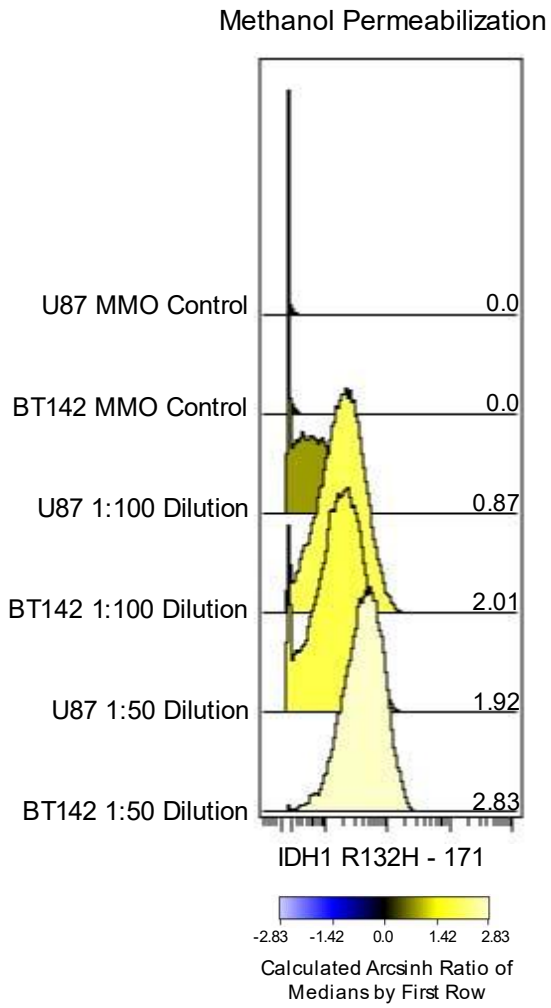


Figure 3.6: **IDH1 R132H can be detected via mass cytometry after methanol permeabilization**. Histogram depicts U87 and BT142 cells permeabilized by ice-cold 100% methanol and stained with IDH1 R132H-171. The first two rows are a mass minus one control. The next rows depict two different dilutions of IDH1-R132H-171. Heat indicates the arcsinh transformed values of median mass intensity based on the first row.

IDH1 R132H was not specifically detected in dissociated primary glioma samples

A primary goal of the development of the IDH1 R132H mutation specific antibody for use in flow cytometry was application to patient samples to identify single cells in which the mutant protein is expressed. To this end, the antibody, conjugated to either a fluorophore or a heavy metal, was used to label freshly dissociated glioma patient samples. In $DAPI^{+}$, live events from

IDH1-wild type (top row) or *IDH1* R132H-tumors (bottom row), staining for the mutant protein was not increased in the tumor confirmed to bear the mutation (Figure 3.7). Furthermore, in both tumors, $CD45^+$ events, which should not contain the mutation, also demonstrated staining by the IDH1 R132H antibody (Figure 3.7). To avoid autofluorescence issues and spectral overlap concerns, wild type and mutant tumors were stained with the heavy metal tagged antibody. Four *IDH1*-wild type and four *IDH1* R132H-tumors were stained for both CD45 (y-axis) and IDH1 R132H (x-axis, Figure 3.8). As in fluorescent flow cytometry, known wild type tumors demonstrated signal in the 171 channel and $CD45^+$ events also displayed 171 staining (Figure 3.8).

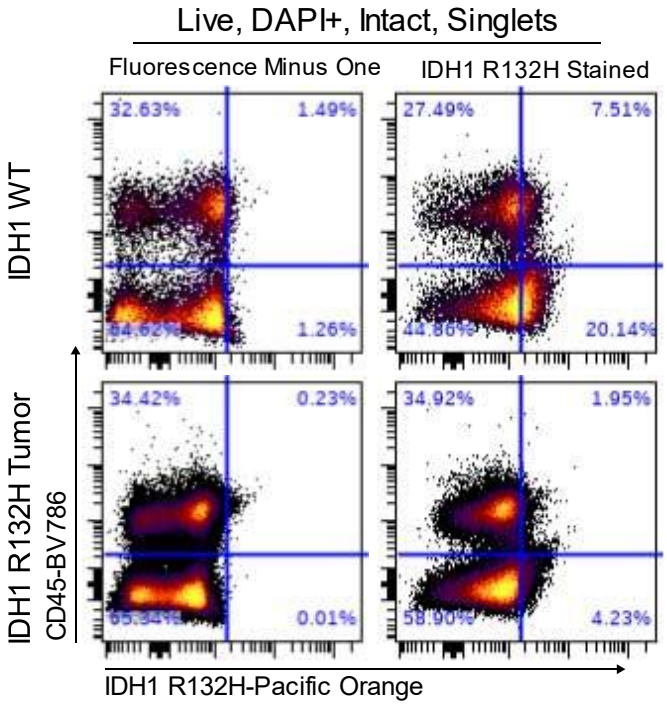


Figure 3.7: **IDH1 R132H -Pacific Orange signal is detected in wild type and mutant glioma tumors**. Dot plots, colored by cell density, depict an *IDH1*-wild type (WT, top row) and *IDH1* R132H-tumor (bottom row) stained with CD45-BV786 (y axis) and IDH1 R132H-Pacific Orange (x axis). Blue lines are quadrant gates to determine percent positive $CD45^+$/IDH1 $R132H^-$ (upper left), $CD45^+$/IDH1 $R132H^+$ (upper right), $CD45^-$/IDH1 $R132H^+$ (lower right), $CD45^-$/IDH1 $R132H^-$ (lower left).

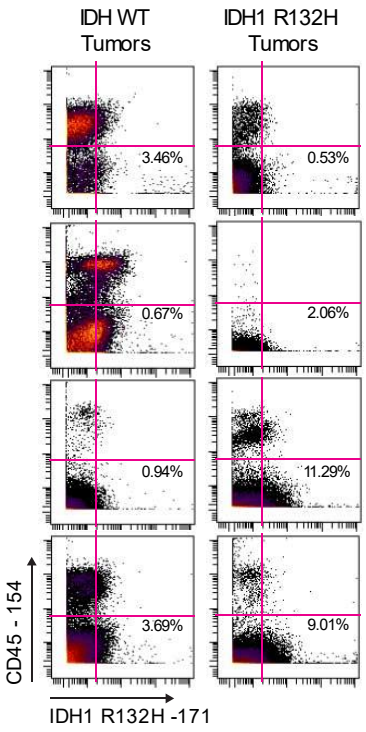


Figure 3.8: **IDH1 R132H-171 signal is detected in *IDH1*-wild type and *IDH1* R132H-mutant glioma samples.** Dot plots, with heat depicting cell density, demonstrate CD45-154 (y axis) and IDH1 R132H-171 (x axis) are shown for *IDH1*-wild type (left) and *IDH1* R132H-gliomas (right). Pink lines indicate a quadrant gate drawn to differentiate double negative, double positive, and single positive populations. The percent indicated is the number of cells in the lower right quadrant, which are IDH1 R123H-171 positive.

In order to determine if *IDH1*-wild type tumors were picking up IDH1 R132H antibody on dead or dying cells, or if these same cells expressed other markers observed in gliomas, a larger mass cytometry experiment was performed on four *IDH1*-wild type and four *IDH1* R132H-tumors. 26 features were measured on 4,000,000 cells (Table 3.1). After gating for intact, live events, a t-SNE analysis was performed on an equal number of cells from each sample using 19 of the measured features. The t-SNE maps indicated that IDH1 R132H is detected in cells that express other proteins previously described to be expressed in glioma cells, such as S100B, EGFR, and GFAP (Figure 3.9). Additionally, the percent of cells expressing IDH1 R132H varied between tumors, with no enrichment in *IDH1* R132H-mutant tumors (Figure 3.9).

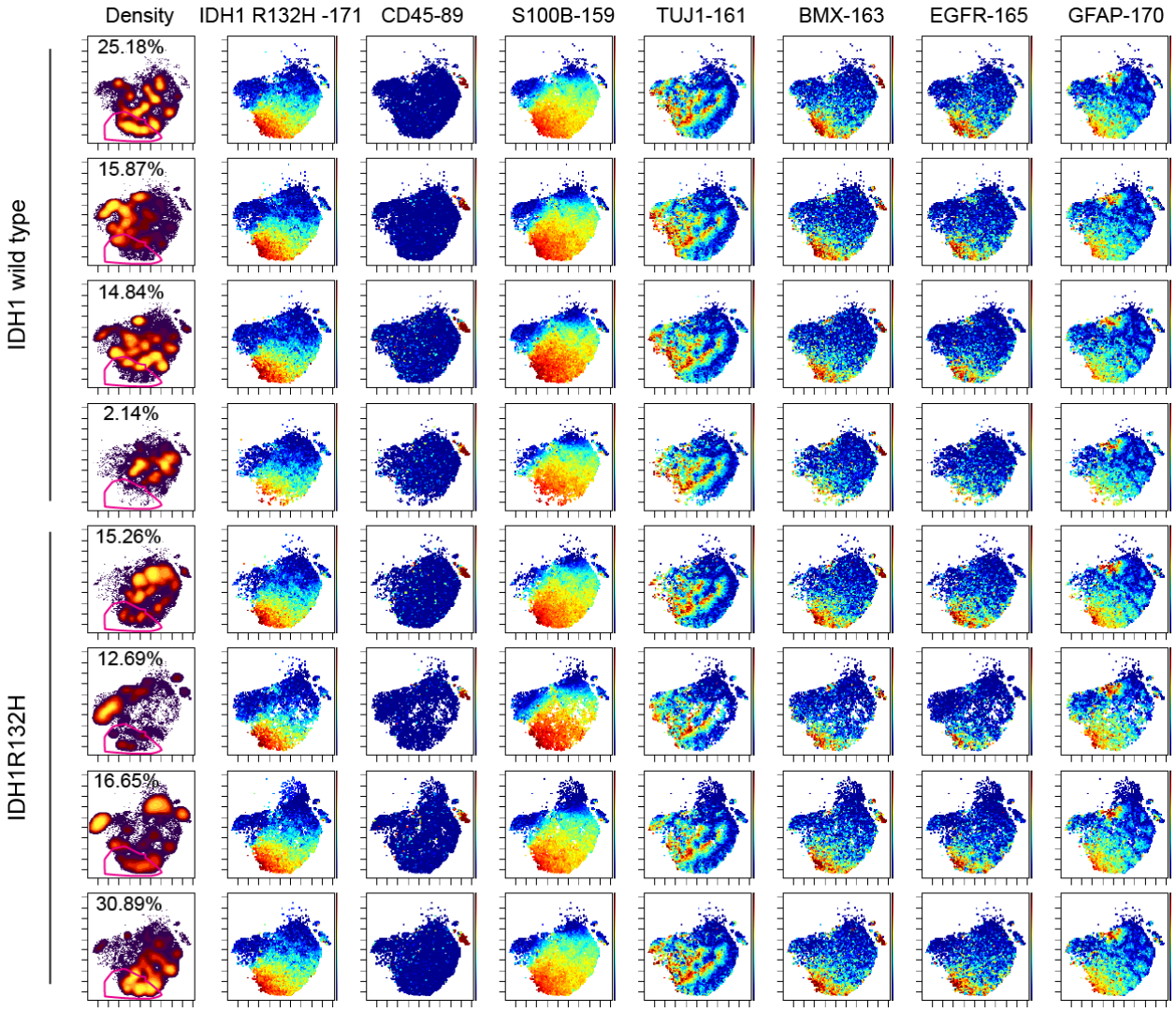


Figure 3.9: **IDH1 R132H-171 positive cells are positive for other glioma proteins in wild type and mutant samples.** t-SNE plots for 8 glioma samples (rows) depicting 8 features (columns). Contour plots in first column are colored by density. Dot plots in remaining columns have heat on each marker listed. The percent of IDH1 R132H – positive cells was determined based on the pink gate depicted in the first column (same for all samples). The percent of cells in this gate is listed within the density plot (first row) for all samples. A t-test reveals these values are not statistically different between *IDH1*-wild type samples (first four rows) and *IDH1* R132H-mutant samples (last four rows) (p-value = 0.51).

Non-specific staining in glioma samples is not affected by antibody conjugation

Non-specific staining in glioma samples and healthy PBMC could arise due to a change in antibody structure induced during antibody reduction during the conjugation protocol. To test

whether staining of primary samples was affected by antibody conjugation, U87, BT142, PBMC, and an *IDH1* R132H-mutant glioma sample were stained with either IDH1 R132H-171 conjugated antibody or IDH1 R132H unconjugated antibody plus an anti-mouse secondary antibody (172). In secondary only controls (in which no mouse antibodies are added but the anti-mouse-172 antibody is included) there was demonstrable background, especially in U87 and BT142 cells (Figure 3.10). However, using a secondary antibody to detect either the unconjugated primary antibody or the conjugated version resulted in very similar staining patterns, with non-specific staining on a subset of PBMC cells when the antibody against IDH1 R132H was included.

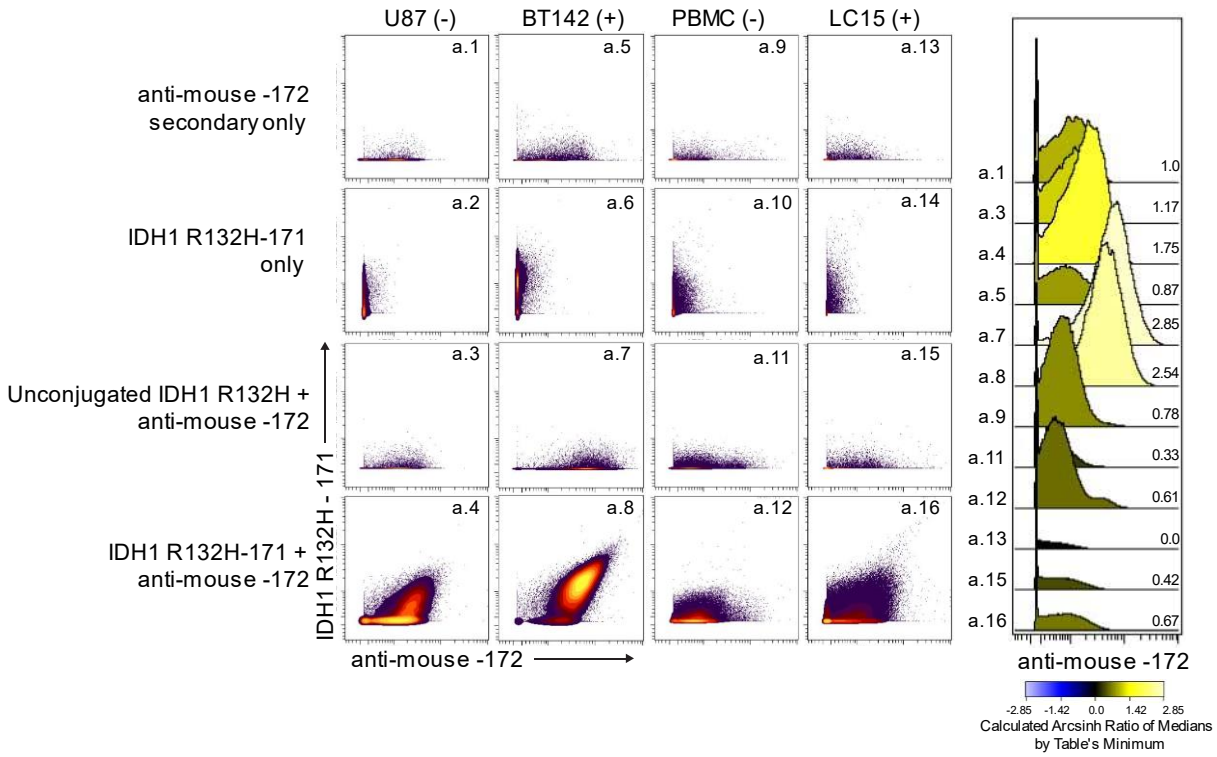


Figure 3.10: **Anti-mouse secondary antibody detects the same features in different cell types as IDH1 R132H-171 antibody**. Contour plots (left) depict IDH1 R132H-171 (y axis) and anti-mouse-172 (x axis) signal on four different cell types (columns). First row is a secondary only control (no IDH1 R132H antibody). The second row depicts cells stained only with the conjugated IDH1 R132H-171. The third row contains cells stained with a primary, unconjugated IDH1 R132H antibody and the anti-mouse secondary. The fourth row contains both the IDH1 R132H-171 and

the anti-mouse-172. Selected conditions are shown as histograms (right), with heat (and x axis) showing median mass intensity of measured samples relative to table's minimum (condition a.13).

IDH1 R132H can be detected in disaggregated FFPE patient samples

As *IDH*-mutant tumors are a smaller percentage of glioma cases, many research studies use paraffin-embedded archival specimens [220-222]. Therefore, the antibody was titrated for use on archived tissue using FFPE-DISSECT, a method that involves staining FFPE tissue prior to disaggregation for flow cytometry (Figure 3.11) [155]. After choosing 1:200 as the condition that maximally separated U87 cells from BT142 (Figure 3.11), six glioma patient FFPE samples were examined for IDH1 R132H expression (three *IDH1* R132H-mutant gliomas and three *IDH1*-wild type glioblastomas). Expression of mutant protein and genotype was confirmed via immunohistochemistry (Figure 3.12) and by SNaPshot PCR multiplex base extension assay, respectively (data not shown). Portions of each FFPE block were isolated for FFPE-DISSECT. When FFPE samples from the six glioma test samples were dewaxed, rehydrated, stained for IDH1 R132H and then dissociated, mutant samples stained positively for IDH1 R132H while the wild type GBM samples did not (Figure 3.12). Tumors with more abundant IDH-positive cells by IHC staining exhibited similar patterns using flow cytometry. Importantly, the implementation of the DISSECT protocol did not increase non-specific staining with the IDH1 R32H antibody or a secondary anti-mouse antibody (Figure 3.12). These results indicate that the FFPE-DISSECT protocol can be used on glioma tissue and in concert with the IDH1 R132H antibody.

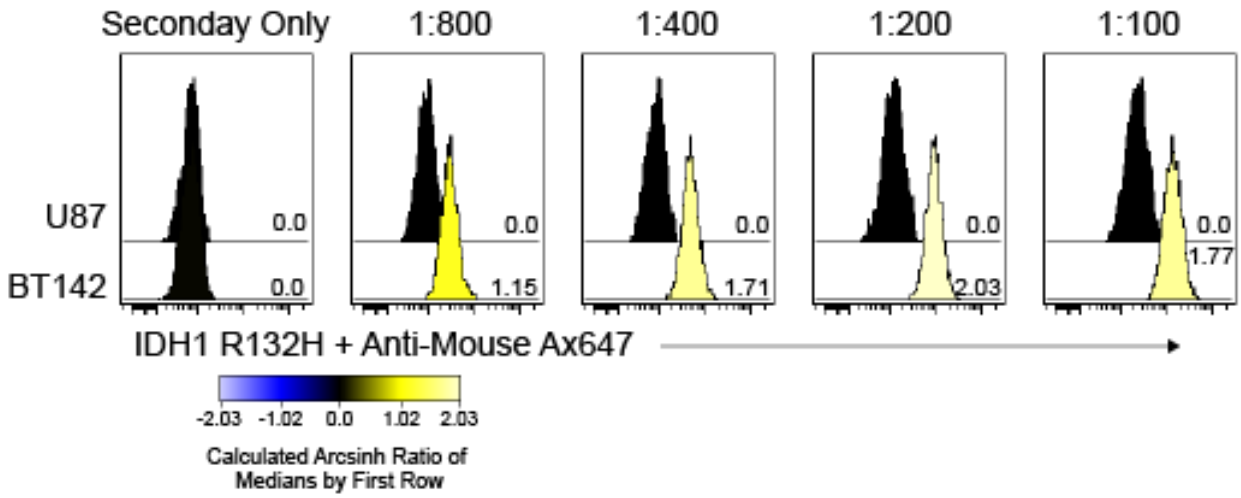


Figure 3.11: **IDH1 R132H was titrated for FFPE-DISSECT.** Each overlaid histogram plot depicts U87 or BT142 cells stained with IDH1 R132H primary antibody and anti-mouse Ax647 secondary antibody (x axis). Dilution of the antibody is shown above each graph. Heat and numbers indicate median mass intensity for Ax647, relative to first row.

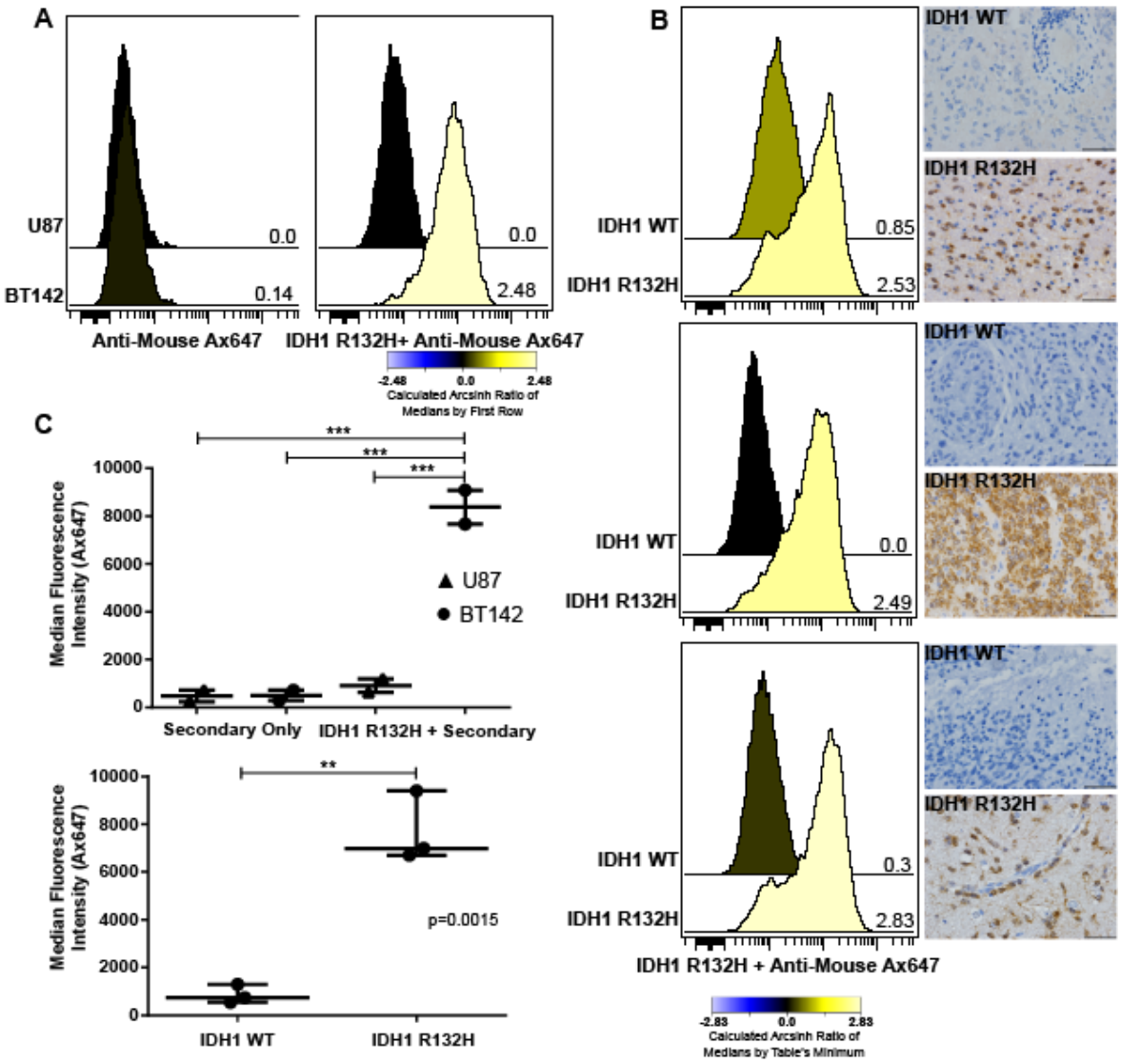


Figure 3.12: **IDH1 R132H is detected in archival FFPE tissue via FFPE-DISSECT and flow cytometry** (a) IDH1 R132H detection in U87 and BT142 cells from FFPE blocks of embedded spheres. Histograms depict relative cell count for each condition (Y axis) and signal from IDH1 R132H antibody staining or anti-mouse Ax488 secondary antibody (x axis). Heat = arcsinh transformed median fluorescence values, first row set to zero. (b) IDH1 R132H detection in six GBM patient samples, 3 *IDH1*-mutant and 3 *IDH1*-wild type (see methods for patient details) from FFPE blocks of tissue (left). Immunohistochemistry for IDH1 R132H on same 6 *IDH1*-mutant and -wild type patient gliomas (right). Histograms depict relative cell count for each condition (Y axis) and signal from IDH1 R132H antibody staining (x axis). Heat = arcsinh transformed median fluorescence values, minimum value set to zero. Scale bar 50μm. (c) Median fluorescence intensities for IDH1 R132H staining on U87 and BT142 cells (top) and 6 glioma patient samples (bottom). BT142 and *IDH1*-mutant glioma samples have significantly higher fluorescent signal from IDH1 R132H antibody staining via flow cytometry.

Discussion

This chapter describes the use of a mutation-specific antibody against IDH1 R132H in glioma cell lines and patient FFPE samples via flow cytometry. A retrieval protocol consisting of heat exposure in citrate buffer was employed to detect IDH1 R132H after finding PFA fixation alone, citrate buffer alone, and 0.02% saponin permeabilization to be insufficient for specific epitope detection. The use of heat-induced epitope retrieval (HIER) for intracellular antigen detection has been described previously [209, 210] and is similar to the protocol for protein detection using the IDH1 R132H clone H09 antibody in the clinical laboratory.

Traditional fluorescence flow cytometry now has the capacity to simultaneously and quantitatively measure 12-18 antigens on hundreds of thousands of cells [223] and is increasingly being incorporated into clinical workflows [224, 225]. Mass cytometry, with the capacity to measure even more features, can be used to identify multidimensional cell phenotypes based on cell surface proteins, transcription factors, and signaling states [103-105]. Cells identified in this manner may be correlated with patient outcome or tumor type, expose vulnerable cell signaling states that can be targeted with small molecules, or provide evidence for tumor cell of origin [114, 151]. Identification of cancer-lineage cells via detection of mutant proteins would be a useful addition to such protocols. To this end, the IDH1 R132H antibody was tested using both fluorescence and mass cytometry on dissociated gliomaspheres.

Given the success of using both the fluorophore- and mass-tagged IDH1 R132H antibody on glioma cell lines (U87-/BT142+), it was surprising that the antibody did not appear to specifically label cancer-lineage cells in *IDH1* R132H-mutant glioma tumors. This study and others [226, 227] have observed autofluorescence in glioma cells, potentially confounding results from the fluorescence flow cytometry experiments. However, the non-specific staining of *IDH1*-wild type samples in mass cytometry experiments suggested that this non-specific labeling is not solely due to variable autofluorescence across samples or cell types. Using a secondary against

the mass-tagged antibody suggested that the primary conjugated IDH1 R132H-171 antibody was functioning as expected on U87 and BT142 cells (Figure 3.10). Thus, it is likely that a property either inherent to glioma cells or conferred during the retrieval process is exposing an epitope that the IDH1 R132H antibody recognizes in both *IDH1*-wild type and -mutant tumors. The consistent use of known control cells (U87, BT142, PBMC), testing of permeabilization techniques (MeOH and HIER), and validation of conjugations with secondary antibodies against the primary conjugate (anti-mouse secondaries against either IDH1 R132H-Pacific Orange or IDH1 R132H-171) were all important in identifying likely sources of this artifact, and underscore the importance of proper antibody validation for each platform where a reagent is used [228].

While dissociated, viably cryopreserved cells exhibited unexpected false positive staining, the use of a protocol designed for retrieval from paraffin, staining, and finally dissociation yielded events consistent with expected outcomes. FFPE-DISSECT was initially developed for interrogation of cell signaling in intact epithelial tissues by flow cytometry [120, 155]. The data reported here on the success of IDH1 R132H detection using FFPE-DISSECT with FFPE glioma tissue suggest that the antibody studied may be useful in future applications of FFPE-DISSECT to larger, retrospective studies of *IDH1*-mutant patient cohorts. In such studies, protein expression and phosphorylation events could be measured, without the disruption of tissue architecture or cell-cell contacts via dissociation. Further, it is an avenue to investigate spatial differences in IDH1 R132H positive cells from a single patient. By using these techniques, researchers can achieve several important objectives: 1) Cancer cells can be positively identified in *IDH1*-mutant gliomas; 2) Basal and potentiated signaling can be investigated specifically in *IDH1*-mutant cancer cells; 3) IDH1 R132H association with expression of other proteins of interest can be measured; 4) retrospective analyses of archived FFPE tissue can be expanded beyond the scope of standard immunohistochemistry.

This work describes the optimal retrieval conditions for IDH1 R132H detection via flow cytometry and demonstrates the utility of this protocol for examining cell line models and FFPE

tissue. This expands the types of samples that can be considered for use in flow cytometry, suspension mass cytometry, and high-dimensional imaging cytometry experiments to include archived *IDH1*-mutant tissue. These studies will allow researchers to more deeply probe the phenotypes, signaling states, and metabolic changes in *IDH1*-mutant glioma cells to understand the biology of these tumors.

Table 3.1: **Glioma CyTOF Panel**

Target	Mass	Clone	Signaling & proteins		Stain		
			Panel	t-SNE	Live	Sap	MeOH
CD45	89	HI30	●	■	✓		
Rhodium	103	-	●		✓		
HLA-ABC	141	W3-32	●	■	✓		
cCasp3	142	D3E9	●				✓
CD117	143	104D2	●		✓		
SSEA1/CD15	144	W6D3	●		✓		
CD31	145	WM59	●	■	✓		
SYP	148	SP17	●	■			✓
SOX10	149	20B7	●	■			✓
SOX2	150	O30-678	●	■		✓	
Nestin	151	25/Nestin	●				✓
CD133	152	AC133	●		✓		
CD171	154	5G3	●	■	✓		
NCAM/CD56	155	B159	●	■	✓		
PDGFRβ	156	18A2	●	■	✓		
S100B	159	19/S100B	●	■			✓
TUJ1	160	TUBB3/TUJ1	●	■			✓
CD44	161	IM7	●	■	✓		
BMX	163	40/BMX	●	■			✓
ITGA6/CD49	164	GoH3	●	■	✓		
EGFR	169	D38B1	●	■			✓
GFAP	170	1B4	●	■			✓
IDH1 R132H	171	H09	●	■			✓
CD90	173	5E10	●	■	✓		
HLA-DR	174	L243	●	■	✓		
Histone H3	176	D1H2	●				✓
CD11b	209	ICRF44	●		✓		

● = included in the panel
■ = included for generation of t-SNE map
Live = live surface stain
Sap = 0.02% saponin stain
MeOH = stain after ice-cold methanol permeabilization

CHAPTER IV

ANALYSIS OF PROTEIN EXPRESSION PHENOTYPES FOUND IN TUMORS PRESENTING IN DIFFERENT LOCATIONS WITHIN THE BRAIN

The following authors contributed to this work: Justine Sinnaeve, Nalin Leelatian, Akshitkumar M. Mistry, Allison R. Greenplate, Kyle D. Weaver, Reid C. Thompson, Lola B. Chambless, Bret C. Mobley, Jonathan M. Irish, and Rebecca A. Ihrie.

Excerpt from this chapter were previously published in Sinnaeve, J., B.C. Mobley, and R.A. Ihrie, *Space Invaders: Brain Tumor Exploitation of the Stem Cell Niche.* Am J Pathol, 2018. **188**(1): p. 29-38. https://doi.org/10.1016/j.ajpath.2017.08.029

Preface

An unanswered question in the field of glioblastoma is role of the normal neural stem cell niche, the ventricular-subventricular zone, on glioma cell biology. It is well documented that tumors that have radiologic contact with the ventricular-subventricular zone (V-SVZ) result in worse outcomes for patients compared to patients with no apparent contact with this region. However, extensive studies to date have failed to identify a mechanism for this observation. I sought to approach this question by looking at differences in protein expression, enrichment, and

complex cellular phenotypes of glioblastoma cells in V-SVZ contacting and non-contacting tumors. However, despite extensive analysis of the data collected in Chapter II and explored further in this chapter, no differences were identified, in line with previous studies. It is important to note that this work does not explore the role of immune cell subsets, which preliminary data suggest vary between the tumor types explored here. The work described below could, nevertheless, inform future directions for studies of the two tumor groups, explored in this chapter and in Chapter V.

Abstract

Increasing evidence indicates that the adult neurogenic niche of the ventricular sub-ventricular zone (V-SVZ) affects the biology of patients with malignant gliomas. Radiographic glioblastoma (GBM) contact with this niche predicts worse prognosis, suggesting a supportive role for the V-SVZ environment in tumor initiation or progression. Previously reported analyses of bulk gene expression and DNA alteration data do not demonstrate differences between tumors that contact the niche compared to tumors that do not, suggesting a need to measure features in single cells from these tumors. In this chapter, protein expression, enrichment, and basal signaling features were explored in glioblastoma cells from a cohort of patients in which V-SVZ niche contact stratified survival. Citrus, a supervised analysis tool, uncovered cell phenotypes that were differentially abundant between contacting and non-contacting tumors. Based on these observations, future studies should include investigation of cell-cell interactions, soluble factors, and extracellular matrix composition of the V-SVZ that may contribute to this tumor-permissive environment.

Introduction

Extensive work over the past decades has demonstrated the existence of two neurogenic niches in the adult mammalian brain: the ventricular-subventricular zone (V-SVZ, sometimes referred to as the subventricular zone (SVZ) or the sub-ependymal zone (SEZ)) and the subgranular zone (SGZ). The cellular constituents, intercellular interactions, and extracellular components of these niches support stem cell maintenance and differentiation (reviewed in [50-52]). The V-SVZ is the larger of the two niches and is located immediately adjacent to the lateral ventricles in the cerebrum. Recently there has been increased focus on the role of this niche in high-grade (III and IV) gliomas, the most common primary malignant neoplasms of the adult brain, as well as its effects on pediatric pontine gliomas [61, 229, 230].

Evaluation of clinical data indicates that glioblastoma patients whose tumors contact the V-SVZ have worse outcomes. A meta-analysis of multiple studies demonstrated that radiographic contact of GBM with the V-SVZ is associated with significantly decreased overall survival [61], independent of extent of tumor resection [60]. V-SVZ contacting glioblastomas also display earlier recurrence after treatment compared to V-SVZ non-contacting GBMs [60, 61], and recurrent tumors are more likely to contact the V-SVZ [231, 232]. Some studies further describe a tendency for V-SVZ contacting glioblastomas to be multifocal at the time of diagnosis [60, 232, 233], and to recur after treatment at sites distant from the initial tumor site [233-235], although the latter observation is debated [60, 236]. Strikingly, GBM contact with the SGZ has not been found to influence survival [60], suggesting that features unique to the V-SVZ contribute to outcome. Access to the ventricles and cerebrospinal fluid (CSF), the presence of a gap layer, and the proximity of the V-SVZ to major white matter tracks are all features unique to the V-SVZ. At the molecular level, niche enriched factors may be derived from different cellular sources, present at different levels, or delivered through specific cell-cell contacts.

A subset of glioma tumors appears to spread specifically within the ventricular-subventricular zone (also referred to as subependymal spread) [63-65]. V-SVZ spread is evaluated by magnetic resonance imaging (MRI) in pediatric and adult gliomas and is interpreted as the presence of contrast enhancement and/or abnormally elevated T2-weighted signal within the subependymal region [237-239]. Though microscopic analysis of this region is uncommon due to the rarity of resections which include the V-SVZ, histologic sections show increased cell density at time of resection, including cytologically atypical glial cells with enlarged, hyperchromatic, angular nuclei expressing the cell-cycle associated antigen, Ki67 [53, 69, 78]. Both *isocitrate dehydrogenase* (*IDH*)- wild type and *IDH*-mutant GBMs as well as brain metastases have demonstrated such spread through the subependyma [63, 64]. Although *IDH*-mutant gliomas occur more frequently in the frontal lobe than *IDH*-wild type tumors, there appears to be no difference between the two groups in V-SVZ contact [240].

Analyses of public datasets using bulk tissue analyses have not identified transcriptional signatures unique to V-SVZ contacting tumors [67]. Only a limited number of candidate signatures have been found, which may not be cancer cell derived [241]. Given that neural stem cell niche components may enhance glioma initiation, maintenance, and/or recurrence, the interaction between the V-SVZ and tumor cells warrants investigation. In this chapter, single cell mass cytometry data on 28 *IDH*-wild type primary patient samples was used to investigate differences between tumors that contact the V-SVZ and tumors that do not contact the niche, based on radiographic imaging. A variety of high dimensional data analysis tools were applied to the single cell data to parse phenotypes, cell clusters, and protein enrichment. There were no significant differences between the two tumor groups in this dataset, as assessed by multiple different metrics. Citrus, a supervised data analysis tool, did uncover two cell subpopulations that were differentially abundant in V-SVZ contacting and non-contacting tumors. These findings suggest that future studies should explore probe the capacity of cells from these tumors to respond to

stimuli that might be differentially available in the V-SVZ compared to other sites in the brain or contributions of non-cancer lineage cells to differences in tumor behaviors and patient outcomes.

Methods

Patient samples

Surgical resection specimens of 28 *IDH*-wild type glioblastomas collected at Vanderbilt University Medical Center between 2014 and 2016 were processed into single cell suspensions following an established protocol (Appendix A and [121]). Only samples that were confirmed to be *IDH*-wild type glioblastomas by standard pathological diagnosis were used. Samples were categorized as V-SVZ contacting or V-SVZ non-contacting as described in [67]. Briefly, magnetic resonance images of the brain were available for all patients. The preoperative brain images were assessed for V-SVZ contact by a neurosurgeon and neuroradiologist without knowledge of patient outcome. Using OsiriX Lite software (version 9.4, Pixmeo, Geneva, Switzerland), VSVZ-contacting GBMs were identified by the contact or involvement of the post-contrast tumor enhancement with the lateral ventricular ependyma. All samples were collected with patient informed consent in compliance with the Vanderbilt Institutional Review Board (IRBs #030372, #131870, #181970), and in accordance with the declaration of Helsinki.

Patient characteristics and collection of clinical data

Additional patient characteristics are included in Chapter II (Table 2.3) for all samples in this study. All patients were adults (≥ 18 years of age) at the time of their maximal safe surgical resection of their cerebral (supratentorial) glioblastomas. Extent of surgical resection was independently classified as either gross total or subtotal resection by a neurosurgeon and a

neuroradiologist. Gross total resection was defined as agreement by both viewers of no significant residual tumor enhancement on patients' gadolinium-enhanced magnetic resonance imaging (MRI) of the brain obtained within 24 hours after surgery. All patients were considered for treatment with postoperative chemotherapy (temozolomide) and radiation according to the standard of care [2], after determination of *MGMT* promoter methylation status by pyrosequencing (Cancer Genetics, Inc., Los Angeles, CA, USA). Multiplex polymerase chain reaction (PCR) was used to determine *IDH1/2* mutational status. Patients' postoperative course was followed until February 2019, noting time to first, definitive radiographic progression or recurrence of glioblastoma as agreed upon by the treating neuro-oncologist and neuroradiologist, and the time to patients' death. All deaths were deemed to be due to the natural course of patients' glioblastoma. Median overall survival of the analyzed 28 patients with *IDH*-wild type glioblastoma was 388.5 days (13 months) and median PFS was 187.5 days (6.3 months), which is typical for the disease [2, 3].

Mass cytometry analysis

Data pre-processing: Cells derived from patient samples were prepared as previously described [121]. A multi-step staining protocol was used, which included 1) live surface stain, 2) 0.02% saponin permeabilization intracellular stain, and 3) intracellular stain after permeabilization with ice-cold methanol. All antibodies used, including clone information, and the steps when used are given in Table 2.4. After staining, cells were resuspended in deionized water containing standard normalization beads (Fluidigm) [166], and collected on a CyTOF 1.0 instrument located in the Cancer and Immunology Core facility at Vanderbilt University. Mass cytometry standardization beads were used to remove batch effects and to set the variance stabilizing arcsinh scale transformation for each channel following field-standard protocols [121, 146, 167]. Rhodium viability stain and cleaved caspase-3 antibody were included in staining to exclude non-

viable and apoptotic cells, respectively. Detection of total histone H3 was used to identify intact, nucleated cells [98]. A 34-dimensional mass cytometry antibody panel was used to analyze over 2 million viable cells from 28 tumors (ranging from 5,875 to 336,284 cells per tumor). Data were normalized with MATLAB-based normalization software [166], and were arcsinh transformed (cofactor 5), prior to analysis using the Cytobank platform and statistical programming environment R via R Studio [168].

Dimensionality Reduction: A patient-specific t-SNE view was generated on Cytobank, using 26 of the measured markers for all tumor and stromal cells from each patient's tumor [125]. Immune ($CD45^+$) and endothelial cells ($CD31^+$) were computationally excluded, via manual gating, from each individual patient prior to subsequent downstream analysis. Remaining $CD45^-$ $CD31^-$ cells were included in a common t-SNE analysis, generated using 24 of 34 measured markers (equal sampling of 4,710 cells, 20,000 iterations, perplexity 30, theta 0.5). Unless otherwise noted, these cells and the common t-SNE analysis were used for the remaining data analysis.

Marker Enrichment Modeling: MEM analysis was performed in R, using the previously published R package [158]. In short, MEM captured and quantified cell subset-specific feature enrichment by scaling the magnitude (median) differences between clusters, depending on the spread (IQR) of the data. These values were then computed in comparison to the remaining cells in a given dataset. MEM values were interpreted as either being positively enriched (▲, UP positive values) or negatively enriched (▼, DN negative values). The variation of a given cellular feature across cell subsets was quantified as ± standard deviations (SD).

Citrus: Citrus analysis [127] was implemented in Cytobank. Events from the common t-SNE analysis described above were uploaded to a new experiment and input to the Citrus algorithm. Minimum cluster size was set to 1%, cross validation folds was set to 5%, and the false discovery rate was set to 1%. The Nearest Shrunken Centroid (PAMR) association model was used to assess differences in cluster abundance between two groups, V-SVZ contacting and V-

SVZ non-contacting (n=14 for each group). Figure 4.8 depicts outputs from this analysis including comparison of cluster abundance, histograms of marker expression for each cluster, and a hierarchical clustering tree.

Earth Movers Distance: Earth Movers Distance was performed as described in [146]. Briefly, events from each tumor were analyzed on a common t-SNE (equal sampling of 5875 cells per tumor, 10,000 iterations, perplexity 30, and theta 0.5). These events were downloaded from Cytobank. EMD was calculated between each pair of files using the "transport" library for R (https://cran.r-project.org/web/packages/transport/citation.html). The optimizations described in [146] were applied as described. R was used to create a heatmap to visualize the calculated EMD score. Statistical comparisons of EMD values between groups were calculated using an ANOVA in Graphpad Prism.

Population Abundance Analysis: The abundance of cell populations identified in Chapter II from RAPID was compared for all tumors. Percent abundance between V-SVZ contacting and V-SVZ non-contacting was compared using a two tailed t-test in Microsoft Excel.

Survival and statistical analysis

Statistics were calculated in Microsoft Excel or Graphpad Prism software (noted in methods for each analysis). Statistical significance α was set at 0.05 for all statistical analyses, one- or two-tailed noted in figure legends. Time from surgical resection to death (overall survival, OS) was depicted using right-censored Kaplan-Meier curves and analyzed in R. Survival time points were censored if, at last follow up, the patient was known to be alive. Differences in the survival curves of groups were compared using a univariate Cox regression model to estimate the effect size (hazard ratio, HR, of death) on survival with 95% confidence intervals between the survival curves, and quantify its statistical significance with a p-value.

Results

Earth Mover's Distance comparison of glioblastoma mass cytometry data

28 primary glioblastoma samples were collected during the time of first resection and confirmed to be *IDH*-wild type. The median progression-free-survival (PFS) and overall survival (OS) after diagnosis were 6.3 and 13 months, respectively, typical of the trajectory of this disease [2]. A Kaplan Meier curve and univariate cox regression analysis indicated that V-SVZ contact stratified overall patient survival (Figure 4.1) as it had in larger cohorts [61].

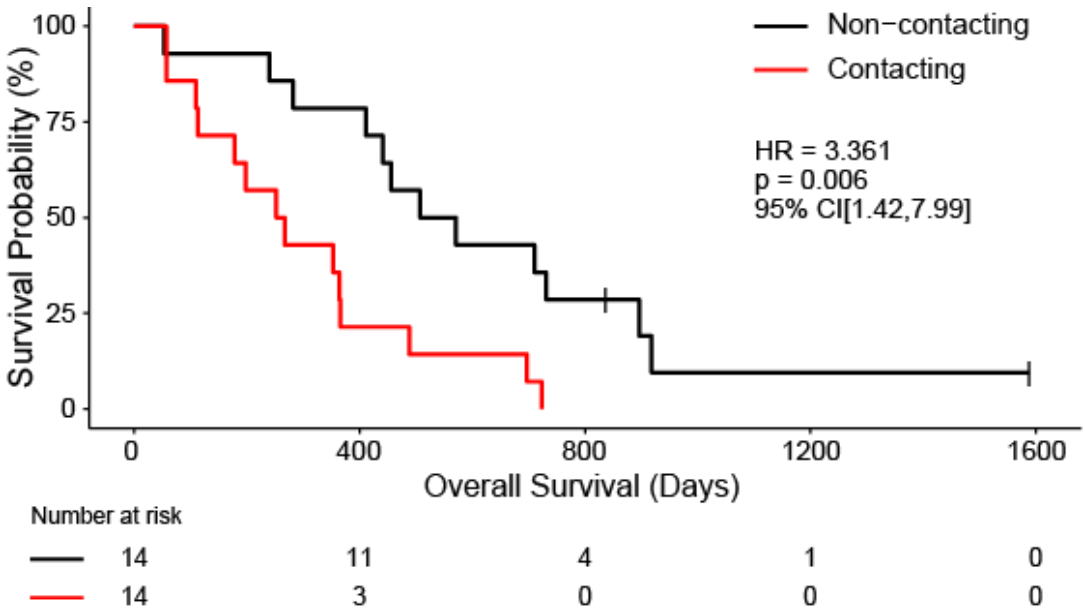


Figure 4.1: **V-SVZ contact stratifies overall survival.** A Kaplan Meier curve depicts overall survival in days (x-axis) for patients with a V-SVZ contacting tumor (red line) or a tumor that does not contact the niche (black line). Percent survival is indicated on the y-axis. Vertical ticks represent censored patients who were alive at the indicated time. The hazard ratio of death for V-SVZ contact is 3.361 with a 95% confidence interval of [1.42-7.99] and a p-value of 0.006.

Mass cytometry data was generated for the patient samples using a custom antibody panel designed to probe brain lineage proteins, stem cell markers, and phospho-signaling events. Data from each patient sample was analyzed to identify live events and remove debris and non-cellular events. Patient-specific t-SNE analyses were performed using 26 of the measured features on all of the cells from the sample (Table 2.4). Immune (CD45+) and endothelial (CD31+) cells were evident in varying abundances in each tumor, consistent with prior studies of gliomas [33, 98]. CD45+ or CD31+ cells were computationally excluded from each patient prior to subsequent analyses. All non-immune, non-endothelial events (CD45-, CD31-) were analyzed in the same t-SNE analysis using 24 of the measured features.

Earth Mover's Distance (EMD) was used to quantify the differences between the t-SNE maps (Figure 4.2). Earth Mover's Distance can be described as the energy required to take one plot of cells and move it to match another [242]. Therefore, the more divergent two t-SNE plots are, the higher the EMD score, and the more dissimilar these two samples are in the measured feature space. A heatmap was generated to depict the differences between individual patient t-SNE maps. Dark blue represents an EMD score of zero, indicating that the maps are identical. The diagonal dark blue line down the heatmap represents each patient sample compared to itself, and the heatmap is reflected over this line. Dark red represents the highest EMD scores and thus shows that the two samples are very different. Looking at the EMD heatmap, it is evident that non-contacting tumors are not overall more similar to other non-contacting tumors compared to contacting tumors in this comparison. This is reflected in the quantification of EMD scores, where non-contacting tumors were compared to other non-contacting tumors (intra-group), contacting tumors were compared to contacting tumors (intra-group), and non-contacting tumors were compared to contacting tumors (inter-group). If tumors within the same category were more similar to each other in high dimensional phenotypic space, smaller EMD scores would be expected for intra-group comparisons than inter-group comparisons. However, there are no significant differences between the three different EMD comparisons (Figure 4.2). This suggests

that non-contacting and contacting tumors are equivalently variable in high dimensional space for the features measured.

