Stratification of Adult Glioblastoma with Signaling and Single Cell Biology

By

Nalin Leelatian, M.D.

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

May 11, 2018

Nashville, Tennessee

Approved:

Jonathan M. Irish, Ph.D.
Alissa M. Weaver, M.D., Ph.D.
Jin Chen, M.D., Ph.D.
Bruce D. Carter, Ph.D.
Rebecca A. Ihrie, Ph.D.

ACKNOWLEDGEMENTS

My deepest gratitude is extended to the following individuals, who helped making my dream of becoming a scientist turn into a reality:

Dr. Jonathan Irish, my thesis advisor, for his continued scientific and career mentorship, and for teaching me that it is okay to be assertive and to voice my opinion.

Dr. Rebecca Ihrie, my co-mentor, for her encouragements and guidance, and for catching every single typo in my abstracts and manuscripts.

Drs. Alissa Weaver, Jin Chen, and Bruce Carter, my thesis committee members, for their invaluable discussions that had shaped the direction of the work in this dissertation.

Kirsten, my introvert buddy, for spending quiet afternoons not talking, sipping coffee, and reading all the books together.

Justine, my glioblastoma partner in crime, for giving me strength through the mass cytometry experiment marathon, and for inventing the nanodrop dance.

Allie, my rotation and thesis-writing buddy, for always being there to lend an ear, and for giving me reassurances a thousand times a day.

My beloved mom and dad, for all the love, sacrifice, patience, prayers, for being the voice of reason, for seeing my potential, and for never stop believing in me.

My dearest husband, for keeping me grounded, strong, humble, and for all the lovingly silly moments that make me breathe a little easier each day. With all my heart, I love you.

This work would not have been possible without the financial support from the Vanderbilt International Scholars Program (VISP), the Vanderbilt University Discovery Grant, the Vanderbilt-Ingram Cancer Center (VICC, P30 CA68485), and the Michael David Greene Brain Cancer Fund.

TABLE OF CONTENTS

Page

LIST OF FIGURES

Figure Page

LIST OF TABLES

CHAPTER 1

Scope of Dissertation

Glioblastoma is a rapidly fatal brain tumor. Despite the recognizable need for novel treatments, the only major update in standard clinical care over the past decade was the incorporation of a non-targeted, highly toxic, chemotherapeutic agent, temozolomide (*1*). The major efforts to dissect the underlying mechanistic aberrancies of adult glioblastomas have relied on genomic and transcriptomic profiling (*2-5*). Even though these approaches have revealed the inter-tumor and intra-tumor diversity of glioblastoma, they have neither successfully stratified patient outcome nor determined specific, individualized therapeutic strategies.

The dissection of glioblastoma pathogenesis must go beyond the understanding of genomic and transcriptomic alterations. Although many functional hallmarks that enable cellular malignant properties can be affected by genomic changes, additional components, including epigenetic and extracellular influences, can also shift the properties of cancer cells (*6-8*). Cancer cell signaling profiles provide a quantitative outlook on the effects secondary to cell-intrinsic changes, as well as the mechanisms by which cancer cells respond to the external environment (*9, 10*). Clinical stratification based on activated signaling profiles has proven successful in blood malignancies and has pinpointed the signaling bottlenecks that correlated with adverse patient outcome (*10-15*). The goal of this dissertation is to establish a novel stratification approach for adult glioblastoma, which couples signaling quantification and delineation of cellular diversity. This will be accomplished by using advanced and automated technical approaches, to dissect the molecular features of single cells in an unbiased fashion. The overarching goal of this work is to identify cellular characteristics that can be immediately investigated as targets for innovative and individualized therapies based on the unique cellular composition within each tumor.

Chapter 2, entitled “The diversity of human malignancies”, describes the different levels of tumor diversity, including the heterogeneity between tumors from different individuals, between tumors that are spatially distinct, as well as intra-tumor single-cell diversity. Additionally, this chapter explains the different biological components that contribute to the diversity between individual cancer cells. Overall, this chapter builds the rationale for dissecting the alterations of signaling activity in single cancer cells, which is the central goal of this dissertation.

Chapter 3, entitled “Adult glioblastoma”, covers the current glioblastoma clinical diagnostics, standard therapeutics, and previous attempts to dissect and target its molecular alterations. This chapter continues to elaborate on the concept of tumor diversity, explained in Chapter 2, in the context of adult glioblastoma. A significant portion of this chapter describes the signaling network alterations in glioblastoma and why they should be considered hallmarks of glioblastoma pathogenesis. Most importantly, this chapter explores the reasons glioblastoma remains incurable and proposes some of the potential research and therapeutic approaches to tackle this deadly disease, including single-cell signaling quantification using flow cytometry-based approaches and computational analysis tools.

Chapter 4, entitled “Characterizing phenotypes and signaling networks of single human cells by mass cytometry”, describes the application of single-cell mass cytometry analysis to human peripheral blood mononuclear cells. Specifically, it describes a step-by-step workflow that includes sample preparation, immunostaining, mass cytometry data collection and analysis. Multiple computational tools, including biaxial analysis, t-distributed stochastic neighbor embedding (t-SNE), and spanning-tree progression analysis of density-normalized events (SPADE), were used for the analysis and data display for various purposes. Moreover, the methods for performing and characterizing potentiated signaling in single cells are also described in detail. This chapter provides a basic analytical workflow for single-cell analysis that can be further applied in more complex and less well-characterized tissue types, such as glioblastoma, as described in subsequent chapters.

Chapter 5, entitled “Single-cell analysis of human tissues and solid tumors with mass cytometry”, describes the development of a combined mechanical and enzymatic dissociation protocol that allows derivation of viable single cells from human tissues and solid tumors for cytomic analysis, including fluorescence and mass cytometry. The impact of enzyme combinations and dissociation duration on cell viability and diversity were systematically tested, leading to a development of a universal cell preparation protocol that is applicable across multiple human tissue types. Fluorescence and mass cytometry analyses were performed in parallel on the obtained single cells. Immunohistochemistry stains were used to identify cell types within the original tissue, to confirm that the dissociation protocol effectively preserved the cellular diversity. Finally, computational tools, introduced in Chapter 4, were used to characterize the phenotype and cellular diversity of human tissues and

tumors. This work lay a crucial foundation for the discovery of novel glioblastoma cells that stratified patient survival, as described in Chapter 6, and enables the use of fluorescence and mass flow cytometry for the characterization of cellular diversity in other human tissues and tumors.

Chapter 6, entitled “Single-cell signaling profiles reveal clinically and biologically distinct glioblastoma cells”, describes the discovery of novel cell subsets, defined by unique signaling phenotypes and lineage aberrancies that stratified clinical survival of adult glioblastoma patients. Single-cell mass cytometry was used to characterize the signaling profile of individua; glioblastoma cells and was coupled with an automated risk-stratification approach to uncover clinically significant cell subsets. Two novel glioblastoma cell subsets, “Glioblastoma Negative Prognostic” (GNP) cells, and “Glioblastoma Positive Prognostic” (GPP) cells, were discovered and showed contrasting association with patient clinical outcomes. Feature enrichment quantification of the outcome stratifying cell subsets, using Marker Enrichment Modeling, was then used to quantify the novel biological characteristics of the cells, and revealed that the stratifying cell subsets were defined by contrasting signaling profiles. Furthermore, connections between the clinically distinct glioblastoma cells, tumor infiltrating immune cells, and microglial activation phenotype were discovered. The contrasting signaling and protein identity phenotype of GNP cells and GPP cells, as well as the connectedness of glioblastoma cells and the immune cells in the tumor, is a significant advancement towards an establishment of clinical stratification that may dictate individualized therapy, which is discussed in Chapter 7.

Chapter 7 summarizes the tools, analysis approaches, and biological discoveries described in Chapters 4-6, and elaborates on the future directions of this dissertation. Specifically, this chapter emphasizes the evaluation of signaling states as novel targets for therapy. It also explores the idea of using single-cell quantification to dissect the signaling kinetics, as well as the resistance mechanisms of innovative treatments. It also details how single-cell signaling profiles can aid clinical diagnostics and the development of novel surgical approaches in patients with glioblastoma. Moreover, this chapter discusses possible mechanistic relationships between cancer and immune cells in the tumor microenvironment, as well as the benefit of dissecting this connection as an alternative therapeutic approach.

Appendix A describes a step-by-step tissue dissociation protocol for obtaining viable single cells for fluorescence and mass cytometry analysis from human solid tissues and tumors. This protocol is complementary to the technical development described in Chapter 5.

Appendix B describes fluorescence cell barcoding, a method that enhances throughput and maximizes the efficacy of inter-sample comparison by fluorescence flow cytometry. Key considerations for barcode optimization and data analysis are discussed in detail. A step-by-step fluorescence cell barcoding protocol is also provided.

Appendix C describes the use of fluorescence flow cytometry to tracking cell type-specific uptake of novel therapeutic compounds. This work resulted from a collaboration with a previous graduate student, Dr. Katherine Chong, from Dr. Gary Sulikowski's group at Vanderbilt University. Different human cancer cell lines and healthy peripheral blood mononuclear cells were used to compare their differential uptake of fluorescence-tagged apoptolidin, a compound previously shown to preferentially target transformed cells compared to healthy cells. Additionally, immunofluorescence imaging was used in parallel with fluorescence flow cytometry to visualize and confirm the differential drug uptake.

Appendix D describes the application of fluorescence cell barcoding, explained in Appendix B, for a high-throughput drug response assay. This work is a collaboration with a previous graduate student, Dr. David Earl, from Dr. Brian Bachmann's group at Vanderbilt University, and Dr. P. Brent Ferrell Jr., a previous member from Dr. Jonathan Irish's group, now a faculty member at Vanderbilt University. Acute myeloid leukemia patient samples were treated *ex vivo* with crude biological extracts. Samples were then fluorescently barcoded, simultaneously analyzed, leading to the identification of cell type-specific biological impact of the compounds.

Appendix E describes Marker Enrichment Modeling (MEM), a quantitative measure for feature enrichment used for the characterization of novel glioblastoma cells described in Chapter 6. This work was developed by Dr. Kirsten Diggins, a previous graduate student and now a postdoctoral fellow in Dr. Jonathan Irish's lab at Vanderbilt University. This collaboration led to the application of MEM in a glioblastoma mass cytometry dataset, which successfully distinguished tumor infiltrating immune cells from glioblastoma cells using only non-canonical cell identity markers.

CHAPTER 2

The Diversity of Human Malignancies

Introduction

Cellular diversity is a phenomenon that exists in both normal and cancerous tissues (*16*). Eukaryotic organisms are created from single cells that differentiate into cells and organs that carry diverse functions, while remain connected via genetic lineages. A single organ is composed of different cell types that make up the complex architecture, including the vasculature, epithelium, immune cells and connective tissues. The phenotypic differences of genetically identical cells, which are influenced by non-genetic traits, contribute to the intra-organ cellular diversity (*16*). When the programs that control cellular homeostasis are disrupted, cellular imbalance and dysregulated cell growth can occur, and can lead to tumor formation.

Human malignancies exhibit multiple layers of diversity (*17, 18*). Differences in clinical progression and distinct morphological features of cells observed by pathologists has led to the standard categorizations of cancers. Even though the diversity of human malignancies had long been recognized, such as variations in clinical outcomes of patients who had been diagnosed with the same type of cancer, recent advanced technologies have further revealed its extent to be beyond what was previously comprehended. Generally, tumor diversity can be divided into two main categories: inter-tumor diversity and intra-tumor diversity. However, these categories are still relatively broad. The diversity of human malignancies can be further observed spatially and temporally, and even at the single-cell level. Therefore, a thorough understanding of the scope of tumor diversity, as well as its contributing biological features, is essential for the development of clinical diagnostics, as well as for the discovery of novel therapeutics.

The scope of tumor diversity

Diversity between individuals

Diversity of tumor progression and patient survival has been noted in many types of human cancers that, historically, influenced establishment of diagnostic guidelines (*19-21*). The distinct underlying biology of individual tumor cells as well as other patient-specific factors are plausible explanations of the diversity in

therapeutic responses and patient outcomes. Histopathological criteria are historically and routinely used to classify human malignancies into categories that generally, although not perfectly, reflect diverse outcomes and occasionally predict benefits from specific therapies. These distinctions are usually based on microscopic observations of cellular morphology, the architectural arrangements of cells, as well as other unique histological features. Often, immunohistochemistry stains of specific cellular targets are also incorporated into routine diagnostic regimens and disease classifications, which are usually semi-quantitatively assessed by experts. Although clinically informative, many disease categories are distinguished based on arbitrary cut-off values or qualitative histopathologic descriptions. Advanced technologies have enabled large-scale quantitative characterization of human malignancies at genomic, transcriptomic, and proteomic levels, which has revealed extensive diversity between tumors, even among those that were classified as the same disease based on histopathological and immunohistochemical criteria alone (Figure 2-1A) (*2, 3, 22-25*).

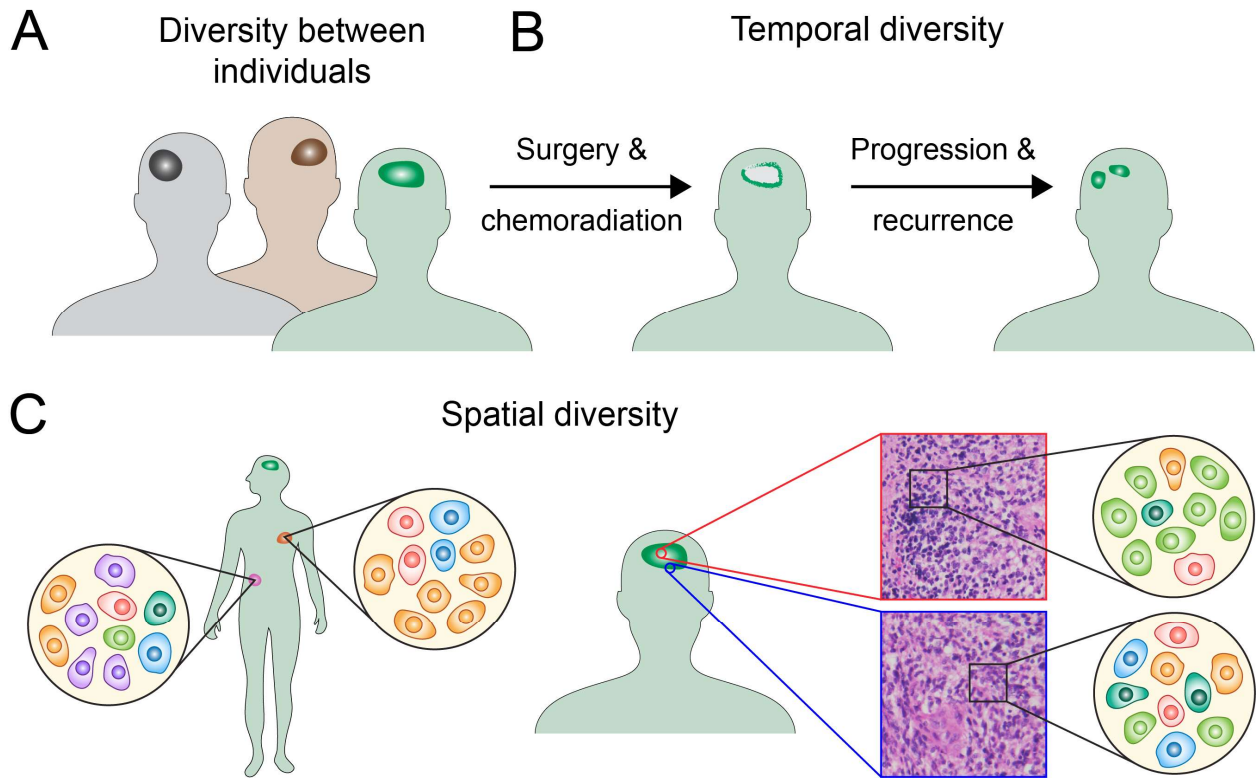


Figure 2-1 The scope of tumor diversity. (A) Tumors from different patients that are categorized as the same disease, based on traditional histopathological criteria, often have molecular differences that result in a diversity between individuals. (B) After treatment, or even as the tumor progress through its natural course of progression, the underlying molecular characteristics of the tumor can change over time. (C) For tumors that can metastasize to distant locations, or tumors that can occupy different sites at the initial diagnosis, the molecular features and cellular composition of tumors from different organs often differ. (D) Even at a single disease site, various cancer cell subsets can occupy different regions of the tumor.

Due to the increasing insights of the molecular alterations that confer diversity between tumors, inter-tumor molecular heterogeneity is now part of the more recent classifications of many types of cancer (*19, 21, 26*). Importantly, many of these classifications direct how patient care should be tailored to be most clinically beneficial based on the tumors' molecular profiles. An example of this is the classification of breast cancer, which is based on the expression hormonal receptors (estrogen and progesterone receptors) and a cell surface receptor (ERBB2, also known as HER2) (*27, 28*). This classification reflects distinct biological mechanisms and predicts clinical responses to different targeted therapies. In general, the goal of cancer classification is to group patients into discrete, well-defined, disease categories. However, quantitative assessments of the molecular lesions in many types of malignancies have revealed that the differences between patients are usually subtle and often display continuous spectra of abnormalities rather than sharply distinct anomalies (*23, 29*). Therefore, approaches that can integrate objective and accurate quantification of features that drive even the subtle differences between different patients are needed to improve cancer stratifications.

Temporal diversity

Despite aggressive therapies, human malignancies occasionally resist treatment and eventually recur (Figure 2-1B). Often, some cancer cells are more susceptible to certain treatment than others and are more likely to be eradicated during the initial course of therapy (*30*). Clinical treatment is a form of external selective pressure that can enrich for cells that are inherently resistant to such therapy. These resistant cells, which can be present at low abundance prior to therapy, can eventually expand and take over the tumor mass (*31*). This process of clonal selection can, therefore, significantly alter the homeostasis within the tumor microenvironment, resulting in a tumor that no longer responds to the initial primary treatment (*32*).

Additionally, exposure to chemotherapeutics and targeted therapies can alter the intrinsic properties of cancer cells. For example, blockade of specific signaling molecules can force the cells to evolve and to re-route their functional network in order to maintain their malignant capacity (*33*). Moreover, cancer cells can acquire new molecular alterations, such as new genetic mutations, that lead to a resistance to the initial therapy (such as the acquisition of T790M mutation in lung adenocarcinoma secondary to inhibition of epidermal growth factor receptor, EGFR) (*34*). Therefore, the mechanisms essential for cancer cell survival can continuously evolve over

time, resulting in a temporal diversity even in an individual patient. Notably, this can occur despite the primary and recurrent tumors being categorized as the same disease based on the standard histopathological criteria. Even without treatment, the cellular makeup within a tumor can evolve as a result of the natural disease progression. An example of this process is the progression of pre-cancerous lesions (such as colonic polyps or pulmonary nodules) to full-blown malignancies (such as colonic adenocarcinoma or lung cancer) (*35*).

Spatial diversity

Tumors that occupy different disease sites or organs can contain uneven distribution of cancer cells (Figure 2-1C, left) (*36, 37*). Most of the previous studies have reported discrepancies between the molecular features of primary tumors and tumors from different metastatic sites (*38-41*). Even in tumors that rarely metastasize beyond the organ of origin, such as malignant brain tumors, tumor samples obtained from different regions within the same tumor mass can have distinct molecular profiles (Figure 2-1C, right) (*42-44*). Hypothetically, the spatial diversity within a single tumor site can be due to cell-intrinsic factors, leading to preferential cell localization, or even extrinsic factors, such as region-specific growth factors, that inevitably impact or alter cellular characteristics and functions. A critical concern that arose from this observation is whether the primary tumor alone, especially a small sample obtained from a single tumor site, can sufficiently serve as a representation for clinical diagnostics and disease monitoring.