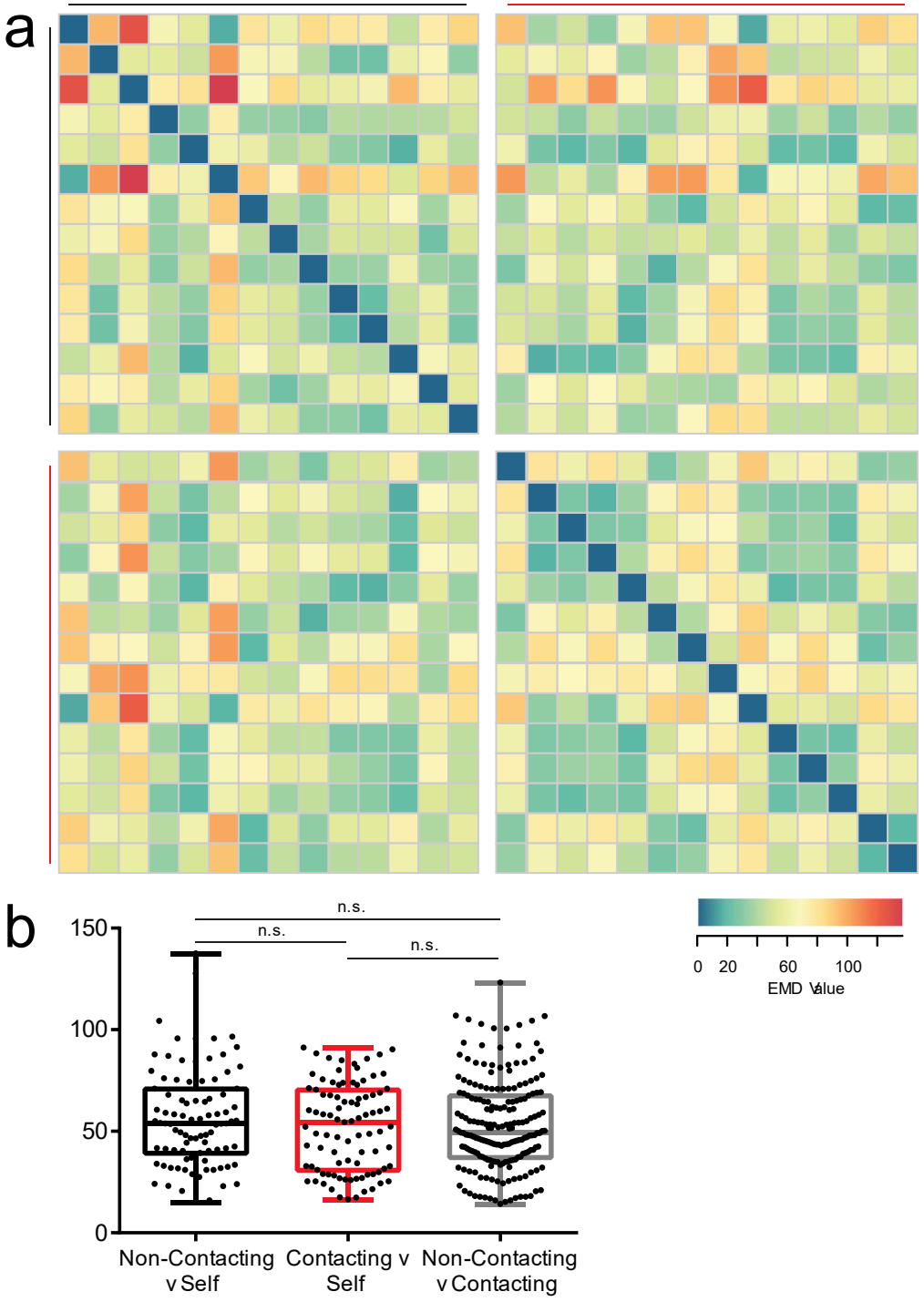


Figure 4.2: **Earth movers distance evaluation of patient t-SNEs demonstrates that contacting and non-contacting tumors are no more similar to tumors of the same category than tumors in the other category**. (a) Heatmap depicting EMD scores. 28 patient samples are arranged according to their contact status (non-contacting tumors: black line; contacting tumors: red line). The heatmap is reflected over the diagonal. (b) Box and whiskers plot of EMD values when non-contacting tumors are compared to other non-contacting tumors (black), contacting tumors are compared to other contacting tumors (red), or non-contacting tumors are compared to contacting tumors (gray). Each dot represents a comparison between two tumors. The box encompasses the 25^{th} to 75^{th} percentile. Bars extend to the minimum and maximum. The horizontal bar denotes the median. An ANOVA analysis indicated no significant difference between the EMD values for these three comparisons.

Comparison of median protein expression using mass cytometry

In order to investigate global differences in protein expression between contacting and non-contacting tumors, median protein expression was compared for all measured features (Figure 4.3). Box and whisker plots depict the median protein expression for each feature on each sample. In some cases, like that of EGFR, non-contacting samples had more tumors with higher values compared to contacting tumors. The opposite pattern was observed for Nestin, with several contacting tumors having the highest median expression of this protein. However, comparing the median mass intensity across all features on all samples, no measured feature was significantly different between the two groups (t-test).

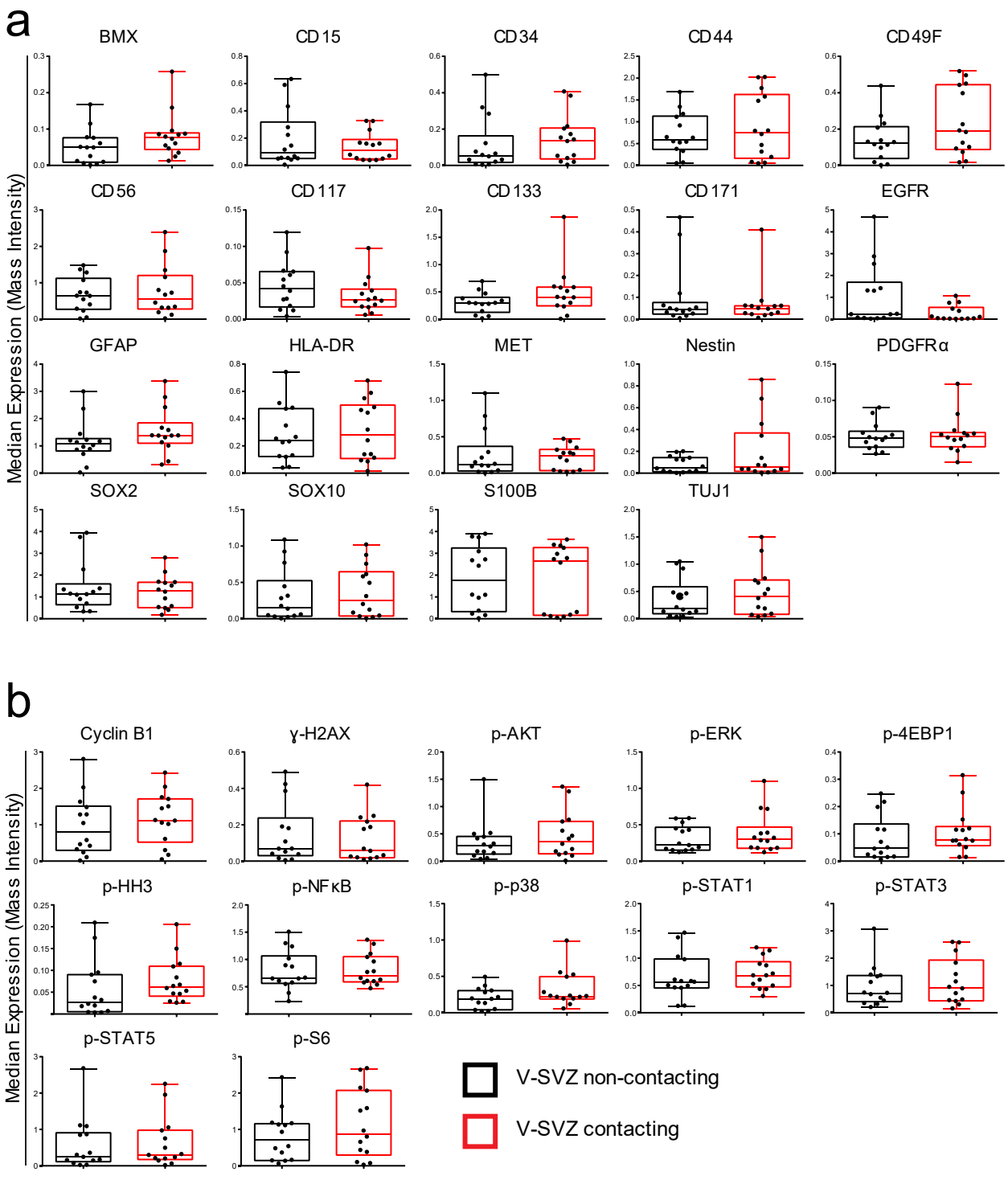


Figure 4.3: **Median mass intensity does not differ for any measured feature between contacting and non-contacting tumors. (**a) Dot plots comparing median mass intensity (y-axis) between non-contacting tumors (black) and contacting tumors (red). Each dot is a single patient tumor, the box outlines the 25th to 75th percentile, while bars extend to the minimum and maximum. The horizontal line indicates median value. (b) Comparison of basal phospho-signaling, cell cycle

protein Cyclin B1, and DNA damage protein, ɣ-H2AX as depicted in a. Each plot was compared using an unpaired, two-tailed t-test in Graphpad Prism.

Marker Enrichment Modeling identifies features enriched on each patient sample

Median protein expression, while insightful, does not encompass all the important differences that might be contained in single cell data. In order to gain insight into marker expression and variation across a sample, marker enrichment modeling (MEM) was used [158]. MEM calculates a score, on a scale from -10 to 10, with -10 indicating that a feature is specifically not expressed on a given sample, while a score of 10 indicates that it is highly and specifically expressed on that sample. A score of zero indicates that either this feature is heterogeneously expressed across the sample or that it is expressed at the same level across all samples. The MEM algorithm generates a heatmap (Figure 4.4) of enrichment scores, where light blue represents -10 and bright yellow represents +10. It also gives a MEM score for each feature on each patient sample (Figure 4.5). Individual MEM analyses were performed respectively on lineage and stem-like proteins together, and on phospho-signaling proteins. By analyzing these types of features separately, smaller differences in signaling proteins could be detected without the risk of being obscured by high enriched features like S100B. Patient samples are displayed by category on the MEM heatmap, with non-contacting tumors (black squares) on the left and contacting tumors (red squares) on the right. The heatmap visualization does not illuminate any consistent pattern of enrichment between the two groups. For example, there are both non-contacting and contacting samples that are positively enriched for GFAP and CD44 while others are negatively enriched for these proteins (Figure 4.4). To quantitively compare the MEM scores, box and whisker plots were generated for each feature, comparing non-contacting tumors to contacting tumors (Figure 4.5). Statistical analysis of these plots, using a t-test for each marker, resulted in no significant differences between the two types of tumors with respect to marker enrichment scores.

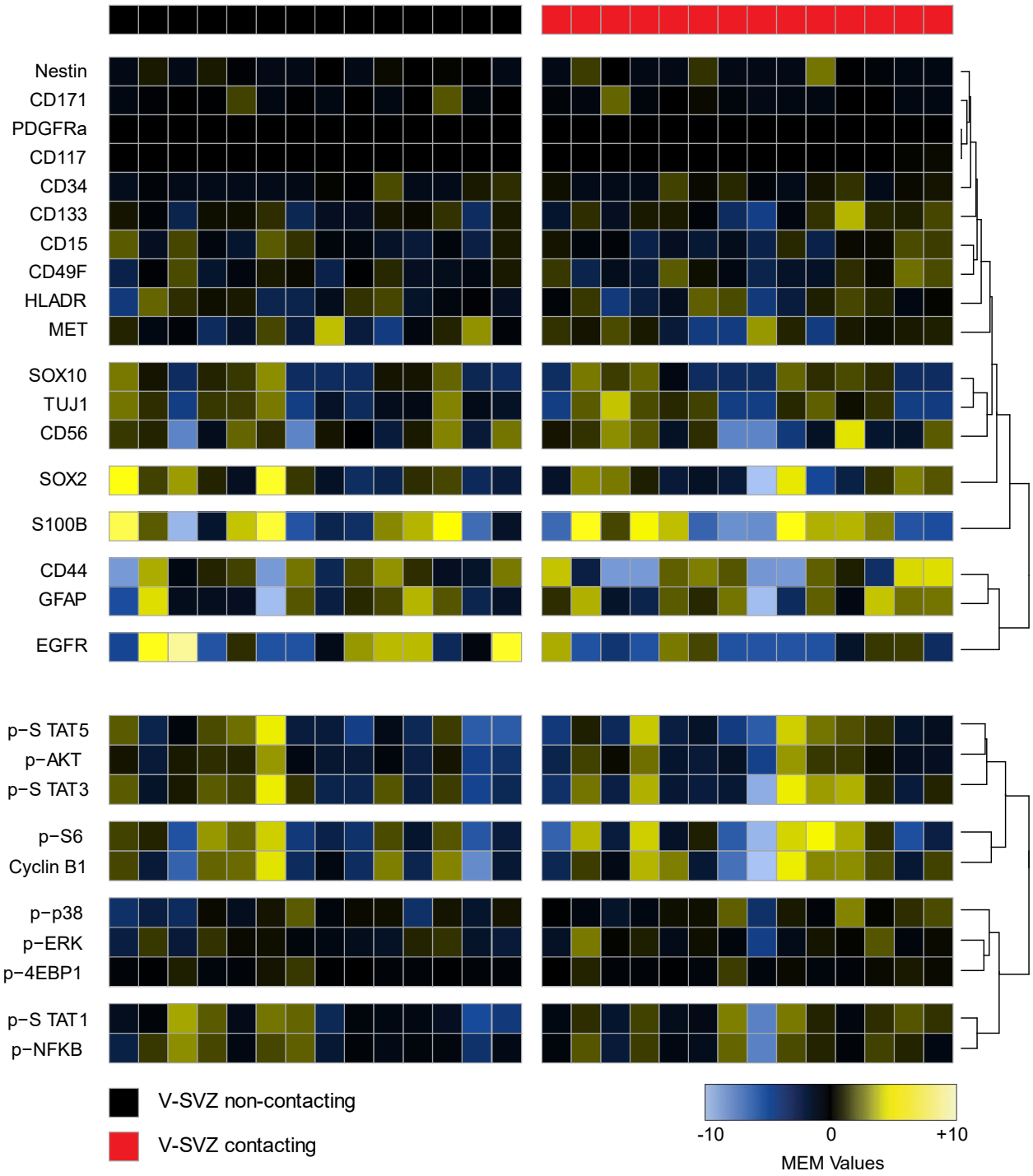


Figure 4.4: **MEM heatmap demonstrates no consistent trends between non-contacting samples and contacting samples**. 28 patient samples are arranged according to their contact status (non-contacting tumors: black squares, left; contacting tumors: red squares, right). A heatmap represents the MEM values of glioblastoma tumors (columns). Hierarchical clustering was performed based on MEM values and is depicted on the right of the heatmap for measured features. Protein features are depicted in the top heatmap and signaling features are depicted underneath.

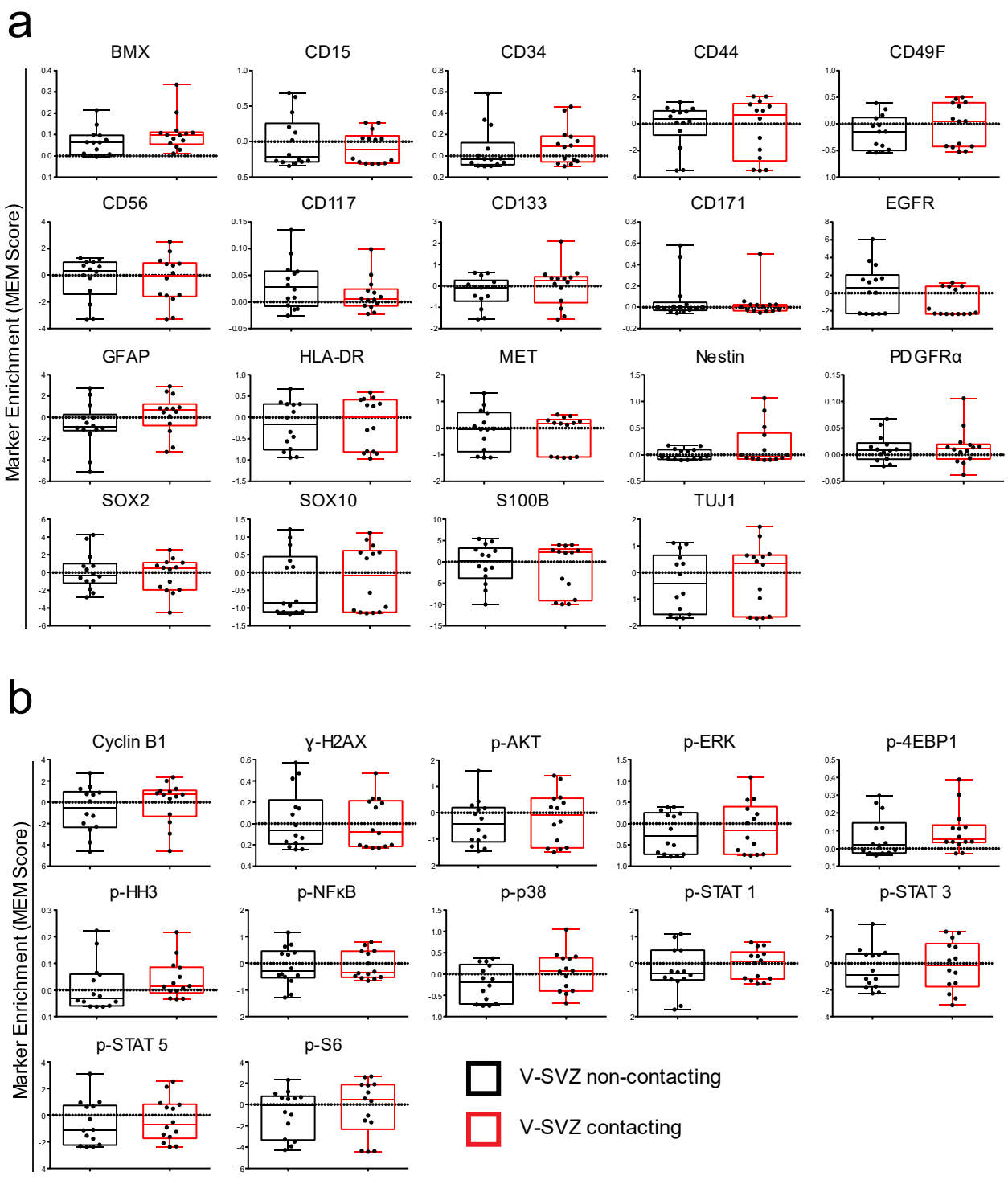


Figure 4.5: **Marker enrichment modeling scores do not significantly differ for any measured feature between contacting and non-contacting tumors.** (a) Dot plots comparing MEM score (y-axis) between non-contacting tumors (black) and contacting tumors (red). Each dot is a single patient tumor, the box outlines the 25th to 75th percentile, while bars extend to the minimum and maximum. The horizontal line indicates median value. (b) Comparison of phospho-signaling, cell

cycle protein Cyclin B1, and DNA damage protein, ɣ-H2AX as depicted in a. Each plot was compared using an unpaired, two-tailed t-test in Graphpad Prism.

Although there were no significant differences between the MEM scores from contacting and non-contacting samples, another way to identify broad differences between the two groups is to compare enrichment of features in combined files. In this case, all cells from non-contacting tumors were concatenated into one file and all cells from contacting tumors were concatenated into a second file. MEM was used to analyze these two files and several interesting features were identified in the heatmap and reported MEM scores (Figure 4.6). First, neither group was positively enriched for any specific feature. This is perhaps not surprising, given the fact that there were no significant differences in the MEM scores between patient samples. However, the contacting samples are highly negatively enriched for EGFR, meaning that these samples specifically lack EGFR compared to non-contacting tumors. Furthermore, when phosphorylation events are considered, there are striking trends. Cells from contacting tumors are enriched for almost every phospho-protein measured, while cells from non-contacting tumors are lacking active signaling events. This is depicted in the MEM labels calculated for each file (Figure 4.6).

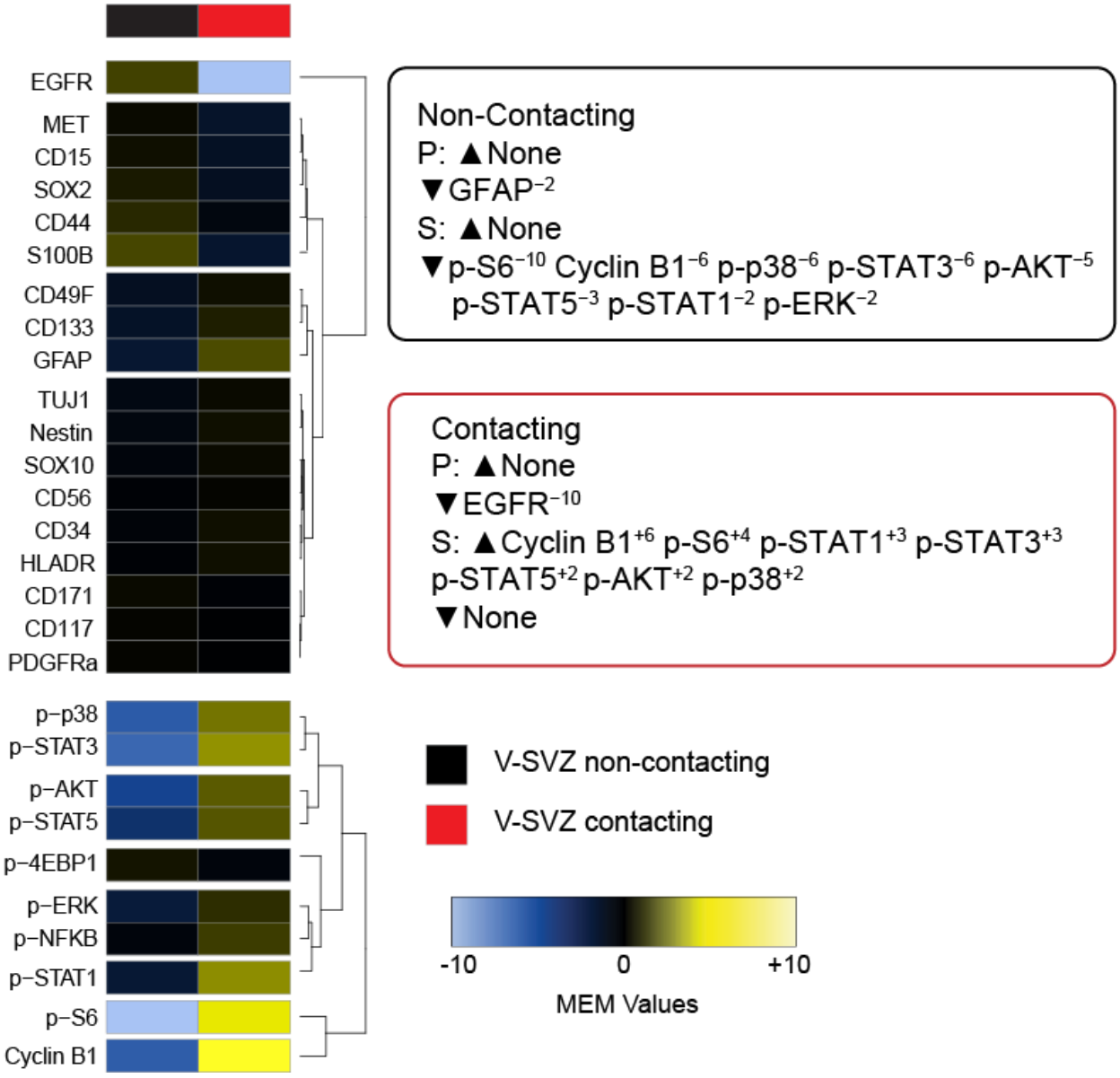


Figure 4.6: **MEM analysis of concatenated files reveals trends in EGFR expression and phospho-signaling**. 14 non-contacting samples were combined into one file and are represented by the first column (black square). 14 contacting samples were combined into one file and are depicted in the second column (red square). A heatmap represents the MEM values of these files. Hierarchical clustering was performed based on MEM values and is depicted on the right of the heatmap for measured features. Protein features are depicted in the top heatmap and signaling features are depicted underneath. MEM scores for each file are in rectangular boxes. Non-contacting file is in black and the contacting file is in red. Positively enriched proteins follow the ▲ and negatively enriched, or absent proteins follow the ▼. The MEM score is depicted in the superscript following each listed feature.

Glioma Positive Prognostic phenotypes are more abundant in non-contacting tumors

The observation that non-contacting tumors contain more EGFR and generally lack active phospho-signaling is reminiscent of the phenotypes observed in glioma positive prognostic cells described in Chapter II. To determine if glioma positive prognostic (GPP) cells are more abundant in non-contacting tumors, the percentage of glioma negative prognostic (GNP) or GPP cells were calculated for both non-contacting (black) and contacting (red) tumors (Figure 4.7). Neither GNP nor GPP cells were more abundant in contacting or non-contacting tumors. Interestingly, the two tumors with the highest percentage of GPP cells and the two tumors with the highest percentage of GNP cells were all non-contacting samples (Figure 4.7). The abundance of all 43 of the cell populations identified by RAPID in Chapter II were also compared between V-SVZ contacting and non-contacting tumors (data not shown). A t-test indicated that no single population significantly differed in abundance between these two groups.

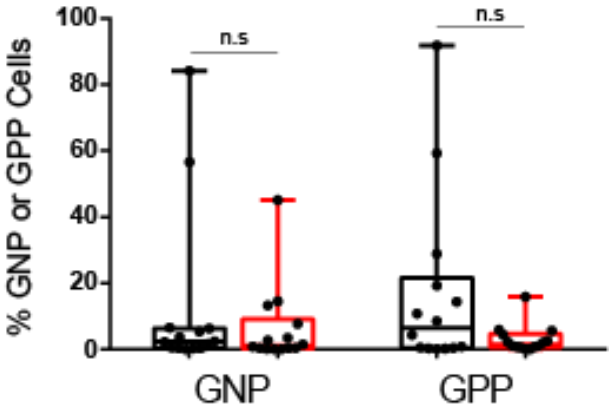


Figure 4.7: **GNP and GPP cell abundances are not significantly different between contacting and non-contacting tumors**. Box and whisker plots depict percent of GNP (left) or GPP (right) in each tumor, represented by the points. Non-contacting tumors (black) are not significantly different from contacting tumors (red) for either GNP or GPP cell abundance (two-tailed t-test). Box outlines the 25th to 75th percentile and whiskers extend to the minimum and maximum.

GNP and GPP cells were identified using RAPID (Chapter II) which was specifically developed to handle continuous variables, like overall survival. However, contact status as measured in this study, is not a continuous variable, and is therefore amenable to approaches that require samples to be categorized. For this reason, Citrus (cluster identification, characterization, and regression) was used to identify any cell populations that were differentially abundant between non-contacting and contacting samples [127]. Using this tool, two cell clusters were significantly associated with contact status (Figure 4.8). Cluster 131867 was more abundant in non-contacting tumors while cluster 131871 was more abundant in contacting tumors. Interestingly, these clusters had very similar phenotypes across multiple markers used to denote GNP cells (i.e. S100B, SOX2, p-S6, p-STAT5, p-STAT3) but diverged with respect to EGFR, GFAP, and CD44, all of which are enriched in the GPP phenotype. Cluster 131867 expressed higher levels of EGFR compared to all other clusters including 131871. Cluster 131867 also expressed higher levels of GFAP and CD44 compared to most other cells, however, cluster 131871 cells demonstrated higher GFAP and CD44 expression compared to 131867 (Figure 4.8). Overall, this is additional evidence that there are cells with a GPP-like phenotype (high EGFR, GFAP, and CD44 and low phospho-signaling) that are enriched in non-contacting tumors. Interestingly, there are also cells enriched in contacting tumors that demonstrate some, but not all, features of GPP cells (high GFAP and CD44 expression but low EGFR expression).

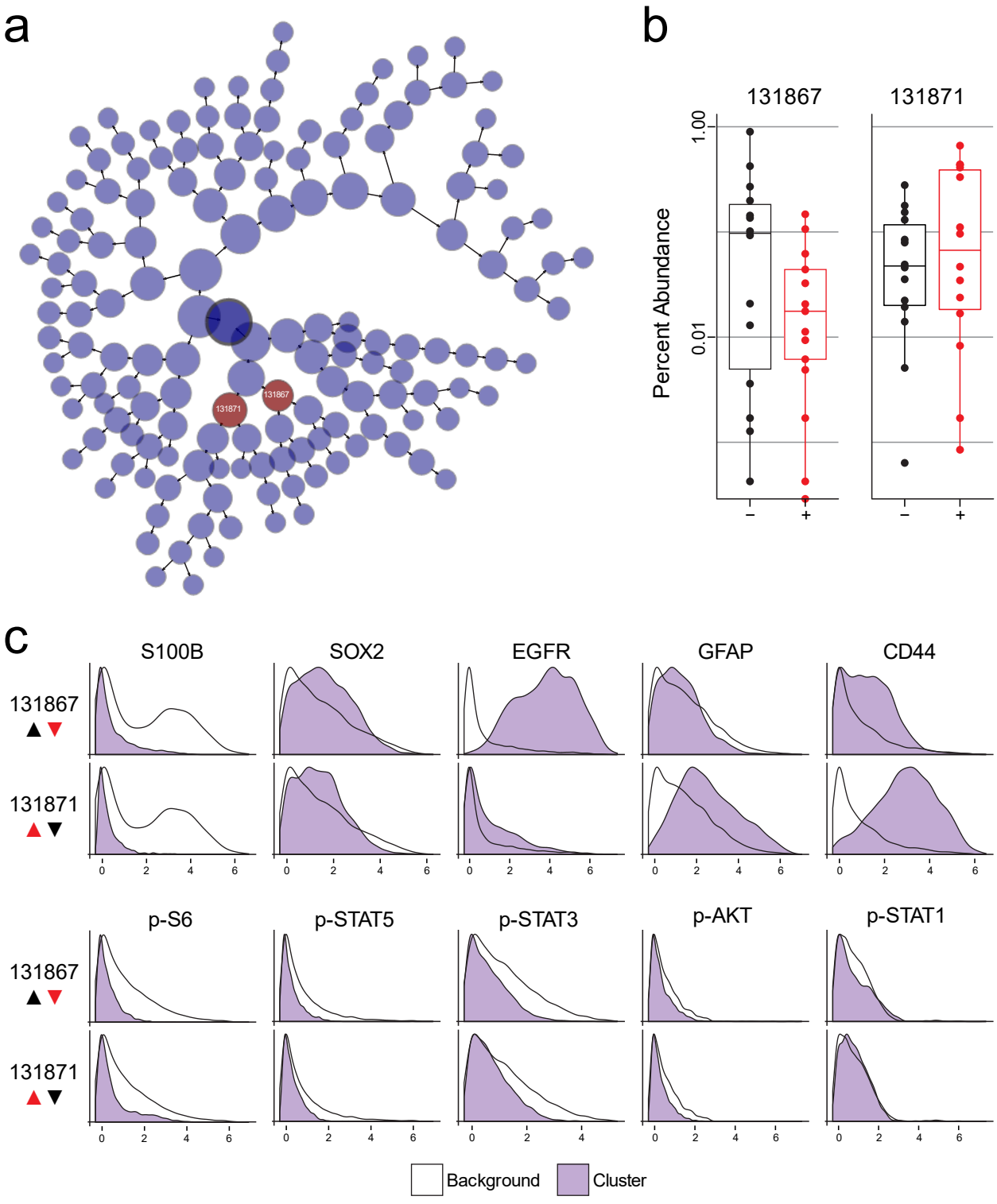


Figure 4.8: **Citrus identified two cell clusters that were associated with contact status**. (a) Cell clusters identified by the Citrus algorithm implemented in Cytobank. Significant clusters (131867 and 131871) are colored in red. (b) Abundance (y-axis) of each significant cluster in non-contacting (black) or contacting tumors (red). (c) Histograms of protein expression for a subset of the measured features. Purple histogram depicts data for the indicated cluster and white

histogram depicts data for all remaining cells not in that cluster. Red arrowhead (▲) indicates if the cluster is more or less abundant in contacting tumors while the black arrowhead (▲) indicates if the cluster is more or less abundant in non-contacting tumors.

In previous analyses reported in Chapter II, tumors with more GPP cells also had higher levels of immune cells. To determine if immune cell abundance correlated with tumor contact status, the percent of CD45+ cells was compared between contacting and non-contacting tumors (Figure 4.9). A t-test revealed that there was no statistical difference in the percent of CD45+ immune cells between the two groups.

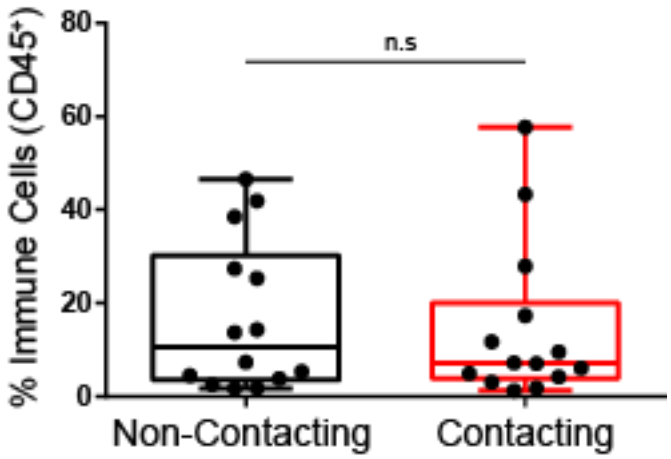


Figure 4.9: **The abundance of immune cells is not significantly different between non-contacting and contacting tumors**. Box and whisker plots depict percent of immune cells in each tumor, represented by the points. Non-contacting tumors (black) are not significantly different from contacting tumors (red) for either GNP or GPP cell abundance (two-tailed t-test). Box outlines the 25th to 75th percentile and whiskers extend to the minimum and maximum.

Discussion

The study of neurogenic niches in the adult brain is a rich and active field of research. Beyond the function of these sites as generators of new neurons, emerging data reveal significant effects of the ventricular-subventricular zone niche on the behavior of malignant gliomas. Treatment options for these aggressive neoplasms are limited and patient survival remains dismal; therefore, the potential role of this niche in tumor initiation, maintenance, or recurrence merits further research. The purpose of this work was to investigate basal protein expression, signaling features, or glioblastoma cell subsets that were differentially associated with V-SVZ contact. Despite a significant difference in patient outcomes between V-SVZ contacting and non-contacting (Figure 4.1), no differences were detected in median protein expression, MEM protein enrichment values, or specific cell subsets like immune cells or populations automatically identified in Chapter II. Furthermore, high dimensional representations of the data (t-SNE plots) were not significantly different between the two groups. Interestingly, Citrus uncovered two populations that were differentially abundant in contacting and non-contacting tumors.

These findings are consistent with previous findings, describing a surprising lack of large differences in gene expression of single genes or gene networks as well as previously defined molecular subclasses between contacting and non-contacting tumors [28, 67, 68]. Investigations of expression of common stem-cell genes indicated no correlation between these genes or gene sets and V-SVZ contact [68, 243, 244]. Some studies have proposed molecular features that may contribute to the phenotypes or development of V-SVZ contacting tumors including proteins involved in centromere assembly, cell cycle, chromosome segregation, epigenetic regulation, metabolism, and invasion [245, 246]. However, these studies compare the V-SVZ contacting tumors to non-malignant brain samples or cultured neural stem cells, not non-contacting tumors as is described in this chapter.

However, other groups have identified putative differences that were not specifically interrogated in this work. Analysis of TCGA gene expression and imaging data from GBM samples revealed that proneural and neural tumor subtypes were more frequently found closer to the V-SVZ compared to classical or mesenchymal tumors when distance to the V-SVZ was measured on a continuous scale (rather than a binary contacting versus non-contacting categorization) [247]. In a cohort of 23 patients, V-SVZ contact was associated with vascularization gene sets and decreased hypoxia signatures while non-contacting tumors had reduced stem cell gene expression and an association with inflammation pathways [248]. A later study found an association of notch pathway signaling features with V-SVZ contact in gene expression analysis of 36 patient samples [241].

Although this work does not immediately point to a mechanism of action for the negative prognostic correlation with neural stem cell niche contact, the clinical data suggest a possible distinct, aggressive biology of V-SVZ contacting tumors and it is a prognostic feature that should be investigated further. The prognostic value of V-SVZ contact may be due, in part, to the specialized contacts and multiple cell types found in the niche, including a prominent vascular component, contact with the ventricles and the CSF within, and proximity to the corpus callosum (the major white matter tract of the brain). These features might impact signaling between niche and tumor cells. Immediate areas for study include *in situ* measurements of glioma-associated proteins and phospho-signaling molecules, to avoid confounding factors like the disruption of tissue architecture during dissociation. The abundance of formalin-fixed, paraffin-embedded tissue that is collected and prepared from glioma samples and improving imaging technologies that can measure many features in a single cell make this an achievable goal [249-251]. Additionally, these modalities would allow researchers to determine if there is a distinct cellular spatial organization associated with the V-SVZ that may be obscured in dissociated samples. Further questions include the capacity for tumor cells to respond to known niche factors like those identified in [70] or constituents such as cerebrospinal fluid or media conditioned by stem cells or

choroid plexus [79, 230]. Removing tumor cells from the brain microenvironment and thus, potentially important sources of stimuli, may stunt key signaling differences *ex vivo*.

The ultimate goal of this work is to change patient outcomes, and to this end, patients presenting with tumors in different locations may require different treatments. Improvements in tumor therapy could include targeting niche factors, like soluble growth factors, cell-cell junction proteins, or cell surface ligands, or disrupting niche-tumor cell interactions, with radiotherapeutic targeting of the V-SVZ representing a first step along this course [80, 90, 252, 253]. As our knowledge of the normal niche continues to expand, newly revealed features may also drive better understanding of tumor etiology and therapy response. Areas of interest include the impact of cell-to-cell heterogeneity and lineage priming within normal neural stem cells (NSCs) on the disease state, and the contribution of microglia, the major innate immune population within the brain, to normal and tumor-bearing V-SVZ. Studies from the mouse brain indicate that neural stem cells are spatially diverse, meaning that stem cells from different regions of the V-SVZ give rise to different progeny [51, 254-259]. Examination of spatial differences in tumor-forming or tumor-homing capabilities of heterogeneous NSCs could further inform the design of targeted therapies. Detailed investigations of the role of innate and adaptive immune cells in this niche will also be critical to understanding how these tumors may evade immune detection or targeted immunotherapy approaches (e.g. anti-PD-1/CTLA-4 agents). Microglia help to define and maintain the neurogenic niche, and recent studies indicate that the V-SVZ resident population is functionally distinct and temporally dynamic, exhibiting an immature phenotype which changes with organismal age[76, 260].

Finally, glioma research currently benefits from the ample available patient tissue from primary tumor resections, as well as a plethora of imaging data collected during routine care. One persistent challenge is the integration of molecular information (bulk and single-cell genomic and proteomic approaches) with spatial information obtained from clinical imaging. Recent advances in both the preparation of single-cell suspensions and the collection of high-dimensional data will

enhance our ability to map specific populations of cancer and niche cells, providing a better understanding of the impact of V-SVZ niche diversity on tumor behavior [98, 99].

CHAPTER V

CONCLUSIONS AND FUTURE DIRECTIONS

Summary and significance

Gliomas are complicated and serious health concerns that, to date, have not benefitted from the targeted or personalized medicine revolution that has changed outcomes for malignancies like lung cancer, melanoma, leukemia, or breast cancer [261-265]. As single cell technologies and sequencing approaches improve, the scientific community continues to learn more about the cell types present in glioma tumors, their interactions with the immune system, and the differences between patients [37, 47, 266]. In this dissertation, I contribute to this growing and important field by leveraging mass cytometry to measure proteins and signaling features in single glioma cells. These measurements were used to identify cell types associated with patient outcomes and investigate differences in glioblastomas in different brain regions.

Along with colleagues from the Ihrie and Irish labs, I helped to establish an efficient and effective way to isolate single, viable cells from glioma tumors, which directly led to the ability to measure proteins and signaling events at the single cell level (Appendix A) [98, 121]. In Chapter II, I describe the collection and analysis of mass cytometry data resulting from these advances, as well as the development of a new data analysis algorithm to identify prognostic cell subsets, called Risk Assessment Population IDentification (RAPID). This algorithm is a valuable tool for the field, as it provides a modular, unsupervised, and automatic pipeline for understanding single cell data in many disease contexts. Using RAPID, I identified two novel types of glioma cells,

glioma negative prognostic cells (GNP) and glioma positive prognostic cells (GPP). The phenotypes observed in these cells lend insight into potentially effective therapies for patients whose tumors contain these subsets. Further, patients with low percentages of both GNP and GPP cells can still benefit from this analysis approach, as single cell phenotypes for the cells within their tumors are also characterized by RAPID. This information could change how patients are chosen or assigned for clinical trials.

In Chapter III, I expand the tools that can be used to measure proteins in single cells by describing approaches for detection of the mutated isocitrate dehydrogenase (IDH1) R132H protein. Currently, there are no definitive methods to detect cancer-lineage cells using proteins in glioma tumors. Over 80% of grade II and III tumors contain a mutation in *IDH1*, and of those, 90% contain the *IDH1* R132H mutation [14, 15, 22]. The ability to detect this protein in experimental samples would be incredibly powerful for the study of cancer-lineage cells specifically. Furthermore, understanding the differences in the metabolism and epigenetics of *IDH*-wild type and *IDH*-mutant tumors will be crucial to establish effective treatment strategies for these two groups of patients [267-269]. As described in Chapter III and further elaborated in Appendix A, new and modified flow cytometry approaches were critical to detecting IDH1 R132H. However, there were significant challenges in detecting the protein in patient samples. This technical challenge is an area ripe for future studies. Not only is IDH1 R132H an incredibly important protein in the study of glioma biology, but additionally other targets and reagents are being developed at an exciting pace. Having the infrastructure to address technical difficulties in the detection of these targets will improve researchers' ability to answer impactful scientific questions.

Finally, in Chapter IV, I use mass cytometry data to understand the role of the neurogenic ventricular-subventricular zone (V-SVZ) in glioma biology and how GNP and GPP cells might be related to V-SVZ niche contact. It is well established that the V-SVZ has a unique and nuanced role in the behavior of gliomas [60, 66, 70, 270]. Patients whose tumors contact the V-SVZ have worse outcomes than patients whose tumors do not contact the V-SVZ [61, 232, 239]. Tumors

that contact the V-SVZ at presentation are more likely to have multifocal recurrences, suggesting that the impact of niche contact persists even after surgical resection [61, 233, 236]. Furthermore, even in tumors where radiographic evidence suggests the tumors are not in direct contact with the V-SVZ, individual cells can be found residing in the V-SVZ, just one example of V-SVZ tropism exhibited by tumor cells [63-65, 86, 271]. Although we found no differences in protein expression or enrichment, cell subset abundance, or total immune cell infiltration in the tumors studied here, these findings lead to alternative hypotheses that can be explored using existing tools, developed by our group and others [62, 67, 100, 155]. These ideas are explored below.

In summary, this dissertation advances the understanding of brain tumor composition at the cellular level, the biology of cells specifically involved in patient prognosis, and describes a new tool that can be used to parse different types of data from many different disease contexts. Furthermore, building on the methods described here could standardize or improve the field of tissue processing and single cell data acquisition. The ultimate goal of this work is to positively impact human health, whether it is through better categorization of patients for inclusion in clinical trials of targeted therapies or identifying new targets or new combinations of targets for therapeutic investigation.

Machine learning and the application of RAPID to new datasets

In Chapter II, I describe the development and implementation of RAPID, a novel and automatic data analysis algorithm for high dimensional single cell data analysis. Chapter II outlines many of the validation and testing approaches used to strengthen and legitimize RAPID and to generate confidence in its use as a tool for other applications outside of our research group. It is my goal that other groups will use this tool on their datasets, to identify prognostic cell subsets in other disease settings or cells associated with other continuous variables of interest.

Although other workflows and algorithms have been developed to identify cell populations of interest in cancer samples (reviewed in Table 1.1), many require a level of prior knowledge, which may not always be available, especially for solid tumors. For example, Phenograph and developmentally dependent predictor of relapse (DDPR) depend on maps of developmental lineage in healthy tissue, while Citrus and Cytofast are both supervised methods that require pre-clustered data and samples prospectively assigned to a group, but RAPID enables analysis with continuous, ungrouped data [113, 114, 127, 178]. In studies of diseased human tissue, especially primary glial tumors where healthy samples are infrequently obtained for reference and the developmental lineage is largely quiescent in adults, it is difficult to anticipate the number of expected unique phenotypic subsets, making it particularly valuable for the analysis to be independent of prior knowledge. RAPID is designed to be free from supervision in the identification of the number of clusters and in the assessment of cluster abundance in tumors.

Additional exploration of RAPID that was outside the scope of this work includes substituting additional dimensionality reduction algorithms or clustering algorithms in the place of t-SNE and FlowSOM. UMAP is described above as an effective replacement for t-SNE, but other iterations of the t-SNE tool, such as opt-SNE, fit-SNE, or H-SNE were not explored [134, 138-140]. Each of these dimensionality reduction tools has benefits that may be suitable for different datasets (Table 1.1). FlowSOM was chosen as the clustering tool in RAPID because it is widely recognized to be a useful and efficacious clustering algorithm [129, 130]. However, there are many other options that may be more appropriate for different datasets or in different applications. dbscan, Phenograph, and X-shift might all be reasonable next steps for substitution within the workflow [113, 123, 130, 135].

It was a challenge to find additional datasets on which to test RAPID, though I do not believe this is due to a lack of existing data. Rather, publications infrequently report both patient outcomes and make raw data files publicly available. Although I describe the results from a B-cell precursor acute lymphoblastic leukemia dataset analyzed using RAPID, I was unable to procure

additional mass cytometry datasets for further validation [114]. However, as RAPID is freely available on the GitHub, research groups that have such datasets and access to variables of interest (e.g., patient outcome) can use this tool immediately. Additionally, RAPID was designed around a test case of mass cytometry data but is intended to work well for any single cell data. Such datasets include single cell RNA-seq studies and high dimensional imaging studies, including imaging mass cytometry (IMC), multiplexed ion beam imaging (MIBI), or highly multiplexed immune-fluorescence imaging [118, 119, 249]. These types of imaging modalities have been increasingly used in recent years, especially in cancer studies. One of the benefits of these technologies is the spatial information that is retained by studying intact tissue. Distance from features of interest (for example the V-SVZ) is a continuous variable that would be very amenable to analysis by RAPID. In this way, phenotypes of cells enriched either near to or far from the ventricular surface in brain tumors could be assessed [118]. In other cancers, features of interest may include distance from blood vessels, a hypoxic or necrotic core, or the stem cell niche of the tissue of origin. Therefore, the application potential of RAPID is quite broad, encompassing various diseases, clinical features, and data types.

GNP and GPP cells

RAPID was applied to glioblastoma mass cytometry data to identify positive and negative prognostic cell types, named GPP and GNP respectively. These cells and their respective phenotypes were novel findings that had not previously been described in gliomas. The GNP and GPP subsets correlated with survival, independent of the effects of other widely accepted prognostic factors (age [171, 172], *MGMT* promoter methylation status [10, 173], and treatment including extent of surgical resection [174, 175], therapy with temozolomide [2], and radiation [176, 177]). As noted in Chapter II, some aspects of the GNP phenotype were consistent with previous findings. S100B and SOX2 are widely expressed in gliomas, and as might be expected,

the combination of a stem cell marker and an astrocyte protein is indicative of an aberrant phenotype [181, 203, 204, 272]. Additionally, the combination of p-S6, p-STAT3, and p-STAT5 signaling was of note, as it provides a potential starting point for therapeutic intervention [99]. P-STAT5, specifically, has not previously been identified as a negative prognostic feature in patient samples. An important part of the GNP cell identity was that the cells specifically lacked EGFR, GFAP, and CD44. On the other hand, EGFR, and to a lesser extent CD44 and GFAP, were important defining features of the GPP phenotype, in combination with lower phospho-signaling patterns. This was interesting for several reasons. EGFR and CD44 were previously identified to be critical to defining different molecular subclasses based on bulk tumor sequencing analysis [28]. Further, the question of the prognostic significance of EGFR in glioma outcomes has generated a large and inconclusive body of work [194-196]. Studies report conflicting results, including no association with outcome, association with worse outcome when measured at recurrence, and moderate negative impact on outcome in some studies of primary samples. These findings are at odds with the results reported here, though that is not completely surprising. Many of the aforementioned studies were conducted using bulk tumor samples, or focus on copy number analysis and gene expression analysis [28, 164]. The study reported in Chapter II is the first of its kind to examine protein expression in single glioma cells and to consider multi-dimensional phenotypes when identifying prognostic features. However, this discrepancy is an excellent starting point for further investigation into the GNP and GPP cells identified by the work in this dissertation.

First, much of the prior work in gliomas and glioblastoma has been focused on gene expression and DNA alterations. Therefore, it would be very informative to close the information loop of DNA, RNA, and protein in tumor samples generally, as well as specifically with the GNP and GPP populations. Using RNA-seq information, the molecular subclass to which tumors with high GNP or high GPP content belong could be determined. Even more important would be to sequence single cells in the GNP and GPP clusters to determine whether these cells fall into any

of the previously defined transcriptional classes [28, 37, 47, 48]. This would also help resolve whether alterations in key genes, such as *EGFR*, or gene expression levels are translating to the protein level. Furthermore, in my work, EGFR protein was detected on the cell surface (in mass cytometry) or via the c-terminal domain (immunohistochemistry). This does not include specific measurement of the common EGFR mutant, EGFRvIII [273]. Sequencing studies could lend insight into the expression of this protein in either the GNP or GPP compartment. Given its previously described co-expression with wild type EGFR, one would expect to find it in GPP cells, however, its expression is generally associated with poor outcomes, suggesting EGFRvIII may be enriched in GNP cells [273, 274]. Sequencing studies would also help tease out the impact of SOX2, a stem cell transcription factor, and phospho-STAT signaling on gene expression. These proteins, through their transcription factor functions, could be driving changes in the proteome and thus contribute to the maintenance or transition of GNP cells in a specific cellular state. Such cell states, and transitions between them have been increasingly described in the field over the past few years [47, 48, 183].If this is the case, that SOX2 or STAT proteins are driving a more aggressive or treatment resistant phenotype, downstream effectors whose expression are controlled by these transcription factors may be ideal therapeutic targets.

Another advance that would push this work forward would be the development of a sorting method for isolating GNP and GPP cells. It may be possible to sort live GPP cells using EGFR and CD44, but the distinctive features of GNP cells described here are all intracellular, precluding live cell sorting at this time. Sequencing to determine other enriched features may highlight cell surface targets that could be used to isolate these cells, a strategy employed in other studies [47]. Once live cells are isolated, several interesting questions can be interrogated. First, it is known that cells grown in culture can vary dramatically from their tumor of origin or the phenotypes displayed *in vivo* [191, 275, 276]. The stability of the GNP and GPP phenotypes *in vitro* would be important information for determining the utility of *in vitro* experiments. If one were to assume that the phenotypes could be preserved in culture, the impact of therapeutics on cell survival,

proliferation, and phenotypes could be interrogated. For example, an interesting question would be to uncover the plasticity of the GNP phenotype, especially when GNP cells are challenged with small molecule inhibitors against p-STAT3, p-STAT5, and p-S6. GNP cells may shift to other phospho-signaling cascades to survive such an onslaught or alternatively, be eliminated [99]. If cells cannot be outright eliminated, perhaps such a treatment can shift the cells to a less aggressive or less harmful phenotype. A culture system that allows researchers to test new therapeutics, or new combinations of available reagents would be incredibly powerful. Furthermore, this type of system could be employed to co-culture glioma cells with immune cells, an increasingly important area of investigation in glioma biology, especially in light of the result that immune cell abundance was correlated with GNP and GPP abundance. I would expect that GNP cells would have a negative impact on the health of immune cells, while GPP cells would tolerate immune presence. These types of interactions would be incredibly important for predicting the efficacy of immunotherapies in different patients.