Understanding the disparity of the cellular composition of tumors at different sites, as well as those from different regions within the same tumor, is critical for cancer diagnostics, for disease monitoring, as well as for designing treatments that can prevent or overcome disease progression. However, most of the current diagnostic and therapeutic approaches are usually based on the characterization of only the primary tumors and frequently on a small number of cells obtained via biopsy (*17*). Ultimately, studies that are based on matched primary-metastasis samples from the same patients and samples from different regions within the same tumors are key to understanding the spatial diversity of human malignancies.

Single-cell diversity

Multiple subsets of cancer cells can co-exist within individual tumors, even at a single disease site (Figure 2-1C, right) (*5, 14, 15, 44*). These cells are often functionally distinct (*45, 46*), carry contrasting molecular profiles

(*5, 44, 47*), and can respond distinctly to both non-specific and targeted treatments (*30, 48, 49*), leading to therapy resistance or even disease recurrence. In part, the unique combinations of these cancer cells contribute to the inter-tumor diversity described above, as observed via sample-level analyses (i.e. the degree or the average level of a specific molecular alteration changes with the presence, as well as the different proportions, of cancer cell subsets) (*11, 50*).

The presence of single-cell diversity is a challenge for tumor characterization and can significantly interfere with clinical disease stratification that is based on bulk sample evaluation or a small representation of single cells (such as biopsies or needle aspirations). Without high-throughput sensitive single-cell assessments, rare cancer cell subsets can easily be overlooked. This is especially challenging with the current standard use of histopathological diagnosis, especially in solid tumors, which are often semi-quantitative and are expert-dependent. Even though individual cancer cells can be evaluated via morphological and immunohistochemical observations, tissue imaging only captures a small fraction of cells from the tumor. Considering the spatial diversity that can be present within a tumor, assessments of a few hundred or a few thousand cells by histopathological approaches may not adequately capture the clinically-relevant cancer cells, or the significant features in small cell subsets, that may immensely impact patient clinical outcomes.

In addition to cancer cell subsets, tumors are also composed of other non-cancer stromal cells, including normal cells that are specific to the organ of origin (such as neurons in the brain or hepatocytes in the liver), a variety of tumor-infiltrating immune cells, as well as cells forming the vascular structure including endothelial cells and pericytes. Single cell characterization is, therefore, essential to enable accurate distinction between stromal features and cancer-specific traits, which is critical for the identification of novel therapeutic targets.

Factors that contribute to the diversity between individual cancer cells

Genomic diversity

Genomic instability is one of the hallmarks of human malignancies (*8, 51, 52*). Genetic mutations initiate tumor formation by 1) activating oncogenic pathways, and/or 2) inhibiting tumor suppressing mechanisms (*53*). This is followed by subsequent acquisitions of additional genetic alterations secondary to genomic instability. Advanced technological approaches and analyses have revealed that cancers contain cells that differ in their

genomic makeup (Figure 2-2A) (*17, 54*). Since these aberrancies are inherently transmissible, the temporal developmental phylogeny of a given cancer cell can be systematically traced and reconstructed based on its genomic fingerprint (*42, 54*).

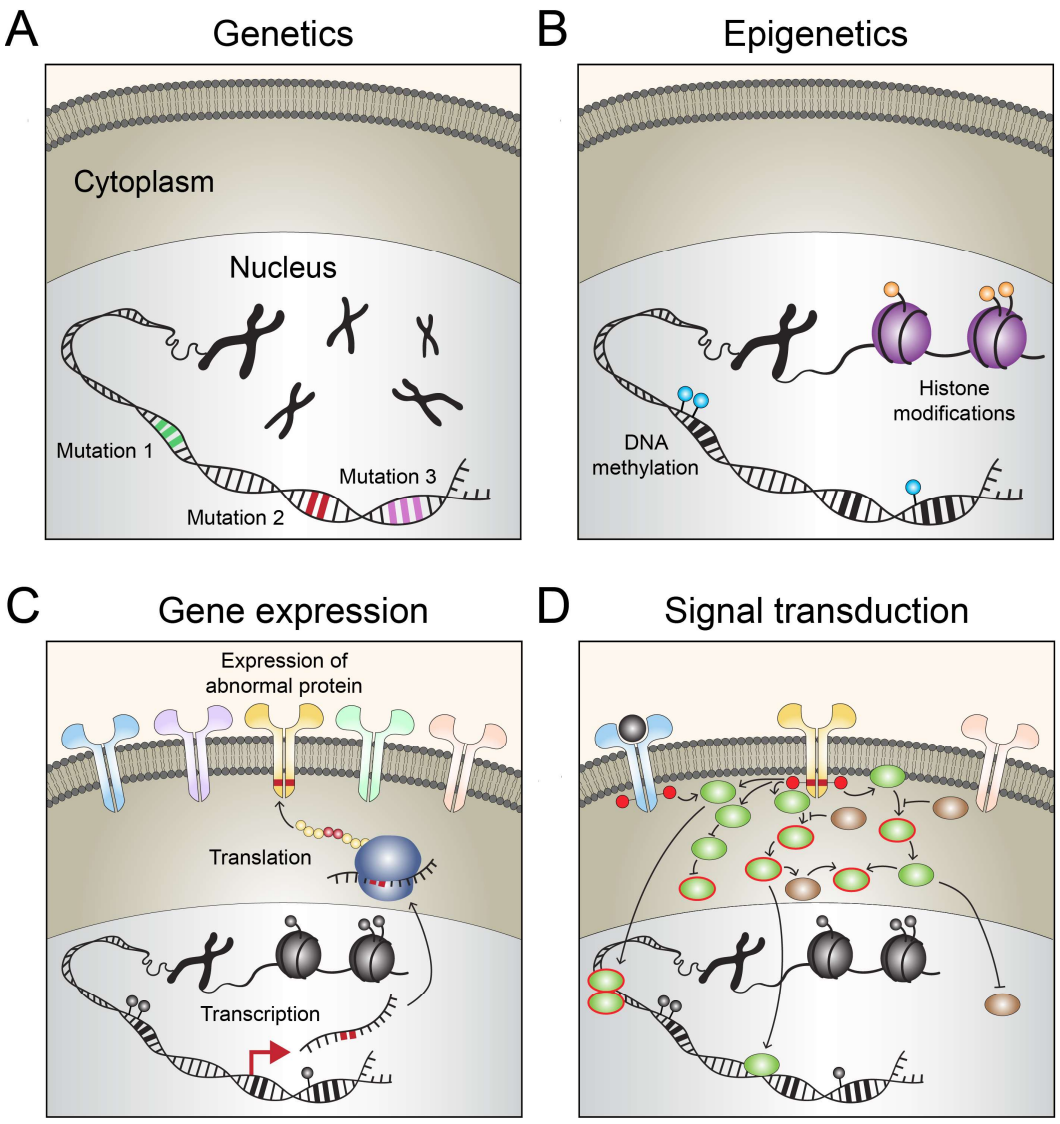


Figure 2-2 Factors that contribute to the diversity between individual cancer cells. Different cellular components contribute to the diversity between individual cancer cells. (A) Genetic changes indicate alterations that occur specifically at the DNA nucleotide level. (B) Epigenetic changes indicate alterations secondary to enzymatic reactions that modify gene expression not resulting from changes of the DNA sequence. Specifically, epigenetic changes include other forms of DNA modifications, as well as histone and chromatin alterations. (C) Gene expression is a multi-step process that eventually leads to a specific cellular phenotype. Here, gene expression is broadly divided into transcription (DNA to RNA), and translation (RNA to protein). (D) Signal transduction indicate interactions between molecules that lead to their biochemical modifications, which can alter their subsequent activity and functions. Signal transduction can be a response to external stimulation, internal cues, and can even occur autonomously in some circumstances such as in the presence of an abnormal protein.

The genomic diversity within an individual tumor variably accumulates with every round of cell proliferation, resulting in the presence of different subsets of cancer cells that can vary at any given time in disease progression. For the last few decades, the characterization of cancer diversity largely focused on dissecting genomic alterations, as well as identifying and targeting the driver mutations (*3, 55-57*). This has proven to be successful in some, but not all types of cancer. Accumulation of additional passenger mutations can possibly impact cellular phenotypes and confer malignant capabilities, which may contribute to therapy resistance (*17, 58*). Moreover, the balance of cellular genetic diversity of human malignancies can also fluctuate throughout the course of treatment secondary to genomic instability and environmental selection.

Cellular diversity is, however, not simply a result of genetic mutation alone. In a normal physiologic environment, the distinction of the different types of cells in a human body is based on non-genomic components since all cells are genetically identical (except for some immune cells). The only way to distinguish the different cell types is to characterize their cellular phenotype (such as protein expression) and functions, which are results of non-genetic influences. Critically, not all genomic changes lead to phenotypic differences, let alone clinically relevant features. Non-genomic alterations in cancer cells, such as epigenetic regulations and influences from the tumor microenvironment, can significantly impact gene expression, phenotype, and subsequent cellular functions. Thus, deciphering the diversity of human malignancies should not only be limited to the understanding of the diverse genetic component of the tumor, but should also encompass other cellular components that dictate clinical outcome and therapy resistance.

Epigenetics

One of the explanations of why genomic information is insufficient to fully capture the degree of cancer diversity is the variation of gene expression that confer subsequent cellular phenotype and function. Epigenetic modifications of DNA, chromatin structures, and histones, are results of reversible enzymatic reactions, which control the process of gene expression (Figure 2-2B). This mechanism is tightly regulated in normal physiologic conditions (*59*). Epigenetic changes can lead to an increase or decrease in gene expression, depending on the types (i.e. methylation, acetylation, etc.) and the targets (i.e. DNA, histones, etc.) of alterations. Mutations of genes encoding epigenomic regulatory enzymes can alter the epigenomic landscape of the cancer cells. Thus,

genetics and epigenetics can be closely connected. Critically, epigenetic regulation of a given gene is not strictly equivalent its genetic mutation (*18*). Disruptions of cancer cell epigenetics can also be secondary to the abnormal communications between cancer cells and other cells, as well as factors in the tumor microenvironment (*17, 59, 60*).

Overall, epigenetics is commonly altered in human malignancies. DNA hypo- and hyper-methylation have both been observed in different types of human malignancies (*61-64*). Alterations of the cancer cell epigenetics can be a consequence of oncogenic transformation. For example, the induction of mutated IDH (isocitrate dehydrogenase) enzyme in human gliomas, a prognostic mutation that is believed to occur very early during oncogenesis, leads to a global hypermethylation phenotype in the cancer cells (*65*). The connection between the mutant IDH and DNA hypermethylation is also observed in glioma patients (*22*). In contrast, epigenetic alterations can be the initiator of oncogenesis as the induction of genome-wide hypomethylation in mice was sufficient for the formation of aggressive T cell lymphomas (*66*). Moreover, DNA methylation profiles are tied to the clinical outcomes of patients with different types of cancer (*22, 67*). Drugs that enzymatically reverse the epigenomic landscape have also been approved for cancer treatments (*68, 69*). Epigenetics, however, can only partially explain gene expression, transcriptional changes, and subsequent cellular functions (*8, 17, 51*).

Gene expression: Transcriptomics and proteomics

A parallel sequencing of genomic DNA (gDNA) and mRNA from the same single cells revealed that a given genetic alteration can lead to variable transcript abundance between individual cells (*70*). This observation had reinforced the notion that genomic information alone cannot sufficiently predict the level of gene expression, and therefore, should not be used as a sole predictor of therapy responsiveness in cancer. Recently, the effort to dissect single-cell diversity in human tissues and cancers has shifted towards the characterization of transcripts, one of the cellular components that reflect gene expression. Particularly, the appealing single-cell RNA-sequencing approach enables a global, unbiased view of the entire transcriptome, without the need to select for specific cellular targets for quantification. As a result, single-cell RNA-sequencing has been used to quantify transcript abundance. Additionally, the abundance of specific transcripts has been used to infer the identity of single cells (Figure 2-2C) (*5, 71-74*).

Proteins are products of the translation of mRNA and are the components of cells that execute many of the key cellular functions (Figure 2-2C). Proteins are traditionally used as the defining features of cell identity. The complex regulatory mechanisms that govern translation result in an imperfect correlation between per-cell level of a given transcript and its protein product. Even in non-disease states, the rates of mRNA and protein turnover substantially impact their levels and, therefore, significantly contribute to their discrepancy (*75, 76*). The rate of translation is another contributing factor to the abundance of a given protein and is tightly regulated throughout development (*75*). However, the rate of translation can be significantly altered in disease states, including in human malignancies, secondary to intrinsic changes (such as genetic mutations) as well as to signaling transduction that responds to internal and external cues (*77*). The common discrepancies between the transcripts and proteins suggest that neither should be used as the sole representative of the other, let alone a representative of the eventual cellular mechanisms and functions. Even though proteins are considered the functioning component of cells, it is important to note that although quantification of total protein expression may indicate the cell's "potential" mechanism, it still does not indicate the actual function of the protein at any given state.

Post-translational modification of proteins: The signal transduction

Signal transduction is a mechanism by which the molecules in the cell biochemically interact, either autonomously or in response to external and internal cues, to execute cellular functions (*9, 78*). The connections between these effector molecules are often mapped as signaling cascades or pathways, showing their potential interactions, which are inter-connected to form a complex signaling network (Figure 2-2D). A variety of genetic mutations can subsequently lead to alterations of a similar set of signaling effectors. Therefore, core signaling molecules can be viewed as the nexus or the functional regulators of the cells (*79*).

The complex signaling network enables the cells to adapt under various pressure or stimulating conditions (*33*). Quantification of post-translational modifications (i.e. phosphorylation, acetylation, etc.) of signaling effectors provides a "snapshot" of the ongoing activity within a cell, under specific circumstances (*10, 12-15, 80-83*). A parallel evaluation of total protein expression and signaling activity in acute myeloid leukemia revealed that signaling profiles in hematopoietic, leukemic, progenitors remained consistent despite the

irregularity of the cell phenotypes determined by conventional surface protein molecules (*12*). Moreover, potentiated signaling states strongly correlated with patient outcome as well as other clinically relevant biological traits (*12, 15, 80*). Even the unstimulated, basal, signaling states alone effectively stratified the clinical outcome of patients with hematologic malignancies (*13*).

Post-translational modifications of signaling molecules, both surface receptors and intracellular signaling effectors, can be easily detected by antibody-based techniques (i.e. using an antibody that can detect phosphorylation at a specific amino acid residue on a protein) (*80, 84, 85*). By coupling this detection with high-throughput, flow cytometry, the signaling states can be mapped at the single-cell level, reflecting the mechanistic information of millions of cells at a given point in time (*50*). Importantly, the integration of protein identity assessments can also distinguish between cancer and stromal features. This enables simultaneous dissection of the cellular activity and relevant functions, and to monitor the targetable features that are specific to cancer cells. This approach would immensely aid the selection of drugs and compounds that can interfere with the activation states of the cancer cells, which significantly impact patient outcomes.

Towards quantitative single-cell analysis in solid tumors

Cellular diversity is a critical challenge towards successful cancer diagnostics and treatment. Previous studies have indicated that individual cells can distinctly respond to therapies and, therefore, can differentially confer to clinical resistance (*30, 32*). Critically, the detection of clinically unfavorable cancer cells can facilitate disease monitoring and clinical stratification of hematologic malignancies (*86*). Considering the degree of cellular diversity that is appreciated in the different types of cancer, the benefit of single-cell quantification and monitoring in other types of human malignancies should be thoroughly investigated.

With the exception of hematologic malignancies, quantification and tracking of single cells are often not part of the current routine clinical diagnosis and stratification of cancers due to two main challenges. First, sample preparation protocols and quantitative analytical procedures for the characterization of single cells from non-hematologic (i.e. solid) cancers are not widely available and/or are not standardized. Second, expert consensus on the cellular parameters that can couple clinical outcome prediction and identification of molecular lesions that can be targeted to improve patient prognosis is often lacking. Thus, the immediate task is to develop a

standardized, quantitative, single-cell analysis approach that 1) can be easily adopted in many types of solid tumors, and 2) enables automated discovery of clinical stratifiers that can be tested as therapeutic targets.

Ideally, a high-throughput approach that would enable characterization of a large number of cancer cells should be used, to maximize sample representation. Flow cytometry can quantify features of millions of cells in a very short period of time (Chapter 4). In addition, fluorescence flow cytometry is already routinely implemented for the analysis of blood and bone marrow samples, making it an ideal platform to be immediately tested for clinical solid tumor analysis. The key challenge is to identify a standard protocol that can derive single cells from patient sample that would preserve the cellular diversity present in the original tissue, enabling the identification of clinically meaningful cell types (Chapter 5).

Signaling quantification of single cancer cells would introduce a novel approach to define cellular identity based on functional characteristics. Risk-stratification analysis can further reveal the connections between the functionally distinct cancer cells and clinical outcomes (tumor progression and patient survival; Chapter 6). Notably, the association of a given signaling state with patient outcomes would indicate its role as a prognostic biomarker; whether it should be used as a target for therapy must be further elucidated via experimentations and clinical trials (Chapter 7). Importantly, using single-cell signaling profiles as surrogates for identifying novel therapeutic targets is highly advantageous. First, different layers of cellular alterations (i.e. genetic mutations, epigenetic changes, etc.) often converge through signaling effector modifications, the critical mechanistic converging point within a cell that leads to the execution of many cellular functions (*79, 87*); it would be more efficient to target one or few signaling effectors as opposed to attempting to modify thousands of redundant upstream features. Second, since most drugs and small molecules act against proteins, signaling effectors can be easily tested as direct targets for novel therapeutics. Taken together, the characterization of signaling states in single cancer cells would potentially reveal the key regulators or the mechanistic bottleneck that govern the cancer hallmark-promoting cellular functions (*8, 51*), which can potentially be targeted to improve patient outcomes.

CHAPTER 3

Adult Glioblastoma

Introduction

Glioblastoma is a grade IV and the most aggressive form of human gliomas (*21*). Glioblastoma comprises 46.6% of all the adult malignant brain tumors that arise primarily within the brain parenchyma. The incidence of glioblastoma increases with age, with the median age at initial diagnosis being 64 years (*88*). Glioblastoma has a slightly higher incidence in males compared to females (3.2 vs 2.5 cases per 100,000 populations), in the United States (*88*). The median overall survival of glioblastoma patients is slightly less than 15 months after initial diagnosis, despite a very aggressive standard therapeutic regimen that includes maximal, safe surgical resection, radiation therapy, and administration of a highly toxic chemotherapeutic agent, temozolomide (*1*). The exact etiology of glioblastoma remains unknown and most cases occur sporadically. However, certain familial cancer syndromes that carry genetic alterations of signaling-related genes are associated with an increased incidence of glioblastoma (*89*). Examples of these syndromes include Li-Fraumeni syndrome (germline *TP53* mutation) (*90*), neurofibromatosis types 1 and 2 (germline *NF1* and *NF2* mutation, respectively) (*91, 92*), and Turcot's syndrome (germline *APC* mutation) (*93*). However, these hereditary cases are relatively rare compared to glioblastomas that arise *de novo*.

Patients with glioblastoma often present with various neurological symptoms (*94*). These can either be 1) generalized neurological deficits secondary to an increase in intracranial pressure (resulting from mass occupying lesions and/or an obstruction of the ventricular system), or 2) localized neurological symptoms, which depend largely on the anatomical location of the tumor within the brain. Adult glioblastomas often arise in the cerebral cortex, in contrast to gliomas in the pediatric population which tend to occur in midline locations, such as the pons (*25*). Gadolinium-enhanced magnetic resonance imaging (MRI) is the common imaging modality that is initially used when a brain space-occupying lesion is suspected clinically. This is to pinpoint the anatomical location of the mass or tumor and to gauge its extensiveness for planning a surgical approach (*95, 96*). Although not sufficient to establish a final diagnosis, the radiologic appearances of the mass can often suggest a differential diagnosis, or a tentative diagnosis, which is immensely crucial for pre-operative evaluation.