If the GNP or GPP cells cannot be maintained in cultures (either they perish outside of the *in vivo* environment or they change phenotypes on plastic culture dishes), they may be able to be propagated in patient-derived xenograft models (PDXs). In this case, questions about tumorigenicity, tropism to the V-SVZ, and response to therapies *in vivo* may be addressed. It would also permit analysis of the stem-like potential of each of these populations. If xenografts derived from GNP cells are made entirely out of GNP cells, it suggests that this population does not contain a cell that can repopulate the multitude of phenotypes observed in human patient tumors. The same would be true for a GPP-derived xenograph that consisted entirely of GPP cells. However, if many cell types are observed, it would lend credence to the idea that these populations are plastic and able to generate a number of diverse phenotypes. This capacity has been shown for many glioblastoma cell subsets identified in previous work [37, 39, 47].

A critical question regarding the GNP and GPP cells is whether they are preserved in recurrent tumors. Currently, recurrent GBMs are rarely resected because these resections have

not been demonstrated to be particularly beneficial for most patients, and because recurrent tumors are more likely to be multifocal and infiltrative, complicating the surgical process [13, 277, 278]. Additionally, at recurrence, patients have a lower Karnofsky Performance Status (an indicator of overall health), making additional surgeries riskier. For this reason, second, third, or even fourth resections are much less common than primary, untreated tumors, and the composition of such resected samples is poorly understood. GNP and GPP cells were identified in pre-therapy, untreated patient samples, suggesting that these phenotypes are linked to biological mechanisms of therapy response or tumor detection by the immune system. It would be very informative to collect recurrent samples, perhaps in collaboration across multiple institutions to achieve a meaningful sample size, to determine if the GNP and GPP phenotypes persist in recurrent, post-treatment samples. If GNP subsets have the capacity to evade therapy and retain their active proliferation properties, recurrent tumors would be expected to contain higher proportions of GNP cells and have a more uniform phenotype. Although loss or gain of genetic aberrations post-temozolomide and radiation therapies has been investigated, little is known about signaling in recurrent tumor cells and thus it is unclear if clonal evolution and/or a shift in activated phospho-proteins is necessary for tumor cell survival and repopulation. I would anticipate that the GNP phenotype would persist and be more prevalent in recurrent samples compared to primary samples. However, there is certainly the possibility that an entirely new phenotype emerges due to the influence of tumor debulking, radiation, or chemotherapy [279, 280]. If the GNP phenotype does persist, it may reinforce the value of aggressively targeting this cell type. If new phenotypes emerge, the role of SOX2 and the STAT proteins as mediators of transcription may be responsible for the shift in cell state.

In Chapter II, I used RAPID to uncover the GNP and GPP cells in patient samples. However, there is much to be learned about these populations and how knowledge of their existence can change how patients are cared for. I am optimistic that this work is the foundation for future studies that can answer some of the questions above.

Mutant protein detection in brain tumors

In Chapter III, I describe the detection of the mutant IDH1 protein, IDH1 R132H. Using a combination of techniques common in flow cytometry and immunohistochemistry, the protein of interest was detected in control cell lines and FFPE tissue from patient samples. However, in dissociated patient samples, the antibody displayed unexpected and non-specific staining patterns. Extensive validation and testing suggest that this is not due to alterations in the antibody structure itself after conjugation to a heavy metal (for mass cytometry) or to a fluorophore (for fluorescence flow cytometry). Using secondary antibodies to detect the conjugated versions provides evidence that they are fully functional in control cell lines. Therefore, additional explanations for its non-specific performance in patient samples must be considered.

First, the preparation of the glioma patient samples may be altering an epitope that subsequently becomes available for antibody binding. Thus, although optimization of patient sample preparation was an important part of my dissertation work, our current protocol may not be ideal for detection of this specific mutant protein (Appendix A) [121]. Alternative preparation strategies should be considered, such as the commercially available gentleMACS™, avoiding cryopreservation, or additional enzymatic conditions. Another consideration is the difficulty in assessing unsorted patient samples via fluorescence flow cytometry. In our hands, we see an abundance of debris in glioma samples run through the fluorescence cytometer and, compared to cell lines, the cells in tumors are irregular in shape and size, both of which contribute to intrinsic fluorescent signal. This makes the platform challenging to use to analyze samples that have not been previously enriched for cells of interest.

The current inability to use the IDH1 R132H antibody on dissociated samples is contrasted by the success of FFPE-DISSECT (FFPE-disaggregation for intracellular signaling in single epithelial cells from tissue) on glioma tissue samples. FFPE-DISSECT was designed with epithelial tissues in mind, as indicated in the name, so its application to brain tissue is a significant advance [155]. In the work described above, FFPE-DISSECT was leveraged to detect IDH1

R132H using fluorescence flow cytometry. However, FFPE tissue is also highly amenable to IMC, allowing many antigens to be detected at one time in single cells [250]. Knowing that the antibody can be conjugated to a heavy metal and that it is specific for mutant tumors in FFPE tissue suggests that IMC would be a viable option in this situation. IMC provides many of the same advantages as mass cytometry, albeit, at a lower throughput of cells. However, spatial information would be retained unlike in dissociated samples.

Careful examination of possibly rare cancer cells in the infiltrative edges of patient samples has the potential to reveal novel biology. By measuring IDH1 R132H in FFPE tissue, the cancer cells could be distinguished from the non-malignant surrounding. The abundance of IDH1 R132H positive cells that have infiltrated beyond the radiological border is currently unknown, as is the potentially distinct biology of these more infiltrative cells. Measuring multiple parameters such as stem cell-like markers, metabolic factors, or intracellular signaling molecules using FFPE-DISSECT or IMC could illuminate how these cells differ from tumor bulk and how they may be specifically targeted.

The work described in Chapter III highlights the importance of extensive and thorough validation of antibody staining conditions, titrations, and testing on a variety of controls (in our case, cell lines, peripheral blood mononuclear cells, and glioma patient samples). Although many in the flow cytometry community have been aware of the challenges of antibody-based work, recent attention has been directed specifically at antibody validation and titration [228]. By reporting both the positive and negative results in this dissertation, I hope to provide a roadmap for others seeking to validate antibodies in new contexts and for those observing unexpected results.

V-SVZ contact and glioma biology

In Chapter IV, I explore the mass cytometry data first introduced in Chapter II, to understand a phenomenon observed in many groups, including our own: patients presenting with a glioblastoma that has radiographic contact with the V-SVZ perform worse than patients whose tumors present with no evidence of radiographic V-SVZ contact [61, 232]. Although this phenomenon is well accepted in the field, a mechanism for this observation remains elusive. Previous work by our group indicated that there are no discernable differences in gene expression profiles or DNA alternations between these two patient groups [67]. Since gene expression does not always reflect protein expression and mass cytometry provides an opportunity to measure post-translational modifications, I believe mass cytometry data analysis is a valuable addition to this field. I was surprised that despite the myriad of analysis approaches applied, no statistically significant differences between V-SVZ contacting and non-contacting tumors was observed in this dataset. Individual features were not significantly different, nor were abundances of cell clusters first introduced in Chapter II. One interesting finding was that the data analysis tool Citrus was able to identify two clusters that were different between the two groups, with one cluster in particular resembling the previously described GPP phenotype. This cluster was more abundant in non-contacting tumors, perhaps hinting at a mechanism that explains the better outcomes observed for non-contacting tumors.

An interesting trend uncovered by MEM, demonstrated that V-SVZ contacting tumors generally had higher basal phosphorylation than non-contacting tumors. This did not rise to the level of statistical significance, but it does suggest that the ability to signal through pro-survival, proliferation, or migration pathways is perhaps enhanced by contact with the V-SVZ [70]. As outlined in Chapter I and Chapter IV, the V-SVZ contains a multitude of unique cell types that, in healthy brain tissue, are important for maintaining the neural stem cell niche [281]. The factors that these cells express or secrete could be co-opted by tumor cells in close proximity to facilitate escape from therapy and repopulation of the tumor after treatment [50]. One way to address the

potential for these cells to signal in response to V-SVZ-derived factors is to stimulate cells *ex vivo* with factors of interest. In such a scenario, glioblastoma cells from patient samples would be exposed to either specific factors in isolation or combinations in conditioned media or cerebrospinal fluid. Following stimulation, measurement of phospho-proteins of interest could be performed using flow cytometry. Mass cytometry is likely to be best suited for this goal, due to the absence of autofluorescence issues and the capacity to measure more features at a single time. One hypothesis is that cells from tumors contacting the V-SVZ will have increased signaling through key phospho-proteins including p-S6, p-STAT3, p-STAT5, and p-Src after stimulation with niche factors than cells from non-contacting tumors. Alternatively, cells from both tumors may respond equivalently, suggesting that it is access to these factors that is variable between contacting and non-contacting tumors. One way to address this question would be to implant cells from contacting or non-contacting tumors either proximal or distal to the V-SVZ in mouse, orthotopic, patient-derived xenografts. By isolating the resulting tumor cells from such xenografts, one would expect to find that cells closer to the V-SVZ would be actively signaling and that cells further from the niche to would be less active through the measured signaling pathways.

This type of analysis could be facilitated by IMC or highly multiplexed immunofluorescence imaging [249, 250]. As mentioned above, mass cytometry, though very useful for some applications, does not preserve location information. Cell position could be very informative in several contexts, not the least of which is in examining V-SVZ influence on GBM. First, IMC may help confirm that phenotypes observed in mass cytometry data are representative of *in vivo* phenotypes. Samples for IMC need not be dissociated, a process that almost certainly alters the biology of protein expression and phosphorylation in some cells. Therefore, data derived from IMC may better reflect the state of cells within the tumor prior to surgery. Location specific information would also allow us to investigate whether the cells that are in direct contact with the V-SVZ are different from cells in the same tumor that are not in contact with the niche. If access to factors is truly important, we would expect to see that cells far away from the V-SVZ resemble

cells from non-contacting tumors. However, if contact with the V-SVZ fundamentally alters the biology of the tumor, even cells further from the V-SVZ within the sample may exhibit similar phenotypes.

Another important area for exploration is the role of the immune system in glioblastoma tumors, specifically in the context of V-SVZ contact. While the results reported in Chapter IV demonstrate that total $CD45^+$ immune cells are not enriched in either contacting or non-contacting tumors, I did not explore the abundance or activity of immune cell subsets. Preliminary data from our group suggests that specific immune cell subsets are preferentially abundant in contacting tumors compared to non-contacting tumors, and that like the glioblastoma cells, these immune cells may have differential capacities for signaling. In depth analysis of such subpopulations can illuminate which types of cells are present in each tumor, the pro- or anti-tumor activity of each subset, and avenues for manipulation that may result in more efficacious applications of immunotherapies.

Further investigation into the biologic mechanism of differing outcomes between patients with contacting tumors compared to non-contacting tumors will be critical to appropriately treat these groups. The work described in this dissertation builds on previous work from our own group and others that describe these two tumor groups and generates data to support the conclusion that basal differences in DNA, RNA, and proteins are not are the root of the differences [62, 67]. This opens the door to more complex interactions between tumor cells and non-malignant niche cells that can be explored using the tools developed here and elsewhere.

Concluding remarks

The field of glioblastoma research is ripe for application of new technologies and approaches to innovate solutions that will improve patient lives. Collectively, the field has been implementing single cell technologies and *ex vivo* models to better understand cell phenotypes

and behaviors [37, 47, 100, 189, 190]. This dissertation builds on this work by developing protocols to prepare glioblastoma for mass cytometry analysis, a custom antibody panel for interrogation of brain lineage proteins, stem cell makers, and post-translational modifications, a new data analysis algorithm to uncover prognostic cells, and techniques for detecting cancer-lineage specific features [98, 121]. Using these tools, I describe glioma negative and positive prognostic cells, as well as interrogate differences in tumors that contact the neural stem cell niche (the V-SVZ) compared to tumors that do not contact the V-SVZ. Not only will these discoveries advance further glioblastoma research, but the tools can also be implemented across a variety of diseases and data types. RAPID specifically has the capacity to change how individuals view their high dimensional data from flow cytometry, RNA sequencing, or imaging studies. Future studies investigating single cell phenotypes and responses to current or novel therapies will be crucial for development of new strategies to treat patients.

APPENDIX A

PREPARING VIABLE SINGLE CELLS FROM HUMAN TISSUE AND TUMORS FOR CYTOMIC ANALYSIS

Authors: Nalin Leelatian, Deon B. Doxie, Allison R. Greenplate, Justine Sinnaeve, Bret C. Mobley, Rebecca A. Ihrie, and Jonathan M. Irish

This work appears in manuscript form in Leelatian, N., et al. (2017). "Preparing Viable Single Cells from Human Tissue and Tumors for Cytomic Analysis." Curr Protoc Mol Biol 118: 25C 21 21-25C 21 23. https://doi.org/10.1002/cpmb.37

Preface

This work has been modified from its published form [121] in this Appendix to include staining protocols developed and described in Chapter III. All of the tumors used for mass or fluorescence flow cytometry in this dissertation (excluding FFPE samples as noted) were prepared as described in this chapter. Furthermore, the steps outlined in basic protocol 2 were those for all of the fluorescence and mass cytometry experiments.

Abstract

Flow cytometry, encompassing fluorescence and mass cytometry, is a single cell biology technique that can sample >500 cells per second and measure 2-45 features per cell. This combination of technical assets has powered a series of recent cytomic studies where investigators used fluorescence or mass cytometry to measure protein and phospho-protein expression in millions of cells, characterize cell types in healthy and diseased tissues, and reveal novel, unexpected cells. However, these advances largely occurred in studies of blood, lymphoid tissues, and bone marrow, since the cells in these tissues are readily obtained in single cell suspensions. This Appendix establishes a primer for single cell analysis of solid tumors and tissues and has been tested with mass cytometry, and whose fundamentals are applicable to fluorescence-based studies as well. The cells obtained from these protocols can be fixed for study, cryopreserved for long-term storage, or perturbed *ex vivo* to dissect responses to stimuli and inhibitors.

Introduction

One key method for understanding a tissue or organ is to dissect and identify the diverse cells that comprise it. Flow cytometry excels at quantifying the abundance and protein expression signatures of hundreds to thousands of cells per second [282] and holds great promise for understanding diseases like cancer, where altered protein expression and signaling activity in rare cell subsets can contribute to oncogenesis and drive treatment resistance [149, 150]. The ability of flow cytometry to quantify proteins on each of millions of cells and reveal signaling in rare, 1-in-10,000, cells has made it indispensable to modern immunology and clinical hematopathology, where cells in suspension are readily obtained. Fluorescence flow cytometry

(FFC) is a decades old technology based on detection of fluorescent molecules in single cells [283]. Early experiments measured 2-4 features, while newer cytometers aim to reach up to 50 parameters. Mass cytometry (MC) is a newly developed form of flow cytometry with the ability to measure 35 or more features at a rate of 500 or more cells per second [112, 147, 159, 282, 284]. This expanded detection capacity is ideal for characterizing the diverse cells present in human tumors, which typically include endothelial cells, epithelial cells, fibroblasts, immune cells, and malignant cells [98].

Flow cytometry provides outstanding statistical power to detect rare cells and to quantify the cellular identity of millions of cells, compared to other techniques that are limited to hundreds or thousands of cells [149, 282, 284]. This technology has been implemented in studies of donor and patient cells that are obtained as a suspension, such as blood and bone marrow [112, 124, 167, 285-288] or that can be disaggregated from lymphoid structures by mechanical force alone [115, 152, 289-291]. Clinical diagnoses of blood malignancies use fluorescence flow cytometry characterization of cell surface marker expression, as well as cell subset quantification [292-295]. Additionally, flow cytometry has been used clinically to identify minimal residual disease and to detect disease progression in leukemia [125, 296, 297]. Fluorescence flow cytometry has also been applied to studies of solid tissues and tumors for research purposes [39, 298-302].

In addition to their ability to characterize cell surface markers, flow cytometry technologies allow simultaneous detection and quantification of intracellular targets in individual cells [150, 151, 219]. Commercially available fluorescence flow cytometers generally measure 8 to 12 targets per cell using target-specific antibodies conjugated to individual fluorophores [223]. The number of targets is limited due to the overlap of emission spectra of different fluorophores, although advances in flow cytometry technology continue to increase this number [303, 304]. Mass cytometry is a newer flow cytometry–based technology that allows detection of more than 35 targets in individual cells. Instead of conjugation to fluorescent dyes, mass cytometry antibodies are conjugated to isotopically pure heavy metals. Specifically, fundamental elements of mass

cytometry include 1) the staining of individual cells with isotope-tagged antibodies to detect specific cellular targets; and 2) quantification of the isotopic signal via time-of flight, as in other forms of mass spectrometry, which indicates specific antibody binding [103, 105, 112]. Therefore, the abundance of a specific metal isotope in each cell corresponds to the abundance of a specific cellular target detected by the antibody. The use of metal isotopes and time-of-flight quantification in mass cytometry results in relatively little spectral overlap between the channels distinguished by isotopes [286, 305, 306]. Additionally, multiple cellular targets of interest can be measured simultaneously, and the numbers are greater than those routinely measured in current fluorescence-based cytometry [105, 159, 282, 284]. Mass cytometry has the potential to track evolving cell subsets and to measure features typically associated with one cell type (e.g., mature immune cell or stem cell associated proteins) on all the cells in a sample [147, 307, 308]. This type of single-cell systems biology has the potential to reveal unexpected, clinically relevant cell types and measure a wealth of features on cells without the need to return to a sample for repeat measurements [308] [113, 284, 309, 310].

Mass cytometry–based characterization of human bone marrow [112], blood [311], and tonsil [290] cell subsets has been accomplished in prior studies and described in protocols [167]. However, mass cytometry has just recently been developed and applied in solid tissues and organs [158]. One of the major limitations for flow cytometry is the need to generate a suspension of viable single cells derived from the tissue of interest. Although fluorescence flow cytometry has been used to study some solid tissues and cancers, the protocols used to derive viable single cells, even from the same organs, can vary significantly between studies [312-314]. Basic Protocol 1 below has been optimized to yield viable cells and to preserve known cell subsets from a variety of human tissues, including lymph nodes, gliomas, melanomas, and small cell lung cancer (SCLC) patient-derived xenografts (PDXs) [98]. It is thus suitable for preparing single cells for fluorescence cytometry, mass cytometry, and other applications requiring isolated single cells. We also provide a protocol detailing cellular immunostaining for detection of cell-surface and

intracellular epitopes in fluorescence and mass cytometry analysis of cells from human tonsils, gliomas, and melanomas (Basic Protocol 2), and a section in the Commentary describing troubleshooting approaches and other considerations for these experiments [124, 125, 145].

Caution: When working with human cells, appropriate biosafety practices must be followed.
Note: All solutions and equipment coming into contact with living cells must be sterile, and aseptic technique should be used accordingly.

Basic protocol 1: Preparation of viable single cells from human tissue and tumors

This protocol describes preparation of single-cell suspensions from human tissues. It has been experimentally tested to preserve cell subsets detected using imaging platforms and maximize cell viability for cells from human tonsils, glioma tumors, melanoma tumors, and small cell lung cancer (SCLC) patient-derived xenografts (PDX) [98]. Human tonsils, glioma tumors, and melanoma tumors were resected from patients and transported directly to the laboratory (within 1 hr after collection for human gliomas and melanomas, and within 4 hr after collection for human tonsils). SCLC PDXs were flank xenografts in immunocompromised mice, generated from patient specimens. When grown as flank tumors, these xenografts form a solid tissue about 1 to 2 cm in diameter. SCLC PDXs were transported to lab within 1 hr after collection. We expect that this protocol will work in other human tissue and cancer types, as well as solid tissues from other species. However, it is important to note that 1) choice of enzymes, and 2) total dissociation time need to be optimized before routine use of the protocol in tissues not indicated here.

Materials

1. Tissue sample
2. Phosphate-buffered saline (PBS; Corning/Mediatech, cat. no. 21040CV,), room temperature
3. Experimental media:
    a. For glioma: DMEM/F12 + GlutaMax (Gibco/Life Technologies, cat. no. 10565018) with a defined hormone and salt mix (Reynolds, Tetzlaff, & Weiss, 1992) and 50 µg/ml gentamicin sulfate (Corning/Mediatech, cat. no. 30-005-CR)
    b. For melanoma: MEM (Corning/Mediatech, cat. no. 10010CV) with 10% FBS (Thermo Fisher Scientific, cat. no. 26140079) and 100 U/ml penicillin/100 µg/ml streptomycin (add from 100X penicillin-streptomycin solution, GE Healthcare, cat. no. SV30010)
    c. For tonsils: RPMI 1640 (Corning/Mediatech, cat. no. 10040CV) with 10% FBS (Thermo Fisher Scientific, cat. no. 26140079) and 100 U/ml penicillin/100 µg/ml streptomycin (add from 100X penicillin-streptomycin solution, GE Healthcare, cat. no. SV30010)
4. 20X collagenase II: dilute collagenase from *Clostridium histolyticum* (Sigma, cat. no. C6885) to 2500 CDU/ml (20 mg/ml) in PBS (store at –80°C)
5. 100X DNase I: dilute DNase I from bovine pancreas (Sigma-Aldrich, cat. no. DN25) to 10,000 Kunitz Units/ml in PBS (store at –80ºC)
6. ACK lysing buffer (Lonza, cat. no. 10-548E)
7. Trypan blue (Hyclone, cat. no. SV30084.01, prepared as recommended by manufacturer)
8. DMSO (Catalog no. BP231-1, Fisher Scientific, MA)
9. 15-ml (Corning Falcon, cat. no. 430055) and 50-ml (Corning Falcon, cat. no. 430829) conical tubes

10. Benchtop centrifuge with swing-out rotor (Sorvall model ST 16; Thermo Scientific)
11. 60-mm petri dish (Fisher Scientific, cat. no. FB0875713)
12. P1000 plastic pipet tips with narrow end cut to make a wide opening (diameter 2 to 3 mm)
13. Scalpels with blade no.10 (Fisher Scientific, cat. no. 12-460-451)
14. Incubator set at 37°C, 5% CO_2
15. Nutating platform placed inside incubator, set to 18 rpm (Fisher Scientific, cat. no. 05-450-213)
16. 70-µm (Corning Falcon, cat. no. 431751) and 40-µm (Corning Falcon, cat. no. 431750) cell strainers sized to fit 50-ml conical tubes
17. Inverted phase contrast microscope for cell culture (use 10X objective magnification for quantifying cell viability)
18. 1.8-ml cryogenic tubes with cap (Thermo Fisher Scientific, cat. no. 377267)
19. Additional reagents and equipment for counting viable cells by trypan blue exclusion

Perform mechanical dissociation

This protocol is for preparation of single cells from human tissues from surgical resections. Samples should be placed in appropriate experimental medium (see below), phosphate-buffered saline (PBS), or normal saline (0.9% NaCl), immediately after surgical resection. The volume of medium or normal saline should be enough to immerse the entire sample (Figure A.1). Ideally, samples should be transported directly to lab for preparation at room temperature (23°C).

1. Transfer pieces of human tissue from surgery to a cell preparation laboratory while keeping the sample submerged in room temperature PBS (see Time Considerations).
2. Once in lab, transfer tissue pieces and PBS to one or more 50-ml conical tubes using cut P1000 tips, ensure tubes are well balanced, and centrifuge 5 min at 100 x g, room temperature, to pellet cells and tissue pieces.

3. Carefully discard supernatant by pipetting and resuspend tissue in 5 ml or more of warm (37°C) experimental medium, as needed to cover tissue.
    a. For larger pieces of tissue (larger than 1 cm^3), use multiple rounds of mincing as in steps 4 and 5.
    b. Dead cells will not pellet effectively at 100 x g and will be present in the supernatant with other, non-cellular tissue components and secreted factors.
    c. Experimental medium may vary by cell type, as different cell types may have distinct nutrient and supplement requirements. For this protocol, media were selected based on established cell culture protocols for each cell type. Furthermore, if additional assays, such as a signaling response assay using phospho-specific flow cytometry [218, 315], are to be performed, it is important to test different types of medium for those specific assays. For example, to preserve lymphocyte signaling capability for subsequent detection by phospho-specific flow cytometry, medium containing FBS is superior to serum-free medium [152, 289, 290]. Conversely, multiple growth factor supplements are added to the neurosphere culture medium to ensure growth of human glioma cells [275, 316].

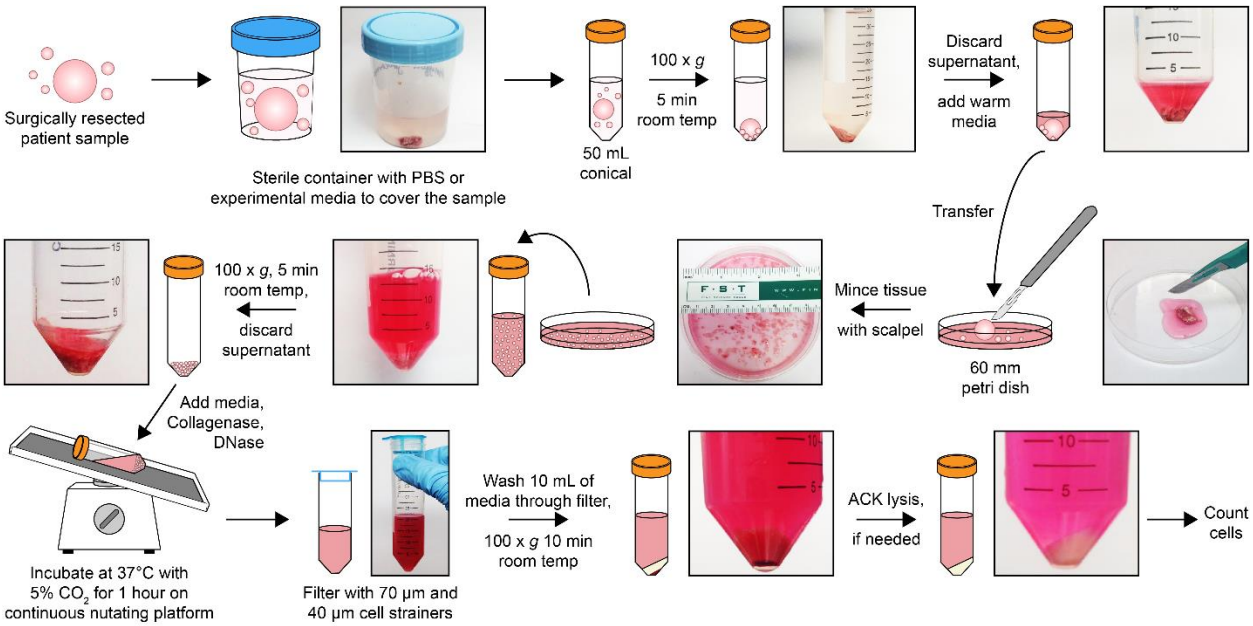


Figure A.1: **Step-by-step illustration of tissue dissociation protocol.** Surgically resected patient samples were transported in PBS, normal saline, or experimental medium at room temperature. Mechanical dissociation was followed by 1-hr enzymatic dissociation using collagenase II and DNase I (see text). ACK lysis was used to eliminate red blood cell contamination, prior to cell counting and cryopreservation or experiment.

4. Transfer tissue and experimental medium into a 60-mm petri dish using cut P1000 tips.
5. Mince tissue in experimental medium with scalpel to obtain 1 to 3 mm^3 pieces.
6. Transfer minced tissue and cells in experimental medium into 15- or 50-ml conical tubes, as dictated by the total volume of the cell and medium suspension, using cut P1000 tips.
7. Centrifuge tissue and cells in experimental medium 5 min at 100 x g, room temperature.
8. Discard supernatant by pipetting and add 4.7 ml of warm experimental medium.
    a. This volume of experimental medium leaves room for 300 µl of enzyme solutions in the next step, and is recommended for tissue that was originally 1 cm^3 in size. For larger pieces of tissue, the volumes in steps 8 and 9 should be increased proportionately to match tissue size. For example, 9.4 ml of warm experimental medium would be used in step 8 for tissue that was originally 2 cm^3 in size.

Perform Enzymatic Dissociation

9. Add 250 µl of 20X collagenase II and 50 µl of 100X DNase I, and mix with serological pipet.
    a. The final concentrations of collagenase II and DNase I should be 1 mg/ml and 100 Kunitz Units/ml, respectively.
    b. Collagenases II, IV, V, and XI displayed equivalent activity on tumor and tissue types tested [98].
10. Incubate the tube on a nutating platform (18 rpm) in an incubator (37°C, 5% CO_2) for 60 min.
11. Remove tubes from the incubator and carefully triturate the cell suspension by pipetting up and down 25 to 50 times using a 10-ml plastic serological pipet. When complete, the cell suspension should look homogeneous and have no visible tissue pieces.
12. Strain with a 70-µm cell strainer into a new 50-ml conical tube.
13. Strain flow-through from step 12 with a 40-µm cell strainer into a new 50-ml conical tube.
14. Wash by passing 10 ml of warm (37°C) experimental medium through the 40-µm strainer into the same tube.
15. Centrifuge the collected strained cell suspension 10 min at 100 x g, room temperature, and discard supernatant by pipetting.

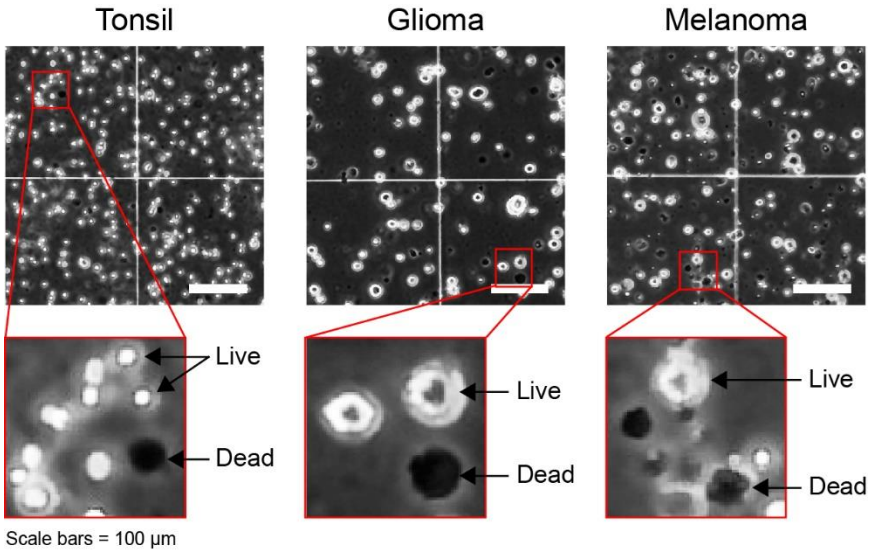


Figure A.2: **Trypan blue stain for viable cell quantification.** Trypan blue stain was used to quantify cell viability after mechanical and enzymatic dissociation. Representative images of dissociated human tissues including tonsil, glioma, and melanoma are shown. Red boxes show higher resolution of live (trypan blue–negative, white) and dead cells (trypan blue–positive, black) of each tissue type. Note that some pigmented cell types, such as melanocytes or neurons of the substantia nigra, can be brown or red and therefore appear dark in monochrome phase-contrast images. These cells should be distinguished from dead cells in counting. Scale bars = 100 µm.

Remove RBC and quantify Viable Cells

16. If pellet contains red blood cells or platelets, add 5 ml or more of ACK lysis buffer according to the manufacturer's protocols, mix with a serological pipet, and leave at room temperature for 60 sec to allow for hypotonic lysis.
17. Add 5 ml or more of warm experimental medium (the same volume used in step 16 for ACK lysis buffer, for a final 1:1 ratio), centrifuge 10 min at 100 x g, room temperature, and discard supernatant.
18. Resuspend cells in warm experimental medium and count viable cells using trypan blue exclusion (Figure A.2).
    a. Cells are now ready to be prepared for mass cytometry analysis. If mass cytometry analysis is to be performed on a different day, or if the cells need to be preserved

for long-term storage, cryopreservation is required. This can be performed per a previously established protocol (Leelatian et al., 2015). In brief, pelleted cells should be resuspended in freezing medium (experimental medium with 10% DMSO) at a concentration of 10−15 x 10^6 cells/ml, aliquoted into cryopreservation tubes at 1 ml per tube, and frozen slowly at the rate of −1°C/min in a −80°C freezer before transfer into liquid nitrogen the following day.

Basic Protocol 2: Preparation of cells for flow cytometry

This section describes a protocol for immunostaining of single-cell suspensions derived from cell lines, human tissues, and tumors. Variations of the protocol presented in this Appendix was used for the studies presented in Chapters II, III, and IV. In this Appendix, tonsils, glioma tumors, and melanoma tumors are presented as examples. Using antibodies listed in Table A.1, this protocol allows characterization of immune cell subsets ($CD45^+$) in tonsils, as well as infiltrating immune cells in glioma tumors and melanoma tumors. These antibodies allow characterization of immune cells into distinct groups: myeloid lineage (CD11b, CD11c, CD14, CD16, CD64, CD68, HLA-DR), B cell and plasma cell lineage (CD19, CD38, CD27, IgM, IgD, HLA-DR), and T cell lineage (CD3, CD4, CD8, CD8a, CD45RA, CD45RO). Additionally, antibodies that were specifically selected for identifying non-immune cell subsets in glioma (CD31, TUJ1, S100B, PDGFRα, c-MET, SOX2, CD24, Nestin, CD44, GFAP, αSMA, and CD56) and melanoma (CD31, β-catenin, S100B, vimentin, CD49F, cytokeratin, SOX2, Nestin, CD44, αSMA, and CD56) tumors were included. The antibodies described here are isotope-tagged antibodies for mass cytometry analysis. This protocol was adapted from previously established fluorescence flow cytometry protocols [167, 219, 317, 318].

All new antibodies should be titrated prior to use with appropriate positive control cells that express the target of interest and negative control cells that are known to not express the target of interest. The goal of an antibody titration is to determine the optimal concentration of an antibody that separates the true signal of the positive control cells from any background or nonspecific signal observed in the negative control cells. Antibody titration is required for every combination of antibody clone, tissue preparation technique, and antibody conjugation. This extensive validation is necessary because clones can perform differently under different antigen-exposure conditions, such as permeabilization of cells by detergent or alcohol, and protocols to conjugate fluorochrome or metal reporter tags to antibodies can change their binding properties. Examples of appropriate validation and titration have been published [218, 315, 317, 319]. For each cellular target that the user aims to detect, it must first be determined if the target is exposed on the cell-surface (i.e., extracellular) or present within the cell (i.e., intracellular). Most extracellular targets are detected with live cell staining (see “Stain viable cells to detect extracellular targets” below). However, if the target of interest is an intracellular target, it is especially important to optimize the permeabilization technique and reagents [218]. Examples of permeabilization reagents include citrate buffer with heat (Perm 1), saponin (Perm 2), methanol (Perm 3), ethanol, and Triton X-100, among many others. For the protocol described here, citrate buffer with heat (Perm 1; for IDH1 R132H antibody staining) or saponin permeabilization (Perm 2; for SOX2 antibody staining) is used prior to methanol permeabilization (Perm 3; for staining of the remaining intracellular targets) in cell lines, human PBMC, glioma, and melanoma.

Materials

1. Perm 1: 10mM citrate buffer, pH 6.0 (heat-induced epitope retrieval: HIER) (Fisher Scientific, cat. no. BP327-500) in ddH_2O

2. Perm 2: room-temperature 0.02% (w/v) saponin (Calbiochem, cat. no. 558255) in PBS
3. Perm 3: ice-cold 100% methanol (Fisher Scientific, cat. no. A412-4) kept at −20°C until immediately prior to adding to cells
4. Dissociated single cells (Basic protocol 1, or prepared as described below)
5. Experimental media:
    a. For primary gliomas and gliomaspheres: DMEM/F12 plus GlutaMax (Gibco/Life Technologies, cat. no. 10565018) with a defined hormone and salt mix [211] and 50 µg/ml gentamicin sulfate (Corning/Mediatech, cat. no. 30-005-CR)
    b. For melanoma: MEM (Corning/Mediatech, cat. no. 10010CV) with 10% FBS (Thermo Fisher Scientific, cat. no. 26140079) and 100 U/ml penicillin/100 µg/ml streptomycin (add from 100X penicillin-streptomycin stock solution, GE Healthcare, cat. no. SV30010)
    c. For tonsils and cell lines Ramos and Jurkat: RPMI 1640 (Corning/Mediatech, cat. no. 10040CV) with 10% FBS (Thermo Fisher Scientific, cat. no. 26140079) and 100 U/ml penicillin/100 µg/ml streptomycin (add from 100X penicillin-streptomycin stock solution; GE Healthcare, cat. no. SV30010)
    d. For cell lines: Experimental media varies between cell types and sources, cell types used in this work are noted above. It is recommended to consult the ATCC for additional cell types.
6. Base media: varies between cell types and sources. It is the media formulation without any added growth factors, supplements, or antibiotics. For example, DMEM/F12 plus GlutaMax for glioma cell lines. Alternatively, conditioned media, collected from the flask prior to Accutase® treatment can be used in the place of base media.
7. Accutase® (Innovative Cell Technologies, Inc., cat. no. AT104)
8. Optional: 100X DNase I: dilute DNase I from bovine pancreas (Sigma-Aldrich, cat. no. DN25) to 10,000 Kunitz Units/ml in PBS (store at –80ºC)

9. Staining medium: 1% (w/v) bovine serum albumin (BSA; Fisher Scientific, cat. no. BP9703100) in PBS
10. Live Stain reagent mix: A combined solution of all relevant antibodies (antibody list in Table A.1)
11. Phosphate-buffered saline (PBS; Corning/Mediatech, cat. no. 21040CV)
12. 16% paraformaldehyde (PFA; Electron Microscopy Sciences, cat. no. 15710)
13. HIER Stain reagent mix: A combined solution of all relevant antibodies (antibody list in Table A.1)
14. Saponin Stain reagent mix: A combined solution of all relevant antibodies (antibody list in Table A.1)
15. Methanol Stain reagent mix: A combined solution of all relevant antibodies (antibody list in Table A.1)
16. 1X Four Elements Calibration Beads (Fluidigm, cat. no. 201078)
17. 15-ml conical tubes (Corning Falcon, cat. no. 430055)
18. Benchtop centrifuge with swing-out rotor (Sorvall model ST 16; Thermo Scientific)
19. 5-ml round-bottom FACS tubes without cap (Corning Falcon, cat. no. 352052)
20. Rotor adapters with round buckets that accommodate 5 ml FACS tubes (Thermo Fisher Scientific, cat. no. 75003680)
21. 5-ml round-bottom FACS tubes with filter caps (Corning Falcon, cat. no. 352235)

Table A.1: **Tissue-specific antibody panels**

Metal	Antibody	Clone	Working conc (µg/mL)	Dilution	Volume to use in 100 µL stain (µL)	Sample type			Staining condition			
						Ton	Glio	Mel	Surf	HIER	Sap	MeOH
141Pr	HLA-ABC	W3-32	Fluidigm	1:200	0.5			✓	✓			
142Nd	cCasp3	D3E9	Fluidigm	1:200	0.5	✓	✓	✓				✓
144Nd	CD11b	ICRF44	Fluidigm	1:200	0.5		✓		✓			
145Nd	CD4	RPA-T4	Fluidigm	1:200	0.5	✓			✓			
	CD31	WM59	Fluidigm	1:200	0.5		✓	✓	✓			
146Nd	IgD	IA6-2	Fluidigm	1:200	0.5	✓			✓			
	CD64	10.1	Fluidigm	1:200	0.5		✓		✓			
	CD8a	RPA-T8	Fluidigm	1:200	0.5			✓	✓			
147Sm	β-catenin	D10A8	Fluidigm	1:200	0.5			✓				✓
148Nd	CD16	3G8	Fluidigm	1:200	0.5	✓			✓			
149Sm	CD45RO	UCHL1	Fluidigm	1:200	0.5	✓	✓	✓	✓			
152Sm	TUJ1	TUBB3	50	1:100	1		✓					✓
153Eu	CD45RA	HI100	Fluidigm	1:200	0.5	✓			✓			
	S100B	19-S100B	100	1:100	1		✓	✓				✓
154Gd	CD45	HI30	Fluidigm	1:400	0.25	✓	✓	✓	✓			
155Gd	CD27	L128	Fluidigm	1:100	1	✓			✓			
156Dy	Vimentin	RV202	Fluidigm	1:200	0.5			✓				✓
159Tb	CD11c	Bu15	Fluidigm	1:200	0.5	✓			✓			
	CD49F	GoH3	100	1:100	1			✓	✓			
160Gd	CD14	M5E2	Fluidigm	1:200	0.5	✓			✓			
161Dy	CD19	HIB19	100	1:100	1	✓			✓			
	PDGFRα	16A1	200	1:100	1		✓		✓			
	Cytokeratin	C-11	Fluidigm	1:200	0.5			✓				✓
162Dy	c-MET	L6E7	100	1:100	1		✓		✓			
163Dy	SOX2	O30-678	100	1:100	1		✓	✓			✓	
166Er	CD24	ML5	Fluidigm	1:200	0.5		✓		✓			
167Er	CD38	HIT2	Fluidigm	1:200	0.5	✓			✓			
168Er	CD8	SK1	Fluidigm	1:200	0.5	✓			✓			
	Nestin	10C2	100	1:100	1			✓				✓
169Tm	CD44	BJ18	100	1:100	1		✓	✓	✓			
170Er	CD3	SP	Fluidigm	1:200	0.5	✓	✓	✓	✓			
171Yb	CD68	Y1/82A	Fluidigm	1:200	0.5	✓			✓			
	GFAP	1B4	25	1:100	1		✓					✓
	IDH1R132H	H09	100	1:100	1		✓			✓		
172Yb	IgM	MHM-88	Fluidigm	1:200	0.5	✓			✓			
173Yb	αSMA	Ab54723	50	1:100	1		✓	✓				✓
174Yb	HLA-DR	L243	Fluidigm	1:200	0.5	✓	✓	✓	✓			
175Lu	CD56	HCD56	50	1:100	1	✓	✓	✓	✓			
176Yb	Histone H3	D1H2	Fluidigm	1:200	0.5	✓	✓	✓				✓

Conc = concentration; Ton = tonsil; Glio = glioma; Mel = melanoma; Surf = surface; HIER = heat-induced epitope retrieval Sap = saponin; MeOH = post-methanol; Fluidigm = use antibodies provided by Fluidigm

Prepare Antibody

1. Example mass cytometry reagent mixes for healthy human tonsil tissue, glioma tumors, and melanoma tumors are shown in Table A.1 and separated according to staining step. Prepare reagent mixes separately for each staining step: live cell staining (Live), staining after heat-induced epitope retrieval with citrate buffer (HIER Stain) staining in 0.02% saponin (Saponin Stain), and staining after methanol treatment (Methanol Stain).
    a. Antibodies for mass cytometry are pre-labeled with metal isotopes. Many of these antibodies are commercially available (see Table A.1). For antibodies that are not commercially available in isotope-tagged formats, they can be labeled with a metal isotope using a commercial conjugation kit [320]. The isotope-labeled antibodies can then be used for immunostaining by following the protocol described below
    b. Adaptation of this protocol for phospho-flow should detect cell surface proteins following fixation as described in “Live cell staining” even though the cells are no longer viable, as described by [167, 218, 315].
    c. This protocol does not use barcoding, but that technique can be useful in addressing potential batch effects from staining and collecting data at different times [321-323].

Collect viable cells for flow cytometry

2. Samples of different origins will require different collection protocols.
    a. For tissue samples, follow Basic Protocol 1 above

b. For adherent cells, remove media by aspirating or manual pipetting and add 2 ml Accutase®[1] per T75 flask or 100 mm dish and incubate at 37° (5% CO_2) for 10 minutes. Triturate Accutase® up and down with a P1000 or 5 to 10 ml pipet to dislodge cells and collect cells in a 15 ml conical. Rinse plate with experimental base media to remove residual cells using a P1000 or 5 to 10 ml pipet. Do not exceed total volume of 15 ml. Spin down collected cells at 200 x g for 10 minutes. Aspirate off supernatant and resuspend in desired volume of base media and count cells using a hemocytometer and Trypan Blue

c. For suspension cells, collect cells in media using pipettor and transfer to 15 ml or 50 ml conical (use an appropriate conical to fit entire volume). Spin down collected cells at 200 x g for 10 minutes, aspirate off supernatant and resuspend in desired volume of base media. Count cells using a hemocytometer and Trypan Blue.

d. For spheroids, collect spheres in media using pipettor and transfer to 15 ml or 50 ml conical (use an appropriate conical to fit entire volume). Spin down collected spheres at 200 x g for 10 minutes. Aspirate off supernatant and resuspend the pelleted spheres in 2 ml of Accutase® per T75 flask or 100 mm dish and incubate at 37° (5% CO_2) for 10 minutes. Triturate the pellet in Accutase® vigorously, approximately 20 times, with a P1000 to disrupt spheres. Resuspend to total volume of 10 ml in base media and spin down at 200 x g for 10 minutes. Aspirate off supernatant and resuspend cells in desired volume of base media. Count cells using a hemocytometer and Trypan Blue

[1] Accutase® is preferred for gliomaspheres or in cases where extracellular targets are to be detected. If only intracellular targets are to be measured, Trypsin or other general proteases are suitable.

e. For cryopreserved cells, remove cells from liquid nitrogen and immediate transfer to 37° bead or water bath to quickly thaw (approximately 2-4 minutes). Transfer cells from cryovial to 15 ml conical and add base media to a total volume of 10 ml. The goal is to dilute and remove DMSO as quickly as possible after thawing. Addition of DNase I may be helpful upon sample thawing to preserve viability. Use the same concentration of DNase as used during tissue dissociation (see Basic Protocol 1, step 9). Specifically, resuspend the cells in this step in 9.9 ml of warm experimental medium, add 100 µl of 100X DNase I to the cell suspension, and proceed to the next step. Spin down at 200 x g for 10 minutes. Aspirate off supernatant and resuspend cells in desired volume of base media. Count cells using a hemocytometer and Trypan Blue.

Prepare cells for flow cytometry staining

3. Aliquot single cells to individual FACS tubes[2]. If stimulation or drug treatments are to be included, they should be performed at this time according to user experimental design and as noted in step 1a.
4. Perform live/dead staining by adding the appropriate staining agent to FACS tubes containing single cells. For FFC, Alexa dyes with a succinimidyl-ester group are commonly used and for MC, rhodium (103Rh) or cisplatin (195Pt) are commonly used.

[2] The optimal number of cells per tube will vary by experiment based on the total number of cells the user aims to collect. Assume a 50% loss of cells over the course of experimentation, though the actual loss is likely less. For MC, an additional 50-60% loss of cells can be expected during data collection. Furthermore, starting the protocol with fewer cells results in *worse* cell retention over the course of the protocol. Users should also consider how many different staining steps or tube transfers will be performed when considering loss of cells.

For example, Alexa 700-SE (at 50ng/ml) incubated for 10 minutes at room temperature. For MC, add rhodium (1μM) and incubate for 1-10 minutes[3] at room temperature

5. Wash out live/dead stain with PBS by adding 2 ml of PBS per wash[4]. Spin down at 200 x g for 7 minutes and remove supernatant by pipetting off liquid or decanting. Vortex and repeat for second wash.
    a. For all centrifugation steps involving a cell pellet, invert and decant only once. After placing the tube upright again, cells typically enter suspension and the pellet can detach. Thus, additional decanting significantly lowers viable cell yield.
6. Proceed to either Live cell staining (Step 7) for surface epitopes or cell fixation for intracellular staining (Step 12)

Stain viable cells to detect extracellular targets ("Live cell staining")

7. Cells are currently in residual PBS from live/dead washes. Add 1ml staining medium and vortex. Spin down at 200 x g for 7 minutes. Decant supernatant, vortex, and repeat wash with staining medium.
8. Resuspend cell pellet in staining medium to achieve the transfer volume.
    a. The "transfer volume" is calculated by subtracting the summed volume of staining antibodies (the "antibody volume") from the total volume in which staining will occur in step 9. For example, consider a protocol where approximately 10 µl of cells in staining medium from step 8 are to be stained with 1 µl each of 30 antibodies in a total volume of 100 µl in step 9. In this case, at step 8, at least 60 µl of staining medium should be added to the 10 µl of cells in staining medium to achieve a

[3] Rhodium incubation should be optimized for individual samples. Dissociated glioma samples had improved quality with shorter (1 minute) incubations while other samples were not affected.
[4] As space allows. FACS tubes hold 5ml volume max but it is not recommended to go above 4ml. Use 1ml PBS for the first wash if required.

transfer volume of 70 µl for step 9. For steps 8 and 9, the volume of antibodies varies based on experimental design. For the examples depicted here, volumes are specified in Table A.1 for tonsil, glioma, and melanoma. The total volume in step 9 is 100 µl (including cells in transfer volume and all antibodies).

9. Add Live Stain reagent mix (Table A.1) to new FACS tubes labeled for surface staining and add cells, in staining medium, to antibody mixture. Vortex to mix cells and antibodies and incubate for 30 minutes at room temperature. For FFC, incubate in the dark.
    a. In the work presented here, the total staining volume (cells in staining medium plus antibodies) should be exactly 100 µl.
    b. Live Stain reagents in Table A.1 can be combined in a 1.5-ml microcentrifuge tube prior to mixing with cells in FACS tubes.
10. After 30 minutes, add 1 ml of staining medium to the FACS tube and spin down at 200 x g for 7 minutes. Decant supernatant, vortex, and repeat wash with PBS. Decant supernatant and vortex
11. Proceed to either cell fixation (step 12) or preparing sample for the cytometer (step 46).

Fix cells for intracellular staining ("Cell fixation")

12. Cells are currently in residual PBS. Resuspend in 900 ml PBS. Add 100 µl of 16% electron microscopy grade PFA (for a final concentration of 1.6% PFA in PBS) and vortex to mix. Incubate for 10 minutes at RT
13. After 10 minutes, add 1 ml of PBS and spin down at 800 x g for 5 minutes. Decant supernatant and vortex pellet (see step 5).
14. Determine the optimal permeabilization conditions required for each intracellular target. If a subset of intracellular antigens requires permeabilization with heat-induced epitope retrieval, as is the case with IDH1 R132H, include steps 15-20. If some intracellular

antigens require saponin, as is the case with antibodies used to stain glioma and melanoma, include steps 21-31 If all intracellular antibodies have been shown to effectively detect target antigens after permeabilization with ice-cold methanol (as is the case for the antibodies that were used to stain tonsils), skip to step 32.