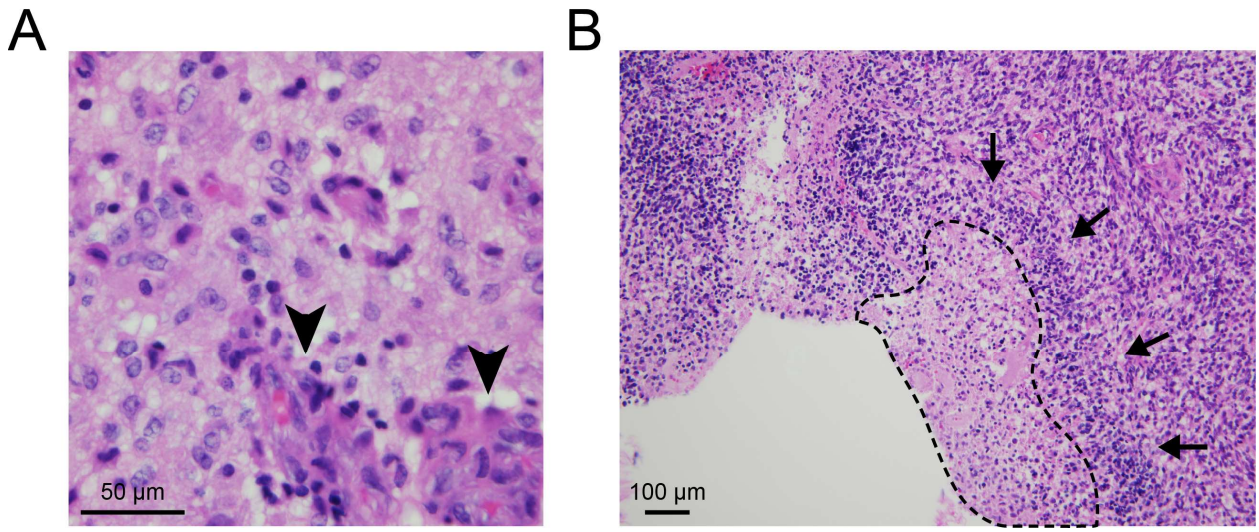


Figure 3-1 Histopathological findings of glioblastomas. (A) Abnormal vascular proliferation is one of the histopathological features of glioblastoma (arrowheads). (B) Palisading necrosis is occasionally observed in glioblastoma and is defined an area of necrosis (dashed outline) surrounded by abnormal tumor cells with nuclei that are organized perpendicularly to the necrotic region (black arrows). Images were taken by Dr. Bret C. Mobley, Vanderbilt University Medical Center.

The standard diagnosis of glioblastoma is based on histopathological findings (*21, 97*). Tissue samples can be obtained via stereotactic biopsies, open biopsies, or surgical resections. The modality to obtain tissue is generally selected based on the patient's clinical performance (whether the patient can withstand aggressive surgery), which is often assessed by the Karnofsky Performance Status (KPS) (*98*), as well as the anatomical location and the extent of the tumor itself (whether the tumor is located at a location that is safe for surgical removal). Glioblastomas are characteristically composed of poorly differentiated cells that can morphologically resemble brain resident glial cells. These cells demonstrate nuclear atypia, cellular pleomorphism, and mitotic activity (*99*). Importantly, the histological features that discern glioblastomas from lower-grade diffuse gliomas are the presence of necrosis and/or abnormal microvascular proliferation (Figure 3-1A). Although not always present, palisading necrosis is a pathognomonic feature of glioblastomas, and is characterized by a necrotic region surrounded by an accumulation of cancer cells whose nuclei are arranged perpendicularly to the necrotic border, in a parallel fashion (Figure 3-1B). These tumors were historically named "glioblastoma multiforme", which reflected the diverse morphological appearances of cells within individual tumors; even by cell morphology alone, the cellular heterogeneity of glioblastomas can be greatly appreciated (*100*). Another fundamental feature of glioblastomas is their highly invasive and diffuse growth pattern. Unlike many other types of malignancies, the

tumor borders of glioblastomas are usually not well-defined radiologically or even microscopically. Based on histopathological findings, pleomorphic abnormal cells can be seen invading through the seemingly normal brain parenchyma, often along the white matter tracts. Not surprisingly, glioblastomas almost always cannot be completely surgically resected, and inevitably, eventually recur.

2016 WHO classification of gliomas: Inclusion of inter-tumor molecular diversity in the diagnostic scheme

The 2016 World Health Organization (WHO) classification of diffused gliomas requires that assessment of genetic and cytogenetic abnormalities be part of the routine clinical diagnosis and disease sub-classification (*99*). This is especially important for lower-grade diffused gliomas where the distinction between astrocytic and oligodendroglial tumors by histopathological appearances alone can be challenging. Previously, this distinction was based primarily on the morphological appearances of cells (*101*); astrocytic tumors contained cells with more “angular” nuclei, whereas oligodendroglial tumors contained cells with nuclei that appeared round (“fried egg” appearance) and often contained “chicken-wire” (angulated and branched) vascular structures. There was also an intermediate entity, oligoastrocytic tumors, that appeared to have mixed features between the two categories. This latter category, however, no longer exist in the latest standard classification scheme. These morphologic descriptions were found to be highly subjective and were highly variable even among experts (*102*). Presently, morphological information is not sufficient to distinguish between lower-grade astrocytic and oligodendroglial tumors.

The molecular alterations that are part of the diagnostic protocol are molecules that are closely associated with the signaling network that enables the Cancer Hallmarks (*8, 51*). Loss or mutations of *ATRX* gene (alpha-thalassemia X-linked intellectual disability), often accompanied by *TP53* mutations, are characteristic of astrocytic tumors (although not required for the diagnosis), whereas 1p/19q co-deletion establishes the diagnosis of oligodendroglial tumors. Mechanistically, altered ATRX 1) impairs cellular DNA repair mechanisms and 2) lengthens telomeres via the ALT (alternative lengthening of telomeres) pathway, leading to an accumulation of genetic mutations in the immortal cancer cells (*103-105*). In contrast, patients with 1p/19q co-deleted oligodendroglial tumors have overall better clinical prognosis than those with astrocytic tumors (*106*). These

molecular alterations, however, are not as prevalent in glioblastomas compared to lower-grade diffused gliomas (*104*).

The critical molecular component that is required for the diagnosis of all grades of diffuse gliomas, including glioblastomas, is the determination of the mutational status of *IDH1/2* gene (encoding isocitrate dehydrogenase 1/2) (*99*). This distinguishes patients who have significantly better prognosis (*IDH*-mutant) from those with inherently worse outcome (*IDH*-wildtype). *IDH*-wildtype glioblastomas should only be defined after assessments of *IDH1* codon 132 and *IDH2* codon 172, the most common forms of *IDH* mutation (*99*). And since *IDH1* R132H is significantly more common than mutations in *IDH2*, many institutions only rely on the assessment of the former, especially in older patients where *IDH* mutations are uncommon (*21*).

MGMT (O^6-methylguanine-DNA-methyltransferase) promoter methylation is a predictive biomarker for response to alkylating chemotherapeutics (*107*). Currently, besides the assessment of *MGMT* promoter methylation, there are no other molecular biomarkers that can guide therapy decisions. However, it is important to note that 1) determination of *MGMT* promoter methylation status is not required to establish a diagnosis of gliomas or glioblastomas according to the most recent WHO guideline, and 2) its status only aids the decision to administer the non-targeted chemotherapeutic agent, temozolomide (TMZ), in patients who are more likely to be susceptible to its cytotoxic effects (e.g. elderly, clinically unstable patients). To date, there are no approved clinical biomarkers that can guide individualization of targeted therapies based on the underlying pathogenesis of individual tumors. Moreover, none of the current, clinically approved targeted therapies have consistently or significantly prolonged the progression or survival of patients with glioblastoma beyond a few months compared to baseline (*108-110*).

Current states of glioblastoma therapy

Surgery, radiation, and chemotherapy are the current standard therapies for glioblastoma patients and are often used in combination. The standard glioblastoma treatment regimen that includes temozolomide, the chemotherapeutic agent that has shown the greatest survival benefit for glioblastoma patients to date, prolongs the median overall survival of glioblastoma patients by approximately 2.5 months (*1*). However, the median overall survival of glioblastoma patients is still only 14.6 months, and the two-year survival rate is only 26.5

percent (*1*). Moreover, patient 5-year survival rarely exceeds 5 percent (*111*). Although younger patients tend to have a better prognosis that older patients (*112*), this highly aggressive brain tumor is almost always fatal.

Surgery and radiation

For surgically assessible glioblastomas, the initial treatment is maximal, safe tumor resection. The tissue for histopathological diagnosis is also obtained at the time of surgery. Glioblastomas are locally invasive and tend to rapidly invade along the white matter tracts in the brain. However, the tumors can also involve cortical and deep gray structures. Although part of the tumors may be surgically resectable, these aggressive tumors usually have already infiltrated a large portion of the brain at the time of initial diagnosis. Data have shown that gross total resection of the tumor, based on post-operative MRI evaluation of imaging enhancement, significantly prolongs the overall survival of glioblastoma patients, compared to sub-total surgical resection (*113, 114*). However, post-operative neurological functions must be sufficiently preserved when planning the surgical approach; this can often significantly limit the extent of surgery. Intra-operative radiologic evaluations (computed tomography (CT) or MRI scanners) can aid the assessment of the anatomic location and the extent of safe surgical resection of tumors in real-time. However, it is important to note that microscopic invasions of cancer cells are often present despite the lack of radiologic enhancement as assessed either in real-time or post-operatively, making surgery extremely challenging.

Newer surgical approaches have enabled rapid or even real-time assessments and visualization of tumor margins. The administration of 5-aminolevulinic acid (5-ALA) to the patients pre-operatively results in a higher sensitivity in identifying malignant tissues at the invading margins of glioblastomas compared to conventional gadolinium-based MRI. Unlike gadolinium, 5-ALA is not dependent on the disruption of blood-brain barrier, but rather the intrinsically accelerated metabolic rate of cancer cells. The fluorescent porphyrin, a metabolite of 5-ALA, is more rapidly accumulated in the glioblastoma cells compared to non-malignant cells and can be visualized intraoperatively using a fluorescence operating microscope (*115*). This method appears to improve the progression-free survival of glioblastoma patients (*116*). However, this technique can lead to false-positive detection of fluorescence signals in some non-malignant cells, such as reactive astrocytes, since it is dependent upon the differential cellular metabolic rate (*117*) Additionally, many factors can also attribute to false negative

signals, such as timing from 5-ALA administration to surgery, photobleaching, necrosis, as well as structural obstructions that prohibit fluorescence visualization (*118*).

Mass spectrometric quantification of 2-HG (2-hydroxyglutarate, a by-product of mutant IDH enzyme) is an evaluation of a cancer-specific metabolite, which enables the distinction between infiltrative tumor cells and the invaded brain structures (*119*). This can be evaluated on either sampled tissue (performed within the operating room), or in real-time using a suctioning scalpel attached to a mass spectrometer (*120*). The latter approach is an appealing platform since it utilizes the vapor or byproduct of the surgery, which is normally discarded, eliminating the need for the collection of additional valuable tissue for pathological evaluation. Furthermore, it might have a role in the treatment of tumors that are inaccessible by conventional surgical approaches. However, mass spectrometric assessment heavily relies on quantification of metabolites that are known to be specifically present in malignant tissues and not in the normal adjacent brain. Since 2-HG is only produced in *IDH*-mutant cells, its detection is only beneficial in a small subset (less than 5-10%) of adult patients with glioblastomas (*26*). Notably, *IDH*-wildtype cancer cells that may co-exist within seemingly *IDH*-mutant tumors could potentially be overlooked. Therefore, for real-time mass spectrometry to be more widely applicable for the majority of glioblastoma patients, the identification of cancer-specific protein targets in *IDH*-wildtype glioblastomas, specifically in *IDH*-wildtype glioblastoma cancer cells, is critical.

Following surgery, radiation is administered as an adjuvant therapy. Whole-brain radiation following surgery substantially improves patient overall survival (*116, 121*). Plan for radiation is often assessed after surgery and generally includes both the gadolinium-enhanced regions and the adjacent, radiographically non-enhancing brain parenchyma. The goal of radiation is to prevent tumor recurrence, which most often occurs locally. Regular radiographic assessment is also needed after the initial treatment for early detection of tumor recurrence, should it occur. However, it is important to distinguish between true tumor recurrence and radiographic pseudo-progression (*122*). The latter is more likely if the new radiographic enhancement occurs during the first few months after completion of radiation treatment, especially if the patient also receives concurrent temozolomide (*123*). Other key distinguishing factors that would favor pseudo-progression include the lack of new symptomatic neurological deficits (*124*), and presence of *MGMT* promoter methylation (*123*).

More recently, the use of tumor-treating fields (TTFields) as an adjunct, maintenance therapy after the initial surgery and radiation was introduced (*125, 126*). TTFields are portable devices that deliver low-intensity, alternating electric fields via transducer arrays that are applied to the scalp. The alternating electric field inhibits the microtubules and disrupts the mitotic spindles, leading to abnormal chromosomal segregation and eventual death of cancer cells via caspase-dependent apoptosis (*127, 128*). Thus, the efficacy of TTFields is dependent on cell division rate with the more rapidly dividing cells being more susceptible to the treatment. The side effects of TTFields were shown to be very mild, the most common being local skin erythema and irritation (*126, 129*). Although TTFields are now approved for used as adjunct therapy in patients with glioblastoma, its estimated cost of $21,000 per month had raised concerns for its cost-effectiveness, especially without the proper price regulation (*130*). Moreover, it requires the patients to use the device on a shaved-scalp, and to continuously carry the device for at least 18 hours per day for up to 24 months, which can be somewhat inconvenient and can significantly affect their quality of life (*126*).

Temozolomide

Prior to the incorporation of TMZ into the standard clinical management of glioblastomas, the key therapeutic modality for the clinical management of glioblastomas involved maximal, safe, surgical resection, followed by radiation therapy and administration of nitrosourea-based chemotherapeutic agents. This had improved the survival of glioblastoma patients to some extent, at least compared to surgery alone. However, the median overall survival of patients was still less than a year after initial diagnosis (*131, 132*).

TMZ has been the most important advancement in glioblastoma therapy during the past decade. After administration, TMZ is metabolized to an active compound, 5-(3-methyltriazen-1-yl) imidazole-4-carboxamide (MTIC), which enzymatically adds alkyl groups to either N^7 or O^6 positions of DNA guanine residues, leading to an induction of DNA damage and subsequent cell death (*133*). Post-surgical administration of TMZ plus radiation prolongs both the progression-free survival and overall survival of glioblastoma patients by 1.9 months and 2.5 months, respectively, compared to radiation alone (*1*). Thus, TMZ is included in the standard treatment whenever patients are evaluated to be clinically stable enough to withstand some of the common side effects of the drug, which include thrombocytopenia and leukopenia (*134*).

The methylation status of *MGMT* promotor predicts clinical responsiveness of patients to TMZ therapy (*107*). Patients whose tumors harbor detectable methylated *MGMT* promotor (about 50% of patients of primary adult glioblastomas) respond slightly better to temozolomide therapy, compared to those with unmethylated *MGMT* promotor, regardless of whether the patients receive radiation. Mechanistically, methylation of *MGMT* promoter blocks its transcription, leading to a decreased level of MGMT enzyme. This leaves the alkylated DNA, produced by TMZ, to remain unperturbed, which subsequently induces cell death (*135*). Therefore, although it is not part of the standard WHO classification, *MGMT* promoter methylation status is often assessed clinically, specifically to determine if the benefit from TMZ would outweigh its potential adverse effects, especially in patients with poor clinical performances. However, this biomarker does not offer any other benefit beyond the decision to administer a non-targeted and highly toxic chemotherapy. Additionally, TMZ only prolongs the overall survival of patients with methylated *MGMT* promotor by only a few months (*107*).

Signaling alterations in glioblastoma cells sustain cancer hallmarks

Receptor tyrosine kinase signaling

Glioblastomas harbor diverse alterations of signaling molecules that regulate and enable one or more Cancer Hallmarks (Figure 3-2) (*3, 7, 8, 22*). RTK (receptor tyrosine kinase), p53, and RB (retinoblastoma) pathways were demonstrated to be the most commonly altered signaling pathways in glioblastomas, based on studies by The Cancer Genome Atlas (TCGA) (*3, 136*). Bulk genomic and transcriptomic profiles of glioblastoma were used to classify this deadly brain tumor into four molecular subclasses, namely 1) Proneural, 2) Classical, 3) Mesenchymal, and 4) Neural (*2*). Alterations of genes encoding RTKs and downstream signaling effectors were proposed to be definitional of the different subclasses (*PDGFRA* alterations = proneural; *EGFR* alterations = classical; *NF1* loss or mutation = mesenchymal) (*2*). Even though these analyses revealed that the overall inter-tumor diversity of glioblastomas is due to alterations of signaling molecules, these subclasses are not predictive of clinically distinct patient outcomes. Moreover, subsequent single-cell transcript analysis revealed that the molecular subclass diversity is present even between individual cancer cells in a single tumor (*5*).