Permeabilize cells using heat-induced epitope retrieval with citrate buffer (Perm 1; Figure A.3)

15. Cells are currently in residual PBS from cell fixation. Resuspend in 1 ml staining medium and vortex to mix. Spin down at 800 x g for 5 minutes. Decant supernatant and vortex pellet. Resuspend in 200 µl Perm 1 and transfer cells to 50 ml conical. Bring total volume in the conical to 10 ml of Perm 1 and cells.
16. Add 300 ml water to a 600 ml beaker. Microwave on high in a 1000W microwave for 2.5 minutes. Swirl each conical tube in the heated water so that the entire volume is submerged for 2 minutes. After 2 minutes, bring the cells to room temperature by incubation on ice (about 10 minutes). Heat the water for an additional 45 seconds prior to each subsequent tube. Repeat swirling (2 minutes) and heating of the water (45 seconds) until all samples have been treated.
17. Once samples have cooled to room temperature, spin at 800 x g for 10 minutes. Aspirate supernatant and resuspend in 1 ml staining medium. Transfer cells in staining medium to FACS tube. Rinse 50 ml conical with additional 1 ml of staining medium and transfer to FACS tube (final volume of approximately 2 ml in tube). Spin down at 800 x g for 5 minutes. Decant supernatant and vortex (see step 5).
18. Add HIER reagent mix (Table A.1) to new FACS tubes labeled for heat-induced epitope retrieval and add cells, in staining medium, to antibody mixture. Vortex to mix cells and antibodies and incubate for 30 minutes at room temperature. For FFC, incubate in the dark.

19. After 30 minutes, add 1 ml of staining medium to the FACS tube and spin down at 800 x g for 5 minutes. Decant supernatant, vortex, and repeat wash with PBS. Decant supernatant and vortex (see step 5).
20. Proceed to additional permeabilization steps as needed: Perm 2, step 21 or Perm 3, step 32. If no further staining is required, proceed to preparing the sample for the cytometer, step 46.

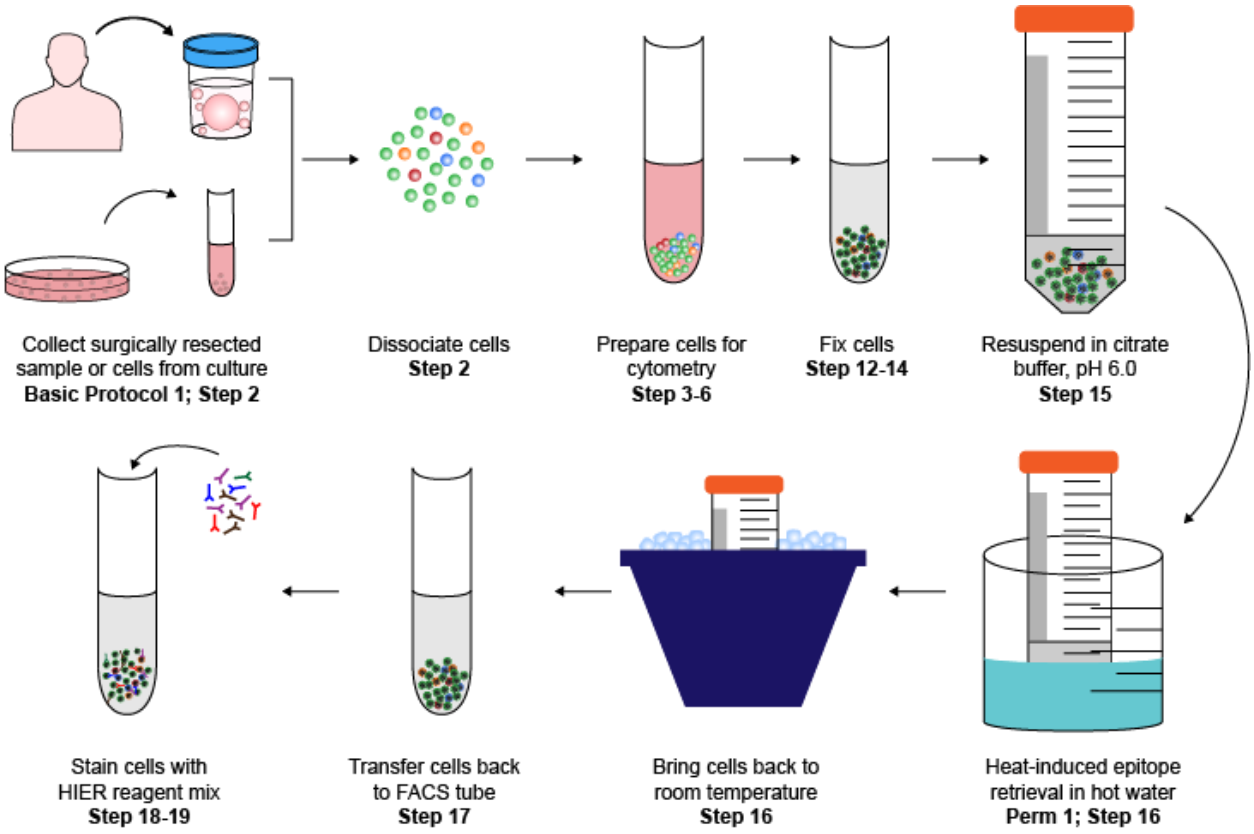


Figure A.3: **Step-by-step illustration of Perm1 (HIER) protocol.** Surgically resected patient samples or gliomaspheres were dissociated and prepared for flow cytometry. Citrate buffer (pH 6.0) and heat were used to permeabilize the cells and expose the IDH1 R132H epitope. Antibodies against the IDH1 R132H antigen were applied after HIER.

Permeabilize cells using 0.02% saponin (Perm 2)

21. Cells are currently in residual PBS from cell fixation or Perm 1. Resuspend cells in 1 ml of Perm 2, pellet 5 min at 800 x g, room temperature, and decant to discard supernatant (see step 5).
22. Repeat step 21 once.
23. Resuspend cell pellet in appropriate volume of Perm 2 (see step 8).
24. Add Saponin Stain (Table A.1) reagent mix to the new FACS tubs.
25. Transfer appropriate volume of cell suspension in Perm 2 to the FACS tubes containing the Saponin Stain reagent mix. Briefly vortex to mix cells with antibodies.
    a. Total staining volume (cells in Perm 2 plus antibodies) should be exactly 100 µl. Note that the Saponin Stain mix must also contain saponin.
26. Leave cells at room temperature for 30 min. For FFC, incubate in the dark
27. Add 1 ml of Perm 2 to the FACS tube, pellet cells 5 min at 800 x g, room temperature, and decant to discard supernatant and vortex pellet (see step 5).
28. Repeat step 27 once.
29. Add 1 ml of PBS to the FACS tube, pellet cells 5 min at 800 x g, room temperature, and decant to discard supernatant (see step 5).
30. Vigorously vortex cell pellet to resuspend cells in void volume.
31. Proceed to Perm 3 permeabilization as needed (step 32) or prepare sample for the cytometer (step 46).

Permeabilize cells using 100% methanol (Perm 3)

32. Cells are currently in residual PBS from cell fixation, Perm 1, or Perm 2.
33. Add 1 ml of Perm 3 to cells in the FACS tube. Vortex to thoroughly resuspend cells with Perm 3.

a. The final concentration of ice-cold methanol after the Perm 3 is added to cells should be > 95%.

34. Keep cells at −20°C for at least 20 min.
    a. Alternatively, cells can be stored in Perm 3 at −20°C overnight, or at −80°C for days or weeks before proceeding, if necessary. Cover tubes in parafilm for storage longer than 20 minutes.
35. Remove FACS tubes from −20°C (or −80°C) and add 1 ml of PBS to each tube.
36. Pellet cells 5 min at 800 x g, room temperature, then decant to discard supernatant and vortex pellet (see step 5).
37. Add 1 ml of staining medium to FACS tubes. Pellet cells 5 min at 800 x g, room temperature, then decant to discard supernatant and vortex pellet (see step 5).
38. Repeat step 37.
39. Resuspend cell pellet in appropriate volume of staining medium (see step 8).
40. Add Methanol Stain reagent mix to new FACS tubes labeled for Perm 3.
    a. See step 8 and Table A.1. Total staining volume (cells in staining medium plus antibodies) should be exactly 100 µl. Note that Methanol Stain reagent mix does not contain methanol.
    b. Methanol Stain reagents in Table A.1 can be combined in a 1.5-ml microcentrifuge tube prior to mixing with cells in FACS tubes.
41. Transfer appropriate volume of cell suspension in staining medium to the FACS tubes containing Methanol Stain reagent mix (see step 8). Briefly vortex to mix cells with antibodies. Incubate cells at room temperature for 30 min. For FFC, incubate in the dark.
42. Add 1 ml of staining medium to the FACS tube and pellet cells 5 min at 800 x g, room temperature, then decant to discard supernatant and vortex pellet (see step 5).
43. Repeat step 42 once.

44. Add 1 ml of PBS to the FACS tube and pellet cells 5 min at 800 x g, room temperature, then decant to discard supernatant (see step 5).
45. Proceed to preparing sample for the cytometer (step 46).

Prepare samples for the cytometer

46. Cells are in residual PBS from staining or fixation steps.
47. For FFC:
    a. Resuspend in 300 µl PBS.
    b. Cover in aluminum foil and store at 4°C for no more than 24 hours.
48. For MC:
    a. Resuspend in PBS + 1.6% PFA and 1x iridium (125 nM). Incubate for 15 minutes, up to overnight, at 4°C.
    b. Add 1 ml of PBS to the FACS tube and pellet cells 5 min at 800 x g, room temperature, then decant to discard supernatant (see step 5). Vortex pellet.
    c. Add 1 ml of deionized water to the FACS tube and pellet cells 5 min at 800 x g, room temperature, then decant to discard supernatant (see step 5).
    d. Resuspend cell pellets in 1X Four Elements Calibration Beads in deionized water prior to mass cytometry analysis. Use 1 ml for every 0.5 x 10^6 cells.
    e. Filter cells using FACS tubes with 0.2µm filter caps.

Cells are now ready for cytometry analysis.

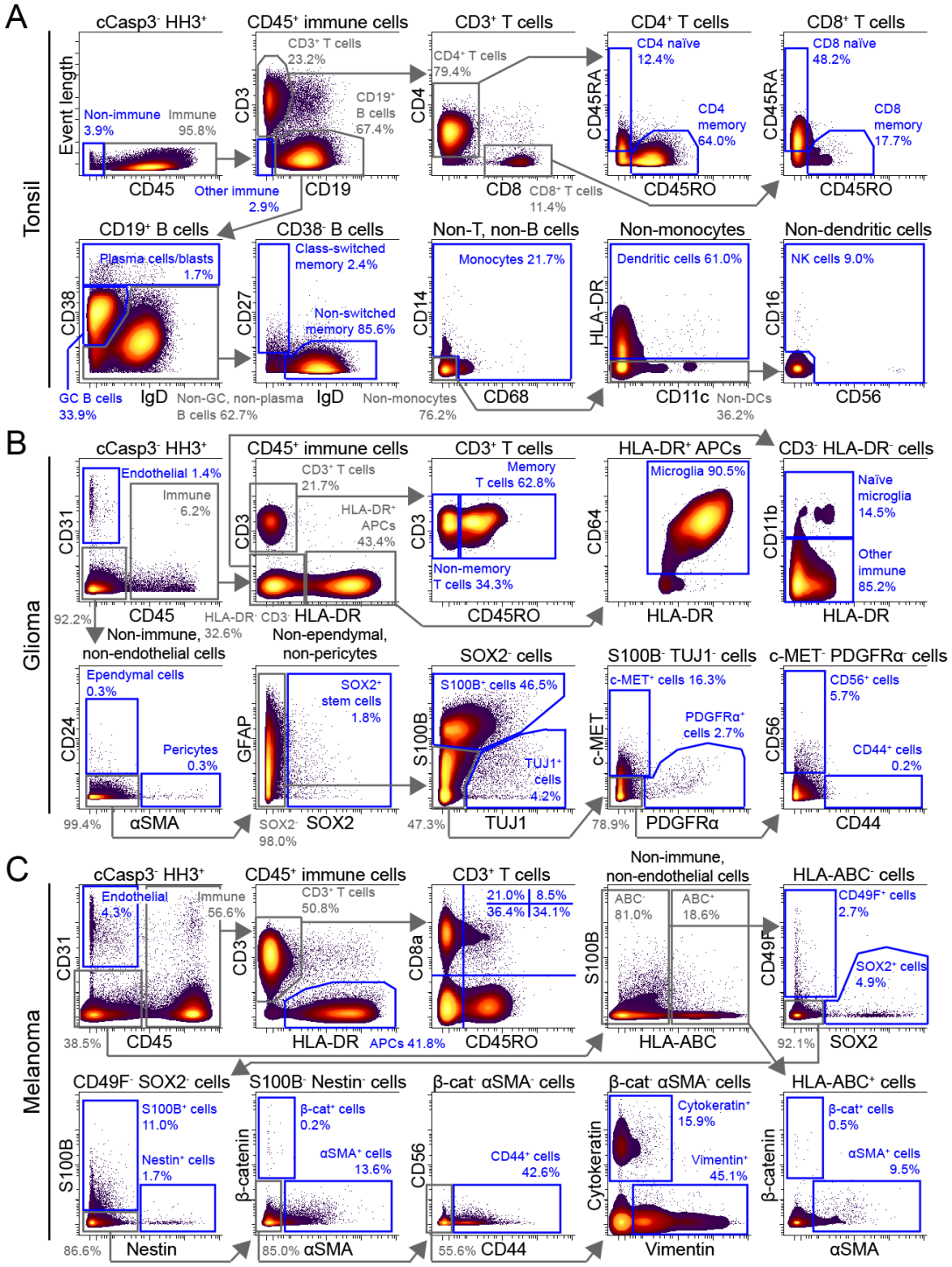


Figure A.4: **Biaxial analysis of cells derived from human tissue and tumors.** Biaxial plots of non-apoptotic (cCasp3⁻), nucleated (HH3⁺) cells from mass cytometry analysis of (A) tonsil, (B) glioma, and (C) melanoma are shown. Intermediate gates are shown in gray, and terminal gates are shown in blue. Cell types or protein identity of cells in each gate are indicated. The percentages of cells in gates are also specified.

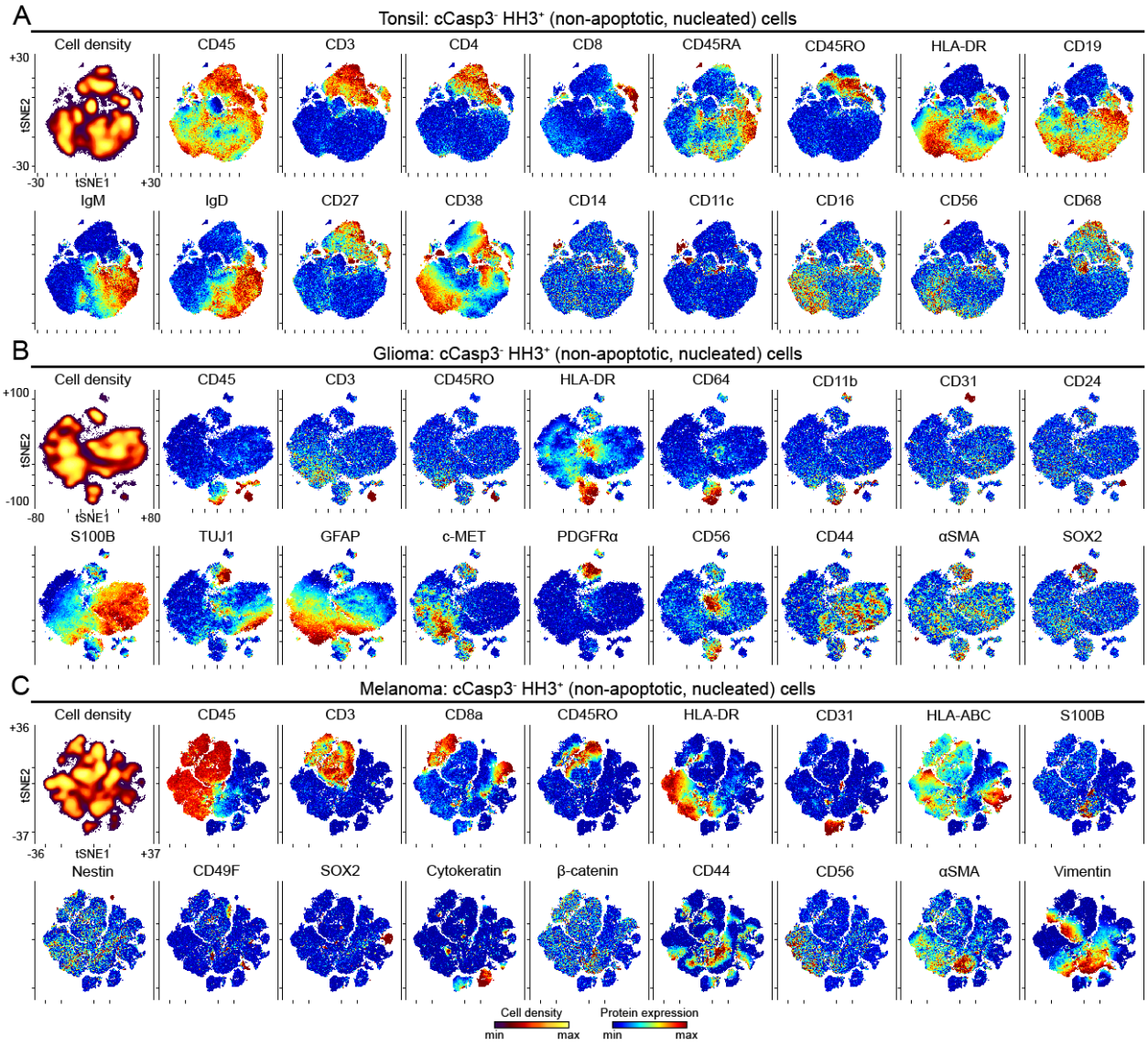


Figure A.5: **High-dimensional analysis of mass cytometry data using viSNE**. Non-apoptotic ($cCasp3^-$), nucleated ($HH3^+$) cells from mass cytometry analysis of (A) tonsil (plots of 106,568 cells), (B) glioma (plots of 65,834 cells), and (C) melanoma (plots of 94,810 cells), are shown. The first plot of each tissue type depicts cell density of the viSNE map. The remaining plots display expression of indicated protein. viSNE maps of each tissue type were generated separately and the markers shown here for each tissue type were used to generate the maps.

Commentary

Background Information

Flow cytometry technologies employing fluorescence- and mass-based reporters have been successfully applied to characterize protein phenotype and to quantify the abundance of diverse human cell types. Flow cytometry protocols commonly use reporter-conjugated antibodies to make relative quantitative measurements for tens of features in each of hundreds of thousands of cells in minutes [104, 320, 324-327]. Mass cytometry, a newer form of flow cytometry based on mass spectrometry, has gained attention for the relative ease with which more than 35 cellular features can be measured [103, 112, 328]. However, as flow cytometry requires individual cells in suspension, applications to solid tumors and tissues has previously been modest compared to its rapid adoption in immunology and blood cancer research, where samples of viably cryopreserved cells have been collected and characterized for decades [128, 285, 287, 296, 308, 329, 330]. The key limitation has been the perceived difficulty in making cells from solid tissues into single-cell suspensions that are viable and representative of different cell types present in the original tissue. Furthermore, antibody sets had not been designed and tested to effectively identify cells outside the immune system. It is only recently that mass cytometry has been tested and applied to solid tissues and tumors [98, 158]. Key to this work was the development of a protocol that preserved the viability and diversity of the tissue cells in a way compatible with detection of cell surface and intracellular features by mass cytometry.

Critical parameters and troubleshooting

Table A.2 lists some possible problems that may arise in preparing cells for mass cytometry, along with their possible causes and solutions.

Tissue quality and transportation

The protocols in this unit are applicable to human tissues extracted by surgery or to animal tissues isolated after dissection. To preserve tissue viability, samples should be transported to the laboratory for further preparation as rapidly as possible. The dissociation protocol presented here was tested on samples that were processed between 30 min and 4 hr after surgical resection [98]. Additionally, samples should be transported in sterile PBS, appropriate experimental medium, or other sterile transport medium that has been tested to preserve cell viability and representative cell subsets for specific tissue types. Samples should be entirely submerged in the transport medium in a closed container. Unless specifically optimized and validated using other conditions, samples should be immediately transported at room temperature (23°C) to the laboratory for further preparation.

Mechanical dissociation

As described in Basic Protocol 1, tissue should be mechanically dissociated into fine (1 to 3 mm^3) pieces to maximize surface contact with dissociation enzymes in subsequent steps. During mechanical dissociation in the petri dish, tissue pieces should be adequately covered in warm (37°C) experimental medium. For larger samples, the tissue should be divided into batches for mincing and combined prior to addition of the dissociation enzymes: collagenase II and DNase I.

Selection of enzymes and duration of dissociation

Dissociation enzymes are incorporated in the protocol to break down the extracellular matrix to yield a single-cell suspension. Many types of enzymes are available, such as collagenases, trypsin, papain, and HyQTase, among others. These enzymes can be used individually or in combination. The types of dissociation enzymes used can affect

the viability of single cells derived from the starting tissue. As described above, a suitable dissociation protocol will maximize cell viability as well as preserve representative cell subsets in the original tissue. Variables to consider when testing the dissociation conditions include the use of a single enzyme or combinations of enzymes, and duration of dissociation [98]. Additionally, inclusion of DNase I in the dissociation protocol described here significantly improves viable cell yield from multiple tissues [98]. Therefore, it is highly recommended to include DNase I in the dissociation solution unless it is specifically demonstrated experimentally that DNase I is not required to improve cell viability. The protocol described here uses 1 hr of enzymatic dissociation, which has been shown to result in the highest viable cell yield for various human tissues and tumors [98]. Specifically, for most tissues, shorter dissociation time led to release of fewer cells, whereas longer dissociation led to increased cell death. It is recommended that the type of dissociation enzyme and duration of enzymatic dissociation be tested and optimized for a new tissue type to achieve optimal viable cell yield. This can be quantified by using trypan blue staining (see Basic protocol 1). In addition to overall cell viability, it is crucial to determine preservation of known cell types and cells of interest as part of the optimization process. When testing the dissociation on a particular tissue, imaging techniques such as colorimetric immunohistochemistry or immunofluorescent detection of known cellular targets can be used to characterize the presence of cell subsets in the original tissue.

Immunostaining for mass cytometry

It is highly recommended to optimize the immunostaining protocol for each antibody in a panel to ensure target-specific staining and to optimize signal-to-background levels [218]. Parameters that need optimization [218] include 1) antibody specificity (which can be tested using positive and negative control cells that are known to express and lack

the target of interest), 2) antibody concentration (to allow maximal distinction between positive and negative cells, and minimize nonspecific background staining), 3) staining order (extracellular staining or intracellular staining), and 4) compatibility of permeabilization reagents for intracellular targets (HIER, saponin, ice-cold methanol, or other reagents). If multiple permeabilization reagents are required for different intracellular targets, specificity and sensitivity of the antibodies should be tested to ensure that the targets are still detectable after multiple permeabilization steps. For the glioma immunostaining presented here, HIER was used for IDH1 R132H detection or saponin permeabilization was used for SOX2 detection, prior to subsequent permeabilization by ice-cold methanol. It has been previously shown that saponin does not destroy intracellular targets normally detectable after methanol permeabilization, and therefore these reagents can be used in the same protocol, with the use of saponin preceding methanol [321]. However, each antibody should be specifically tested and optimized prior to use.

Treatment of cells with reagents for detection of intact cells via mass cytometry

In flow cytometry analyses, an initial step is to identify intact cells and remove cellular debris or enucleated cells from further analyses. Conventional fluorescence flow cytometry relies on measurement of cell size (forward scatter, FSC) and cell granularity (side scatter, SSC) to identify cells. Additionally, cellular debris can be distinguished due to its smaller size (low FSC) and higher granularity (high SSC) and Alexa dyes with a succinimidyl-ester group can be used to identify dead cells. In contrast to fluorescence flow cytometry, mass cytometry does not have direct parameters to distinguish intact cells from cellular debris. Therefore, mass cytometry analysis requires measurement of indirect parameters to identify intact cells. Iridium-conjugated DNA intercalator is commonly used to identify intact cells by mass cytometry analysis [331]. The per-cell quantity of DNA-intercalated iridium provides information about DNA content, which can be used to define

intact cells. However, iridium-conjugated DNA intercalator cannot be detected using fluorescence flow cytometry. In this protocol, we also used anti-histone H3 antibody staining for detection of intact nucleated cells [98]. The advantage of using an antibody-based technique is that it is readily applied across different flow cytometry platforms (fluorescence and mass cytometry).

Anticipated results

The protocols in this unit produce viable single-cell suspensions from solid tumors and tissues and are expected to identify most common cell types, including endothelial cells, immune cells, epithelial cells, neural cells, and fibroblasts. These protocols have been validated for human tonsil tissue, glioma tumors, melanoma tumors, and small cell lung cancer patient–derived xenografts. Maximum viable cell yield per gram of tissue from the dissociation of human tonsils, glioma, and melanoma using collagenase II plus DNase I should be achieved after 1 hr of incubation [98]. Histone H3 staining should allow highly specific identification of nucleated, intact cells. Additionally, cells derived from Basic Protocol 1 are suitable for quantitative measurement of protein expression in individual cells and cell subset abundance using either fluorescence flow or mass cytometry, among other applications. For other tissue types not mentioned above, tissue-specific optimization of the dissociation protocol that takes into consideration the critical parameters described here is highly encouraged. Specifically, a systematic comparison of different dissociation durations, as well as different enzyme combinations, is required. For every condition, it is crucial to quantify cell viability using techniques such as trypan blue staining. Additionally, the relative abundance of known cell subsets after different dissociation conditions should be quantified, as has been done above using flow cytometry. These data should be compared to prior knowledge of cell types present in the tissue, and possibly with immunohistochemistry stains of the original intact tissue (for more details, see [98]).

Data Analysis

Quantification of single-cell protein expression by biaxial analysis

Biaxial plots are a mainstay in cytometry and are typically used to compare the abundance of cells with differing relative intensity of two or more quantified cellular targets. A protocol to generate biaxial and other common plots of mass cytometry data was previously established [167]. In a typical cytometry analysis workflow [144, 145], cells are filtered or assigned to populations based on expression profiles of cellular targets in a process called gating (Figure A.4). Gating can be repeated sequentially on increasingly refined cell subsets, resulting in a nested hierarchy of cell types that traditionally captures a developmental continuum or indicates an increasingly polarized and specific cell identity [112, 144, 332]. Figure A.4 shows examples of sequential biaxial gating of cells derived from mass cytometry analysis of healthy tonsil (Figure A.4A), a patient glioma (Figure A.4B), and a patient melanoma (Figure A.4C; [98, 167]). For samples with known cell types, such as healthy peripheral blood mononuclear cells (PBMCs), established sets of identity markers can be used to distinguish cell types [333, 334]. However, concepts of cell identity are still under active discussion in established single-cell fields like immunology [332]. Furthermore, methods of defining and identifying cell populations are likely to be re- fined as the field of single-cell biology matures. Key areas of growth include measurement platforms like mass cytometry, analysis tools from machine learning [125, 126, 147, 335], reference knowledge bases of established cell identities [168, 336], annotated repositories of single-cell data, and quantitative labels of cell type [158].

Heat plots and viSNE analysis

Data collected from the same patients as shown in biaxial analyses in Figure A.4 were analyzed by viSNE and are shown in Figure A.5 (tonsil, Figure A.5A; glioma, Figure A.5B; and melanoma, Figure A.5C). For each sample, a separate pair of t-SNE axes was

created (i.e., each sample was analyzed separately in viSNE to create sample-specific viSNE plots, each of which has its own, sample-specific t-SNE axes). All computational analysis was performed using Cytobank software (http://www.cytobank.org; [168]).

Time considerations

The protocols in this unit were experimentally tested on human gliomas, human melanomas, and SCLC PDXs that were transported to the laboratory within 1 hr after surgical resection. Human tonsils were transported within 4 hr after resection. After tissue is transported to the laboratory, this dissociation protocol can be completed in 2 to 3 hr, depending on the size of the tissue sample. The size of the tissue sample determines the time that is needed for mechanical dissociation (larger tissue samples take more time to be properly minced, whereas smaller tissue samples take less time). The time for enzymatic dissociation is not affected by tissue size. The approximate timing of the protocol is: 10 to 30 min for mechanical dissociation, 1 hr for enzymatic dissociation, 15 to 30 min for cell straining, 10 to 30 min for red blood cell lysis and counting, and 10 to 30 min for diluting cells for cryopreservation, if needed (see Basic Protocol 1). Once viable single cells are obtained (either from immediate dissociation or from cryopreservation), live surface immunostaining can be completed in 1 hr, followed by 10 to 15 min of cell fixation. Duration of intracellular staining varies depending on whether permeabilization and staining with saponin (1 hr) is required for the panel of interest. Permeabilization with methanol is usually performed overnight, but can be performed for as little as 10 min [218]. Once all cells are permeabilized by ice-cold methanol, an additional 1 hr is required for intracellular immunostaining with isotope-labeled antibodies.

Table A.2: **Potential problems and troubleshooting**

Problems	Potential causes	Troubleshooting
Few viable cells after dissociation	• Poor sample quality	• Use fresh surgically resected specimens (i.e. avoid using fixed samples).
	• Small tissue size	• If possible, use tissue at least 0.5 cm^3 in size. Smaller samples can be used but will yield fewer cells.
	• Transport medium	• Use sterile PBS or appropriate experimental medium. Other isotonic solutions should be tested prior to routine use.
	• Transport temperature	• Specimens should be transported at room temperature. If transportation at other temperature (i.e. on ice) is to be used, it should be compared to room temperature.
	• Duration of transportation	• Duration of sample transportation should be noted. Ideally, samples should be dissociated immediately after surgical collection. Longer transportation times will decrease cell yield.
	• Dissociation enzyme	• Dissociation enzymes should be tested for optimal viable cell yield prior to use in a new tissue type. • DNase I should be included in the dissociation protocol unless shown experimentally to be unnecessary.
	• Duration of enzymatic dissociation	• Some tissue types may require shorter or longer dissociation duration. The dissociation kinetics should be tested for the specific types of enzyme used in the protocol. Testing of dissociation durations ranging from 15 minutes to ≥ 6 hours is recommended.
Red blood cell or platelet contamination	• ACK lysis is not included in the protocol	• ACK lysis should be used when there is visible red blood cell contamination. This should be done prior to cryopreservation.
Few viable cells after cryopreservation	• Freezing medium	• Freezing medium should contain 10-12% DMSO in appropriate medium (e.g. FBS, or experimental media), which should contain serum or BSA, depending on cell types.
	• Freezing temperature	• Temperature of cell cryopreservation should be gradually decreased in a controlled environment at ~ 1°C decrease/min in a -80°C ultralow freezer. Within 1 week, cells should be transferred to liquid nitrogen for long term storage.

	• Thawing conditions	• Cells from cryopreservation should be warmed in a 37°C water bath for 1-2 minutes until completely thawed. Cell suspension should be immediately washed using warm experimental media to remove DMSO
Cell subsets of interest were not detected	• Inappropriate dissociation condition for tissue type	• Antibodies with known specificity (and known optimal staining condition) should be used to test if the dissociation protocol preserves the cell subsets of interest.
	• Antibody specificity	• Antibodies that have been shown to specifically detect targets with other techniques (such as western blot) might not always work for flow cytometry-based technologies. Therefore, new antibodies should always be tested for specificity prior to use. This should be done using a known positive control cell type, and a known negative control cell type (rather than relying on an isotype control antibody).
	• Antibody concentration	• All new antibodies should be titrated to yield optimal concentration prior to use. The ideal concentration should allow maximal separation between the positive and the negative control cells, while minimizing signal of negative control cells (i.e. non-specific staining, background).
	• Immunostaining condition	• Antibodies that detect surface antigens should be used during live surface immunostaining. Note that some antibodies that detect surface transmembrane antigens were created to detect the intracellular portion of the antigen (i.e. cytoplasmic domain) and, therefore, should be used during intracellular immunostaining. • Antibodies that detect intracellular antigens should be tested for appropriate permeabilization conditions. Different permeabilization reagents might be required for different antibodies. Additionally, steps of permeabilization with different reagents, if needed, should also be tested.
Non-specific staining	• Antibody specificity and concentration	• Every antibody should be tested for their specificity and for the appropriate concentration prior to use for staining.

APPENDIX B

MASS CYTOMETRY PHENOTYPES OF GLIOBLASTOMA PATIENT SAMPLES

Authors: Nalin Leelatian*, Justine Sinnaeve*, Akshitkumar M. Mistry, Sierra M. Barone, Asa A. Brockman, Kirsten E. Diggins, Allison R. Greenplate, Kyle D. Weaver, Reid C. Thompson, Lola B. Chambless, Bret C. Mobley, Rebecca A. Ihrie, and Jonathan M. Irish

*Equal contribution

The figures in this appendix appear as supplemental information in a manuscript in press at *eLife* (6/8/2020).

Preface

Extensive work over the past decade has illuminated the diversity of single cell gene expression phenotypes in glioblastoma. However, datasets that explore the diversity of single cell protein and signaling phenotypes within and among glioblastomas are scarce. The development of a mass cytometry panel and resulting dataset for such tumors is a major contribution of this work. Measuring cell surface proteins, intracellular features, transcription factors, and signaling molecules on tumor cells gives scientists the power to observe expression patterns across and within tumors. In Chapter II, mass cytometry data for 28 patient samples was generated and

analyzed. As noted in Chapter II, visualization of such high dimensional data can be overwhelming. There was not sufficient space in the chapter to depict the scope of cell phenotypes within and between tumors. The data presented in this Appendix represents a per-patient view of the mass cytometry data and RAPID analysis for each sample. All of the t-SNE plots have common axes between patients and are consistent data depicted in Chapter II. However, on each page, only the cells for an individual patient are shown. This allows a reader to see which clusters were present in each patient sample, what proteins are expressed on the cells, and how a given patient's cellular distribution compares to another.

Introduction

In recent years, single cell RNA-seq tools have been leveraged to explore the complex cellular milieu of glioblastoma tissues [33, 37, 47, 101]. These studies have uncovered previously undescribed cell types as well as new targets for therapeutic investigation. This dissertation expands on this work by profiling protein expression and signaling events in individual tumor cells. Specifically, RAPID, described in Chapter II, defines cell clusters within a dataset, that allows comparisons of cluster phenotypes and abundance within and between tumors. Chapter II focuses on selected clusters that are shown to be differentially associated with patient outcomes. These clusters are of particular interest because of their association with important clinical outcomes. However, the remaining cell clusters may also provide insight into the targetable features of glioblastomas and how different patients might be differentially vulnerable to targeted therapies. Table 2.2 reports the abundance of each cell cluster per patient. In order to complement this information, a visual representation of cluster abundance and phenotype was prepared for each patient, based on the results from RAPID analysis in Chapter II.

Methods

Patient samples

Patient samples are described in detail in Chapter II methods. Briefly, the data presented here are derived from surgical resection specimens of 28 *IDH*-wild type glioblastomas collected at Vanderbilt University Medical Center between 2014 and 2016, processed into single cell suspensions following an established protocol (Appendix A and [121]). All samples were collected with patient informed consent in compliance with the Vanderbilt Institutional Review Board (IRBs #030372, #131870, #181970), and in accordance with the declaration of Helsinki.

Patient characteristics and collection of clinical data

Patient information can be found in Table 2.3 and Chapter II methods.

Mass cytometry analysis

Cells derived from patient samples were prepared as previously described in Chapter II and Appendix A [121]. All antibodies used, including clone information, are detailed in Table 2.4. Data was collected on a CyTOF 1.0 instrument located in the Cancer and Immunology Core facility at Vanderbilt University. Mass cytometry standardization beads were used to remove batch effects and to set the variance stabilizing arcsinh scale transformation for each channel following field-standard protocols [121, 146, 167]. Rhodium viability stain and cleaved caspase-3 antibody were included in staining to exclude non-viable and apoptotic cells, respectively. Detection of total histone H3 was used to identify intact, nucleated cells [98]. Data were normalized with MATLAB-based normalization software [166], and were arcsinh transformed (cofactor 5), prior to analysis using the Cytobank platform [168]. A total of over 2 million viable cells from 28 tumors (ranging

from 4,860 to 336,284 cells per tumor) were analyzed. A patient-specific t-SNE view was generated, using 26 of the measured markers for all tumor and stromal cells from each patient's tumor [125]. Immune ($CD45^+$) and endothelial cells ($CD31^+$) were computationally excluded from each individual patient prior to subsequent downstream analysis. Remaining $CD45^-CD31^-$ cells were included in a common t-SNE analysis, generated using 24 of 34 measured markers. This common t-SNE analysis was used for automated analysis of risk stratifying cell subsets in RAPID (below). The distribution of each of the 28 patients' cells on the common t-SNE axes and mass intensity for each marker are shown.

Implementation of RAPID in R

RAPID was performed as described in Chapter II. Cluster abundances and IQR values are reported in Table 2.2 For each cluster, patients were divided into Low and High groups, based on the distribution (IQR) of the abundance of a given cell subset across the cohort.

Cluster Stability Testing

Cluster stability was determined as described in Chapter II. Briefly, a F-measure was calculated per cluster, measuring both the precision and recall of cell assignment in 99 iterative FlowSOM analyses.

Results

Resected glioma tissues were immediately dissociated into single cell suspensions as previously reported (Appendix A and [121]) and the resulting cells were stained with a mass

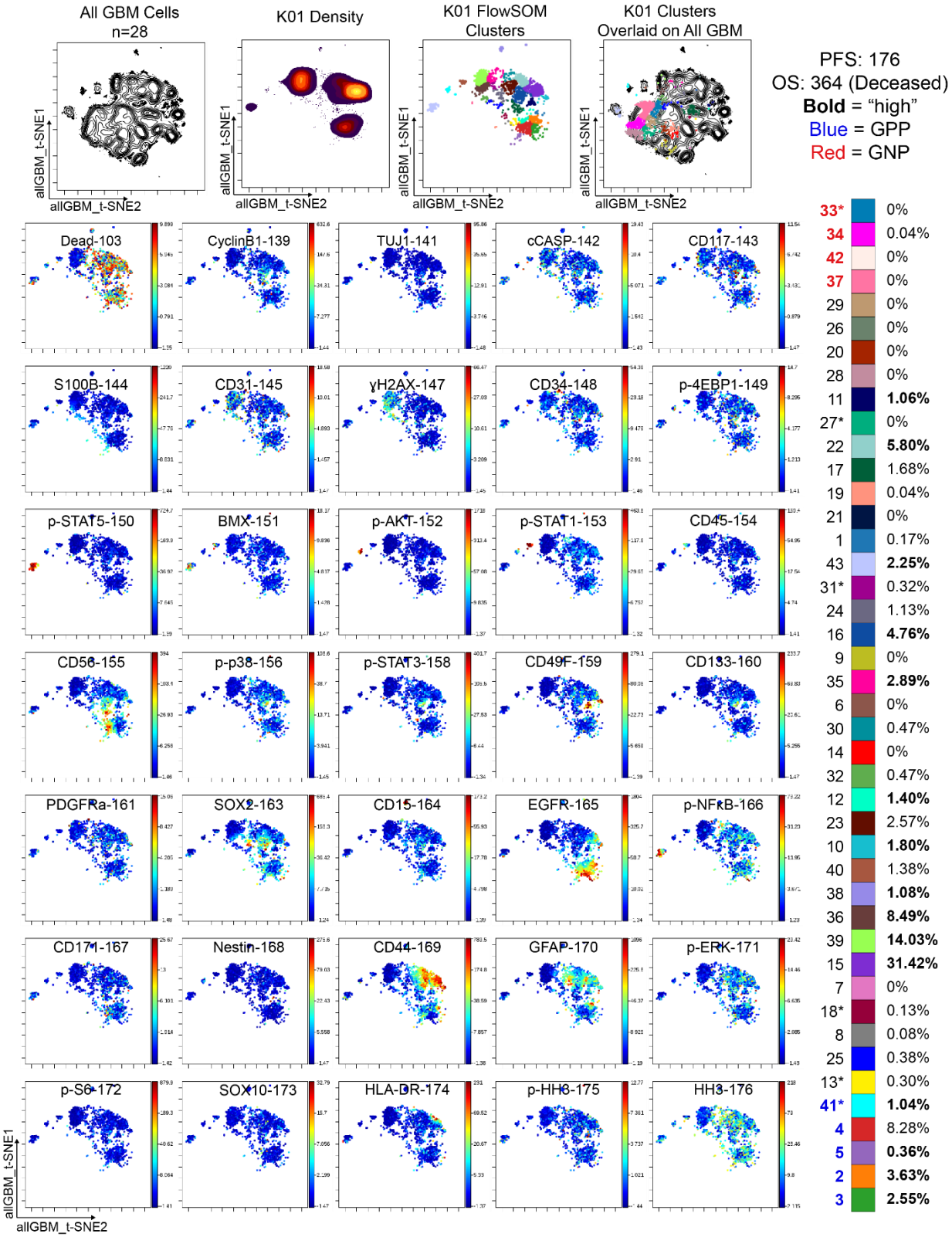
cytometry antibody panel (Table 2.4). For each patient, a t-SNE was generated using 26 of the measured features [125]. On each patient specific t-SNE map, immune ($CD45^+$) and endothelial ($CD45^-CD31^+$) cells were identified and computationally excluded prior to subsequent analysis.

Using 4,710 glioblastoma cells from each patient, a single, common t-SNE map was created (N = 131,880 cells; 4,710 cells x 28 patients, using 24 measured features). The axes of this t-SNE map are those in each panel of Figure B.1. The RAPID algorithm, as described in Chapter II, identified 43 phenotypically distinct cell clusters in this dataset. For each cluster, the abundance (% of tumor cells) was reported for each patient (Table 2.2). Additionally, for each cluster a patient was categorized as high or low abundance, based on the interquartile range of measured abundances across samples. Four of the clusters were determined to be associated with worse overall survival and were called Glioblastoma Negative Prognostic (GNP) clusters (red; clusters 33, 34, 37, and 42). Five of the clusters were associated with longer overall survival and were called Glioblastoma Positive Prognostic (GPP) clusters (blue; clusters 2, 3, 4, 5, and 41). Cluster stability tests, described in Chapter II, based on iterative FlowSOM analyses and repeated cell-subsampling, indicated that 5 of the 43 clusters were not stable, denoted in this work with asterisks.

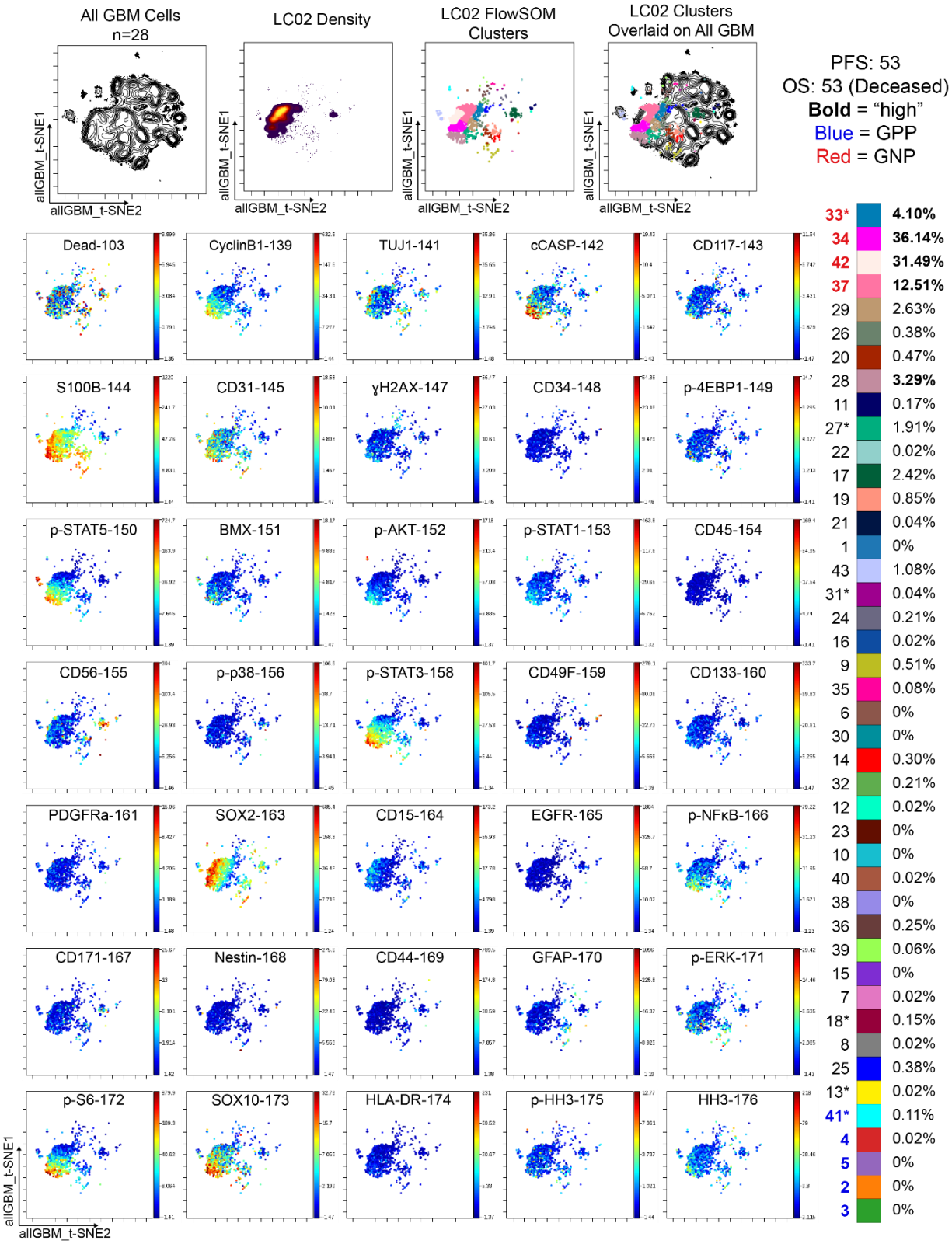
Figure Legend (for pages 222-249)

Figure B.1 **Individual patient tumors have distinct subpopulations of glioma cells and combinations of cell clusters**. B.1a – B.1bb represent data from one patient each (noted in upper left). All t-SNE plots presented have the same axes, from a combined t-SNE analysis of 28 patients. Contour plots indicate cell density. Dot plots indicate cell cluster identity or heat on protein, noted in labels above each plot. To the right, patient information including OS and PFS (top) and cell cluster abundance are provided. Clusters with an * were considered unstable by metrics defined in Chapter II. Blue clusters are associated with better (longer) OS. Red clusters are associated with worse (shorter) OS. Bold clusters indicate that the patient was in the high group for that cluster.