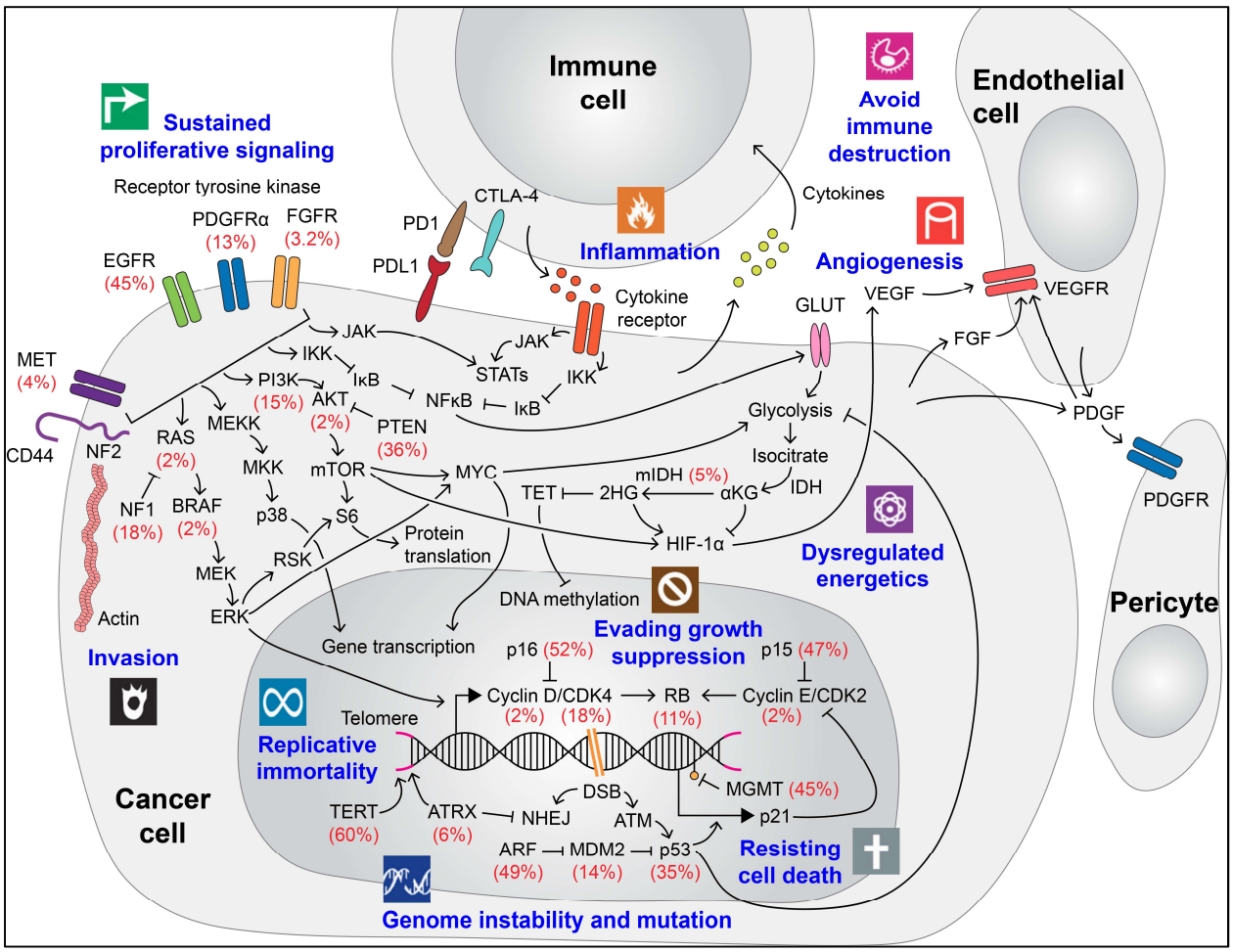


Figure 3-2 Glioblastoma is characterized by alterations of signaling molecules that regulate cancer hallmarks. Signaling network associated with the classic, enabling, and emerging cancer hallmarks (blue text) is portrayed in the context of glioblastoma, focusing on alterations in cancer cells. The incidence of genetic alterations of individual signaling molecules in adult glioblastomas, based on data from The Cancer Genome Atlas (TCGA), is specified (red text). Abbreviations: 2-HG, 2-hydroxyglutarate; αKG, α-ketoglutarate; ATM, ataxia-telangiectasia mutated; ATRX, Alpha-thalassemia X-linked intellectual disability; CDK; cyclin-dependent kinase; CTLA-4, cytotoxic T-lymphocyte-associated protein 4; DSB, double stranded break; EGFR, epidermal growth factor receptor; ERK, extracellular signal-regulated kinase; FGF, fibroblast growth factor; FGFR, FGF receptor; GLUT, glucose transporter; HIF-1α, Hypoxia-inducible factor 1-alpha; IDH, isocitrate dehydrogenase; IκB, nuclear factor of kappa light polypeptide gene enhancer in B-cells inhibitor; IKK, IκB kinase; JAK, Janus kinase; MDM2, mouse double minute 2 homolog; MEKK, mitogen-activated protein (MAP) kinase kinase kinase; MGMT, O^6-methylguanine-DNA-methyltransferase; mIDH, mutant IDH; MKK, MAP kinase kinase; mTOR, mammalian target of rapamycin; NF1, neurofibromin; NF2, merlin; NFκB, nuclear factor kappa-light-chain-enhancer of activated B cells; NEHJ, non-homologous end-joining; PD1, programmed death 1; PDGF, platelet-derived growth factor; PDGFR, PDGF receptor; PDL1, PD-ligand 1; PI3K, phosphoinositide 3-kinase; PTEN, phosphatase and tensin homolog; RB, retinoblastoma; RSK, ribosomal S6 kinase; STAT, signal transducer and activator of transcription; TERT, telomerase reverse transcriptase; TET, ten-eleven translocation; VEGF, vascular endothelial growth factor; VEGFR, VEGF receptor.

Even though they are often thought of as distinct cascades, these pathways are inter-connected by functionally versatile effector molecules that biochemically interact, leading to a formation of a highly complex network (*9*). Therefore, alteration of a single molecule can disrupt the functional output of multiple linked cascades. Alterations of *EGFR* (epidermal growth factor receptor) gene are found in approximately half of adult glioblastomas and are the most common oncogenic lesions in these aggressive tumors (Figure 3-2; Sustained proliferative signaling) (*3*). In most cases, amplification of *EGFR* is also accompanied by its genetic mutations, which can occur as point mutations, in-frame deletions, or fusions of *EGFR* with other genes (*3*). *EGFRvIII*, an *EGFR* mutation that involves an in-frame deletion of exons 2-7, which encode the extracellular domain of the receptor (the ligand binding site), is among its most prevalent mutations. EGFRvIII protein is constitutively active and can initiate autonomous activation of downstream signaling molecules in the absence of extracellular ligand binding and stimulation (sustained proliferative signaling) (*137, 138*). EGFRvIII is also present in approximately half of glioblastomas with amplified *EGFR (139)*. Experiments in animal models suggested that the interaction and dimerization between EGFR and EGFRvIII induced nuclear translocation of EGFRvIII, as well as subsequent increase in STAT3 and STAT5 phosphorylation, which eventually led to an increased in malignant transformation capabilities (*140*).

Mutations of other RTKs in adult glioblastomas, although present, are not as common as *EGFR* alterations. Among these are *PDGFRA* (platelet-derived growth factor receptor, alpha), *MET*, and *FGFR* (fibroblast growth factor receptor) alterations, which are most often accompanied by altered *EGFR* (*3, 136*). Only 7% of adult glioblastomas have amplifications or mutations of one or more non-*EGFR* RTKs without *EGFR* aberrancies (*3*). Activation of different RTKs can lead to activations of similar signaling effectors due to the redundancy of signaling pathways, allowing the cells to maintain functions that sustain their malignant properties via multiple routes (*141, 142*). This has been shown in *ex vivo* experiments that blockade of multiple signaling nodes was required for inhibiting RTK activation (*143*). Moreover, about a third of glioblastoma patients do not have detectable genetic alterations of RTKs at all (*3*), suggesting that alternative oncogenic mechanisms must exist to maintain the malignant properties of glioblastoma cancer cells.

Many of the previously proposed innovative therapies in glioblastoma have focused on targeting the aberrant surface receptors, particularly RTKs, since they are among the most prevalent genetic alterations

observed in adult glioblastomas (*108, 110, 144*). Moreover, disruptions of signaling activities that are connected to RTK potentiation can lead to functional dysregulations, allowing maintenance of cellular malignant characteristics, such as sustained cell proliferation, tumor invasion, and evading growth suppression (*3, 7, 8, 136, 145*). However, drugs that inhibit the enzymatic functions of these receptors, often tested as monotherapies, have shown very limited clinical success (*109, 110, 146, 147*). Several mechanisms have been hypothesized to be the cause of the failures of these seemingly logical innovative therapies, such as receptor co-activation, signaling network re-wiring, and selection of cancer cell subclones that are inherently resistant to such treatments (*44, 48, 142*).

Although surface signaling receptors have been the focus for genetic and transcriptomic characterization, "downstream" signaling effectors are also commonly altered in adult glioblastomas. Mutations of *PTEN* and genes encoding PI3K (phosphoinositide 3-kinase), the components of the AKT signaling pathway downstream to RTK activation, are the most common altered signaling effectors in adult glioblastomas as their combined aberrancies were seen in approximately 50% of patients (*3*). PI3K/AKT activations are associated with regulation of cell proliferation and protein translation (*148*). Increased activation of PI3K/AKT pathway independent of the activation state of the surface receptors can be a result of defective signaling inhibition (such as *PTEN* mutation) or constitutive AKT activation (such as oncogenic PI3K alterations) (Figure 3-2). These changes eventually result in subsequent uncontrolled cell growth observed in glioblastomas.

AKT signaling also plays a major role in cell metabolism, one of the Cancer Hallmarks that is commonly disrupted in human malignancies (*51*). For glioblastoma, dysregulated metabolism has been proposed to be one of its key pathogenesis since the presence of necrosis is one of its pathologic diagnostic criteria. Specifically, hypoxia can induce AKT phosphorylation, which then inactivates the pyruvate dehydrogenase complex leading to an increase in glycolytic metabolism (*149*) (Figure 3-2; Dysregulated energetics). Disruption of AKT signaling can diminish glucose uptake and the subsequent production of lactic acid in glioblastoma cells (*150*). Other mechanisms by which AKT activation increases glycolysis include localizing glucose transporters (GLUT) to the cell membrane to increase glucose uptake, increasing the activity of hexokinase (HK), as well as, activating the glycolytic rate-limiting enzyme phospho-fructokinase-1 (PFK-1) (*151-156*). Increased glycolysis is also associated with poor overall survival in patients with lower-grade gliomas and glioblastomas (*149*). Additionally,

adult glioblastoma is one the human malignancies with the highest level of AKT activation, suggesting that PI3K/AKT signaling has significant biological importance even though mutations of the gene encoding AKT itself are rarely observed in glioblastoma (*157*).

PIK3/AKT signaling is connected to other pathways including MAPK/ERK pathway, as both can activate common signaling effectors, such as S6K (*158*). Inhibition of either pathway can re-wire the signaling network, leading to an induction of the activity of the other pathway (*159*). In contrast to PI3K/AKT signaling, alterations in the MAPK/ERK signaling pathway only occur in approximately 10-12% of adult glioblastoma patients (*3*). Most of the alterations occur in *NF1* gene, which encodes neurofibromatosis 1 (NF1) protein, a negative regulator of RAS (*160*). Deletion of *NF1* is one of the key molecular changes that were proposed to be characteristic for the mesenchymal molecular subclass of adult glioblastomas (*2*). Alterations of RAS and RAF (almost exclusively BRAF), however, were seen in only 1-2% of adult glioblastomas. *BRAF* V600E mutation occurs mainly in pediatric gliomas (*161*). MAPK/ERK signaling can also regulate cell metabolism by increasing the ATP production of the citric acid cycle (*162*).