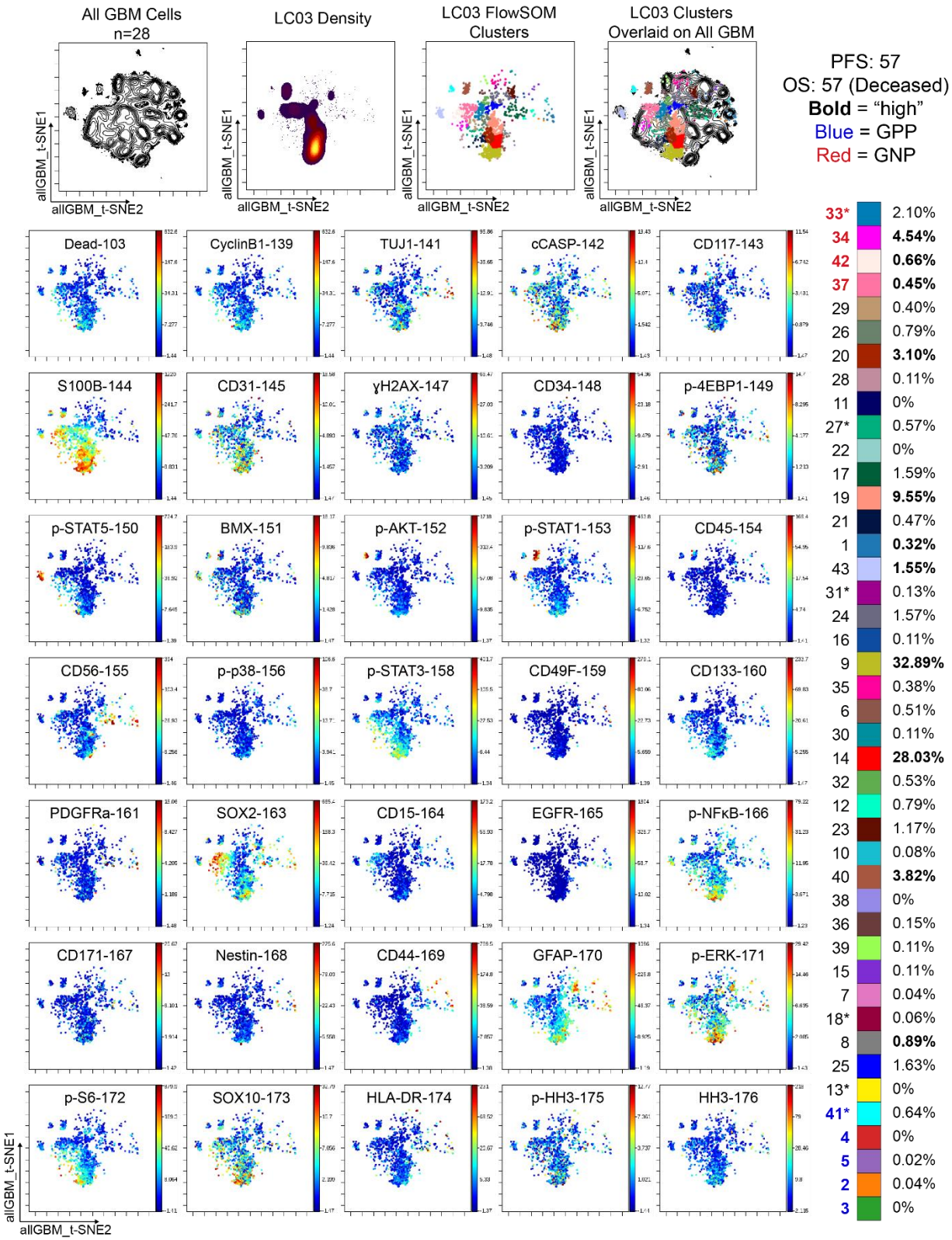
B.1a: K01


All GBM Cells
n=28
K01 Density
K01 FlowSOM
Clusters
K01 Clusters
Overlaid on All GBM
allGBM_t-SNE1
allGBM_t-SNE2
PFS: 176
OS: 364 (Deceased)
Bold = "high"
Blue = GPP
Red = GNP
Dead-103
CyclinB1-139
TUJ1-141
cCASP-142
CD117-143
S100B-144
CD31-145
γH2AX-147
CD34-148
p-4EBP1-149
p-STAT5-150
BMX-151
p-AKT-152
p-STAT1-153
CD45-154
CD56-155
p-p38-156
p-STAT3-158
CD49F-159
CD133-160
PDGFRa-161
SOX2-163
CD15-164
EGFR-165
p-NFkB-166
CD171-167
Nestin-168
CD44-169
GFAP-170
p-ERK-171
p-S6-172
SOX10-173
HLA-DR-174
p-HH3-175
HH3-176
33* 0%
34 0.04%
42 0%
37 0%
29 0%
26 0%
20 0%
28 0%
11 1.06%
27* 0%
22 5.80%
17 1.68%
19 0.04%
21 0%
1 0.17%
43 2.25%
31* 0.32%
24 1.13%
16 4.76%
9 0%
35 2.89%
6 0%
30 0.47%
14 0%
32 0.47%
12 1.40%
23 2.57%
10 1.80%
40 1.38%
38 1.08%
36 8.49%
39 14.03%
15 31.42%
7 0%
18* 0.13%
8 0.08%
25 0.38%
13* 0.30%
41* 1.04%
4 8.28%
5 0.36%
2 3.63%
3 2.55%

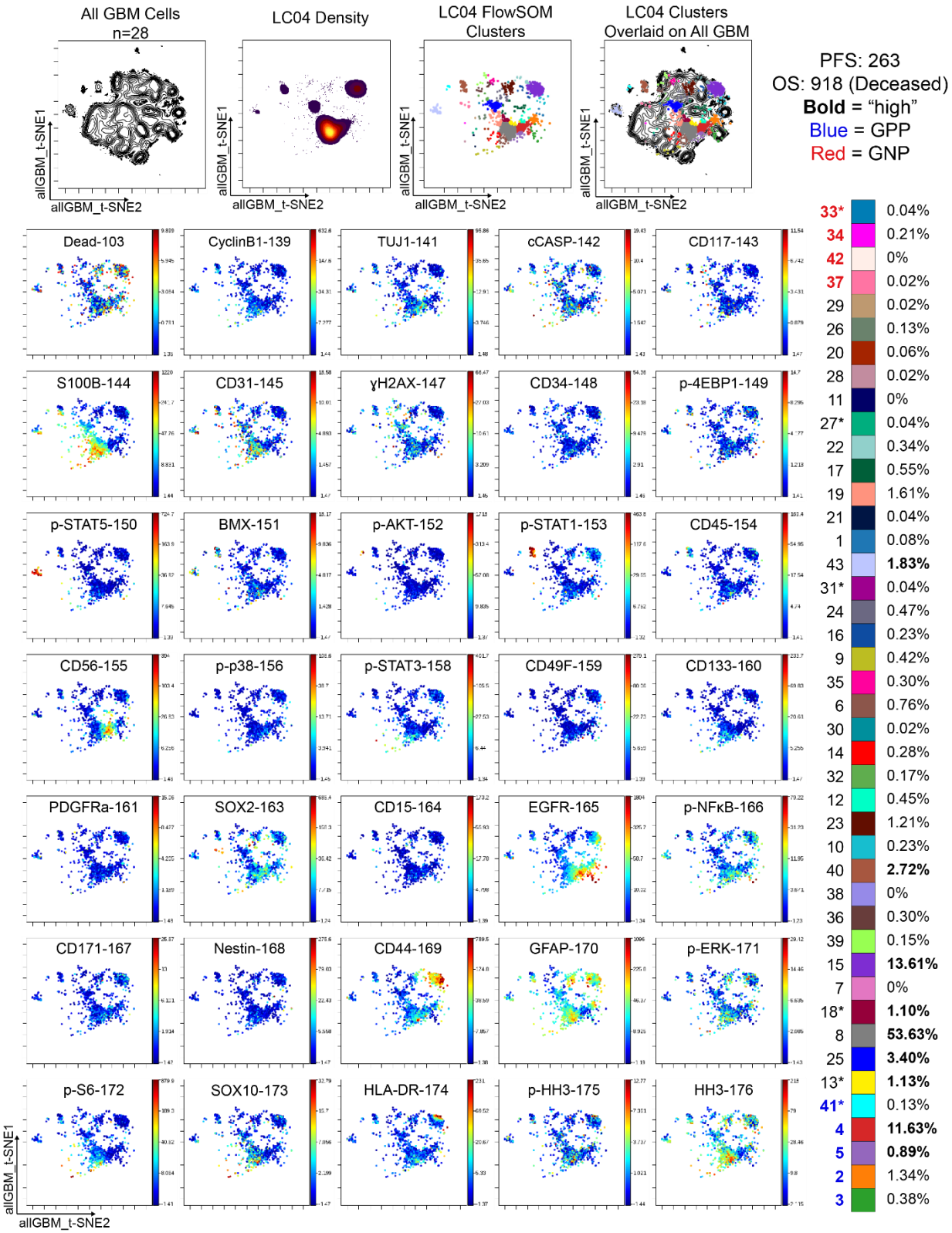
B.1b: LC02


All GBM Cells
n=28
LC02 Density
LC02 FlowSOM
Clusters
LC02 Clusters
Overlaid on All GBM
allGBM_t-SNE1
allGBM_t-SNE2
PFS: 53
OS: 53 (Deceased)
Bold = "high"
Blue = GPP
Red = GNP
Dead-103
CyclinB1-139
TUJ1-141
cCASP-142
CD117-143
S100B-144
CD31-145
γH2AX-147
CD34-148
p-4EBP1-149
p-STAT5-150
BMX-151
p-AKT-152
p-STAT1-153
CD45-154
CD56-155
p-p38-156
p-STAT3-158
CD49F-159
CD133-160
PDGFRa-161
SOX2-163
CD15-164
EGFR-165
p-NFκB-166
CD171-167
Nestin-168
CD44-169
GFAP-170
p-ERK-171
p-S6-172
SOX10-173
HLA-DR-174
p-HH3-175
HH3-176
33* 4.10%
34 36.14%
42 31.49%
37 12.51%
29 2.63%
26 0.38%
20 0.47%
28 3.29%
11 0.17%
27* 1.91%
22 0.02%
17 2.42%
19 0.85%
21 0.04%
1 0%
43 1.08%
31* 0.04%
24 0.21%
16 0.02%
9 0.51%
35 0.08%
6 0%
30 0%
14 0.30%
32 0.21%
12 0.02%
23 0%
10 0%
40 0.02%
38 0%
36 0.25%
39 0.06%
15 0%
7 0.02%
18* 0.15%
8 0.02%
25 0.38%
13* 0.02%
41* 0.11%
4 0.02%
5 0%
2 0%
3 0%

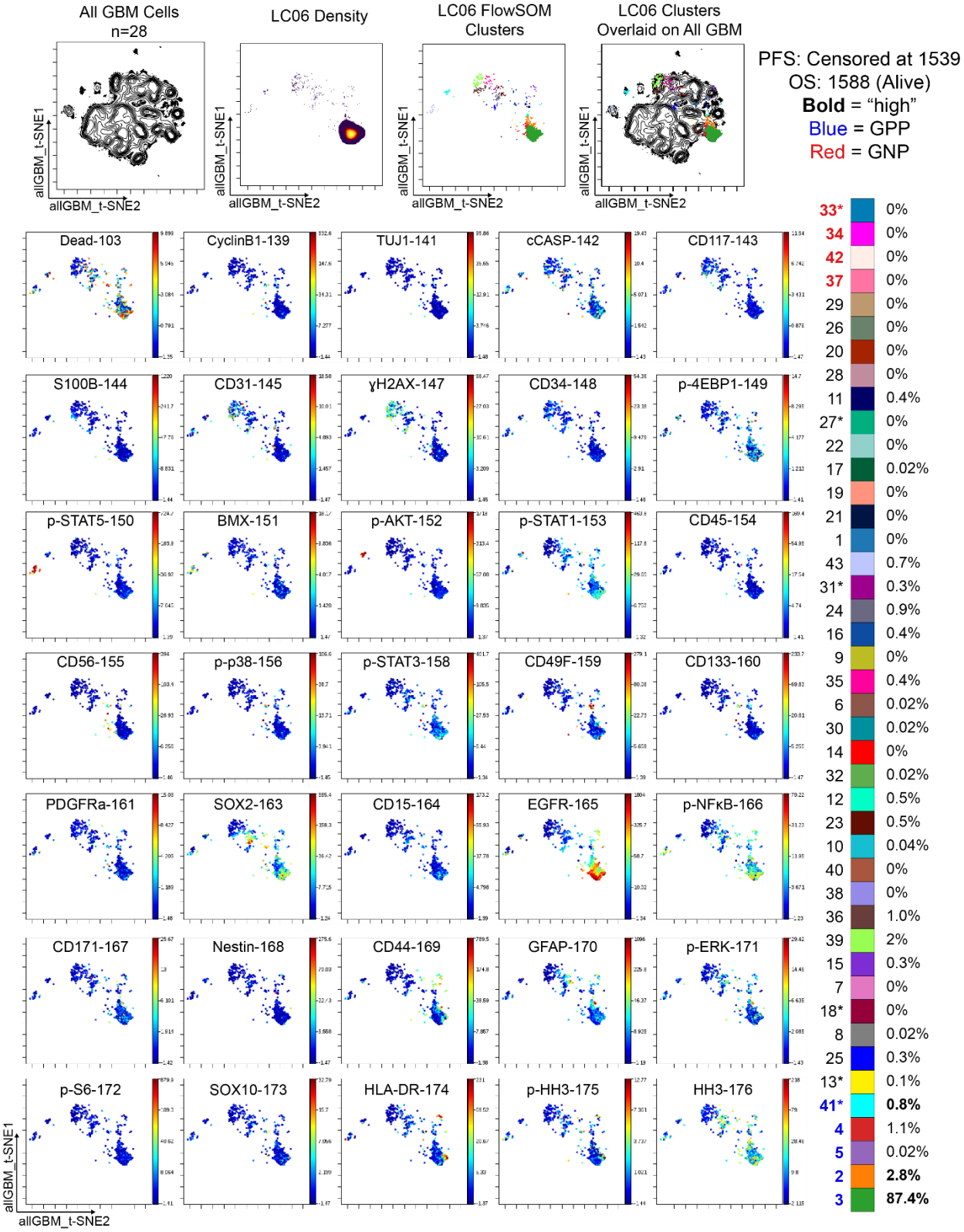
B.1c: LC03


All GBM Cells
n=28
LC03 Density
LC03 FlowSOM
Clusters
LC03 Clusters
Overlaid on All GBM
allGBM_t-SNE1
allGBM_t-SNE2
PFS: 57
OS: 57 (Deceased)
Bold = "high"
Blue = GPP
Red = GNP
Dead-103
CyclinB1-139
TUJ1-141
cCASP-142
CD117-143
S100B-144
CD31-145
γH2AX-147
CD34-148
p-4EBP1-149
p-STAT5-150
BMX-151
p-AKT-152
p-STAT1-153
CD45-154
CD56-155
p-p38-156
p-STAT3-158
CD49F-159
CD133-160
PDGFRa-161
SOX2-163
CD15-164
EGFR-165
p-NFκB-166
CD171-167
Nestin-168
CD44-169
GFAP-170
p-ERK-171
p-S6-172
SOX10-173
HLA-DR-174
p-HH3-175
HH3-176
33* 2.10%
34 4.54%
42 0.66%
37 0.45%
29 0.40%
26 0.79%
20 3.10%
28 0.11%
11 0%
27* 0.57%
22 0%
17 1.59%
19 9.55%
21 0.47%
1 0.32%
43 1.55%
31* 0.13%
24 1.57%
16 0.11%
9 32.89%
35 0.38%
6 0.51%
30 0.11%
14 28.03%
32 0.53%
12 0.79%
23 1.17%
10 0.08%
40 3.82%
38 0%
36 0.15%
39 0.11%
15 0.11%
7 0.04%
18* 0.06%
8 0.89%
25 1.63%
13* 0%
41* 0.64%
4 0%
5 0.02%
2 0.04%
3 0%

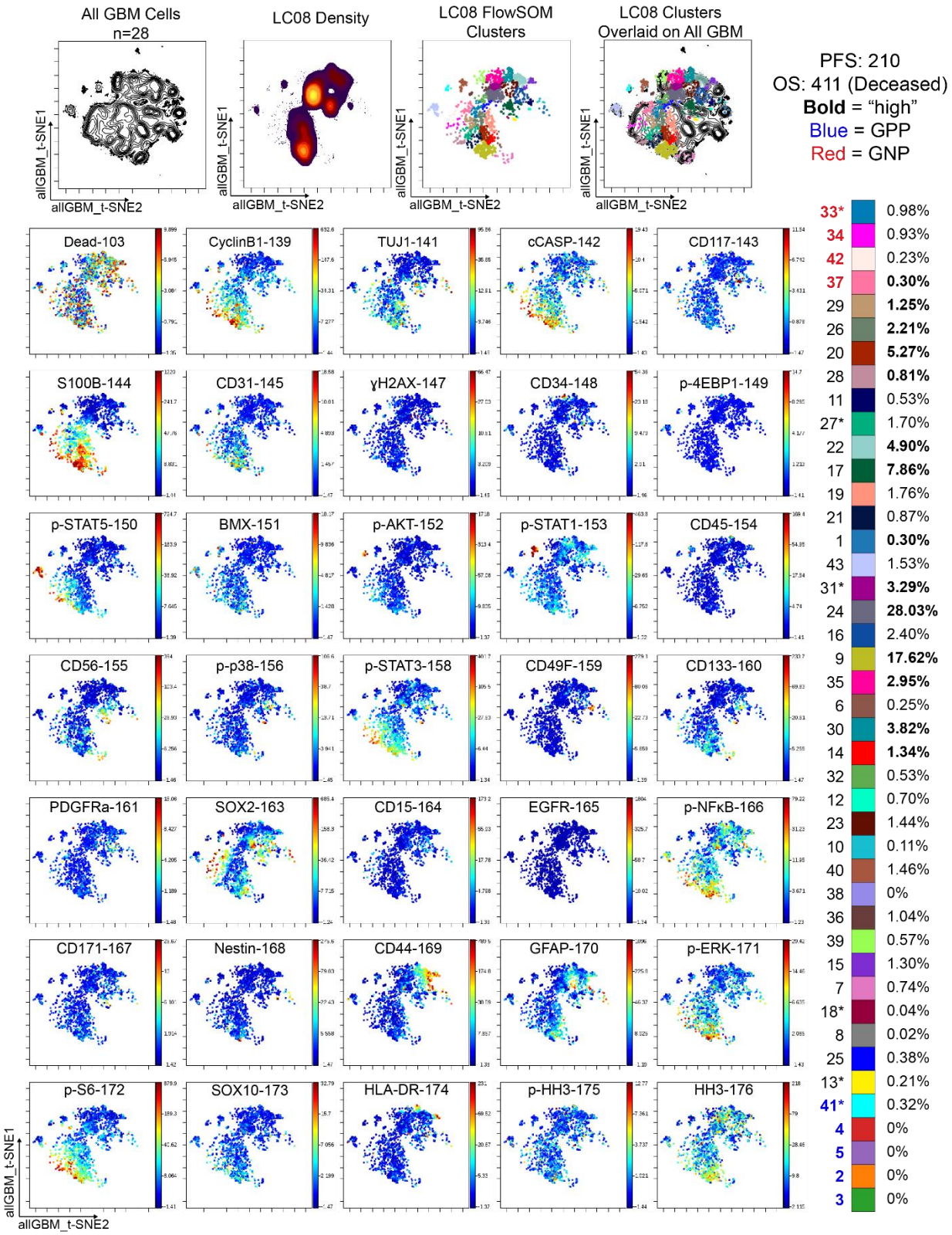
B.1d: LC04


All GBM Cells
n=28
LC04 Density
LC04 FlowSOM
Clusters
LC04 Clusters
Overlaid on All GBM
allGBM_t-SNE1
allGBM_t-SNE2
PFS: 263
OS: 918 (Deceased)
Bold = "high"
Blue = GPP
Red = GNP
Dead-103
CyclinB1-139
TUJ1-141
cCASP-142
CD117-143
S100B-144
CD31-145
γH2AX-147
CD34-148
p-4EBP1-149
p-STAT5-150
BMX-151
p-AKT-152
p-STAT1-153
CD45-154
CD56-155
p-p38-156
p-STAT3-158
CD49F-159
CD133-160
PDGFRa-161
SOX2-163
CD15-164
EGFR-165
p-NFκB-166
CD171-167
Nestin-168
CD44-169
GFAP-170
p-ERK-171
p-S6-172
SOX10-173
HLA-DR-174
p-HH3-175
HH3-176
33* 0.04%
34 0.21%
42 0%
37 0.02%
29 0.02%
26 0.13%
20 0.06%
28 0.02%
11 0%
27* 0.04%
22 0.34%
17 0.55%
19 1.61%
21 0.04%
1 0.08%
43 1.83%
31* 0.04%
24 0.47%
16 0.23%
9 0.42%
35 0.30%
6 0.76%
30 0.02%
14 0.28%
32 0.17%
12 0.45%
23 1.21%
10 0.23%
40 2.72%
38 0%
36 0.30%
39 0.15%
15 13.61%
7 0%
18* 1.10%
8 53.63%
25 3.40%
13* 1.13%
41* 0.13%
4 11.63%
5 0.89%
2 1.34%
3 0.38%

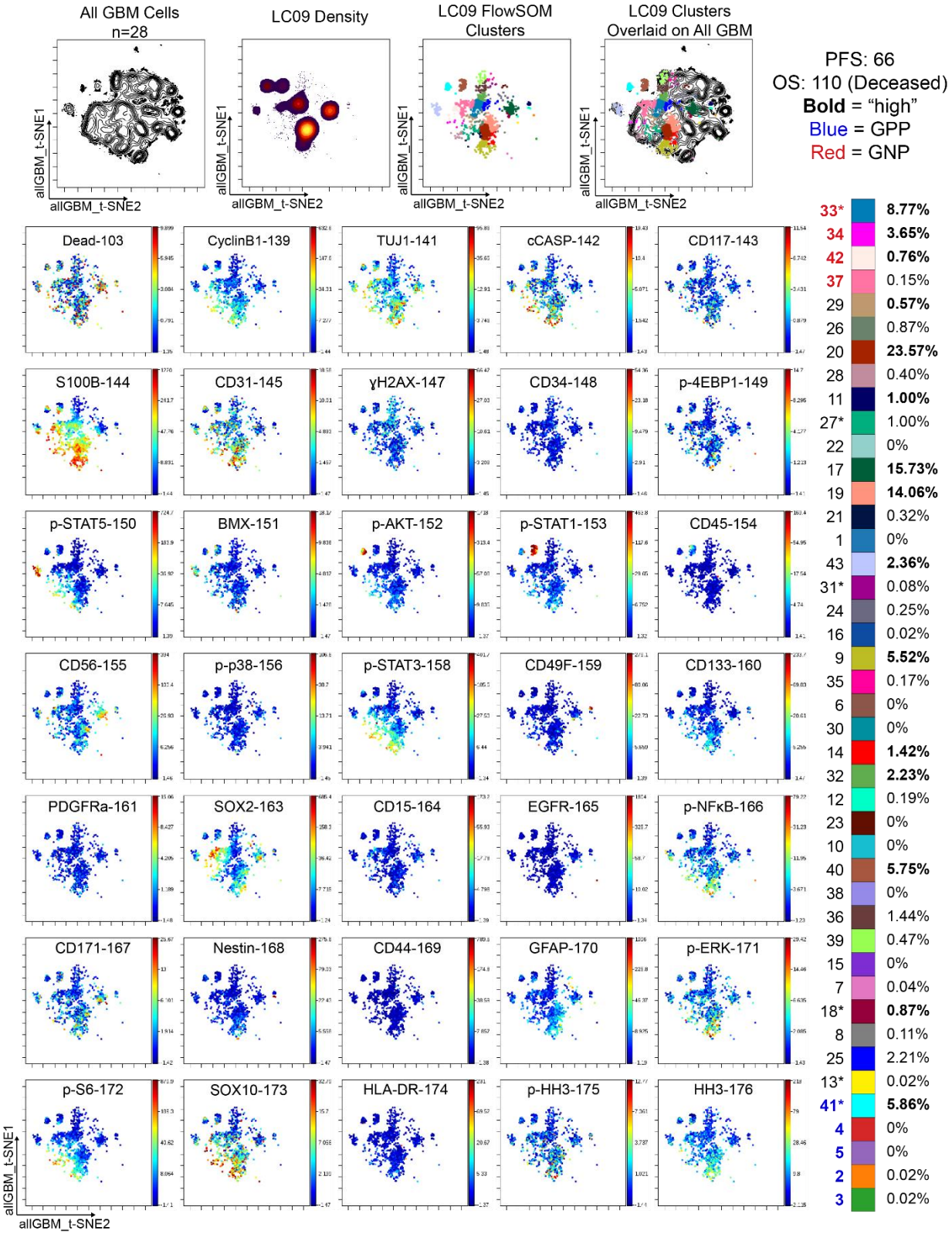
B.1e: LC06


All GBM Cells
n=28
LC06 Density
LC06 FlowSOM
Clusters
LC06 Clusters
Overlaid on All GBM
allGBM_t-SNE1
allGBM_t-SNE2
PFS: Censored at 1539
OS: 1588 (Alive)
Bold = "high"
Blue = GPP
Red = GNP
Dead-103
CyclinB1-139
TUJ1-141
cCASP-142
CD117-143
S100B-144
CD31-145
γH2AX-147
CD34-148
p-4EBP1-149
p-STAT5-150
BMX-151
p-AKT-152
p-STAT1-153
CD45-154
CD56-155
p-p38-156
p-STAT3-158
CD49F-159
CD133-160
PDGFRa-161
SOX2-163
CD15-164
EGFR-165
p-NFκB-166
CD171-167
Nestin-168
CD44-169
GFAP-170
p-ERK-171
p-S6-172
SOX10-173
HLA-DR-174
p-HH3-175
HH3-176
33* 0%
34 0%
42 0%
37 0%
29 0%
26 0%
20 0%
28 0%
11 0.4%
27* 0%
22 0%
17 0.02%
19 0%
21 0%
1 0%
43 0.7%
31* 0.3%
24 0.9%
16 0.4%
9 0%
35 0.4%
6 0.02%
30 0.02%
14 0%
32 0.02%
12 0.5%
23 0.5%
10 0.04%
40 0%
38 0%
36 1.0%
39 2%
15 0.3%
7 0%
18* 0%
8 0.02%
25 0.3%
13* 0.1%
41* 0.8%
4 1.1%
5 0.02%
2 2.8%
3 87.4%

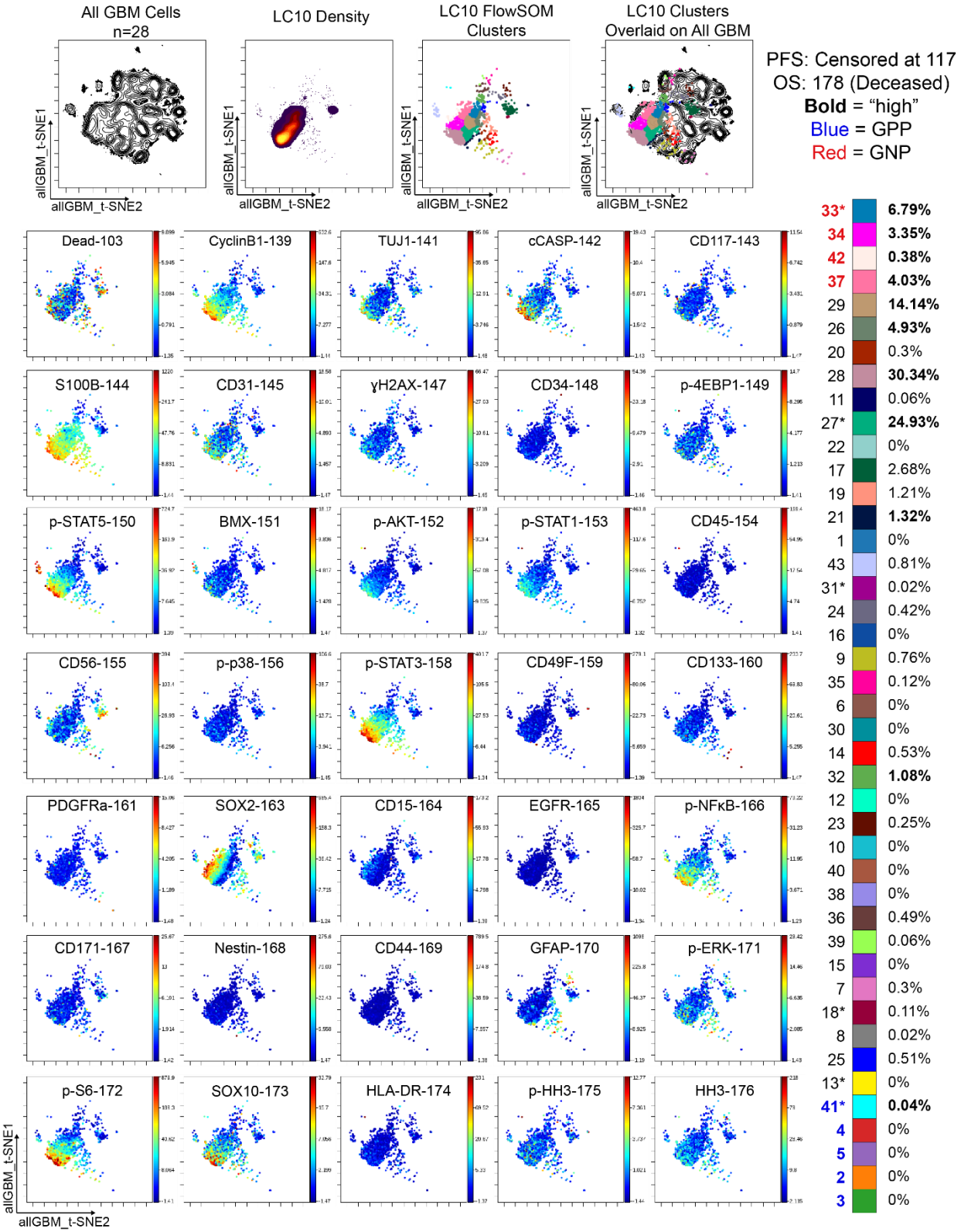
B.1f: LC08


All GBM Cells
n=28
LC08 Density
LC08 FlowSOM
Clusters
LC08 Clusters
Overlaid on All GBM
allGBM_t-SNE1
allGBM_t-SNE2
PFS: 210
OS: 411 (Deceased)
Bold = "high"
Blue = GPP
Red = GNP
Dead-103
CyclinB1-139
TUJ1-141
cCASP-142
CD117-143
S100B-144
CD31-145
γH2AX-147
CD34-148
p-4EBP1-149
p-STAT5-150
BMX-151
p-AKT-152
p-STAT1-153
CD45-154
CD56-155
p-p38-156
p-STAT3-158
CD49F-159
CD133-160
PDGFRa-161
SOX2-163
CD15-164
EGFR-165
p-NFκB-166
CD171-167
Nestin-168
CD44-169
GFAP-170
p-ERK-171
p-S6-172
SOX10-173
HLA-DR-174
p-HH3-175
HH3-176
33* 0.98%
34 0.93%
42 0.23%
37 0.30%
29 1.25%
26 2.21%
20 5.27%
28 0.81%
11 0.53%
27* 1.70%
22 4.90%
17 7.86%
19 1.76%
21 0.87%
1 0.30%
43 1.53%
31* 3.29%
24 28.03%
16 2.40%
9 17.62%
35 2.95%
6 0.25%
30 3.82%
14 1.34%
32 0.53%
12 0.70%
23 1.44%
10 0.11%
40 1.46%
38 0%
36 1.04%
39 0.57%
15 1.30%
7 0.74%
18* 0.04%
8 0.02%
25 0.38%
13* 0.21%
41* 0.32%
4 0%
5 0%
2 0%
3 0%

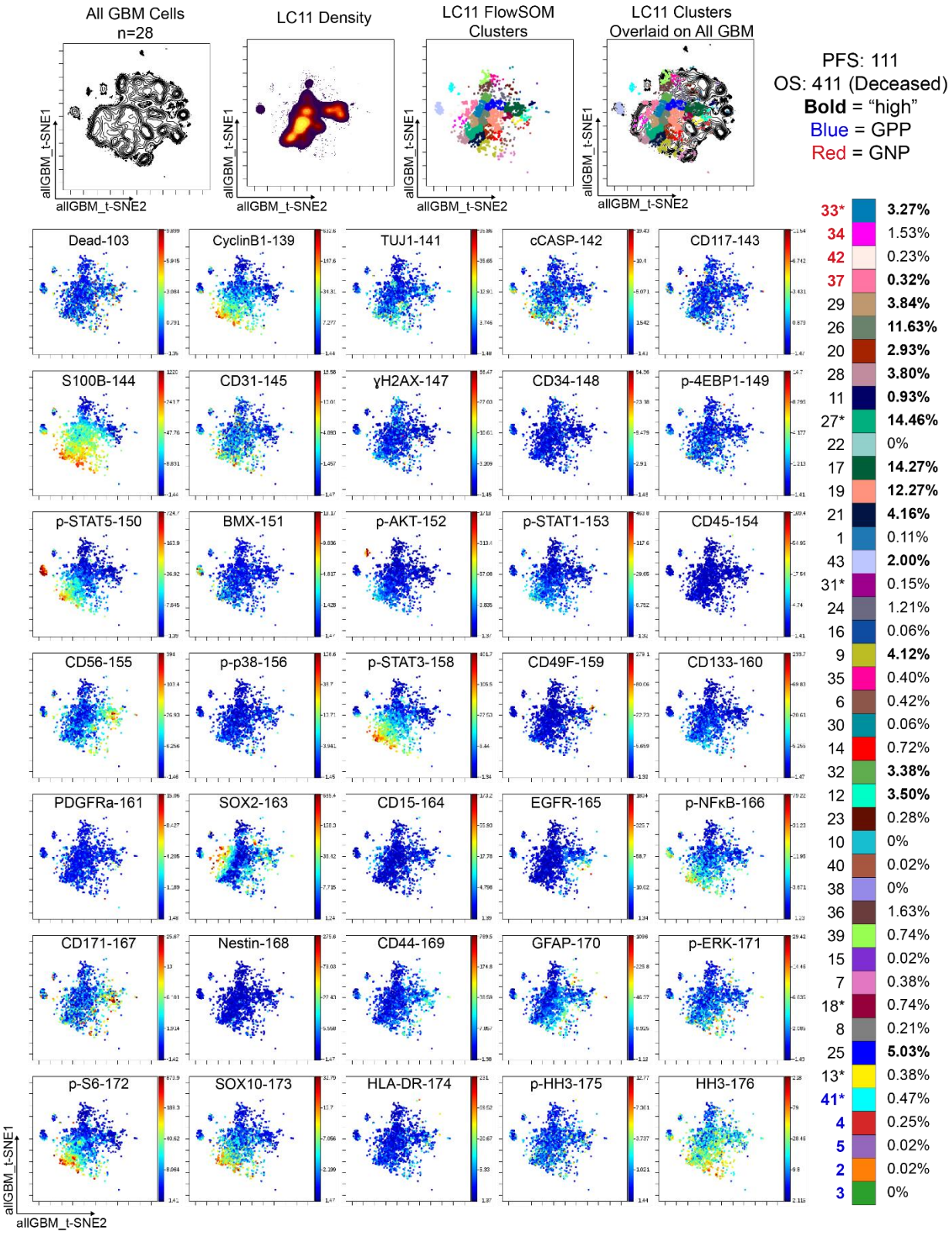
B.1g: LC09


All GBM Cells
n=28
LC09 Density
LC09 FlowSOM
Clusters
LC09 Clusters
Overlaid on All GBM
allGBM_t-SNE1
allGBM_t-SNE2
PFS: 66
OS: 110 (Deceased)
Bold = "high"
Blue = GPP
Red = GNP
Dead-103
CyclinB1-139
TUJ1-141
cCASP-142
CD117-143
S100B-144
CD31-145
γH2AX-147
CD34-148
p-4EBP1-149
p-STAT5-150
BMX-151
p-AKT-152
p-STAT1-153
CD45-154
CD56-155
p-p38-156
p-STAT3-158
CD49F-159
CD133-160
PDGFRa-161
SOX2-163
CD15-164
EGFR-165
p-NFκB-166
CD171-167
Nestin-168
CD44-169
GFAP-170
p-ERK-171
p-S6-172
SOX10-173
HLA-DR-174
p-HH3-175
HH3-176
33* 8.77%
34 3.65%
42 0.76%
37 0.15%
29 0.57%
26 0.87%
20 23.57%
28 0.40%
11 1.00%
27* 1.00%
22 0%
17 15.73%
19 14.06%
21 0.32%
1 0%
43 2.36%
31* 0.08%
24 0.25%
16 0.02%
9 5.52%
35 0.17%
6 0%
30 0%
14 1.42%
32 2.23%
12 0.19%
23 0%
10 0%
40 5.75%
38 0%
36 1.44%
39 0.47%
15 0%
7 0.04%
18* 0.87%
8 0.11%
25 2.21%
13* 0.02%
41* 5.86%
4 0%
5 0%
2 0.02%
3 0.02%

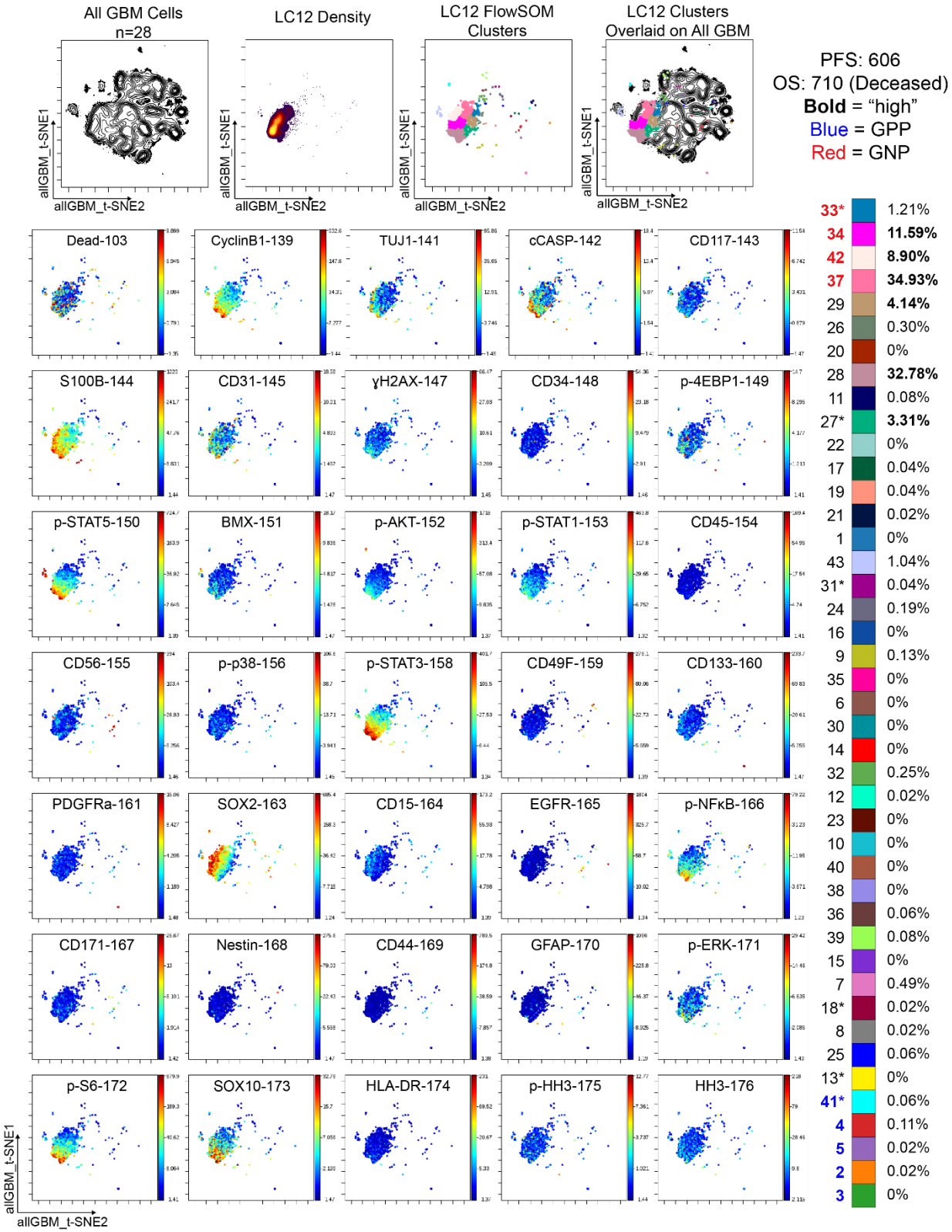
B.1h: LC10


All GBM Cells
n=28
LC10 Density
LC10 FlowSOM
Clusters
LC10 Clusters
Overlaid on All GBM
allGBM_t-SNE1
allGBM_t-SNE2
PFS: Censored at 117
OS: 178 (Deceased)
Bold = "high"
Blue = GPP
Red = GNP
Dead-103
CyclinB1-139
TUJ1-141
cCASP-142
CD117-143
S100B-144
CD31-145
γH2AX-147
CD34-148
p-4EBP1-149
p-STAT5-150
BMX-151
p-AKT-152
p-STAT1-153
CD45-154
CD56-155
p-p38-156
p-STAT3-158
CD49F-159
CD133-160
PDGFRa-161
SOX2-163
CD15-164
EGFR-165
p-NFκB-166
CD171-167
Nestin-168
CD44-169
GFAP-170
p-ERK-171
p-S6-172
SOX10-173
HLA-DR-174
p-HH3-175
HH3-176
33* 6.79%
34 3.35%
42 0.38%
37 4.03%
29 14.14%
26 4.93%
20 0.3%
28 30.34%
11 0.06%
27* 24.93%
22 0%
17 2.68%
19 1.21%
21 1.32%
1 0%
43 0.81%
31* 0.02%
24 0.42%
16 0%
9 0.76%
35 0.12%
6 0%
30 0%
14 0.53%
32 1.08%
12 0%
23 0.25%
10 0%
40 0%
38 0%
36 0.49%
39 0.06%
15 0%
7 0.3%
18* 0.11%
8 0.02%
25 0.51%
13* 0%
41* 0.04%
4 0%
5 0%
2 0%
3 0%

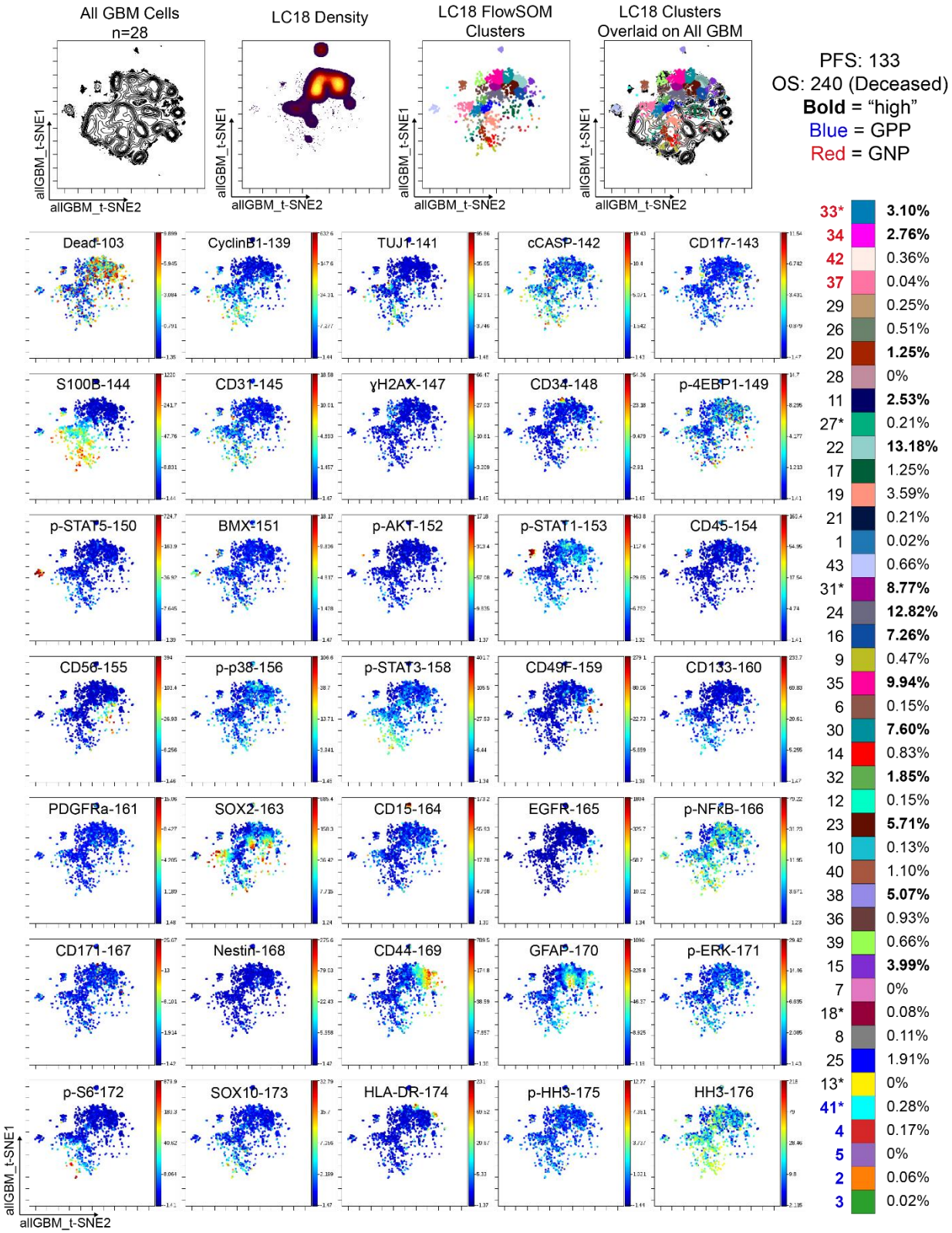
B.1i: LC11


All GBM Cells
n=28
LC11 Density
LC11 FlowSOM
Clusters
LC11 Clusters
Overlaid on All GBM
allGBM_t-SNE1
allGBM_t-SNE2
PFS: 111
OS: 411 (Deceased)
Bold = "high"
Blue = GPP
Red = GNP
Dead-103
CyclinB1-139
TUJ1-141
cCASP-142
CD117-143
S100B-144
CD31-145
γH2AX-147
CD34-148
p-4EBP1-149
p-STAT5-150
BMX-151
p-AKT-152
p-STAT1-153
CD45-154
CD56-155
p-p38-156
p-STAT3-158
CD49F-159
CD133-160
PDGFRa-161
SOX2-163
CD15-164
EGFR-165
p-NFκB-166
CD171-167
Nestin-168
CD44-169
GFAP-170
p-ERK-171
p-S6-172
SOX10-173
HLA-DR-174
p-HH3-175
HH3-176
33* 3.27%
34 1.53%
42 0.23%
37 0.32%
29 3.84%
26 11.63%
20 2.93%
28 3.80%
11 0.93%
27* 14.46%
22 0%
17 14.27%
19 12.27%
21 4.16%
1 0.11%
43 2.00%
31* 0.15%
24 1.21%
16 0.06%
9 4.12%
35 0.40%
6 0.42%
30 0.06%
14 0.72%
32 3.38%
12 3.50%
23 0.28%
10 0%
40 0.02%
38 0%
36 1.63%
39 0.74%
15 0.02%
7 0.38%
18* 0.74%
8 0.21%
25 5.03%
13* 0.38%
41* 0.47%
4 0.25%
5 0.02%
2 0.02%
3 0%

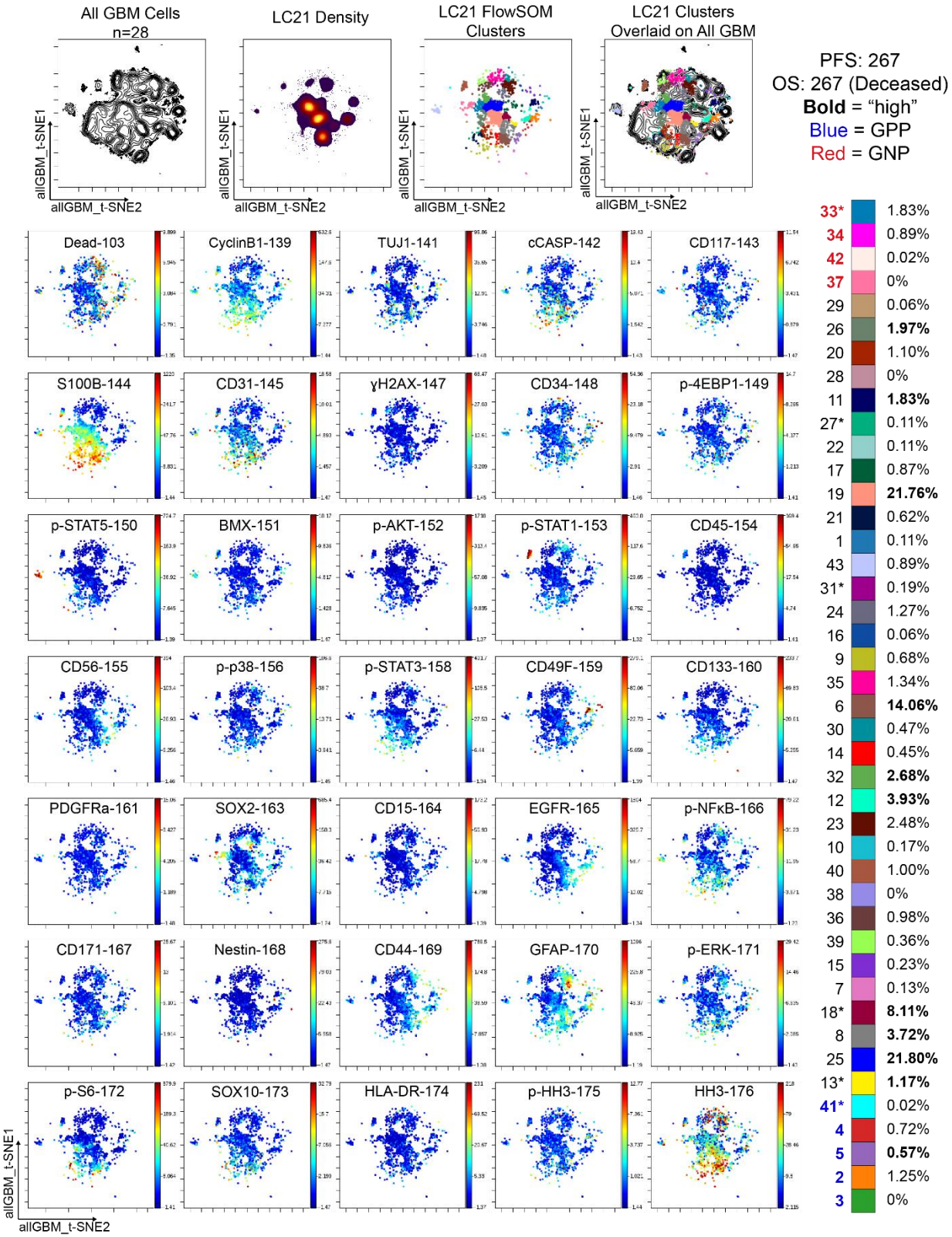
B.1j: LC12


All GBM Cells
n=28
LC12 Density
LC12 FlowSOM
Clusters
LC12 Clusters
Overlaid on All GBM
allGBM_t-SNE1
allGBM_t-SNE2
PFS: 606
OS: 710 (Deceased)
Bold = "high"
Blue = GPP
Red = GNP
Dead-103
CyclinB1-139
TUJ1-141
cCASP-142
CD117-143
S100B-144
CD31-145
γH2AX-147
CD34-148
p-4EBP1-149
p-STAT5-150
BMX-151
p-AKT-152
p-STAT1-153
CD45-154
CD56-155
p-p38-156
p-STAT3-158
CD49F-159
CD133-160
PDGFRa-161
SOX2-163
CD15-164
EGFR-165
p-NFκB-166
CD171-167
Nestin-168
CD44-169
GFAP-170
p-ERK-171
p-S6-172
SOX10-173
HLA-DR-174
p-HH3-175
HH3-176
33* 1.21%
34 11.59%
42 8.90%
37 34.93%
29 4.14%
26 0.30%
20 0%
28 32.78%
11 0.08%
27* 3.31%
22 0%
17 0.04%
19 0.04%
21 0.02%
1 0%
43 1.04%
31* 0.04%
24 0.19%
16 0%
9 0.13%
35 0%
6 0%
30 0%
14 0%
32 0.25%
12 0.02%
23 0%
10 0%
40 0%
38 0%
36 0.06%
39 0.08%
15 0%
7 0.49%
18* 0.02%
8 0.02%
25 0.06%
13* 0%
41* 0.06%
4 0.11%
5 0.02%
2 0.02%
3 0%

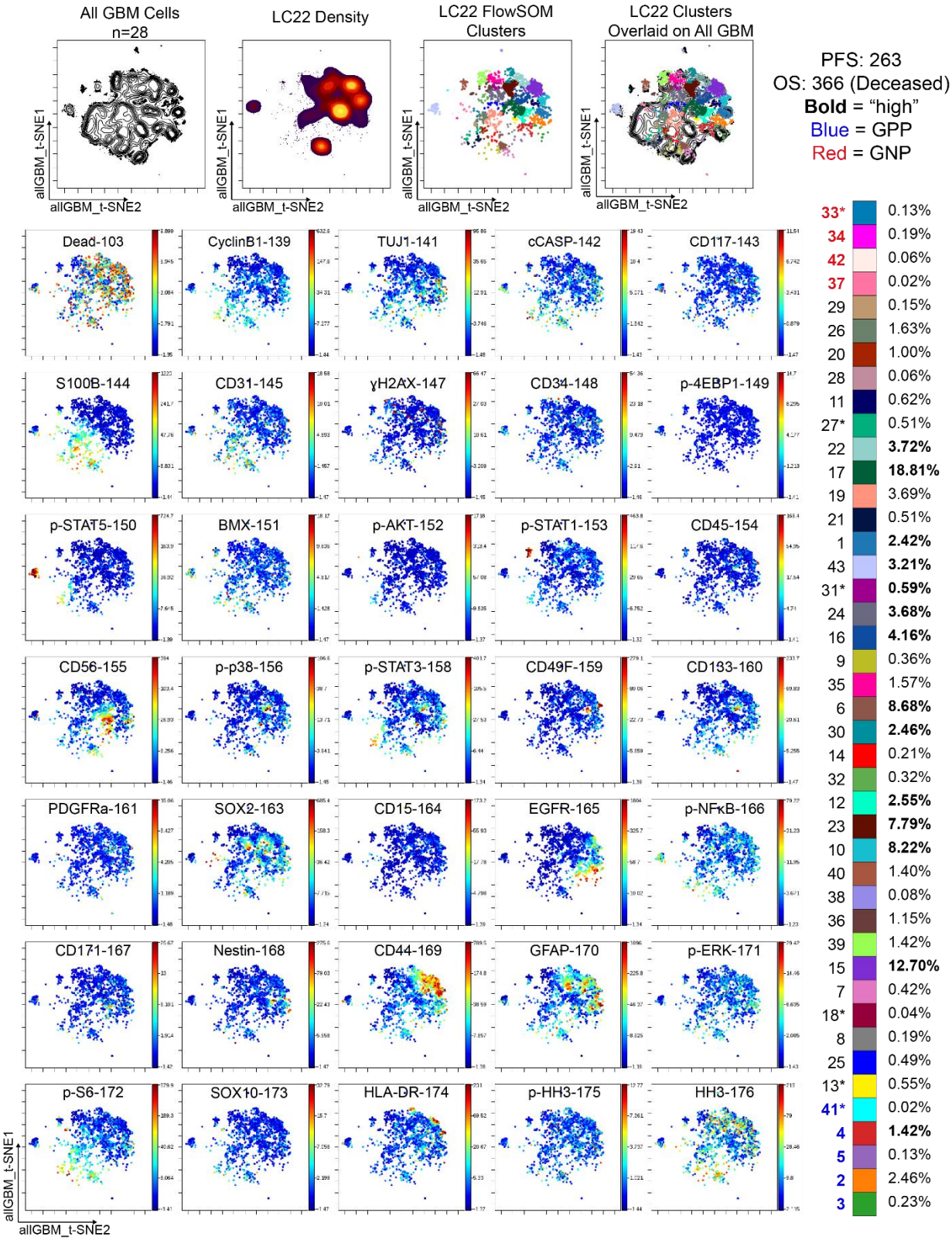
B.1k: LC18


All GBM Cells
n=28
LC18 Density
LC18 FlowSOM
Clusters
LC18 Clusters
Overlaid on All GBM
allGBM_t-SNE1
allGBM_t-SNE2
PFS: 133
OS: 240 (Deceased)
Bold = "high"
Blue = GPP
Red = GNP
Dead-103
CyclinB1-139
TUJ1-141
cCASP-142
CD117-143
S100B-144
CD31-145
γH2AX-147
CD34-148
p-4EBP1-149
p-STAT5-150
BMX-151
p-AKT-152
p-STAT1-153
CD45-154
CD56-155
p-p38-156
p-STAT3-158
CD49F-159
CD133-160
PDGFRa-161
SOX2-163
CD15-164
EGFR-165
p-NFKB-166
CD171-167
Nestin-168
CD44-169
GFAP-170
p-ERK-171
p-S6-172
SOX10-173
HLA-DR-174
p-HH3-175
HH3-176
33* 3.10%
34 2.76%
42 0.36%
37 0.04%
29 0.25%
26 0.51%
20 1.25%
28 0%
11 2.53%
27* 0.21%
22 13.18%
17 1.25%
19 3.59%
21 0.21%
1 0.02%
43 0.66%
31* 8.77%
24 12.82%
16 7.26%
9 0.47%
35 9.94%
6 0.15%
30 7.60%
14 0.83%
32 1.85%
12 0.15%
23 5.71%
10 0.13%
40 1.10%
38 5.07%
36 0.93%
39 0.66%
15 3.99%
7 0%
18* 0.08%
8 0.11%
25 1.91%
13* 0%
41* 0.28%
4 0.17%
5 0%
2 0.06%
3 0.02%

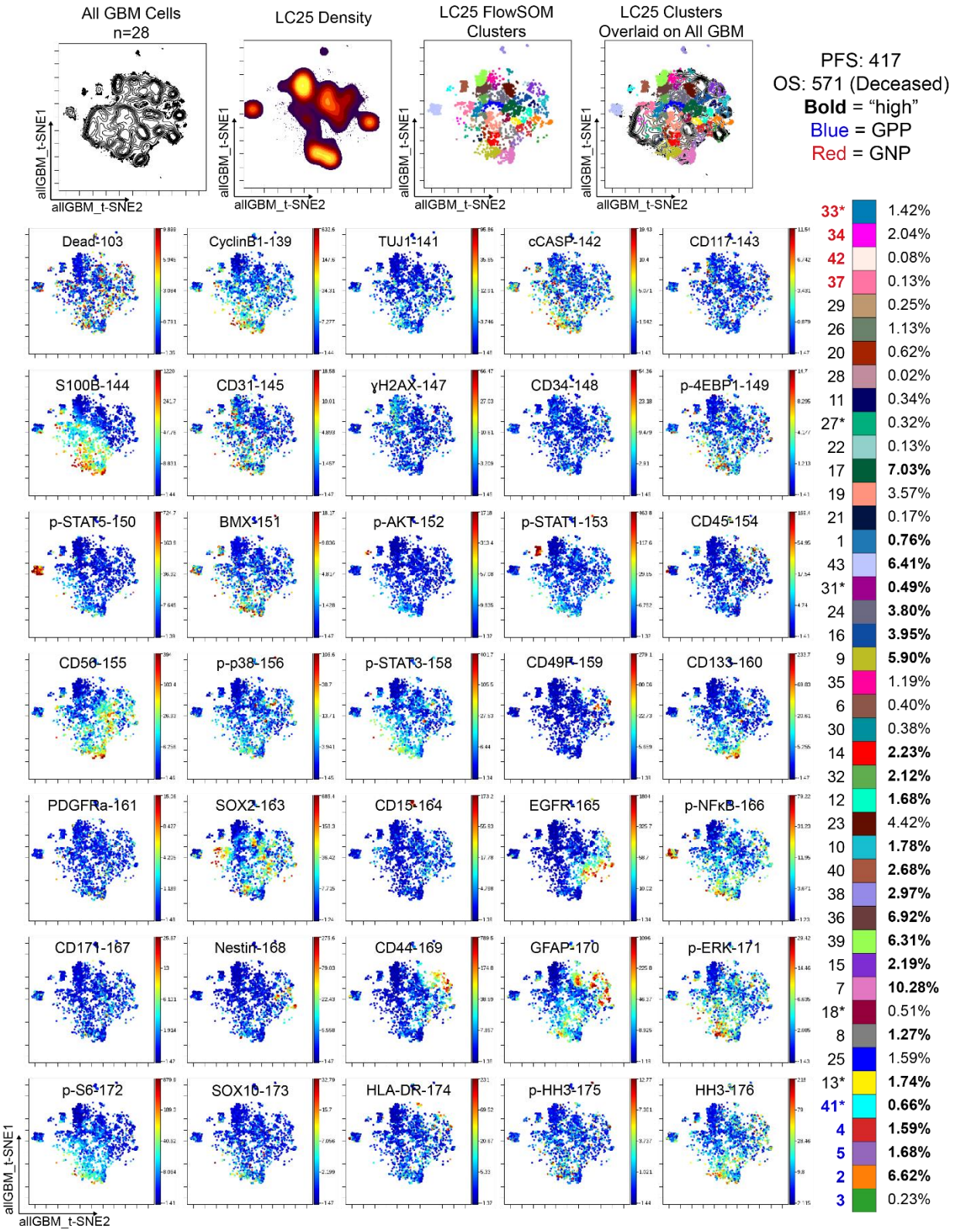
B.1I: LC21


All GBM Cells
n=28
LC21 Density
LC21 FlowSOM
Clusters
LC21 Clusters
Overlaid on All GBM
allGBM_t-SNE1
allGBM_t-SNE2
PFS: 267
OS: 267 (Deceased)
Bold = "high"
Blue = GPP
Red = GNP
Dead-103
CyclinB1-139
TUJ1-141
cCASP-142
CD117-143
S100B-144
CD31-145
γH2AX-147
CD34-148
p-4EBP1-149
p-STAT5-150
BMX-151
p-AKT-152
p-STAT1-153
CD45-154
CD56-155
p-p38-156
p-STAT3-158
CD49F-159
CD133-160
PDGFRa-161
SOX2-163
CD15-164
EGFR-165
p-NFκB-166
CD171-167
Nestin-168
CD44-169
GFAP-170
p-ERK-171
p-S6-172
SOX10-173
HLA-DR-174
p-HH3-175
HH3-176
33* 1.83%
34 0.89%
42 0.02%
37 0%
29 0.06%
26 1.97%
20 1.10%
28 0%
11 1.83%
27* 0.11%
22 0.11%
17 0.87%
19 21.76%
21 0.62%
1 0.11%
43 0.89%
31* 0.19%
24 1.27%
16 0.06%
9 0.68%
35 1.34%
6 14.06%
30 0.47%
14 0.45%
32 2.68%
12 3.93%
23 2.48%
10 0.17%
40 1.00%
38 0%
36 0.98%
39 0.36%
15 0.23%
7 0.13%
18* 8.11%
8 3.72%
25 21.80%
13* 1.17%
41* 0.02%
4 0.72%
5 0.57%
2 1.25%
3 0%

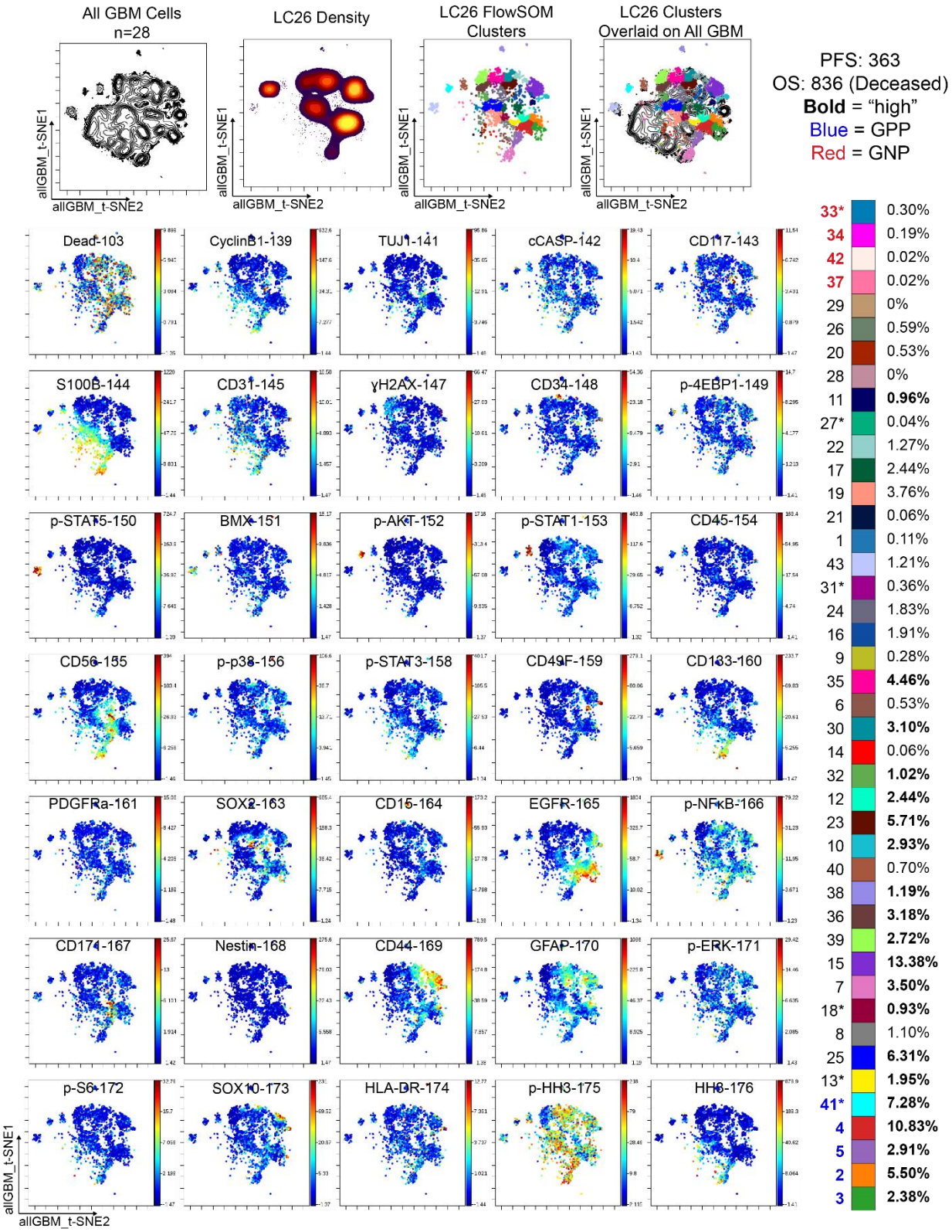
B.1m: LC22


All GBM Cells
n=28
LC22 Density
LC22 FlowSOM
Clusters
LC22 Clusters
Overlaid on All GBM
allGBM_t-SNE1
allGBM_t-SNE2
PFS: 263
OS: 366 (Deceased)
Bold = "high"
Blue = GPP
Red = GNP
Dead-103
CyclinB1-139
TUJ1-141
cCASP-142
CD117-143
S100B-144
CD31-145
γH2AX-147
CD34-148
p-4EBP1-149
p-STAT5-150
BMX-151
p-AKT-152
p-STAT1-153
CD45-154
CD56-155
p-p38-156
p-STAT3-158
CD49F-159
CD133-160
PDGFRa-161
SOX2-163
CD15-164
EGFR-165
p-NFκB-166
CD171-167
Nestin-168
CD44-169
GFAP-170
p-ERK-171
p-S6-172
SOX10-173
HLA-DR-174
p-HH3-175
HH3-176
33* 0.13%
34 0.19%
42 0.06%
37 0.02%
29 0.15%
26 1.63%
20 1.00%
28 0.06%
11 0.62%
27* 0.51%
22 3.72%
17 18.81%
19 3.69%
21 0.51%
1 2.42%
43 3.21%
31* 0.59%
24 3.68%
16 4.16%
9 0.36%
35 1.57%
6 8.68%
30 2.46%
14 0.21%
32 0.32%
12 2.55%
23 7.79%
10 8.22%
40 1.40%
38 0.08%
36 1.15%
39 1.42%
15 12.70%
7 0.42%
18* 0.04%
8 0.19%
25 0.49%
13* 0.55%
41* 0.02%
4 1.42%
5 0.13%
2 2.46%
3 0.23%

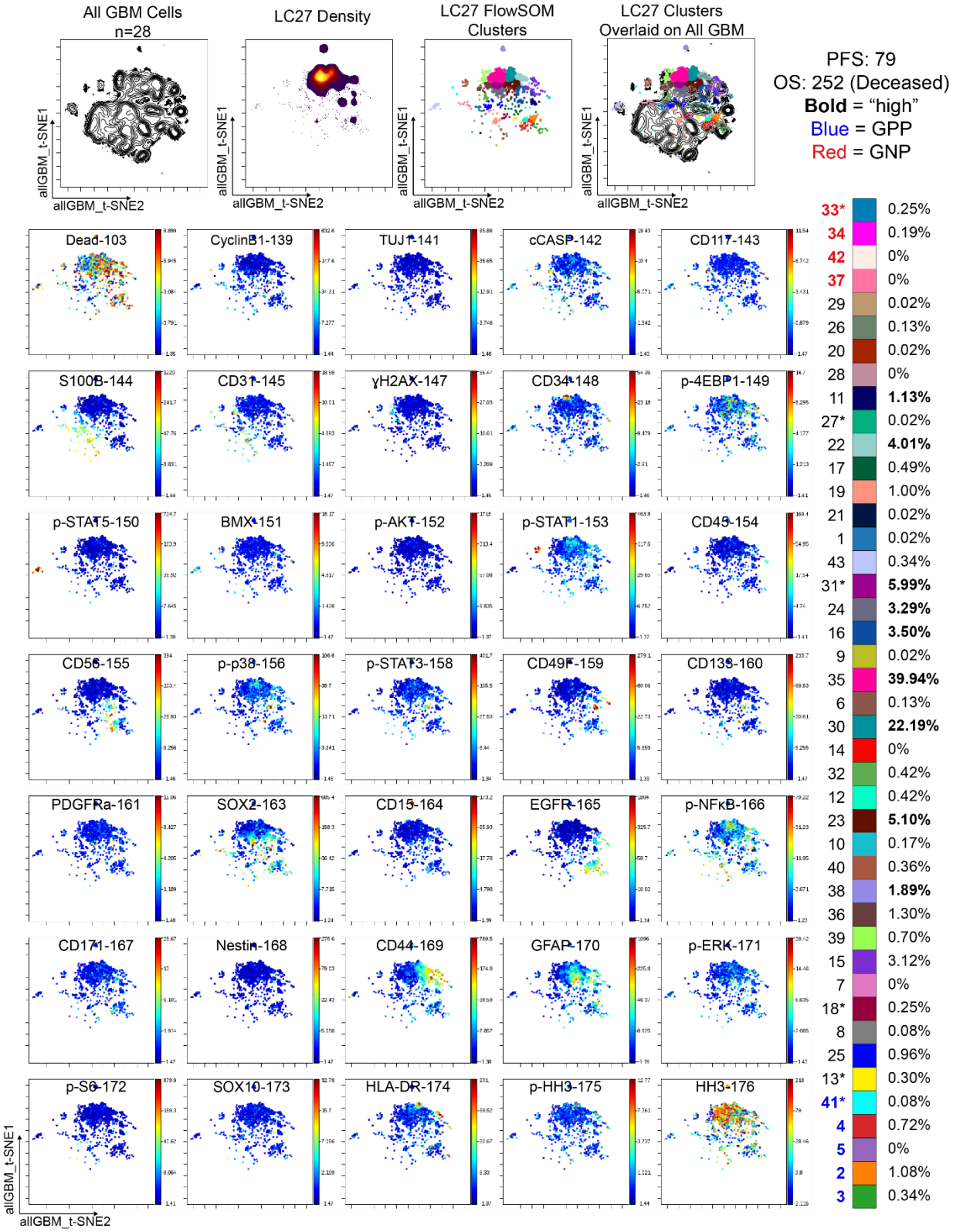
B.1n: LC25


All GBM Cells
n=28
LC25 Density
LC25 FlowSOM
Clusters
LC25 Clusters
Overlaid on All GBM
allGBM_t-SNE1
allGBM_t-SNE2
PFS: 417
OS: 571 (Deceased)
Bold = "high"
Blue = GPP
Red = GNP
Dead-103
CyclinB1-139
TUJ1-141
cCASP-142
CD117-143
S100B-144
CD31-145
γH2AX-147
CD34-148
p-4EBP1-149
p-STAT5-150
BMX-151
p-AKT-152
p-STAT1-153
CD45-154
CD56-155
p-p38-156
p-STAT3-158
CD49F-159
CD133-160
PDGFRa-161
SOX2-163
CD15-164
EGFR-165
p-NFκB-166
CD171-167
Nestin-168
CD44-169
GFAP-170
p-ERK-171
p-S6-172
SOX10-173
HLA-DR-174
p-HH3-175
HH3-176
33* 1.42%
34 2.04%
42 0.08%
37 0.13%
29 0.25%
26 1.13%
20 0.62%
28 0.02%
11 0.34%
27* 0.32%
22 0.13%
17 7.03%
19 3.57%
21 0.17%
1 0.76%
43 6.41%
31* 0.49%
24 3.80%
16 3.95%
9 5.90%
35 1.19%
6 0.40%
30 0.38%
14 2.23%
32 2.12%
12 1.68%
23 4.42%
10 1.78%
40 2.68%
38 2.97%
36 6.92%
39 6.31%
15 2.19%
7 10.28%
18* 0.51%
8 1.27%
25 1.59%
13* 1.74%
41* 0.66%
4 1.59%
5 1.68%
2 6.62%
3 0.23%

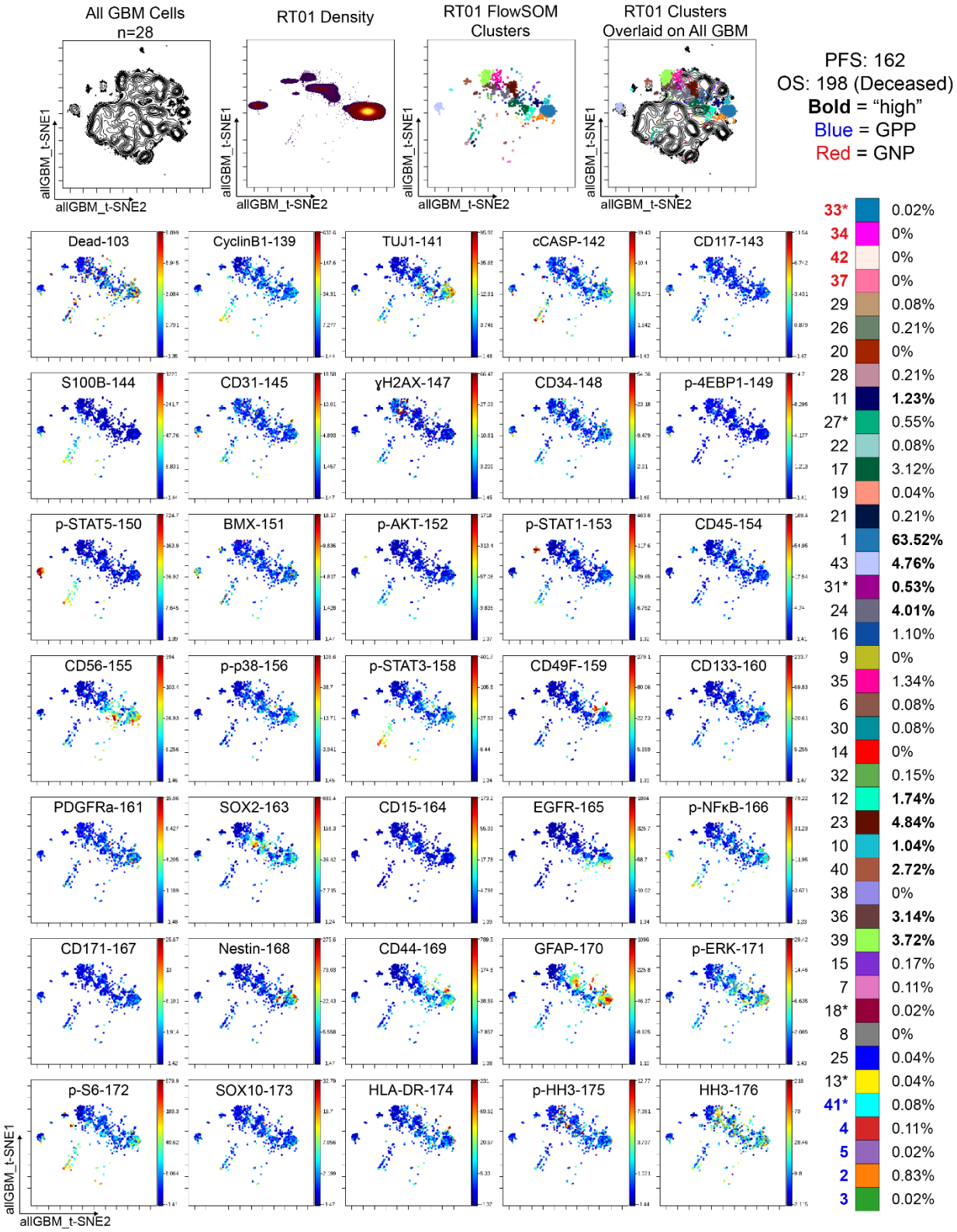
B.1o: LC26


All GBM Cells
n=28
LC26 Density
LC26 FlowSOM
Clusters
LC26 Clusters
Overlaid on All GBM
allGBM_t-SNE1
allGBM_t-SNE2
PFS: 363
OS: 836 (Deceased)
Bold = "high"
Blue = GPP
Red = GNP
Dead-103
CyclinB1-139
TUJ1-141
cCASP-142
CD117-143
S100B-144
CD31-145
γH2AX-147
CD34-148
p-4EBP1-149
p-STAT5-150
BMX-151
p-AKT-152
p-STAT1-153
CD45-154
CD56-155
p-p38-156
p-STAT3-158
CD49F-159
CD133-160
PDGFRa-161
SOX2-163
CD15-164
EGFR-165
p-NFκB-166
CD171-167
Nestin-168
CD44-169
GFAP-170
p-ERK-171
p-S6-172
SOX10-173
HLA-DR-174
p-HH3-175
HH3-176
33* 0.30%
34 0.19%
42 0.02%
37 0.02%
29 0%
26 0.59%
20 0.53%
28 0%
11 0.96%
27* 0.04%
22 1.27%
17 2.44%
19 3.76%
21 0.06%
1 0.11%
43 1.21%
31* 0.36%
24 1.83%
16 1.91%
9 0.28%
35 4.46%
6 0.53%
30 3.10%
14 0.06%
32 1.02%
12 2.44%
23 5.71%
10 2.93%
40 0.70%
38 1.19%
36 3.18%
39 2.72%
15 13.38%
7 3.50%
18* 0.93%
8 1.10%
25 6.31%
13* 1.95%
41* 7.28%
4 10.83%
5 2.91%
2 5.50%
3 2.38%

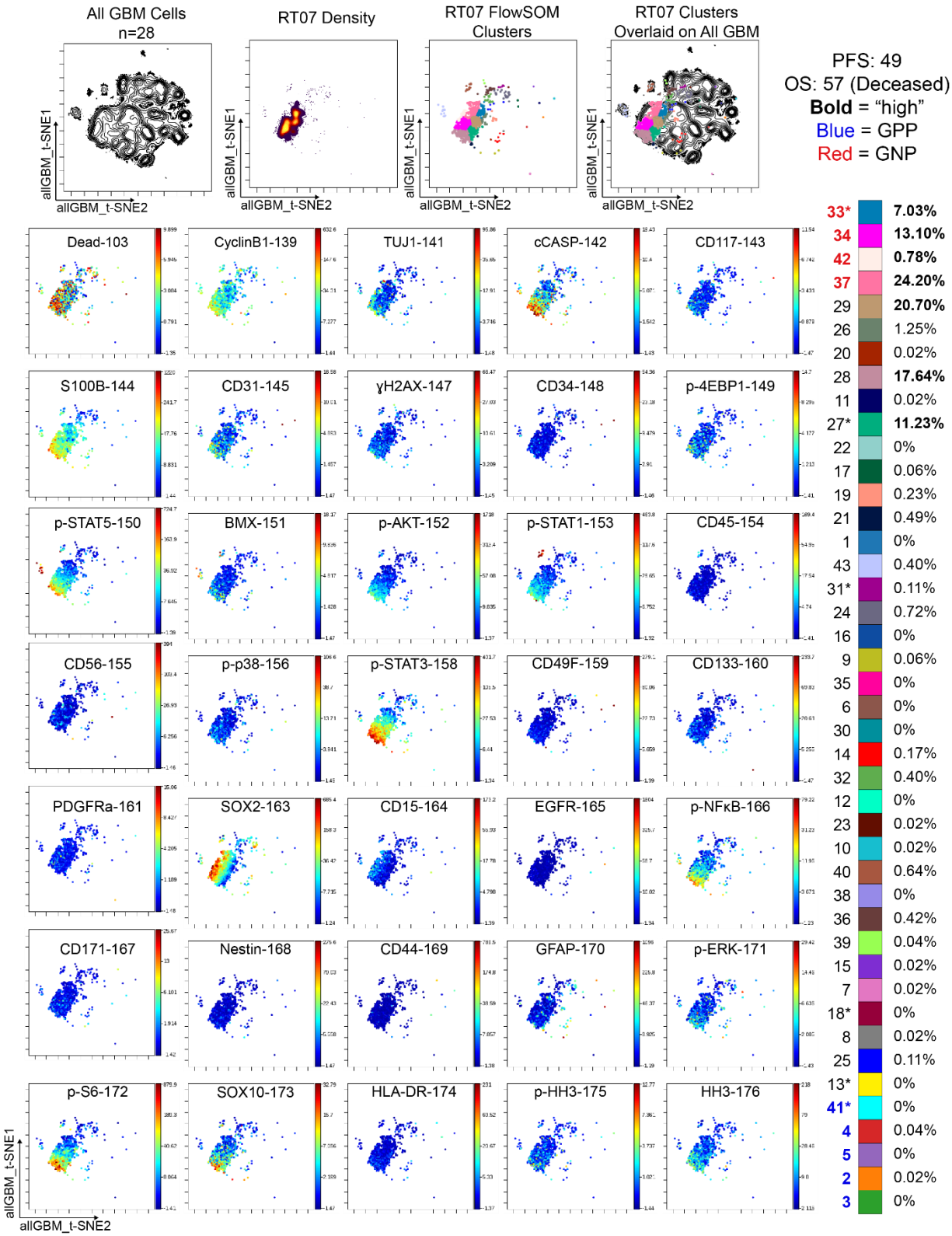
B.1p: LC27


All GBM Cells
n=28
LC27 Density
LC27 FlowSOM
Clusters
LC27 Clusters
Overlaid on All GBM
allGBM_t-SNE1
allGBM_t-SNE2
PFS: 79
OS: 252 (Deceased)
Bold = "high"
Blue = GPP
Red = GNP
Dead-103
CyclinB1-139
TUJ1-141
cCASP-142
CD117-143
S100B-144
CD31-145
γH2AX-147
CD34-148
p-4EBP1-149
p-STAT5-150
BMX-151
p-AKT-152
p-STAT1-153
CD45-154
CD56-155
p-p38-156
p-STAT3-158
CD49F-159
CD133-160
PDGFRa-161
SOX2-163
CD15-164
EGFR-165
p-NFκB-166
CD171-167
Nestin-168
CD44-169
GFAP-170
p-ERK-171
p-S6-172
SOX10-173
HLA-DR-174
p-HH3-175
HH3-176
33* 0.25%
34 0.19%
42 0%
37 0%
29 0.02%
26 0.13%
20 0.02%
28 0%
11 1.13%
27* 0.02%
22 4.01%
17 0.49%
19 1.00%
21 0.02%
1 0.02%
43 0.34%
31* 5.99%
24 3.29%
16 3.50%
9 0.02%
35 39.94%
6 0.13%
30 22.19%
14 0%
32 0.42%
12 0.42%
23 5.10%
10 0.17%
40 0.36%
38 1.89%
36 1.30%
39 0.70%
15 3.12%
7 0%
18* 0.25%
8 0.08%
25 0.96%
13* 0.30%
41* 0.08%
4 0.72%
5 0%
2 1.08%
3 0.34%

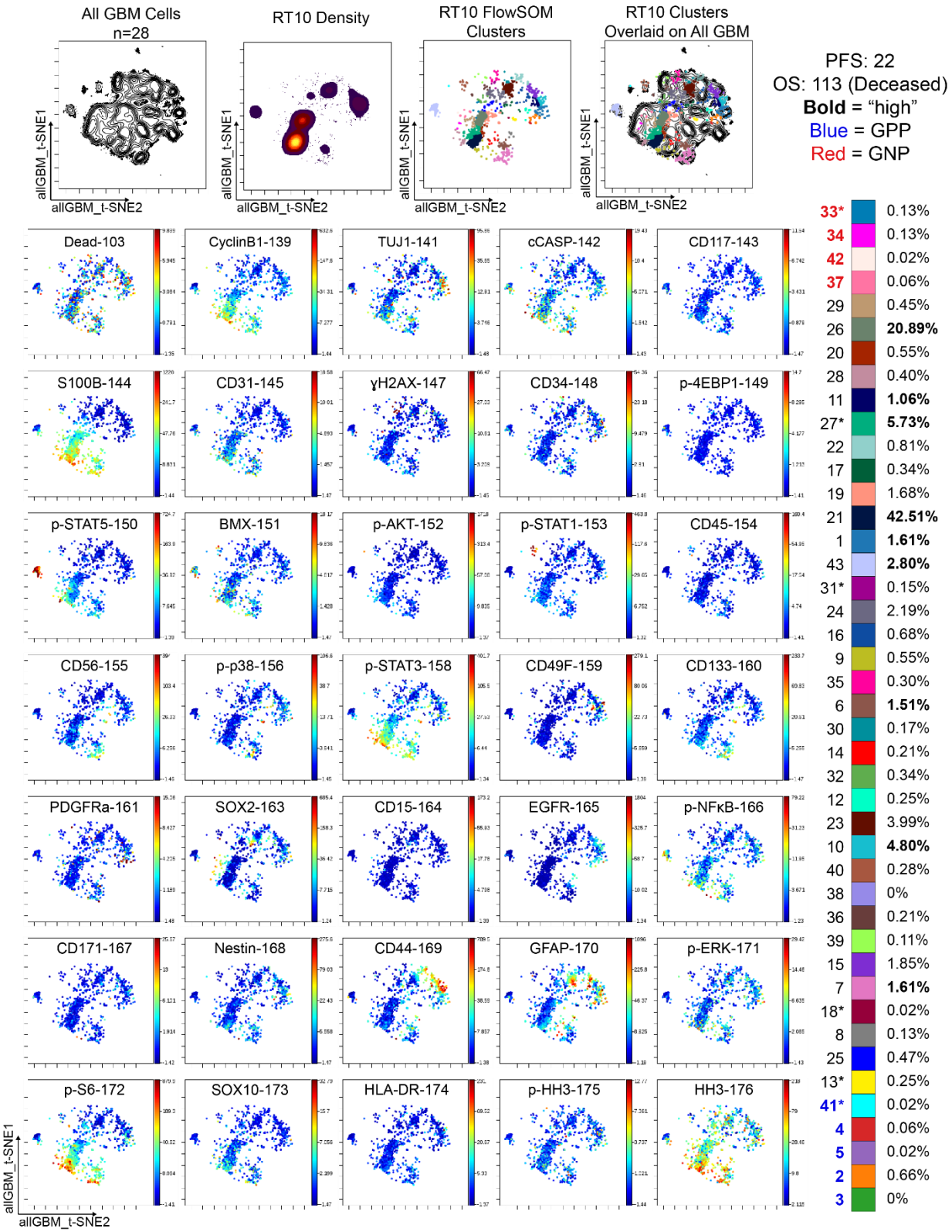
B.1q: RT01


All GBM Cells
n=28
RT01 Density
RT01 FlowSOM
Clusters
RT01 Clusters
Overlaid on All GBM
allGBM_t-SNE1
allGBM_t-SNE2
PFS: 162
OS: 198 (Deceased)
Bold = "high"
Blue = GPP
Red = GNP
Dead-103
CyclinB1-139
TUJ1-141
cCASP-142
CD117-143
S100B-144
CD31-145
γH2AX-147
CD34-148
p-4EBP1-149
p-STAT5-150
BMX-151
p-AKT-152
p-STAT1-153
CD45-154
CD56-155
p-p38-156
p-STAT3-158
CD49F-159
CD133-160
PDGFRa-161
SOX2-163
CD15-164
EGFR-165
p-NFκB-166
CD171-167
Nestin-168
CD44-169
GFAP-170
p-ERK-171
p-S6-172
SOX10-173
HLA-DR-174
p-HH3-175
HH3-176
33* 0.02%
34 0%
42 0%
37 0%
29 0.08%
26 0.21%
20 0%
28 0.21%
11 1.23%
27* 0.55%
22 0.08%
17 3.12%
19 0.04%
21 0.21%
1 63.52%
43 4.76%
31* 0.53%
24 4.01%
16 1.10%
9 0%
35 1.34%
6 0.08%
30 0.08%
14 0%
32 0.15%
12 1.74%
23 4.84%
10 1.04%
40 2.72%
38 0%
36 3.14%
39 3.72%
15 0.17%
7 0.11%
18* 0.02%
8 0%
25 0.04%
13* 0.04%
41* 0.08%
4 0.11%
5 0.02%
2 0.83%
3 0.02%

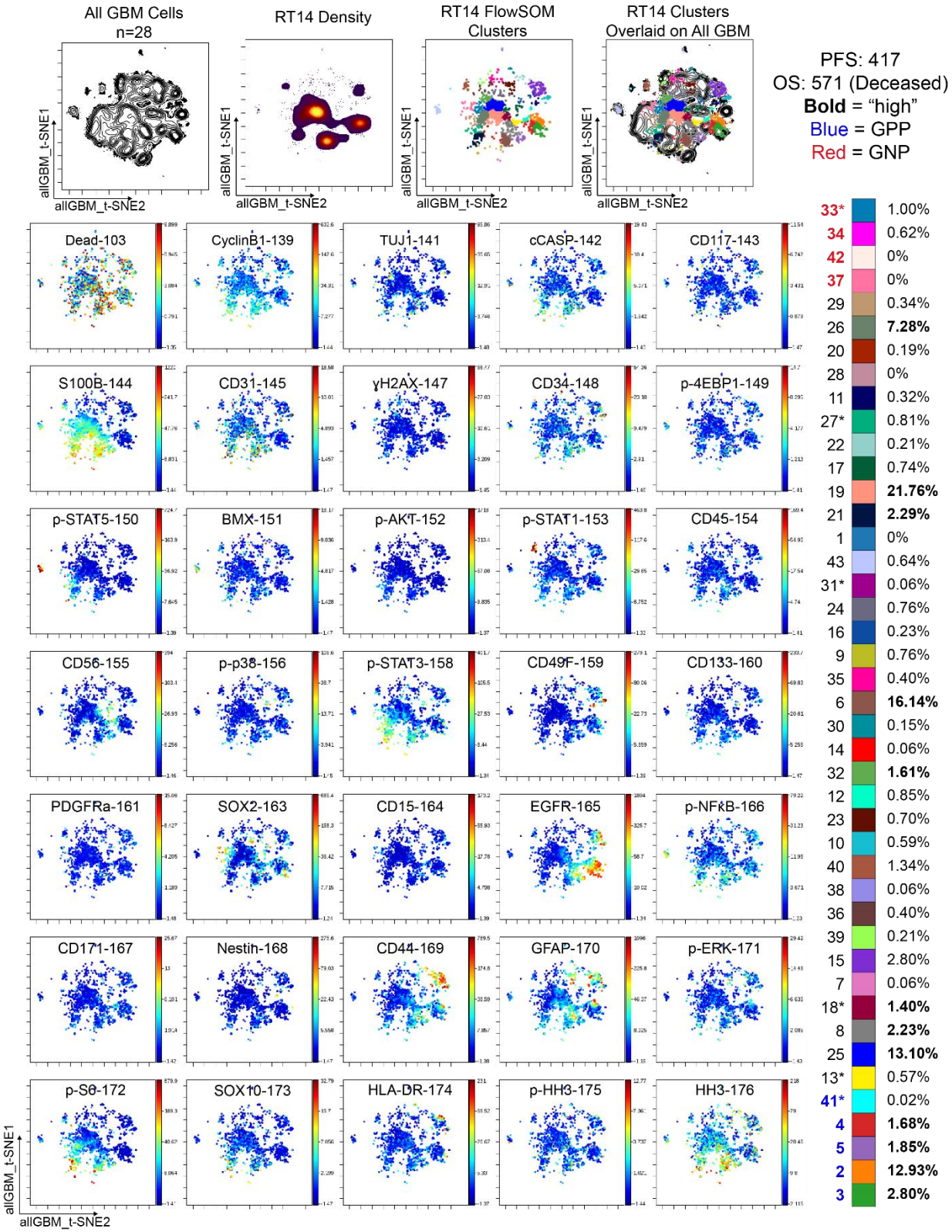
B.1r: RT07


All GBM Cells
n=28
RT07 Density
RT07 FlowSOM
Clusters
RT07 Clusters
Overlaid on All GBM
allGBM_t-SNE1
allGBM_t-SNE2
PFS: 49
OS: 57 (Deceased)
Bold = "high"
Blue = GPP
Red = GNP
Dead-103
CyclinB1-139
TUJ1-141
cCASP-142
CD117-143
S100B-144
CD31-145
γH2AX-147
CD34-148
p-4EBP1-149
p-STAT5-150
BMX-151
p-AKT-152
p-STAT1-153
CD45-154
CD56-155
p-p38-156
p-STAT3-158
CD49F-159
CD133-160
PDGFRa-161
SOX2-163
CD15-164
EGFR-165
p-NFκB-166
CD171-167
Nestin-168
CD44-169
GFAP-170
p-ERK-171
p-S6-172
SOX10-173
HLA-DR-174
p-HH3-175
HH3-176
33* 7.03%
34 13.10%
42 0.78%
37 24.20%
29 20.70%
26 1.25%
20 0.02%
28 17.64%
11 0.02%
27* 11.23%
22 0%
17 0.06%
19 0.23%
21 0.49%
1 0%
43 0.40%
31* 0.11%
24 0.72%
16 0%
9 0.06%
35 0%
6 0%
30 0%
14 0.17%
32 0.40%
12 0%
23 0.02%
10 0.02%
40 0.64%
38 0%
36 0.42%
39 0.04%
15 0.02%
7 0.02%
18* 0%
8 0.02%
25 0.11%
13* 0%
41* 0%
4 0.04%
5 0%
2 0.02%
3 0%

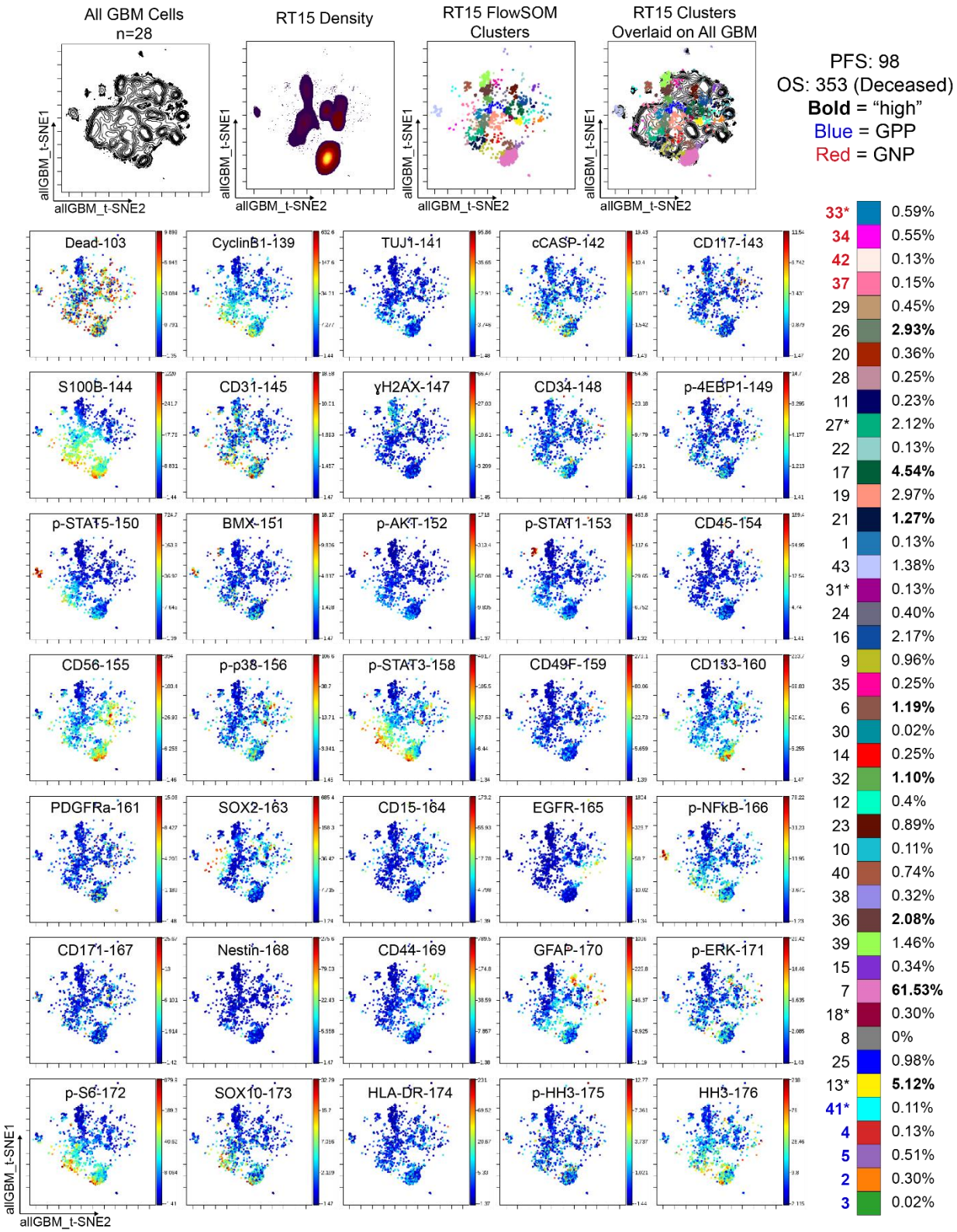
B.1s: RT10


All GBM Cells
n=28
RT10 Density
RT10 FlowSOM
Clusters
RT10 Clusters
Overlaid on All GBM
allGBM_t-SNE1
allGBM_t-SNE2
PFS: 22
OS: 113 (Deceased)
Bold = "high"
Blue = GPP
Red = GNP
Dead-103
CyclinB1-139
TUJ1-141
cCASP-142
CD117-143
S100B-144
CD31-145
γH2AX-147
CD34-148
p-4EBP1-149
p-STAT5-150
BMX-151
p-AKT-152
p-STAT1-153
CD45-154
CD56-155
p-p38-156
p-STAT3-158
CD49F-159
CD133-160
PDGFRa-161
SOX2-163
CD15-164
EGFR-165
p-NFκB-166
CD171-167
Nestin-168
CD44-169
GFAP-170
p-ERK-171
p-S6-172
SOX10-173
HLA-DR-174
p-HH3-175
HH3-176
33* 0.13%
34 0.13%
42 0.02%
37 0.06%
29 0.45%
26 20.89%
20 0.55%
28 0.40%
11 1.06%
27* 5.73%
22 0.81%
17 0.34%
19 1.68%
21 42.51%
1 1.61%
43 2.80%
31* 0.15%
24 2.19%
16 0.68%
9 0.55%
35 0.30%
6 1.51%
30 0.17%
14 0.21%
32 0.34%
12 0.25%
23 3.99%
10 4.80%
40 0.28%
38 0%
36 0.21%
39 0.11%
15 1.85%
7 1.61%
18* 0.02%
8 0.13%
25 0.47%
13* 0.25%
41* 0.02%
4 0.06%
5 0.02%
2 0.66%
3 0%

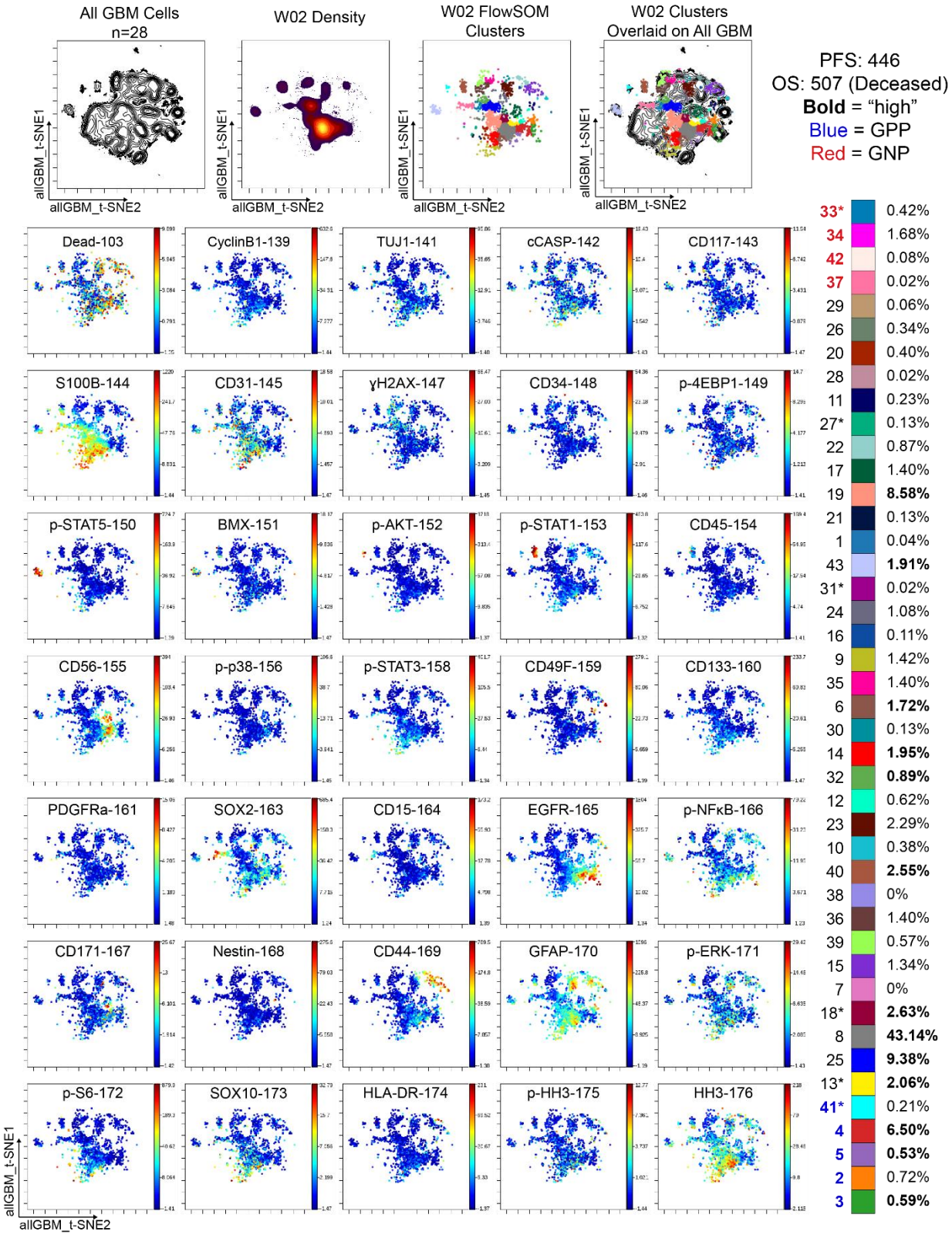
B.1t: RT14


All GBM Cells
n=28
RT14 Density
RT14 FlowSOM
Clusters
RT14 Clusters
Overlaid on All GBM
allGBM_t-SNE1
allGBM_t-SNE2
PFS: 417
OS: 571 (Deceased)
Bold = "high"
Blue = GPP
Red = GNP
Dead-103
CyclinB1-139
TUJ1-141
cCASP-142
CD117-143
S100B-144
CD31-145
γH2AX-147
CD34-148
p-4EBP1-149
p-STAT5-150
BMX-151
p-AKT-152
p-STAT1-153
CD45-154
CD56-155
p-p38-156
p-STAT3-158
CD49F-159
CD133-160
PDGFRa-161
SOX2-163
CD15-164
EGFR-165
p-NFkB-166
CD171-167
Nestin-168
CD44-169
GFAP-170
p-ERK-171
p-S6-172
SOX10-173
HLA-DR-174
p-HH3-175
HH3-176
33* 1.00%
34 0.62%
42 0%
37 0%
29 0.34%
26 7.28%
20 0.19%
28 0%
11 0.32%
27* 0.81%
22 0.21%
17 0.74%
19 21.76%
21 2.29%
1 0%
43 0.64%
31* 0.06%
24 0.76%
16 0.23%
9 0.76%
35 0.40%
6 16.14%
30 0.15%
14 0.06%
32 1.61%
12 0.85%
23 0.70%
10 0.59%
40 1.34%
38 0.06%
36 0.40%
39 0.21%
15 2.80%
7 0.06%
18* 1.40%
8 2.23%
25 13.10%
13* 0.57%
41* 0.02%
4 1.68%
5 1.85%
2 12.93%
3 2.80%

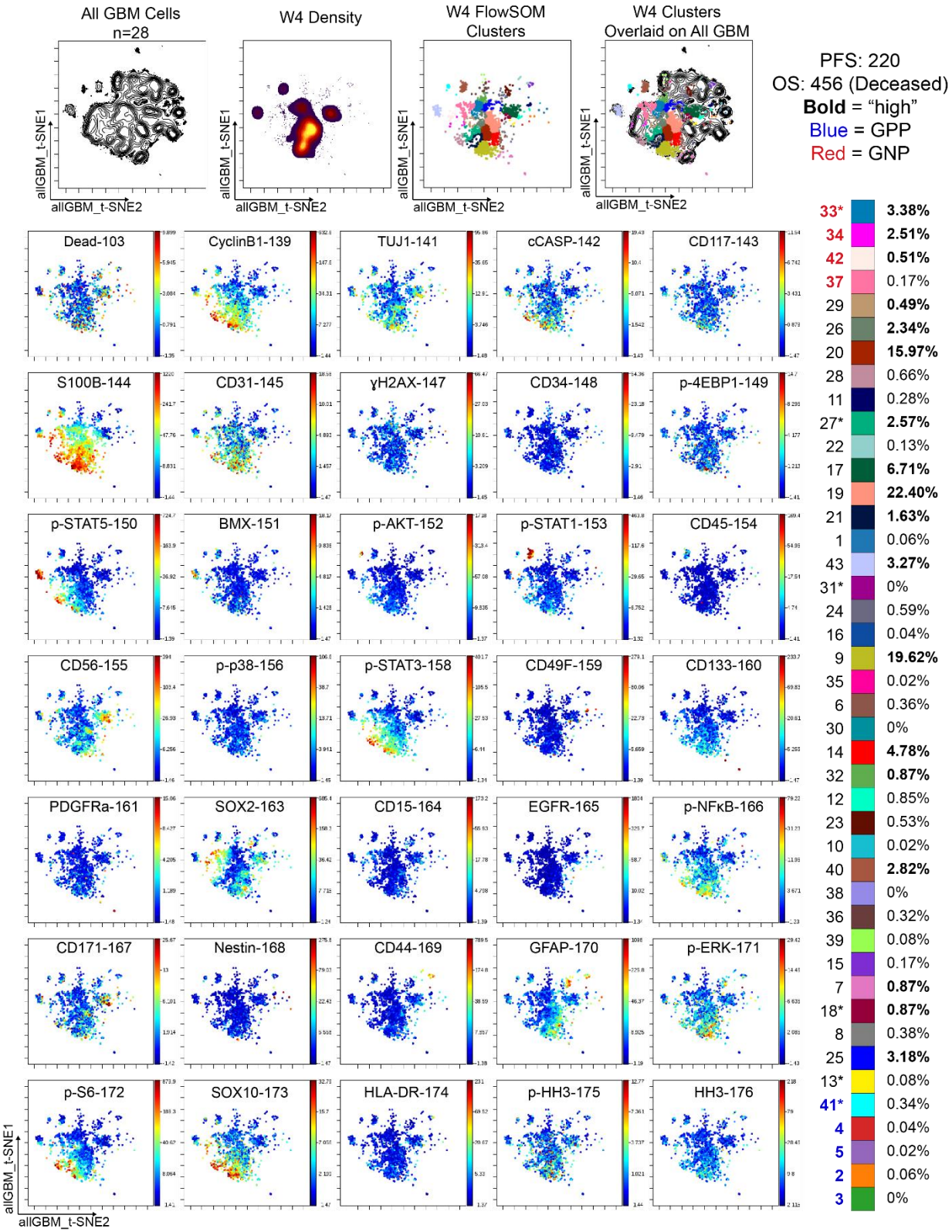
B.1u: RT15


All GBM Cells
n=28
RT15 Density
RT15 FlowSOM
Clusters
RT15 Clusters
Overlaid on All GBM
allGBM_t-SNE1
allGBM_t-SNE2
PFS: 98
OS: 353 (Deceased)
Bold = "high"
Blue = GPP
Red = GNP
Dead-103
CyclinB1-139
TUJ1-141
cCASP-142
CD117-143
S100B-144
CD31-145
γH2AX-147
CD34-148
p-4EBP1-149
p-STAT5-150
BMX-151
p-AKT-152
p-STAT1-153
CD45-154
CD56-155
p-p38-156
p-STAT3-158
CD49F-159
CD133-160
PDGFRa-161
SOX2-163
CD15-164
EGFR-165
p-NFkB-166
CD171-167
Nestin-168
CD44-169
GFAP-170
p-ERK-171
p-S6-172
SOX10-173
HLA-DR-174
p-HH3-175
HH3-176
33* 0.59%
34 0.55%
42 0.13%
37 0.15%
29 0.45%
26 2.93%
20 0.36%
28 0.25%
11 0.23%
27* 2.12%
22 0.13%
17 4.54%
19 2.97%
21 1.27%
1 0.13%
43 1.38%
31* 0.13%
24 0.40%
16 2.17%
9 0.96%
35 0.25%
6 1.19%
30 0.02%
14 0.25%
32 1.10%
12 0.4%
23 0.89%
10 0.11%
40 0.74%
38 0.32%
36 2.08%
39 1.46%
15 0.34%
7 61.53%
18* 0.30%
8 0%
25 0.98%
13* 5.12%
41* 0.11%
4 0.13%
5 0.51%
2 0.30%
3 0.02%

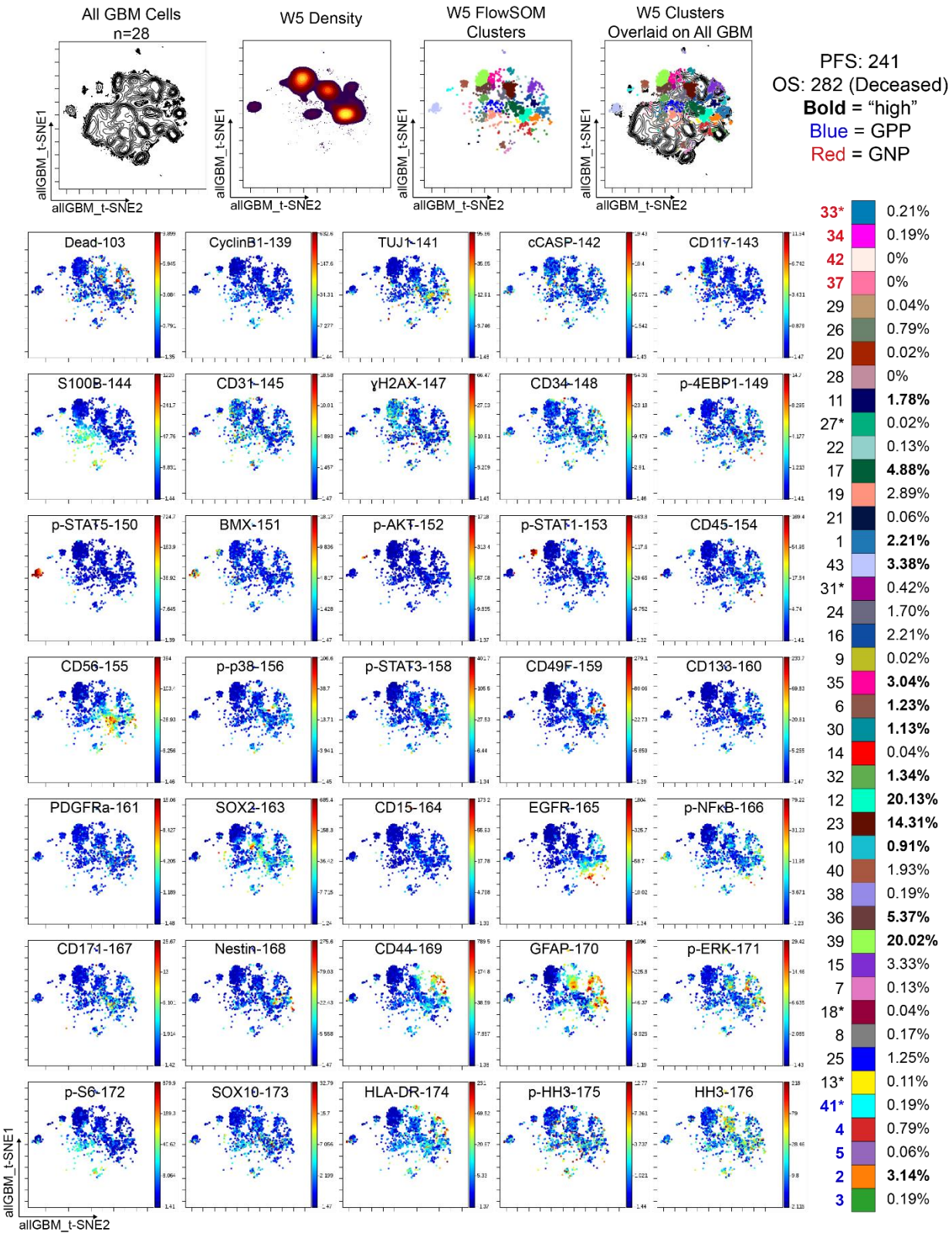
B.1v: W2


All GBM Cells
n=28
W02 Density
W02 FlowSOM
Clusters
W02 Clusters
Overlaid on All GBM
allGBM_t-SNE1
allGBM_t-SNE2
PFS: 446
OS: 507 (Deceased)
Bold = "high"
Blue = GPP
Red = GNP
Dead-103
CyclinB1-139
TUJ1-141
cCASP-142
CD117-143
S100B-144
CD31-145
γH2AX-147
CD34-148
p-4EBP1-149
p-STAT5-150
BMX-151
p-AKT-152
p-STAT1-153
CD45-154
CD56-155
p-p38-156
p-STAT3-158
CD49F-159
CD133-160
PDGFRa-161
SOX2-163
CD15-164
EGFR-165
p-NFκB-166
CD171-167
Nestin-168
CD44-169
GFAP-170
p-ERK-171
p-S6-172
SOX10-173
HLA-DR-174
p-HH3-175
HH3-176
33* 0.42%
34 1.68%
42 0.08%
37 0.02%
29 0.06%
26 0.34%
20 0.40%
28 0.02%
11 0.23%
27* 0.13%
22 0.87%
17 1.40%
19 8.58%
21 0.13%
1 0.04%
43 1.91%
31* 0.02%
24 1.08%
16 0.11%
9 1.42%
35 1.40%
6 1.72%
30 0.13%
14 1.95%
32 0.89%
12 0.62%
23 2.29%
10 0.38%
40 2.55%
38 0%
36 1.40%
39 0.57%
15 1.34%
7 0%
18* 2.63%
8 43.14%
25 9.38%
13* 2.06%
41* 0.21%
4 6.50%
5 0.53%
2 0.72%
3 0.59%

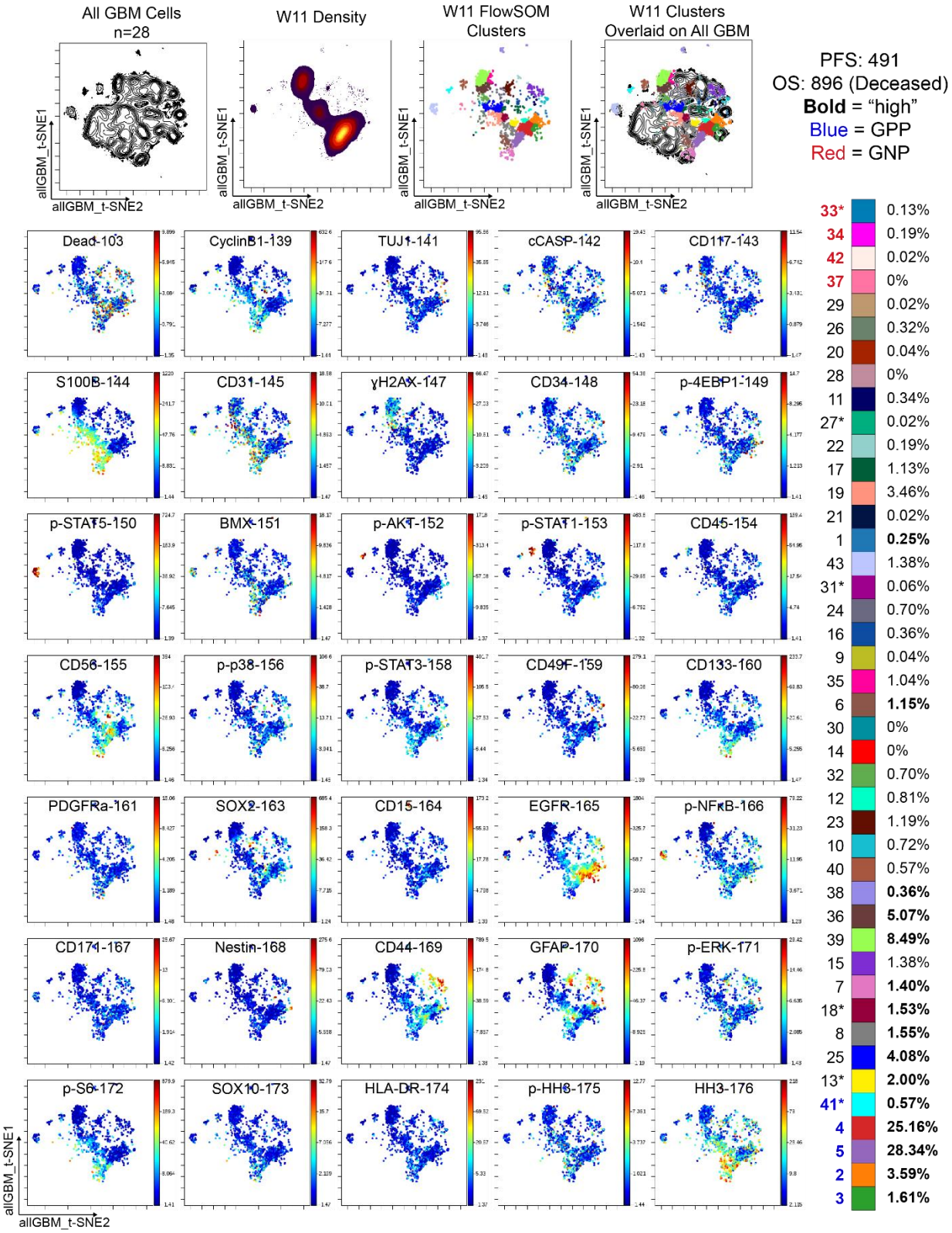
B.1w: W4


All GBM Cells
n=28
W4 Density
W4 FlowSOM
Clusters
W4 Clusters
Overlaid on All GBM
allGBM_t-SNE1
allGBM_t-SNE2
PFS: 220
OS: 456 (Deceased)
Bold = "high"
Blue = GPP
Red = GNP
Dead-103
CyclinB1-139
TUJ1-141
cCASP-142
CD117-143
S100B-144
CD31-145
γH2AX-147
CD34-148
p-4EBP1-149
p-STAT5-150
BMX-151
p-AKT-152
p-STAT1-153
CD45-154
CD56-155
p-p38-156
p-STAT3-158
CD49F-159
CD133-160
PDGFRa-161
SOX2-163
CD15-164
EGFR-165
p-NFκB-166
CD171-167
Nestin-168
CD44-169
GFAP-170
p-ERK-171
p-S6-172
SOX10-173
HLA-DR-174
p-HH3-175
HH3-176
33* 3.38%
34 2.51%
42 0.51%
37 0.17%
29 0.49%
26 2.34%
20 15.97%
28 0.66%
11 0.28%
27* 2.57%
22 0.13%
17 6.71%
19 22.40%
21 1.63%
1 0.06%
43 3.27%
31* 0%
24 0.59%
16 0.04%
9 19.62%
35 0.02%
6 0.36%
30 0%
14 4.78%
32 0.87%
12 0.85%
23 0.53%
10 0.02%
40 2.82%
38 0%
36 0.32%
39 0.08%
15 0.17%
7 0.87%
18* 0.87%
8 0.38%
25 3.18%
13* 0.08%
41* 0.34%
4 0.04%
5 0.02%
2 0.06%
3 0%

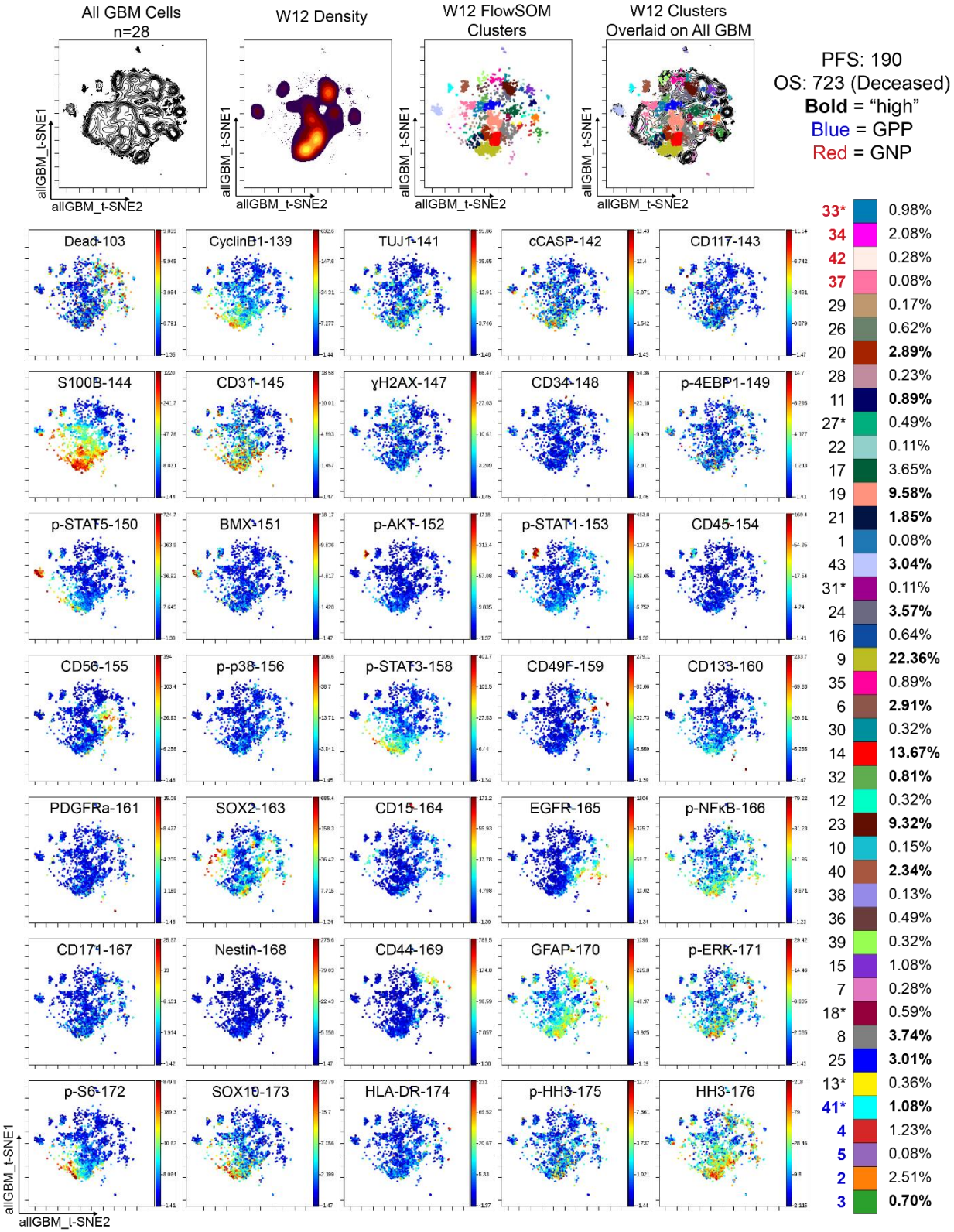
B.1x: W5


All GBM Cells
n=28
W5 Density
W5 FlowSOM
Clusters
W5 Clusters
Overlaid on All GBM
allGBM_t-SNE1
allGBM_t-SNE2
PFS: 241
OS: 282 (Deceased)
Bold = "high"
Blue = GPP
Red = GNP
Dead-103
CyclinB1-139
TUJ1-141
cCASP-142
CD117-143
S100B-144
CD31-145
γH2AX-147
CD34-148
p-4EBP1-149
p-STAT5-150
BMX-151
p-AKT-152
p-STAT1-153
CD45-154
CD56-155
p-p38-156
p-STAT3-158
CD49F-159
CD133-160
PDGFRa-161
SOX2-163
CD15-164
EGFR-165
p-NFκB-166
CD171-167
Nestin-168
CD44-169
GFAP-170
p-ERK-171
p-S6-172
SOX10-173
HLA-DR-174
p-HH3-175
HH3-176
33* 0.21%
34 0.19%
42 0%
37 0%
29 0.04%
26 0.79%
20 0.02%
28 0%
11 1.78%
27* 0.02%
22 0.13%
17 4.88%
19 2.89%
21 0.06%
1 2.21%
43 3.38%
31* 0.42%
24 1.70%
16 2.21%
9 0.02%
35 3.04%
6 1.23%
30 1.13%
14 0.04%
32 1.34%
12 20.13%
23 14.31%
10 0.91%
40 1.93%
38 0.19%
36 5.37%
39 20.02%
15 3.33%
7 0.13%
18* 0.04%
8 0.17%
25 1.25%
13* 0.11%
41* 0.19%
4 0.79%
5 0.06%
2 3.14%
3 0.19%

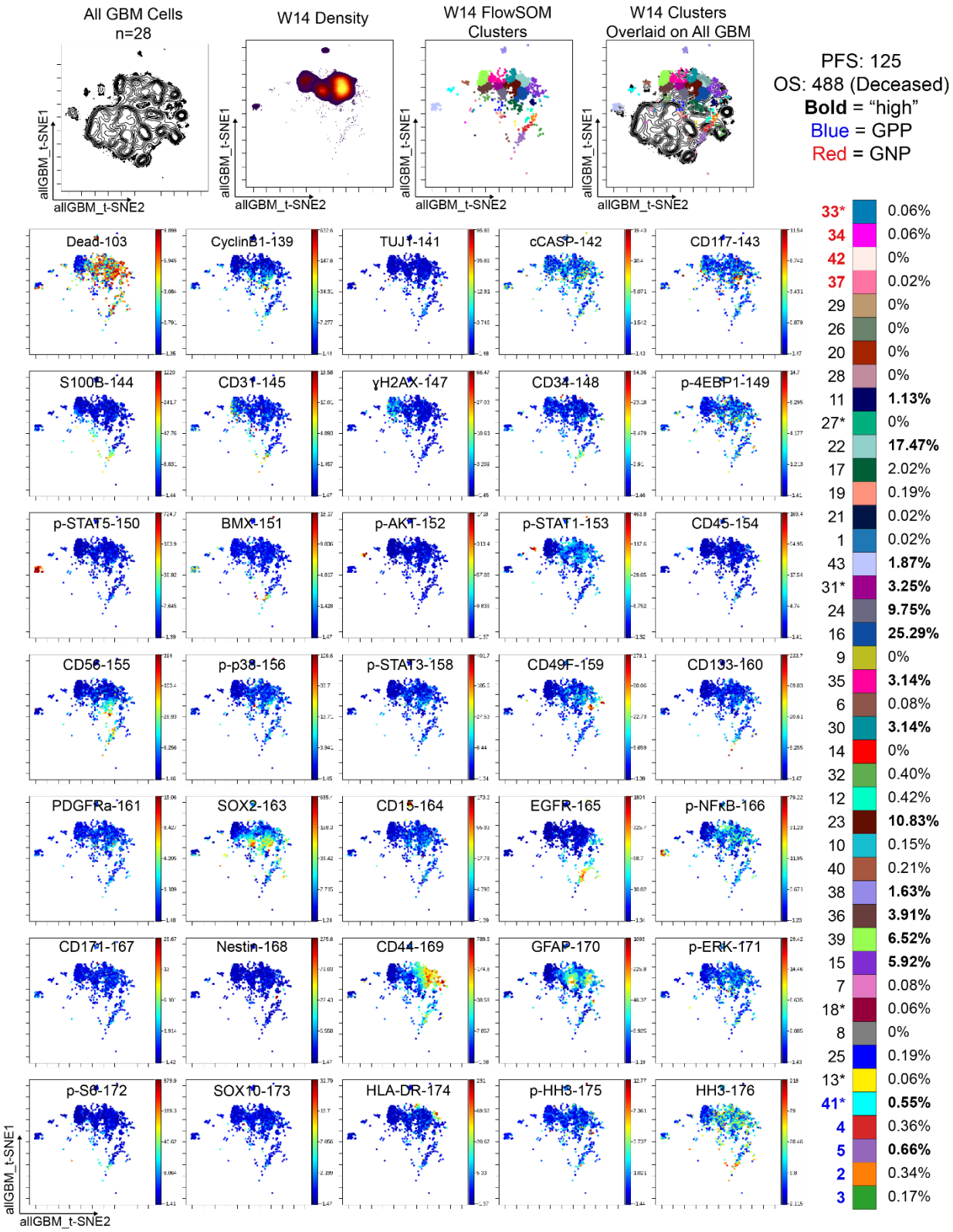
B.1y: W11


All GBM Cells
n=28
W11 Density
W11 FlowSOM
Clusters
W11 Clusters
Overlaid on All GBM
allGBM_t-SNE1
allGBM_t-SNE2
PFS: 491
OS: 896 (Deceased)
Bold = "high"
Blue = GPP
Red = GNP
Dead-103
Cyclin B1-139
TUJ1-141
cCASP-142
CD117-143
S100B-144
CD31-145
γH2AX-147
CD34-148
p-4EBP1-149
p-STAT5-150
BMX-151
p-AKT-152
p-STAT1-153
CD45-154
CD56-155
p-p38-156
p-STAT3-158
CD49F-159
CD133-160
PDGFRa-161
SOX2-163
CD15-164
EGFR-165
p-NFκB-166
CD171-167
Nestin-168
CD44-169
GFAP-170
p-ERK-171
p-S6-172
SOX10-173
HLA-DR-174
p-HH3-175
HH3-176
33* 0.13%
34 0.19%
42 0.02%
37 0%
29 0.02%
26 0.32%
20 0.04%
28 0%
11 0.34%
27* 0.02%
22 0.19%
17 1.13%
19 3.46%
21 0.02%
1 0.25%
43 1.38%
31* 0.06%
24 0.70%
16 0.36%
9 0.04%
35 1.04%
6 1.15%
30 0%
14 0%
32 0.70%
12 0.81%
23 1.19%
10 0.72%
40 0.57%
38 0.36%
36 5.07%
39 8.49%
15 1.38%
7 1.40%
18* 1.53%
8 1.55%
25 4.08%
13* 2.00%
41* 0.57%
4 25.16%
5 28.34%
2 3.59%
3 1.61%

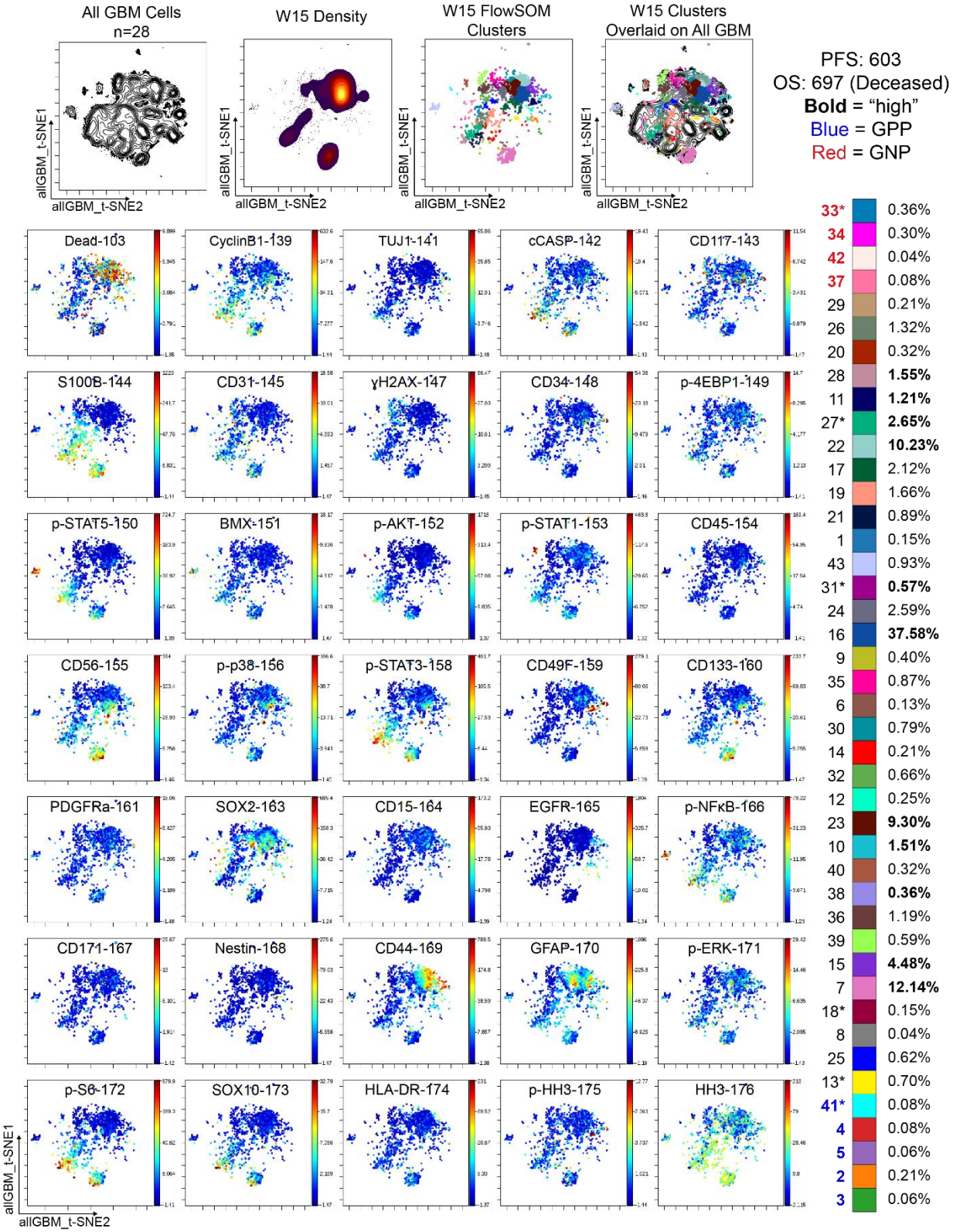
B.1z: W12


All GBM Cells
n=28
W12 Density
W12 FlowSOM
Clusters
W12 Clusters
Overlaid on All GBM
allGBM_t-SNE1
allGBM_t-SNE2
PFS: 190
OS: 723 (Deceased)
Bold = "high"
Blue = GPP
Red = GNP
Dead-103
CyclinB1-139
TUJ1-141
cCASP-142
CD117-143
S100B-144
CD31-145
γH2AX-147
CD34-148
p-4EBP1-149
p-STAT5-150
BMX-151
p-AKT-152
p-STAT1-153
CD45-154
CD56-155
p-p38-156
p-STAT3-158
CD49F-159
CD133-160
PDGFRa-161
SOX2-163
CD15-164
EGFR-165
p-NFκB-166
CD171-167
Nestin-168
CD44-169
GFAP-170
p-ERK-171
p-S6-172
SOX10-173
HLA-DR-174
p-HH3-175
HH3-176
33* 0.98%
34 2.08%
42 0.28%
37 0.08%
29 0.17%
26 0.62%
20 2.89%
28 0.23%
11 0.89%
27* 0.49%
22 0.11%
17 3.65%
19 9.58%
21 1.85%
1 0.08%
43 3.04%
31* 0.11%
24 3.57%
16 0.64%
9 22.36%
35 0.89%
6 2.91%
30 0.32%
14 13.67%
32 0.81%
12 0.32%
23 9.32%
10 0.15%
40 2.34%
38 0.13%
36 0.49%
39 0.32%
15 1.08%
7 0.28%
18* 0.59%
8 3.74%
25 3.01%
13* 0.36%
41* 1.08%
4 1.23%
5 0.08%
2 2.51%
3 0.70%

B.1aa: W14


All GBM Cells
n=28
W14 Density
W14 FlowSOM
Clusters
W14 Clusters
Overlaid on All GBM
allGBM_t-SNE1
allGBM_t-SNE2
PFS: 125
OS: 488 (Deceased)
Bold = "high"
Blue = GPP
Red = GNP
Dead-103
Cyclin B1-139
TUJ1-141
cCASP-142
CD117-143
S100B-144
CD31-145
γH2AX-147
CD34-148
p-4EBP1-149
p-STAT5-150
BMX-151
p-AKT-152
p-STAT1-153
CD45-154
CD56-155
p-p38-156
p-STAT3-158
CD49F-159
CD133-160
PDGFRa-161
SOX2-163
CD15-164
EGFR-165
p-NFkB-166
CD171-167
Nestin-168
CD44-169
GFAP-170
p-ERK-171
p-S6-172
SOX10-173
HLA-DR-174
p-HH3-175
HH3-176
33* 0.06%
34 0.06%
42 0%
37 0.02%
29 0%
26 0%
20 0%
28 0%
11 1.13%
27* 0%
22 17.47%
17 2.02%
19 0.19%
21 0.02%
1 0.02%
43 1.87%
31* 3.25%
24 9.75%
16 25.29%
9 0%
35 3.14%
6 0.08%
30 3.14%
14 0%
32 0.40%
12 0.42%
23 10.83%
10 0.15%
40 0.21%
38 1.63%
36 3.91%
39 6.52%
15 5.92%
7 0.08%
18* 0.06%
8 0%
25 0.19%
13* 0.06%
41* 0.55%
4 0.36%
5 0.66%
2 0.34%
3 0.17%

B.1bb: W15


All GBM Cells
n=28
W15 Density
W15 FlowSOM
Clusters
W15 Clusters
Overlaid on All GBM
allGBM_t-SNE1
allGBM_t-SNE2
PFS: 603
OS: 697 (Deceased)
Bold = "high"
Blue = GPP
Red = GNP
Dead-103
CyclinB1-139
TUJ1-141
cCASP-142
CD117-143
S100B-144
CD31-145
γH2AX-147
CD34-148
p-4EBP1-149
p-STAT5-150
BMX-151
p-AKT-152
p-STAT1-153
CD45-154
CD56-155
p-p38-156
p-STAT3-158
CD49F-159
CD133-160
PDGFRa-161
SOX2-163
CD15-164
EGFR-165
p-NFκB-166
CD171-167
Nestin-168
CD44-169
GFAP-170
p-ERK-171
p-S6-172
SOX10-173
HLA-DR-174
p-HH3-175
HH3-176
33* 0.36%
34 0.30%
42 0.04%
37 0.08%
29 0.21%
26 1.32%
20 0.32%
28 1.55%
11 1.21%
27* 2.65%
22 10.23%
17 2.12%
19 1.66%
21 0.89%
1 0.15%
43 0.93%
31* 0.57%
24 2.59%
16 37.58%
9 0.40%
35 0.87%
6 0.13%
30 0.79%
14 0.21%
32 0.66%
12 0.25%
23 9.30%
10 1.51%
40 0.32%
38 0.36%
36 1.19%
39 0.59%
15 4.48%
7 12.14%
18* 0.15%
8 0.04%
25 0.62%
13* 0.70%
41* 0.08%
4 0.08%
5 0.06%
2 0.21%
3 0.06%

Discussion

Plots of cell density on the t-SNE axes revealed phenotypically distinct subpopulations of glioblastoma cells within a single patient's tumor. Intra-tumoral subsets were distinguished by differences in expression of core neural identity proteins and by aberrant co-expression of neural lineage and stem cell proteins. For example, cells expressing SOX2 and CD133, stem-cell markers, and S100B, a protein found in mature astrocytes, were observed in multiple patients [180, 181]. These plots provide a visual summary of the data generated by mass cytometry analysis in 28 patients. The potential for discovery within this cohort, but especially in expanded cohorts or other disease settings is extensive. By synthesizing information about cell phenotypes, and their presence or absence in individual patients, strategies for eliminating said cells can be refined.

REFERENCES

1. Jain, K.K., *A Critical Overview of Targeted Therapies for Glioblastoma.* Front Oncol, 2018. **8**: p. 419.

2. Stupp, R., et al., *Radiotherapy plus concomitant and adjuvant temozolomide for glioblastoma.* N Engl J Med, 2005. **352**(10): p. 987-96.

3. Ostrom, Q.T., et al., *CBTRUS Statistical Report: Primary brain and other central nervous system tumors diagnosed in the United States in 2010-2014.* Neuro Oncol, 2017. **19**(suppl_5): p. v1-v88.

4. Louis, D.N., et al., *The 2016 World Health Organization Classification of Tumors of the Central Nervous System: a summary.* Acta Neuropathol, 2016. **131**(6): p. 803-20.

5. Louis, D.N., et al., *The 2007 WHO classification of tumours of the central nervous system.* Acta Neuropathol, 2007. **114**(2): p. 97-109.

6. Picca, A., G. Berzero, and M. Sanson, *Current therapeutic approaches to diffuse grade II and III gliomas.* Ther Adv Neurol Disord, 2018. **11**: p. 1756285617752039.

7. Sizoo, E.M., et al., *Symptoms and problems in the end-of-life phase of high-grade glioma patients.* Neuro Oncol, 2010. **12**(11): p. 1162-6.

8. Drumm, M.R., et al., *Extensive brainstem infiltration, not mass effect, is a common feature of end-stage cerebral glioblastomas.* Neuro Oncol, 2020. **22**(4): p. 470-479.

9. Stupp, R., et al., *Current and future developments in the use of temozolomide for the treatment of brain tumours.* Lancet Oncol, 2001. **2**(9): p. 552-60.

10. Hegi, M.E., et al., *MGMT gene silencing and benefit from temozolomide in glioblastoma.* N Engl J Med, 2005. **352**(10): p. 997-1003.

11. Bao, S., et al., *Glioma stem cells promote radioresistance by preferential activation of the DNA damage response.* Nature, 2006. **444**(7120): p. 756-60.

12. Meyer, M., et al., *Single cell-derived clonal analysis of human glioblastoma links functional and genomic heterogeneity.* Proc Natl Acad Sci U S A, 2015. **112**(3): p. 851-6.

13. Barker, F.G., 2nd, et al., *Survival and functional status after resection of recurrent glioblastoma multiforme.* Neurosurgery, 1998. **42**(4): p. 709-20; discussion 720-3.

14. Parsons, D.W., et al., *An integrated genomic analysis of human glioblastoma multiforme.* Science, 2008. **321**(5897): p. 1807-12.

15. Yan, H., et al., *IDH1 and IDH2 mutations in gliomas.* N Engl J Med, 2009. **360**(8): p. 765-73.

16. Wick, W., et al., *MGMT testing--the challenges for biomarker-based glioma treatment.* Nat Rev Neurol, 2014. **10**(7): p. 372-85.

17. Leu, S., et al., *IDH/MGMT-driven molecular classification of low-grade glioma is a strong predictor for long-term survival.* Neuro Oncol, 2013. **15**(4): p. 469-79.

18. Noushmehr, H., et al., *Identification of a CpG island methylator phenotype that defines a distinct subgroup of glioma.* Cancer Cell, 2010. **17**(5): p. 510-22.

19. Mardis, E.R., et al., *Recurring mutations found by sequencing an acute myeloid leukemia genome.* N Engl J Med, 2009. **361**(11): p. 1058-66.

20. Dang, L., et al., *Cancer-associated IDH1 mutations produce 2-hydroxyglutarate.* Nature, 2009. **462**(7274): p. 739-44.

21. Xu, X., et al., *Structures of human cytosolic NADP-dependent isocitrate dehydrogenase reveal a novel self-regulatory mechanism of activity.* J Biol Chem, 2004. **279**(32): p. 33946-57.

22. Hartmann, C., et al., *Type and frequency of IDH1 and IDH2 mutations are related to astrocytic and oligodendroglial differentiation and age: a study of 1,010 diffuse gliomas.* Acta Neuropathol, 2009. **118**(4): p. 469-74.

23. Hartmann, C., et al., *Patients with IDH1 wild type anaplastic astrocytomas exhibit worse prognosis than IDH1-mutated glioblastomas, and IDH1 mutation status accounts for the unfavorable prognostic effect of higher age: implications for classification of gliomas.* Acta Neuropathol, 2010. **120**(6): p. 707-18.

24. Geisbrecht, B.V. and S.J. Gould, *The human PICD gene encodes a cytoplasmic and peroxisomal NADP(+)-dependent isocitrate dehydrogenase.* J Biol Chem, 1999. **274**(43): p. 30527-33.

25. L, M.G., et al., *Oncogenic Activities of IDH1/2 Mutations: From Epigenetics to Cellular Signaling.* Trends Cell Biol, 2017. **27**(10): p. 738-752.

26. Cairns, R.A., I.S. Harris, and T.W. Mak, *Regulation of cancer cell metabolism.* Nat Rev Cancer, 2011. **11**(2): p. 85-95.

27. Cancer Genome Atlas Research, N., *Comprehensive genomic characterization defines human glioblastoma genes and core pathways.* Nature, 2008. **455**(7216): p. 1061-8.

28. Verhaak, R.G., et al., *Integrated genomic analysis identifies clinically relevant subtypes of glioblastoma characterized by abnormalities in PDGFRA, IDH1, EGFR, and NF1.* Cancer Cell, 2010. **17**(1): p. 98-110.

29. Phillips, H.S., et al., *Molecular subclasses of high-grade glioma predict prognosis, delineate a pattern of disease progression, and resemble stages in neurogenesis.* Cancer Cell, 2006. **9**(3): p. 157-73.

30. Sugawa, N., et al., *Identical splicing of aberrant epidermal growth factor receptor transcripts from amplified rearranged genes in human glioblastomas.* Proc Natl Acad Sci U S A, 1990. **87**(21): p. 8602-6.

31. Brennan, C., et al., *Glioblastoma subclasses can be defined by activity among signal transduction pathways and associated genomic alterations.* PLoS One, 2009. **4**(11): p. e7752.

32. Wang, Q., et al., *Tumor Evolution of Glioma-Intrinsic Gene Expression Subtypes Associates with Immunological Changes in the Microenvironment.* Cancer Cell, 2017. **32**(1): p. 42-56 e6.

33. Patel, A.P., et al., *Single-cell RNA-seq highlights intratumoral heterogeneity in primary glioblastoma.* Science, 2014. **344**(6190): p. 1396-401.

34. Johnson, H. and F.M. White, *Quantitative analysis of signaling networks across differentially embedded tumors highlights interpatient heterogeneity in human glioblastoma.* J Proteome Res, 2014. **13**(11): p. 4581-93.

35. Snuderl, M., et al., *Mosaic amplification of multiple receptor tyrosine kinase genes in glioblastoma.* Cancer Cell, 2011. **20**(6): p. 810-7.

36. Szerlip, N.J., et al., *Intratumoral heterogeneity of receptor tyrosine kinases EGFR and PDGFRA amplification in glioblastoma defines subpopulations with distinct growth factor response.* Proc Natl Acad Sci U S A, 2012. **109**(8): p. 3041-6.

37. Bhaduri, A., et al., *Outer Radial Glia-like Cancer Stem Cells Contribute to Heterogeneity of Glioblastoma.* Cell Stem Cell, 2020. **26**(1): p. 48-63 e6.

38. Muller, S., et al., *Single-cell sequencing maps gene expression to mutational phylogenies in PDGF- and EGF-driven gliomas.* Mol Syst Biol, 2016. **12**(11): p. 889.

39. Singh, S.K., et al., *Identification of human brain tumour initiating cells.* Nature, 2004. **432**(7015): p. 396-401.

40. Singh, S.K., et al., *Identification of a cancer stem cell in human brain tumors.* Cancer Res, 2003. **63**(18): p. 5821-8.

41. Galli, R., et al., *Isolation and characterization of tumorigenic, stem-like neural precursors from human glioblastoma.* Cancer Res, 2004. **64**(19): p. 7011-21.

42. Reya, T., et al., *Stem cells, cancer, and cancer stem cells.* Nature, 2001. **414**(6859): p. 105-11.

43. Lapidot, T., et al., *A cell initiating human acute myeloid leukaemia after transplantation into SCID mice.* Nature, 1994. **367**(6464): p. 645-8.

44. Son, M.J., et al., *SSEA-1 is an enrichment marker for tumor-initiating cells in human glioblastoma.* Cell Stem Cell, 2009. **4**(5): p. 440-52.

45. Lathia, J.D., et al., *Integrin alpha 6 regulates glioblastoma stem cells.* Cell Stem Cell, 2010. **6**(5): p. 421-32.

46. Anido, J., et al., *TGF-beta Receptor Inhibitors Target the CD44(high)/Id1(high) Glioma-Initiating Cell Population in Human Glioblastoma.* Cancer Cell, 2010. **18**(6): p. 655-68.

47. Neftel, C., et al., *An Integrative Model of Cellular States, Plasticity, and Genetics for Glioblastoma.* Cell, 2019. **178**(4): p. 835-849 e21.

48. Wang, L., et al., *The Phenotypes of Proliferating Glioblastoma Cells Reside on a Single Axis of Variation.* Cancer Discov, 2019. **9**(12): p. 1708-1719.

49. Suva, M.L., et al., *Reconstructing and reprogramming the tumor-propagating potential of glioblastoma stem-like cells.* Cell, 2014. **157**(3): p. 580-94.

50. Silva-Vargas, V., E.E. Crouch, and F. Doetsch, *Adult neural stem cells and their niche: a dynamic duo during homeostasis, regeneration, and aging.* Curr Opin Neurobiol, 2013. **23**(6): p. 935-42.

51. Fuentealba, L.C., K. Obernier, and A. Alvarez-Buylla, *Adult neural stem cells bridge their niche.* Cell Stem Cell, 2012. **10**(6): p. 698-708.

52. Bonaguidi, M.A., et al., *A unifying hypothesis on mammalian neural stem cell properties in the adult hippocampus.* Curr Opin Neurobiol, 2012. **22**(5): p. 754-61.

53. Sanai, N., et al., *Unique astrocyte ribbon in adult human brain contains neural stem cells but lacks chain migration.* Nature, 2004. **427**(6976): p. 740-4.

54. Ernst, A., et al., *Neurogenesis in the striatum of the adult human brain.* Cell, 2014. **156**(5): p. 1072-83.

55. Kernie, S.G. and J.M. Parent, *Forebrain neurogenesis after focal Ischemic and traumatic brain injury.* Neurobiol Dis, 2010. **37**(2): p. 267-74.

56. Sanai, N., A. Alvarez-Buylla, and M.S. Berger, *Neural stem cells and the origin of gliomas.* N Engl J Med, 2005. **353**(8): p. 811-22.

57. Zong, H., L.F. Parada, and S.J. Baker, *Cell of origin for malignant gliomas and its implication in therapeutic development.* Cold Spring Harb Perspect Biol, 2015. **7**(5).

58. Chen, J., R.M. McKay, and L.F. Parada, *Malignant glioma: lessons from genomics, mouse models, and stem cells.* Cell, 2012. **149**(1): p. 36-47.

59. Wang, R., et al., *Adult Human Glioblastomas Harbor Radial Glia-like Cells.* Stem Cell Reports, 2020. **14**(2): p. 338-350.

60. Mistry, A.M., et al., *Decreased survival in glioblastomas is specific to contact with the ventricular-subventricular zone, not subgranular zone or corpus callosum.* J Neurooncol, 2017. **132**(2): p. 341-349.

61. Mistry, A.M., et al., *Influence of glioblastoma contact with the lateral ventricle on survival: a meta-analysis.* J Neurooncol, 2017. **131**(1): p. 125-133.

62. Berendsen, S., et al., *Adverse prognosis of glioblastoma contacting the subventricular zone: Biological correlates.* PLoS One, 2019. **14**(10): p. e0222717.

63. Iacoangeli, M., et al., *Endoscopy-verified occult subependymal dissemination of glioblastoma and brain metastasis undetected by MRI: prognostic significance.* Onco Targets Ther, 2012. **5**: p. 449-56.

64. Willard, N. and B.K. Kleinschmidt-DeMasters, *Massive dissemination of adult glioblastomas.* Clin Neuropathol, 2015. **34**(6): p. 330-42.

65. Tamura, M., C. Ohye, and Y. Nakazato, *Pathological anatomy of autopsy brain with malignant glioma.* Neurol Med Chir (Tokyo), 1993. **33**(2): p. 77-80.

66. Lee, J.H., et al., *Human glioblastoma arises from subventricular zone cells with low-level driver mutations.* Nature, 2018. **560**(7717): p. 243-247.

67. Mistry, A.M., et al., *Ventricular-Subventricular Zone Contact by Glioblastoma is Not Associated with Molecular Signatures in Bulk Tumor Data.* Sci Rep, 2019. **9**(1): p. 1842.

68. Kappadakunnel, M., et al., *Stem cell associated gene expression in glioblastoma multiforme: relationship to survival and the subventricular zone.* J Neurooncol, 2010. **96**(3): p. 359-67.

69. Piccirillo, S.G., et al., *Contributions to drug resistance in glioblastoma derived from malignant cells in the sub-ependymal zone.* Cancer Res, 2015. **75**(1): p. 194-202.

70. Sinnaeve, J., B.C. Mobley, and R.A. Ihrie, *Space Invaders: Brain Tumor Exploitation of the Stem Cell Niche.* Am J Pathol, 2018. **188**(1): p. 29-38.

71. Mirzadeh, Z., et al., *Neural stem cells confer unique pinwheel architecture to the ventricular surface in neurogenic regions of the adult brain.* Cell Stem Cell, 2008. **3**(3): p. 265-78.

72. Del Bigio, M.R., *The ependyma: a protective barrier between brain and cerebrospinal fluid.* Glia, 1995. **14**(1): p. 1-13.

73. Sawamoto, K., et al., *New neurons follow the flow of cerebrospinal fluid in the adult brain.* Science, 2006. **311**(5761): p. 629-32.

74. Tong, C.K., et al., *Axonal control of the adult neural stem cell niche.* Cell Stem Cell, 2014. **14**(4): p. 500-11.

75. Hoglinger, G.U., et al., *Origin of the dopaminergic innervation of adult neurogenic areas.* J Comp Neurol, 2014. **522**(10): p. 2336-48.

76. Ribeiro Xavier, A.L., et al., *A Distinct Population of Microglia Supports Adult Neurogenesis in the Subventricular Zone.* J Neurosci, 2015. **35**(34): p. 11848-61.

77. Dennis, C.V., et al., *Human adult neurogenesis across the ages: An immunohistochemical study.* Neuropathol Appl Neurobiol, 2016. **42**(7): p. 621-638.

78. Sanai, N., et al., *Corridors of migrating neurons in the human brain and their decline during infancy.* Nature, 2011. **478**(7369): p. 382-6.

79. Silva-Vargas, V., et al., *Age-Dependent Niche Signals from the Choroid Plexus Regulate Adult Neural Stem Cells.* Cell Stem Cell, 2016. **19**(5): p. 643-652.

80. Soroceanu, L., et al., *Identification of IGF2 signaling through phosphoinositide-3-kinase regulatory subunit 3 as a growth-promoting axis in glioblastoma.* Proc Natl Acad Sci U S A, 2007. **104**(9): p. 3466-71.

81. Yin, J., et al., *Pigment Epithelium-Derived Factor (PEDF) Expression Induced by EGFRvIII Promotes Self-renewal and Tumor Progression of Glioma Stem Cells.* PLoS Biol, 2015. **13**(5): p. e1002152.

82. Lorente, M., et al., *Amphiregulin is a factor for resistance of glioma cells to cannabinoid-induced apoptosis.* Glia, 2009. **57**(13): p. 1374-85.

83. Lim, D.A., et al., *Noggin antagonizes BMP signaling to create a niche for adult neurogenesis.* Neuron, 2000. **28**(3): p. 713-26.

84. Piccirillo, S.G., et al., *Bone morphogenetic proteins inhibit the tumorigenic potential of human brain tumour-initiating cells.* Nature, 2006. **444**(7120): p. 761-5.

85. Kokovay, E., et al., *Adult SVZ lineage cells home to and leave the vascular niche via differential responses to SDF1/CXCR4 signaling.* Cell Stem Cell, 2010. **7**(2): p. 163-73.

86. Goffart, N., et al., *Adult mouse subventricular zones stimulate glioblastoma stem cells specific invasion through CXCL12/CXCR4 signaling.* Neuro Oncol, 2015. **17**(1): p. 81-94.

87. Venkatesh, H.S., et al., *Neuronal Activity Promotes Glioma Growth through Neuroligin-3 Secretion.* Cell, 2015. **161**(4): p. 803-16.

88. Venkatesh, H.S., et al., *Electrical and synaptic integration of glioma into neural circuits.* Nature, 2019. **573**(7775): p. 539-545.

89. Venkataramani, V., et al., *Glutamatergic synaptic input to glioma cells drives brain tumour progression.* Nature, 2019. **573**(7775): p. 532-538.

90. Lin, J.H., et al., *Connexin 43 enhances the adhesivity and mediates the invasion of malignant glioma cells.* J Neurosci, 2002. **22**(11): p. 4302-11.

91. Tavazoie, M., et al., *A specialized vascular niche for adult neural stem cells.* Cell Stem Cell, 2008. **3**(3): p. 279-88.

92. Watkins, S., et al., *Disruption of astrocyte-vascular coupling and the blood-brain barrier by invading glioma cells.* Nat Commun, 2014. **5**: p. 4196.

93. Andreu-Agullo, C., et al., *Vascular niche factor PEDF modulates Notch-dependent stemness in the adult subependymal zone.* Nat Neurosci, 2009. **12**(12): p. 1514-23.

94. Ramirez-Castillejo, C., et al., *Pigment epithelium-derived factor is a niche signal for neural stem cell renewal.* Nat Neurosci, 2006. **9**(3): p. 331-9.

95. Leventhal, C., et al., *Endothelial trophic support of neuronal production and recruitment from the adult mammalian subependyma.* Mol Cell Neurosci, 1999. **13**(6): p. 450-64.

96. Jin, K., et al., *Vascular endothelial growth factor (VEGF) stimulates neurogenesis in vitro and in vivo.* Proc Natl Acad Sci U S A, 2002. **99**(18): p. 11946-50.

97. Crouch, E.E., et al., *Regional and stage-specific effects of prospectively purified vascular cells on the adult V-SVZ neural stem cell lineage.* J Neurosci, 2015. **35**(11): p. 4528-39.

98. Leelatian, N., et al., *Single cell analysis of human tissues and solid tumors with mass cytometry.* Cytometry B Clin Cytom, 2017. **92**(1): p. 68-78.

99. Wei, W., et al., *Single-Cell Phosphoproteomics Resolves Adaptive Signaling Dynamics and Informs Targeted Combination Therapy in Glioblastoma.* Cancer Cell, 2016. **29**(4): p. 563-73.

100. Venteicher, A.S., et al., *Decoupling genetics, lineages, and microenvironment in IDH-mutant gliomas by single-cell RNA-seq.* Science, 2017. **355**(6332).

101. Yuan, J., et al., *Single-cell transcriptome analysis of lineage diversity in high-grade glioma.* Genome Med, 2018. **10**(1): p. 57.

102. Tirosh, I., et al., *Single-cell RNA-seq supports a developmental hierarchy in human oligodendroglioma.* Nature, 2016. **539**(7628): p. 309-313.

103. Bandura, D.R., et al., *Mass cytometry: technique for real time single cell multitarget immunoassay based on inductively coupled plasma time-of-flight mass spectrometry.* Anal Chem, 2009. **81**(16): p. 6813-22.

104. Ornatsky, O., et al., *Highly multiparametric analysis by mass cytometry.* J Immunol Methods, 2010. **361**(1-2): p. 1-20.

105. Spitzer, M.H. and G.P. Nolan, *Mass Cytometry: Single Cells, Many Features.* Cell, 2016. **165**(4): p. 780-91.

106. Frei, A.P., et al., *Highly multiplexed simultaneous detection of RNAs and proteins in single cells.* Nat Methods, 2016. **13**(3): p. 269-75.

107. Kimmey, S.C., et al., *Parallel analysis of tri-molecular biosynthesis with cell identity and function in single cells.* Nat Commun, 2019. **10**(1): p. 1185.

108. Roederer, M., *Spectral compensation for flow cytometry: visualization artifacts, limitations, and caveats.* Cytometry, 2001. **45**(3): p. 194-205.

109. Liechti, T. and M. Roederer, *OMIP-060: 30-Parameter Flow Cytometry Panel to Assess T Cell Effector Functions and Regulatory T Cells.* Cytometry A, 2019. **95**(11): p. 1129-1134.

110. Nolan, J.P. and D. Condello, *Spectral flow cytometry.* Curr Protoc Cytom, 2013. **Chapter 1**: p. Unit1 27.

111. Mistry, A.M., et al., *Beyond the message: advantages of snapshot proteomics with single-cell mass cytometry in solid tumors.* FEBS J, 2019. **286**(8): p. 1523-1539.

112. Bendall, S.C., et al., *Single-cell mass cytometry of differential immune and drug responses across a human hematopoietic continuum.* Science, 2011. **332**(6030): p. 687-96.

113. Levine, J.H., et al., *Data-Driven Phenotypic Dissection of AML Reveals Progenitor-like Cells that Correlate with Prognosis.* Cell, 2015. **162**(1): p. 184-97.

114. Good, Z., et al., *Single-cell developmental classification of B cell precursor acute lymphoblastic leukemia at diagnosis reveals predictors of relapse.* Nat Med, 2018. **24**(4): p. 474-483.

115. Myklebust, J.H., et al., *Distinct patterns of B-cell receptor signaling in non-Hodgkin lymphomas identified by single-cell profiling.* Blood, 2017. **129**(6): p. 759-770.

116. Doxie, D.B., et al., *BRAF and MEK inhibitor therapy eliminates Nestin-expressing melanoma cells in human tumors.* Pigment Cell Melanoma Res, 2018. **31**(6): p. 708-719.

117. Gonzalez, V.D., et al., *Commonly Occurring Cell Subsets in High-Grade Serous Ovarian Tumors Identified by Single-Cell Mass Cytometry.* Cell Rep, 2018. **22**(7): p. 1875-1888.

118. Keren, L., et al., *A Structured Tumor-Immune Microenvironment in Triple Negative Breast Cancer Revealed by Multiplexed Ion Beam Imaging.* Cell, 2018. **174**(6): p. 1373-1387 e19.

119. Jackson, H.W., et al., *The single-cell pathology landscape of breast cancer.* Nature, 2020. **578**(7796): p. 615-620.

120. Simmons, A.J., et al., *Impaired coordination between signaling pathways is revealed in human colorectal cancer using single-cell mass cytometry of archival tissue blocks.* Sci Signal, 2016. **9**(449): p. rs11.

121. Leelatian, N., et al., *Preparing Viable Single Cells from Human Tissue and Tumors for Cytomic Analysis.* Curr Protoc Mol Biol, 2017. **118**: p. 25C 1 1-25C 1 23.

122. Hotelling, H., *Analysis of a complex of statistical variables into principal components.* Journal of Educational Psychology, 1933. **24**(6): p. 417-441.

123. Ester, M., et al., *A density-based algorithm for discovering clusters a density-based algorithm for discovering clusters in large spatial databases with noise*, in *Proceedings of the Second International Conference on Knowledge Discovery and Data Mining*. 1996, AAAI Press: Portland, Oregon. p. 226–231.

124. Qiu, P., et al., *Extracting a cellular hierarchy from high-dimensional cytometry data with SPADE.* Nat Biotechnol, 2011. **29**(10): p. 886-91.

125. Amir el, A.D., et al., *viSNE enables visualization of high dimensional single-cell data and reveals phenotypic heterogeneity of leukemia.* Nat Biotechnol, 2013. **31**(6): p. 545-52.

126. Shekhar, K., et al., *Automatic Classification of Cellular Expression by Nonlinear Stochastic Embedding (ACCENSE).* Proc Natl Acad Sci U S A, 2014. **111**(1): p. 202-7.

127. Bruggner, R.V., et al., *Automated identification of stratifying signatures in cellular subpopulations.* Proc Natl Acad Sci U S A, 2014. **111**(26): p. E2770-7.

128. Bendall, S.C., et al., *Single-cell trajectory detection uncovers progression and regulatory coordination in human B cell development.* Cell, 2014. **157**(3): p. 714-25.

129. Van Gassen, S., et al., *FlowSOM: Using self-organizing maps for visualization and interpretation of cytometry data.* Cytometry A, 2015. **87**(7): p. 636-45.

130. Weber, L.M. and M.D. Robinson, *Comparison of clustering methods for high-dimensional single-cell flow and mass cytometry data.* Cytometry A, 2016. **89**(12): p. 1084-1096.

131. Spitzer, M.H., et al., *IMMUNOLOGY. An interactive reference framework for modeling a dynamic immune system.* Science, 2015. **349**(6244): p. 1259425.

132. Höllt, T., et al., *Cytosplore: Interactive Immune Cell Phenotyping for Large Single-Cell Datasets.* Computer Graphics Forum, 2016. **35**(3): p. 171-180.

133. D;, C. and M. P, *Mean shift: a robust approach toward feature space analysis.* IEEE Transactions on Pattern Analysis and Machine Intelligence 2002. **24**(5): p. 603-619.

134. Pezzotti, N., et al., *Hierarchical Stochastic Neighbor Embedding.* Computer Graphics Forum, 2016. **35**(3): p. 21-30.

135. Samusik, N., et al., *Automated mapping of phenotype space with single-cell data.* Nat Methods, 2016. **13**(6): p. 493-6.

136. Pezzotti, N., et al., *Approximated and User Steerable tSNE for Progressive Visual Analytics.* IEEE Trans Vis Comput Graph, 2017. **23**(7): p. 1739-1752.

137. Arvaniti, E. and M. Claassen, *Sensitive detection of rare disease-associated cell subsets via representation learning.* Nat Commun, 2017. **8**: p. 14825.

138. Becht, E., et al., *Dimensionality reduction for visualizing single-cell data using UMAP.* Nat Biotechnol, 2018.

139. Linderman, G.C., et al., *Fast interpolation-based t-SNE for improved visualization of single-cell RNA-seq data.* Nat Methods, 2019. **16**(3): p. 243-245.

140. Belkina, A.C., et al., *Automated optimized parameters for T-distributed stochastic neighbor embedding improve visualization and analysis of large datasets.* Nat Commun, 2019. **10**(1): p. 5415.

141. Liu, X., et al., *A comparison framework and guideline of clustering methods for mass cytometry data.* Genome Biol, 2019. **20**(1): p. 297.

142. Matos, T.R., H. Liu, and J. Ritz, *Research Techniques Made Simple: Mass Cytometry Analysis Tools for Decrypting the Complexity of Biological Systems.* J Invest Dermatol, 2017. **137**(5): p. e43-e51.

143. Palit, S., et al., *Meeting the Challenges of High-Dimensional Single-Cell Data Analysis in Immunology.* Front Immunol, 2019. **10**: p. 1515.

144. Saeys, Y., S.V. Gassen, and B.N. Lambrecht, *Computational flow cytometry: helping to make sense of high-dimensional immunology data.* Nat Rev Immunol, 2016. **16**(7): p. 449-62.

145. Diggins, K.E., P.B. Ferrell, Jr., and J.M. Irish, *Methods for discovery and characterization of cell subsets in high dimensional mass cytometry data.* Methods, 2015. **82**: p. 55-63.

146. Greenplate, A.R., et al., *Computational Immune Monitoring Reveals Abnormal Double-Negative T Cells Present across Human Tumor Types.* Cancer Immunol Res, 2019. **7**(1): p. 86-99.

147. Irish, J.M., *Beyond the age of cellular discovery.* Nat Immunol, 2014. **15**(12): p. 1095-7.

148. van Unen, V., et al., *Visual analysis of mass cytometry data by hierarchical stochastic neighbour embedding reveals rare cell types.* Nat Commun, 2017. **8**(1): p. 1740.

149. Irish, J.M. and D.B. Doxie, *High-dimensional single-cell cancer biology.* Curr Top Microbiol Immunol, 2014. **377**: p. 1-21.

150. Irish, J.M., N. Kotecha, and G.P. Nolan, *Mapping normal and cancer cell signalling networks: towards single-cell proteomics.* Nat Rev Cancer, 2006. **6**(2): p. 146-55.

151. Irish, J.M., et al., *Single cell profiling of potentiated phospho-protein networks in cancer cells.* Cell, 2004. **118**(2): p. 217-28.

152. Irish, J.M., et al., *B-cell signaling networks reveal a negative prognostic human lymphoma cell subset that emerges during tumor progression.* Proc Natl Acad Sci U S A, 2010. **107**(29): p. 12747-54.

153. Capper, D., et al., *Characterization of R132H mutation-specific IDH1 antibody binding in brain tumors.* Brain Pathol, 2010. **20**(1): p. 245-54.

154. Capper, D., et al., *Monoclonal antibody specific for IDH1 R132H mutation.* Acta Neuropathol, 2009. **118**(5): p. 599-601.

155. Simmons, A.J., et al., *Cytometry-based single-cell analysis of intact epithelial signaling reveals MAPK activation divergent from TNF-alpha-induced apoptosis in vivo.* Mol Syst Biol, 2015. **11**(10): p. 835.

156. Leelatian, N., *Stratification of Adult Glioblastoma with Signaling and Single Cell Biology*, in *Cancer Biology*. 2018, Vanderbilt University.

157. Gandelman, J.S., et al., *Machine learning reveals chronic graft-versus-host disease phenotypes and stratifies survival after stem cell transplant for hematologic malignancies.* Haematologica, 2018.

158. Diggins, K.E., et al., *Characterizing cell subsets using marker enrichment modeling.* Nat Methods, 2017. **14**(3): p. 275-278.

159. Newell, E.W. and Y. Cheng, *Mass cytometry: blessed with the curse of dimensionality.* Nat Immunol, 2016. **17**(8): p. 890-5.

160. Kotecha, N., et al., *Single-cell profiling identifies aberrant STAT5 activation in myeloid malignancies with specific clinical and biologic correlates.* Cancer Cell, 2008. **14**(4): p. 335-43.

161. Melchiotti, R., et al., *Cluster stability in the analysis of mass cytometry data.* Cytometry A, 2017. **91**(1): p. 73-84.

162. Akers, J.C., et al., *MiR-21 in the extracellular vesicles (EVs) of cerebrospinal fluid (CSF): a platform for glioblastoma biomarker development.* PLoS One, 2013. **8**(10): p. e78115.

163. Gilbert, M.R., et al., *A randomized trial of bevacizumab for newly diagnosed glioblastoma.* N Engl J Med, 2014. **370**(8): p. 699-708.

164. Brennan, C.W., et al., *The somatic genomic landscape of glioblastoma.* Cell, 2013. **155**(2): p. 462-77.

165. Stommel, J.M., et al., *Coactivation of receptor tyrosine kinases affects the response of tumor cells to targeted therapies.* Science, 2007. **318**(5848): p. 287-90.

166. Finck, R., et al., *Normalization of mass cytometry data with bead standards.* Cytometry A, 2013. **83**(5): p. 483-94.

167. Leelatian, N., K.E. Diggins, and J.M. Irish, *Characterizing Phenotypes and Signaling Networks of Single Human Cells by Mass Cytometry.* Methods Mol Biol, 2015. **1346**: p. 99-113.

168. Kotecha, N., P.O. Krutzik, and J.M. Irish, *Web-based analysis and publication of flow cytometry experiments.* Curr Protoc Cytom, 2010. **Chapter 10**: p. Unit10 17.

169. Hussain, S.F., et al., *The role of human glioma-infiltrating microglia/macrophages in mediating antitumor immune responses.* Neuro Oncol, 2006. **8**(3): p. 261-79.

170. Diggins, K.E., et al., *Generating Quantitative Cell Identity Labels with Marker Enrichment Modeling (MEM).* Curr Protoc Cytom, 2018. **83**: p. 10 21 1-10 21 28.

171. Ohgaki, H., et al., *Genetic pathways to glioblastoma: a population-based study.* Cancer Res, 2004. **64**(19): p. 6892-9.

172. Shapiro, W.R., et al., *Randomized trial of three chemotherapy regimens and two radiotherapy regimens and two radiotherapy regimens in postoperative treatment of*

malignant glioma. Brain Tumor Cooperative Group Trial 8001. J Neurosurg, 1989. **71**(1): p. 1-9.

173. Brown, C.E., et al., *Regression of Glioblastoma after Chimeric Antigen Receptor T-Cell Therapy.* N Engl J Med, 2016. **375**(26): p. 2561-9.

174. Brown, T.J., et al., *Association of the Extent of Resection With Survival in Glioblastoma: A Systematic Review and Meta-analysis.* JAMA Oncol, 2016. **2**(11): p. 1460-1469.

175. Grabowski, M.M., et al., *Residual tumor volume versus extent of resection: predictors of survival after surgery for glioblastoma.* J Neurosurg, 2014. **121**(5): p. 1115-23.

176. Mirimanoff, R.O., et al., *Radiotherapy and temozolomide for newly diagnosed glioblastoma: recursive partitioning analysis of the EORTC 26981/22981-NCIC CE3 phase III randomized trial.* J Clin Oncol, 2006. **24**(16): p. 2563-9.

177. Walker, M.D., et al., *Randomized comparisons of radiotherapy and nitrosoureas for the treatment of malignant glioma after surgery.* N Engl J Med, 1980. **303**(23): p. 1323-9.

178. Beyrend, G., et al., *Cytofast: A workflow for visual and quantitative analysis of flow and mass cytometry data to discover immune signatures and correlations.* Comput Struct Biotechnol J, 2018. **16**: p. 435-442.

179. Lan, X., et al., *Fate mapping of human glioblastoma reveals an invariant stem cell hierarchy.* Nature, 2017. **549**(7671): p. 227-232.

180. Ikushima, H., et al., *Autocrine TGF-beta signaling maintains tumorigenicity of glioma-initiating cells through Sry-related HMG-box factors.* Cell Stem Cell, 2009. **5**(5): p. 504-14.

181. Raponi, E., et al., *S100B expression defines a state in which GFAP-expressing cells lose their neural stem cell potential and acquire a more mature developmental stage.* Glia, 2007. **55**(2): p. 165-77.

182. Wei, J., et al., *miR-124 inhibits STAT3 signaling to enhance T cell-mediated immune clearance of glioma.* Cancer Res, 2013. **73**(13): p. 3913-26.

183. Tan, M.S.Y., et al., *A STAT3-based gene signature stratifies glioma patients for targeted therapy.* Nat Commun, 2019. **10**(1): p. 3601.

184. Bhat, K.P.L., et al., *Mesenchymal differentiation mediated by NF-kappaB promotes radiation resistance in glioblastoma.* Cancer Cell, 2013. **24**(3): p. 331-46.

185. Carro, M.S., et al., *The transcriptional network for mesenchymal transformation of brain tumours.* Nature, 2010. **463**(7279): p. 318-25.

186. Dolma, S., et al., *Inhibition of Dopamine Receptor D4 Impedes Autophagic Flux, Proliferation, and Survival of Glioblastoma Stem Cells.* Cancer Cell, 2016. **29**(6): p. 859-873.

187. Fan, Q., et al., *A Kinase Inhibitor Targeted to mTORC1 Drives Regression in Glioblastoma.* Cancer Cell, 2017. **31**(3): p. 424-435.

188. Hubert, C.G., et al., *A Three-Dimensional Organoid Culture System Derived from Human Glioblastomas Recapitulates the Hypoxic Gradients and Cancer Stem Cell Heterogeneity of Tumors Found In Vivo.* Cancer Res, 2016. **76**(8): p. 2465-77.

189. Jacob, F., et al., *A Patient-Derived Glioblastoma Organoid Model and Biobank Recapitulates Inter- and Intra-tumoral Heterogeneity.* Cell, 2020. **180**(1): p. 188-204 e22.

190. Ogawa, J., et al., *Glioblastoma Model Using Human Cerebral Organoids.* Cell Rep, 2018. **23**(4): p. 1220-1229.

191. Pine, A.R., et al., *Tumor microenvironment is critical for the maintenance of cellular states found in primary glioblastomas.* Cancer Discov, 2020.

192. Yoshimatsu, T., et al., *Non-cell-autonomous action of STAT3 in maintenance of neural precursor cells in the mouse neocortex.* Development, 2006. **133**(13): p. 2553-63.

193. Rushing, G.V., et al., *Location-dependent maintenance of intrinsic susceptibility to mTORC1-driven tumorigenesis.* Life Sci Alliance, 2019. **2**(2).

194. Saadeh, F.S., R. Mahfouz, and H.I. Assi, *EGFR as a clinical marker in glioblastomas and other gliomas.* Int J Biol Markers, 2018. **33**(1): p. 22-32.

195. Xu, H., et al., *Epidermal growth factor receptor in glioblastoma.* Oncol Lett, 2017. **14**(1): p. 512-516.

196. Li, J., et al., *Prognostic significance of epidermal growth factor receptor expression in glioma patients.* Onco Targets Ther, 2018. **11**: p. 731-742.

197. Fan, Q.W., et al., *EGFR signals to mTOR through PKC and independently of Akt in glioma.* Sci Signal, 2009. **2**(55): p. ra4.

198. Fan, Q.W., et al., *EGFR phosphorylates tumor-derived EGFRvIII driving STAT3/5 and progression in glioblastoma.* Cancer Cell, 2013. **24**(4): p. 438-49.

199. Fujioka, S., et al., *Stabilization of p53 is a novel mechanism for proapoptotic function of NF-kappaB.* J Biol Chem, 2004. **279**(26): p. 27549-59.

200. Schneider, A., et al., *NF-kappaB is activated and promotes cell death in focal cerebral ischemia.* Nat Med, 1999. **5**(5): p. 554-9.

201. Baser, A., et al., *Onset of differentiation is post-transcriptionally controlled in adult neural stem cells.* Nature, 2019. **566**(7742): p. 100-104.

202. Chakravarty, D., et al., *EGFR and PDGFRA co-expression and heterodimerization in glioblastoma tumor sphere lines.* Sci Rep, 2017. **7**(1): p. 9043.

203. Holla, F.K., et al., *Prognostic value of the S100B protein in newly diagnosed and recurrent glioma patients: a serial analysis.* J Neurooncol, 2016. **129**(3): p. 525-532.

204. Wang, H., et al., *S100B promotes glioma growth through chemoattraction of myeloid-derived macrophages.* Clin Cancer Res, 2013. **19**(14): p. 3764-75.

205. Cancer Genome Atlas Research, N., et al., *Comprehensive, Integrative Genomic Analysis of Diffuse Lower-Grade Gliomas.* N Engl J Med, 2015. **372**(26): p. 2481-98.

206. Petti, A.A., et al., *A general approach for detecting expressed mutations in AML cells using single cell RNA-sequencing.* Nat Commun, 2019. **10**(1): p. 3660.

207. Liu, F., et al., *Systematic comparative analysis of single-nucleotide variant detection methods from single-cell RNA sequencing data.* Genome Biol, 2019. **20**(1): p. 242.

208. Molenaar, R.J., et al., *Wild-type and mutated IDH1/2 enzymes and therapy responses.* Oncogene, 2018. **37**(15): p. 1949-1960.

209. Millard, I., et al., *Detection of intracellular antigens by flow cytometry: comparison of two chemical methods and microwave heating.* Clin Chem, 1998. **44**(11): p. 2320-30.

210. Ogino, T., X. Wang, and S. Ferrone, *Modified flow cytometry and cell-ELISA methodology to detect HLA class I antigen processing machinery components in cytoplasm and endoplasmic reticulum.* J Immunol Methods, 2003. **278**(1-2): p. 33-44.

211. Reynolds, B.A., W. Tetzlaff, and S. Weiss, *A multipotent EGF-responsive striatal embryonic progenitor cell produces neurons and astrocytes.* J Neurosci, 1992. **12**(11): p. 4565-74.

212. Mistry, A.M., C.L. Vnencak-Jones, and B.C. Mobley, *Clinical prognostic value of the isocitrate dehydrogenase 1 single-nucleotide polymorphism rs11554137 in glioblastoma.* J Neurooncol, 2018. **138**(2): p. 307-313.

213. Scurrah, C.R., A.J. Simmons, and K.S. Lau, *Single-Cell Mass Cytometry of Archived Human Epithelial Tissue for Decoding Cancer Signaling Pathways.* Methods Mol Biol, 2019. **1884**: p. 215-229.

214. Luchman, H.A., et al., *An in vivo patient-derived model of endogenous IDH1-mutant glioma.* Neuro Oncol, 2012. **14**(2): p. 184-91.

215. Nagashima, H., et al., *Diagnostic value of glutamate with 2-hydroxyglutarate in magnetic resonance spectroscopy for IDH1 mutant glioma.* Neuro Oncol, 2016. **18**(11): p. 1559-1568.

216. Long, D.J., 2nd and C. Buggs, *Microwave oven-based technique for immunofluorescent staining of paraffin-embedded tissues.* J Mol Histol, 2008. **39**(1): p. 1-4.

217. Agarwal, S., et al., *Comparative study of IDH1 mutations in gliomas by immunohistochemistry and DNA sequencing.* Neuro Oncol, 2013. **15**(6): p. 718-26.

218. Krutzik, P.O. and G.P. Nolan, *Intracellular phospho-protein staining techniques for flow cytometry: monitoring single cell signaling events.* Cytometry A, 2003. **55**(2): p. 61-70.

219. Krutzik, P.O., et al., *Analysis of protein phosphorylation and cellular signaling events by flow cytometry: techniques and clinical applications.* Clin Immunol, 2004. **110**(3): p. 206-21.

220. Gorovets, D., et al., *IDH mutation and neuroglial developmental features define clinically distinct subclasses of lower grade diffuse astrocytic glioma.* Clin Cancer Res, 2012. **18**(9): p. 2490-501.

221. Olar, A., et al., *IDH mutation status and role of WHO grade and mitotic index in overall survival in grade II-III diffuse gliomas.* Acta Neuropathol, 2015. **129**(4): p. 585-96.

222. Guilhamon, P., et al., *Meta-analysis of IDH-mutant cancers identifies EBF1 as an interaction partner for TET2.* Nat Commun, 2013. **4**: p. 2166.

223. Bendall, S.C., et al., *A deep profiler's guide to cytometry.* Trends Immunol, 2012. **33**(7): p. 323-32.

224. Davis, B.H., et al., *2006 Bethesda International Consensus recommendations on the flow cytometric immunophenotypic analysis of hematolymphoid neoplasia: medical indications.* Cytometry B Clin Cytom, 2007. **72 Suppl 1**: p. S5-13.

225. Roussel, M., et al., *Refining the white blood cell differential: the first flow cytometry routine application.* Cytometry A, 2010. **77**(6): p. 552-63.

226. Yuan, Y., et al., *Autofluorescence of NADH is a new biomarker for sorting and characterizing cancer stem cells in human glioma.* Stem Cell Res Ther, 2019. **10**(1): p. 330.

227. Leppert, J., et al., *Multiphoton excitation of autofluorescence for microscopy of glioma tissue.* Neurosurgery, 2006. **58**(4): p. 759-67; discussion 759-67.

228. Kalina, T., K. Lundsten, and P. Engel, *Relevance of Antibody Validation for Flow Cytometry.* Cytometry A, 2020. **97**(2): p. 126-136.

229. Caretti, V., et al., *Subventricular spread of diffuse intrinsic pontine glioma.* Acta Neuropathol, 2014. **128**(4): p. 605-7.

230. Qin, E.Y., et al., *Neural Precursor-Derived Pleiotrophin Mediates Subventricular Zone Invasion by Glioma.* Cell, 2017. **170**(5): p. 845-859 e19.

231. Chen, L., et al., *Glioblastoma recurrence patterns near neural stem cell regions.* Radiother Oncol, 2015. **116**(2): p. 294-300.

232. Jafri, N.F., et al., *Relationship of glioblastoma multiforme to the subventricular zone is associated with survival.* Neuro Oncol, 2013. **15**(1): p. 91-6.

233. Adeberg, S., et al., *Glioblastoma recurrence patterns after radiation therapy with regard to the subventricular zone.* Int J Radiat Oncol Biol Phys, 2014. **90**(4): p. 886-93.

234. Nestler, U., et al., *Anatomic features of glioblastoma and their potential impact on survival.* Acta Neurochir (Wien), 2015. **157**(2): p. 179-86.

235. Sonoda, Y., et al., *The association of subventricular zone involvement at recurrence with survival after repeat surgery in patients with recurrent glioblastoma.* Neurol Med Chir (Tokyo), 2014. **54**(4): p. 302-9.

236. Kimura, M., et al., *Glioblastoma multiforme: relationship to subventricular zone and recurrence.* Neuroradiol J, 2013. **26**(5): p. 542-7.

237. Lim, D.A., et al., *Relationship of glioblastoma multiforme to neural stem cell regions predicts invasive and multifocal tumor phenotype.* Neuro Oncol, 2007. **9**(4): p. 424-9.

238. Radbruch, A., et al., *Relevance of T2 signal changes in the assessment of progression of glioblastoma according to the Response Assessment in Neurooncology criteria.* Neuro Oncol, 2012. **14**(2): p. 222-9.

239. Liu, S., et al., *Anatomical Involvement of the Subventricular Zone Predicts Poor Survival Outcome in Low-Grade Astrocytomas.* PLoS One, 2016. **11**(4): p. e0154539.

240. Lai, A., et al., *Evidence for sequenced molecular evolution of IDH1 mutant glioblastoma from a distinct cell of origin.* J Clin Oncol, 2011. **29**(34): p. 4482-90.

241. Jungk, C., et al., *Spatial transcriptome analysis reveals Notch pathway-associated prognostic markers in IDH1 wild-type glioblastoma involving the subventricular zone.* BMC Med, 2016. **14**(1): p. 170.

242. Orlova, D.Y., et al., *Earth Mover's Distance (EMD): A True Metric for Comparing Biomarker Expression Levels in Cell Populations.* PLoS One, 2016. **11**(3): p. e0151859.

243. Tomita, T., et al., *Clinicopathological significance of expression of nestin, a neural stem/progenitor cell marker, in human glioma tissue.* Brain Tumor Pathol, 2014. **31**(3): p. 162-71.

244. Han, S., et al., *Can lateral ventricle contact predict the ontogeny and prognosis of glioblastoma?* Journal of Neuro-Oncology, 2015. **124**(1): p. 45-55.

245. Lin, C.A., et al., *Comparative analyses identify molecular signature of MRI-classified SVZ-associated glioblastoma.* Cell Cycle, 2017. **16**(8): p. 765-775.

246. Haskins, W.E., et al., *Molecular Characteristics in MRI-Classified Group 1 Glioblastoma Multiforme.* Front Oncol, 2013. **3**: p. 182.

247. Steed, T.C., et al., *Differential localization of glioblastoma subtype: implications on glioblastoma pathogenesis.* Oncotarget, 2016. **7**(18): p. 24899-907.

248. Jamshidi, N., et al., *Illuminating radiogenomic characteristics of glioblastoma multiforme through integration of MR imaging, messenger RNA expression, and DNA copy number variation.* Radiology, 2014. **270**(1): p. 1-2.

249. McKinley, E.T., et al., *Optimized multiplex immunofluorescence single-cell analysis reveals tuft cell heterogeneity.* JCI Insight, 2017. **2**(11).

250. Giesen, C., et al., *Highly multiplexed imaging of tumor tissues with subcellular resolution by mass cytometry.* Nat Methods, 2014. **11**(4): p. 417-22.

251. Angelo, M., et al., *Multiplexed ion beam imaging of human breast tumors.* Nat Med, 2014. **20**(4): p. 436-42.

252. Guichet, P.O., et al., *Notch1 stimulation induces a vascularization switch with pericyte-like cell differentiation of glioblastoma stem cells.* Stem Cells, 2015. **33**(1): p. 21-34.

253. Nourallah, B., et al., *Irradiating the Subventricular Zone in Glioblastoma Patients: Is there a Case for a Clinical Trial?* Clin Oncol (R Coll Radiol), 2017. **29**(1): p. 26-33.

254. Fuentealba, L.C., et al., *Embryonic Origin of Postnatal Neural Stem Cells.* Cell, 2015. **161**(7): p. 1644-55.

255. Ihrie, R.A., et al., *Persistent sonic hedgehog signaling in adult brain determines neural stem cell positional identity.* Neuron, 2011. **71**(2): p. 250-62.

256. Llorens-Bobadilla, E., et al., *Single-Cell Transcriptomics Reveals a Population of Dormant Neural Stem Cells that Become Activated upon Brain Injury.* Cell Stem Cell, 2015. **17**(3): p. 329-40.

257. Merkle, F.T., Z. Mirzadeh, and A. Alvarez-Buylla, *Mosaic organization of neural stem cells in the adult brain.* Science, 2007. **317**(5836): p. 381-4.

258. Kelsch, W., et al., *Distinct mammalian precursors are committed to generate neurons with defined dendritic projection patterns.* PLoS Biol, 2007. **5**(11): p. e300.

259. Young, K.M., et al., *Subventricular zone stem cells are heterogeneous with respect to their embryonic origins and neurogenic fates in the adult olfactory bulb.* J Neurosci, 2007. **27**(31): p. 8286-96.

260. Solano Fonseca, R., et al., *Neurogenic Niche Microglia Undergo Positional Remodeling and Progressive Activation Contributing to Age-Associated Reductions in Neurogenesis.* Stem Cells Dev, 2016. **25**(7): p. 542-55.

261. Druker, B.J., et al., *Activity of a specific inhibitor of the BCR-ABL tyrosine kinase in the blast crisis of chronic myeloid leukemia and acute lymphoblastic leukemia with the Philadelphia chromosome.* N Engl J Med, 2001. **344**(14): p. 1038-42.

262. Slamon, D.J., et al., *Use of chemotherapy plus a monoclonal antibody against HER2 for metastatic breast cancer that overexpresses HER2.* N Engl J Med, 2001. **344**(11): p. 783-92.

263. Pakkala, S. and S.S. Ramalingam, *Personalized therapy for lung cancer: striking a moving target.* JCI Insight, 2018. **3**(15).

264. Robert, C., et al., *Five-Year Outcomes with Dabrafenib plus Trametinib in Metastatic Melanoma.* N Engl J Med, 2019. **381**(7): p. 626-636.

265. Ribas, A. and K.T. Flaherty, *BRAF targeted therapy changes the treatment paradigm in melanoma.* Nat Rev Clin Oncol, 2011. **8**(7): p. 426-33.

266. Muller, S., et al., *Single-cell profiling of human gliomas reveals macrophage ontogeny as a basis for regional differences in macrophage activation in the tumor microenvironment.* Genome Biol, 2017. **18**(1): p. 234.

267. Fack, F., et al., *Altered metabolic landscape in IDH-mutant gliomas affects phospholipid, energy, and oxidative stress pathways.* EMBO Mol Med, 2017. **9**(12): p. 1681-1695.

268. Parker, S.J. and C.M. Metallo, *Metabolic consequences of oncogenic IDH mutations.* Pharmacol Ther, 2015. **152**: p. 54-62.

269. Garrett, M., et al., *Metabolic characterization of isocitrate dehydrogenase (IDH) mutant and IDH wildtype gliomaspheres uncovers cell type-specific vulnerabilities.* Cancer Metab, 2018. **6**: p. 4.

270. Goffart, N., et al., *CXCL12 mediates glioblastoma resistance to radiotherapy in the subventricular zone.* Neuro Oncol, 2017. **19**(1): p. 66-77.

271. Kroonen, J., et al., *Human glioblastoma-initiating cells invade specifically the subventricular zones and olfactory bulbs of mice after striatal injection.* Int J Cancer, 2011. **129**(3): p. 574-85.

272. Schmitz, M., et al., *Identification of SOX2 as a novel glioma-associated antigen and potential target for T cell-based immunotherapy.* British Journal of Cancer, 2007. **96**(8): p. 1293-1301.

273. Ekstrand, A.J., et al., *Genes for epidermal growth factor receptor, transforming growth factor alpha, and epidermal growth factor and their expression in human gliomas in vivo.* Cancer Res, 1991. **51**(8): p. 2164-72.

274. Aldape, K.D., et al., *Immunohistochemical detection of EGFRvIII in high malignancy grade astrocytomas and evaluation of prognostic significance.* J Neuropathol Exp Neurol, 2004. **63**(7): p. 700-7.

275. Lee, J., et al., *Tumor stem cells derived from glioblastomas cultured in bFGF and EGF more closely mirror the phenotype and genotype of primary tumors than do serum-cultured cell lines.* Cancer Cell, 2006. **9**(5): p. 391-403.

276. Davis, B., et al., *Comparative genomic and genetic analysis of glioblastoma-derived brain tumor-initiating cells and their parent tumors.* Neuro Oncol, 2016. **18**(3): p. 350-60.

277. Bette, S., et al., *Retrospective Analysis of Radiological Recurrence Patterns in Glioblastoma, Their Prognostic Value And Association to Postoperative Infarct Volume.* Sci Rep, 2018. **8**(1): p. 4561.

278. Barbagallo, G.M.V., M.D. Jenkinson, and A.R. Brodbelt, *'Recurrent' glioblastoma multiforme, when should we reoperate?* British Journal of Neurosurgery, 2008. **22**(3): p. 452-455.

279. Kim, H., et al., *Whole-genome and multisector exome sequencing of primary and post-treatment glioblastoma reveals patterns of tumor evolution.* Genome Res, 2015. **25**(3): p. 316-27.

280. Kim, J., et al., *Spatiotemporal Evolution of the Primary Glioblastoma Genome.* Cancer Cell, 2015. **28**(3): p. 318-28.

281. Doetsch, F., J.M. Garcia-Verdugo, and A. Alvarez-Buylla, *Cellular composition and three-dimensional organization of the subventricular germinal zone in the adult mammalian brain.* J Neurosci, 1997. **17**(13): p. 5046-61.

282. Chattopadhyay, P.K., et al., *Single-cell technologies for monitoring immune systems.* Nat Immunol, 2014. **15**(2): p. 128-35.

283. Picot, J., et al., *Flow cytometry: retrospective, fundamentals and recent instrumentation.* Cytotechnology, 2012. **64**(2): p. 109-30.

284. Greenplate, A.R., et al., *Myelodysplastic Syndrome Revealed by Systems Immunology in a Melanoma Patient Undergoing Anti-PD-1 Therapy.* Cancer Immunol Res, 2016. **4**(6): p. 474-480.

285. Hardy, R.R., et al., *Demonstration of B-cell maturation in X-linked immunodeficient mice by simultaneous three-colour immunofluorescence.* Nature, 1983. **306**(5940): p. 270-2.

286. Nicholas, K.J., et al., *Multiparameter analysis of stimulated human peripheral blood mononuclear cells: A comparison of mass and fluorescence cytometry.* Cytometry A, 2015.

287. Parks, D.R., R.R. Hardy, and L.A. Herzenberg, *Three-color immunofluorescence analysis of mouse B-lymphocyte subpopulations.* Cytometry, 1984. **5**(2): p. 159-68.

288. Tung, J.W., et al., *Identification of B-cell subsets: an exposition of 11-color (Hi-D) FACS methods.* Methods Mol Biol, 2004. **271**: p. 37-58.

289. Irish, J.M., et al., *Altered B-cell receptor signaling kinetics distinguish human follicular lymphoma B cells from tumor-infiltrating nonmalignant B cells.* Blood, 2006. **108**(9): p. 3135-42.

290. Polikowsky, H.G., et al., *Cutting Edge: Redox Signaling Hypersensitivity Distinguishes Human Germinal Center B Cells.* J Immunol, 2015. **195**(4): p. 1364-7.

291. Wogsland, C.E., et al., *Mass Cytometry of Follicular Lymphoma Tumors Reveals Intrinsic Heterogeneity in Proteins Including HLA-DR and a Deficit in Nonmalignant Plasmablast and Germinal Center B-Cell Populations.* Cytometry B Clin Cytom, 2017. **92**(1): p. 79-87.

292. Arber, D.A., et al., *The 2016 revision to the World Health Organization classification of myeloid neoplasms and acute leukemia.* Blood, 2016. **127**(20): p. 2391-405.

293. Craig, F.E. and K.A. Foon, *Flow cytometric immunophenotyping for hematologic neoplasms.* Blood, 2008. **111**(8): p. 3941-67.

294. van Dongen, J.J., et al., *EuroFlow antibody panels for standardized n-dimensional flow cytometric immunophenotyping of normal, reactive and malignant leukocytes.* Leukemia, 2012. **26**(9): p. 1908-75.

295. Wood, B.L., et al., *2006 Bethesda International Consensus recommendations on the immunophenotypic analysis of hematolymphoid neoplasia by flow cytometry: optimal reagents and reporting for the flow cytometric diagnosis of hematopoietic neoplasia.* Cytometry B Clin Cytom, 2007. **72 Suppl 1**: p. S14-22.

296. Borowitz, M.J., et al., *Clinical significance of minimal residual disease in childhood acute lymphoblastic leukemia and its relationship to other prognostic factors: a Children's Oncology Group study.* Blood, 2008. **111**(12): p. 5477-85.

297. van Dongen, J.J., et al., *Minimal residual disease diagnostics in acute lymphoblastic leukemia: need for sensitive, fast, and standardized technologies.* Blood, 2015. **125**(26): p. 3996-4009.

298. Al-Hajj, M., et al., *Prospective identification of tumorigenic breast cancer cells.* Proc Natl Acad Sci U S A, 2003. **100**(7): p. 3983-8.

299. Chan, K.S., et al., *Identification, molecular characterization, clinical prognosis, and therapeutic targeting of human bladder tumor-initiating cells.* Proc Natl Acad Sci U S A, 2009. **106**(33): p. 14016-21.

300. Donnenberg, V.S., et al., *Flow cytometric determination of stem/progenitor content in epithelial tissues: an example from nonsmall lung cancer and normal lung.* Cytometry A, 2013. **83**(1): p. 141-9.

301. Richards, J.O., et al., *Flow cytometry assessment of residual melanoma cells in tumor-infiltrating lymphocyte cultures.* Cytometry A, 2012. **81**(5): p. 374-81.

302. Zimmerlin, L., V.S. Donnenberg, and A.D. Donnenberg, *Rare event detection and analysis in flow cytometry: bone marrow mesenchymal stem cells, breast cancer stem/progenitor cells in malignant effusions, and pericytes in disaggregated adipose tissue.* Methods Mol Biol, 2011. **699**: p. 251-73.

303. Brummelman, J., et al., *High-dimensional single cell analysis identifies stem-like cytotoxic CD8(+) T cells infiltrating human tumors.* J Exp Med, 2018. **215**(10): p. 2520-2535.

304. Mair, F. and M. Prlic, *OMIP-044: 28-color immunophenotyping of the human dendritic cell compartment.* Cytometry A, 2018. **93**(4): p. 402-405.

305. Maecker, H.T., et al., *Selecting fluorochrome conjugates for maximum sensitivity.* Cytometry A, 2004. **62**(2): p. 169-73.

306. Wang, L., et al., *Human CD4+ lymphocytes for antigen quantification: characterization using conventional flow cytometry and mass cytometry.* Cytometry A, 2012. **81**(7): p. 567-75.

307. Becher, B., et al., *High-dimensional analysis of the murine myeloid cell system.* Nat Immunol, 2014. **15**(12): p. 1181-9.

308. Ferrell, P.B., Jr., et al., *High-Dimensional Analysis of Acute Myeloid Leukemia Reveals Phenotypic Changes in Persistent Cells during Induction Therapy.* PLoS One, 2016. **11**(4): p. e0153207.

309. Gaudilliere, B., et al., *Clinical recovery from surgery correlates with single-cell immune signatures.* Sci Transl Med, 2014. **6**(255): p. 255ra131.

310. Kordasti, S., et al., *Deep phenotyping of Tregs identifies an immune signature for idiopathic aplastic anemia and predicts response to treatment.* Blood, 2016. **128**(9): p. 1193-205.

311. Newell, E.W., et al., *Cytometry by time-of-flight shows combinatorial cytokine expression and virus-specific cell niches within a continuum of CD8+ T cell phenotypes.* Immunity, 2012. **36**(1): p. 142-52.

312. Codega, P., et al., *Prospective identification and purification of quiescent adult neural stem cells from their in vivo niche.* Neuron, 2014. **82**(3): p. 545-59.

313. Pastrana, E., L.C. Cheng, and F. Doetsch, *Simultaneous prospective purification of adult subventricular zone neural stem cells and their progeny.* Proc Natl Acad Sci U S A, 2009. **106**(15): p. 6387-92.

314. Rahman, M., et al., *Neurosphere and adherent culture conditions are equivalent for malignant glioma stem cell lines.* Anat Cell Biol, 2015. **48**(1): p. 25-35.

315. Schulz, K.R., et al., *Single-cell phospho-protein analysis by flow cytometry.* Curr Protoc Immunol, 2012. **Chapter 8**: p. Unit 8 17 1-20.

316. Azari, H., et al., *Isolation and expansion of human glioblastoma multiforme tumor cells using the neurosphere assay.* J Vis Exp, 2011(56): p. e3633.

317. Krutzik, P.O., M.R. Clutter, and G.P. Nolan, *Coordinate analysis of murine immune cell surface markers and intracellular phosphoproteins by flow cytometry.* J Immunol, 2005. **175**(4): p. 2357-65.

318. McLaughlin, B.E., et al., *Nine-color flow cytometry for accurate measurement of T cell subsets and cytokine responses. Part I: Panel design by an empiric approach.* Cytometry A, 2008. **73**(5): p. 400-10.

319. Hulspas, R., et al., *Considerations for the control of background fluorescence in clinical flow cytometry.* Cytometry B Clin Cytom, 2009. **76**(6): p. 355-64.

320. Leipold, M.D., E.W. Newell, and H.T. Maecker, *Multiparameter Phenotyping of Human PBMCs Using Mass Cytometry.* Methods Mol Biol, 2015. **1343**: p. 81-95.

321. Behbehani, G.K., et al., *Transient partial permeabilization with saponin enables cellular barcoding prior to surface marker staining.* Cytometry A, 2014. **85**(12): p. 1011-9.

322. Zunder, E.R., et al., *Palladium-based mass tag cell barcoding with a doublet-filtering scheme and single-cell deconvolution algorithm.* Nat Protoc, 2015. **10**(2): p. 316-33.

323. Krutzik, P.O. and G.P. Nolan, *Fluorescent cell barcoding in flow cytometry allows high-throughput drug screening and signaling profiling.* Nat Methods, 2006. **3**(5): p. 361-8.

324. Chattopadhyay, P.K. and M. Roederer, *Cytometry: today's technology and tomorrow's horizons.* Methods, 2012. **57**(3): p. 251-8.

325. Mahnke, Y.D. and M. Roederer, *Optimizing a multicolor immunophenotyping assay.* Clin Lab Med, 2007. **27**(3): p. 469-85, v.

326. Perfetto, S.P., P.K. Chattopadhyay, and M. Roederer, *Seventeen-colour flow cytometry: unravelling the immune system.* Nat Rev Immunol, 2004. **4**(8): p. 648-55.

327. Robinson, J.P., *Comparative overview of flow and image cytometry.* Curr Protoc Cytom, 2005. **Chapter 12**: p. Unit 12 1.

328. Bjornson, Z.B., G.P. Nolan, and W.J. Fantl, *Single-cell mass cytometry for analysis of immune system functional states.* Curr Opin Immunol, 2013. **25**(4): p. 484-94.

329. Lee, P.P., et al., *Characterization of circulating T cells specific for tumor-associated antigens in melanoma patients.* Nat Med, 1999. **5**(6): p. 677-85.

330. Roederer, M., et al., *The genetic architecture of the human immune system: a bioresource for autoimmunity and disease pathogenesis.* Cell, 2015. **161**(2): p. 387-403.

331. Ornatsky, O.I., et al., *Study of cell antigens and intracellular DNA by identification of element-containing labels and metallointercalators using inductively coupled plasma mass spectrometry.* Anal Chem, 2008. **80**(7): p. 2539-47.

332. DuPage, M. and J.A. Bluestone, *Harnessing the plasticity of CD4(+) T cells to treat immune-mediated disease.* Nat Rev Immunol, 2016. **16**(3): p. 149-63.

333. *Nomenclature for clusters of differentiation (CD) of antigens defined on human leukocyte populations. IUIS-WHO Nomenclature Subcommittee.* Bull World Health Organ, 1984. **62**(5): p. 809-15.

334. Maecker, H.T., J.P. McCoy, and R. Nussenblatt, *Standardizing immunophenotyping for the Human Immunology Project.* Nat Rev Immunol, 2012. **12**(3): p. 191-200.

335. Wang, B., et al., *Visualization and analysis of single-cell RNA-seq data by kernel-based similarity learning.* Nat Methods, 2017. **14**(4): p. 414-416.

336. Shen-Orr, S.S., et al., *Towards a cytokine-cell interaction knowledgebase of the adaptive immune system.* Pac Symp Biocomput, 2009: p. 439-50.