Dysregulation of cell cycle and cellular mortality

The pathogenesis of glioblastomas is closely associated with disrupted cell cycle regulatory mechanism. Approximately 35% of adult glioblastomas harbor a mutation in *TP53* gene (*3, 136*), leading to the impairment of the cellular DNA repair mechanisms (Figure 3-2; Resisting cell death, and genome instability and mutation). Together, this results in dysregulated proliferation of cells that bear genomic instability (*163, 164*). However, *TP53* mutations are more prevalent in, and are characteristic of, lower-grade diffuse astrocytomas (although they are not required to establish a diagnosis) compared to glioblastomas (*21*). *TP53* mutations are also found more commonly in secondary glioblastomas (those that have progressed from lower-grade tumors) compared to primary glioblastomas (*165*). This indicates a continued progression of cancer cells from pre-existing lesions. Genetic amplification of *MDM2* (mouse double minute 2 homolog), a gene encoding the negative regulator of p53, can also lead to a loss of p53 function (*163, 166*). This is more commonly the case for primary glioblastomas, and often without the genetic mutation of *TP53* itself (*3*). The resulting over expression of MDM2 enables the cancer cell to escape the tightly regulated cell cycle.

Other common mutations in the cell cycle pathways include changes in *CDKN2A* and *CDKN2B* genes, which encode p16 and p15, respectively (Figure 3-2; Evading growth suppression) (*3, 136*). The physiologic roles of these proteins are to control the transition from G1 to S phase and to halt the cell cycle in G1. p16 and p15 are traditionally considered to be part of the RB pathway, one of the major disrupted signaling pathway in glioblastomas. Mutations of *CDKN2A* and/or *CDKN2B* are present in approximately half of *de novo* tumors (*3*). *CDK4* which encodes cyclin dependent kinase 4, a positive G1 regulator, is also commonly amplified resulting in accelerated cell cycle and proliferation (*167*). While these proteins mechanistically affect the function of RB protein, *RB* mutation itself is not as common in glioblastoma (*168*).

In contrast, mutations in the promoter region of *TERT* gene are common in primary *IDH*-wildtype glioblastomas. However, they are less common in secondary tumors (*169, 170*) (Figure 3-2; Replicative immortality). *TERT* encodes telomere reverse transcriptase enzyme, the catalytic subunit of the telomerase complex (*171*). TERT catalyzes the addition of nucleotides to the ends of chromosomal telomeres, a process that prevents overt degradation of the ends of chromosomes that can occur due to repeated cell replications. *TERT* promoter mutations in glioblastomas are associated with increased *TERT* transcript levels, TERT protein levels, and its overall enzymatic activity. This results in telomere lengthening and cellular immortality (*3*). Not surprisingly, patients with glioblastomas or low-grade gliomas with *TERT* promoter mutations have significantly shortened overall survival (*106*).

Lower-grade gliomas and secondary glioblastomas appeared to be more dependent on a pathway that is known as alternative lengthening of telomeres (ALT) (*172, 173*). Specifically, the lack of ATRX is linked to the ALT phenotype since ATRX normally represses ALT (*174*). ATRX is a DNA-dependent ATPase that is required for checkpoint activation and protection of genomic stability (*175*). As mentioned previously, ATRX loss is a hallmark that is often used to distinguish astrocytomas form oligodendrogliomas, by means of immunohistochemistry staining, although this is not absolutely required for diagnosis establishment (*21*). Interestingly, ATRX loss is mostly identified *IDH*-mutant gliomas and is mutually exclusive with 1p/19q co-deletion, a diagnostic feature of oligodendroglial tumors (*104*).

Disruption of cellular energetics

For cancers to progress, they must adapt their cellular metabolisms to keep up with the accelerated rate of cell proliferation, the disrupted tumor microenvironment, and the change in the availability of energy resources. The commonly observed metabolic switch from oxidative phosphorylation to aerobic glycolysis in cancer cells (the Warburg effect, (*151, 153*)) might be due to their need to thrive in low-oxygen environment, which could be secondary to rapid tumor growth. Alterations of cellular metabolism has long been thought to one of the key aberrancies in glioblastoma, partly due to the observation of tumor necrosis (Figure 3-1 and Figure 3-2; Dysregulated energetics) (*21*). Further studies revealed that many genetic changes that impact the signaling network alter the cellular energetics of glioblastomas and lower-grade gliomas, leading to upregulation of surface GLUT, or even biochemical alterations of glycolytic enzymes (*151-156*). For example, the presence of *TP53* mutation can endorse the Warburg effect by promoting anaerobic glycolysis and dysregulating cellular energetics (*176, 177*).

The most important discovery of altered cellular metabolism in glioblastoma is the identification of *IDH1/2* mutations, which disrupt normal cellular metabolomics (*26, 178*). IDH enzyme has multiple isoforms: IDH1 functions mainly in the cytoplasm, whereas IDH2 resides within the mitochondria (*179*). In a normal cell, IDH catalyzes isocitrate into α-ketoglutarate (α-KG), both of which are intermediates of the citric acid cycle. It had been postulated that *IDH* mutations might facilitate the Warburg effect by changing the metabolic pool and subsequently increasing the glycolytic influx (*180*). Alternatively, *IDH* mutations might confer a different metabolic alteration that either occur independently or even in parallel with the classical Warburg effect (*181*). The mutation produces a gain-of-function IDH enzyme that, in addition to its normal metabolic function, also produces an onco-metabolite 2-hydroxyglutarate (2-HG) (*182*). 2-HG mechanistically alters cellular epigenetics and angiogenic signals. Interestingly, *IDH1* and *IDH2* mutations are mutually exclusive, with the majority in diffused gliomas and glioblastomas being IDH1 R132H (*26*). Importantly, these mutations are virtually always heterozygous (affecting only one allele), resulting in the production of mutant and wildtype enzymes within the same cell (*183*). The discovery of *IDH* mutations has reshaped the understanding of glioblastoma pathogenesis, at least in a subset of patients. *IDH* mutation is thought to be one of the earliest oncogenic drivers since it is sufficient to induce several malignant properties, and mosaic somatic *IDH* mutation is associated with many syndromes that carry

an increased risk for gliomas (*65, 184-186*). This is also supported by the much higher prevalence of *IDH* mutation in secondary glioblastomas (50%-80%) and in lower-grade diffuse gliomas (60%-80%), compared to primary glioblastomas (5%-10%) (*187*). These findings suggested that distinct oncogenic mechanisms may be required for the progression or even the initiation of these tumors. Importantly, patients with *IDH* mutant glioblastomas have significantly prolonged progression-free and overall survivals, compared to those that lack the mutations (*26*). This is potentially due to the ability of mutant IDH to alter the epigenomic landscape of cancer cells, enabling a resistance mechanism against treatment with DNA alkylating agents.

Local invasion of glioblastoma cancer cells

The diffuse invasion of glioblastoma cancer cells into the surrounding brain parenchyma at the time of initial diagnosis is a major hurdle for therapeutic interventions. Extensive surgical and radiotherapeutic approaches have been used to radically eradicate the affected brain. However, glioblastomas often recur locally or, not uncommonly, at a different region within the brain (*188*). Unlike many other types of solid tumors that metastasize beyond the organ of origin, glioblastomas rarely metastasize extracranially (*189*). Specifically, glioblastoma cancer cells tend to locally invade and migrate along the white matter tracts and vascular structures (*190*). These observations suggested that the cellular mechanisms that regulate glioblastoma cell migration is both aggressive, yet highly specific. A thorough understanding of this unique feature of glioblastoma can significantly aid the future development of therapeutic approaches that halt cancer cell migration at an early stage.

Many signaling-related molecules that are differentially associated with glioblastoma molecular subclasses are also mechanistically involved in cell motility and cancer cell migration (Figure 3-2; Invasion) (*2, 191, 192*). Hypoxia can also induce cellular migration via activation of HIF1α (*193*). Additionally, *MET* overexpression and *NF1* mutation, which are definitional of mesenchymal glioblastomas, are both associated with increased cell migration (*2, 194, 195*). MET (the receptor of hepatocyte growth factor, HGF, which is highly expressed in glioblastoma) can activate key downstream signaling effectors, including PI3K/AKT and ERK pathways, by heterodimerizing with CD44 (*196, 197*), a marker of mesenchymal phenotype that is closely associated with the cytoskeleton actin (*198*). And since the extracellular matrix in the brain also contain a

significant portion of hyaluronic acid, the ligand of CD44, the interaction between CD44 and MET is thought to play a significant role in glioblastoma cancer cell migration (*191, 199*). Moreover, higher degree of resistance to radiation therapy was observed in an *ex vivo* model of gliomas that underwent mesenchymal differentiation via NFκB signaling activation (*200*). Although these mesenchymal features were thought to be related to an aggressive cancer phenotype that requires more intensive therapies (*2*), the association between CD44 expression and cancer aggressiveness is not a straightforward, linear correlation (*192*). Intermediate expression of CD44 is associated with significantly more aggressive tumors compared to those with either very low or very high expression levels (*192*). Notably, a subset of patients with $CD44^{hi}$ mesenchymal gliomas showed longer survival times compared to patients with $CD44^{int}$ mesenchymal gliomas (*192*). Further characterization of cell subsets that may affect sample-level biphasic CD44 expression, as well as their associated signaling mechanisms, could reveal molecules that, once targeted, can suppress cancer cell migration and tumor progression.

Epigenetic alterations

In addition to genetic alterations, epigenetic aberrancies are vastly diverse among glioblastoma tumors and are closely related to the genomic profiles and even the anatomic locations of the tumors (*22, 25, 201*). *MGMT* promoter methylation is present in approximately 45% of glioblastoma patients (Figure 3-2; Resisting cell death) (*3, 202*) and is often evaluated in routine clinical testing to predict benefit from TMZ administration (*107, 202, 203*). The global DNA methylation landscape classifies glioblastomas into biologically distinct groups that are correlated with genetic alterations, clinical parameters, and patient outcomes (*25*). Tumor-level DNA methylation is also closely connected to transcriptional molecular subclasses (*3, 22*). Specifically, the glioma-CpG island methylator phenotype (G-CIMP) group is more prevalent in proneural glioblastomas. Mechanistically, 2-HG, the onco-metabolite of mutant IDH (more commonly present in proneural tumors), can inhibit many α-KG-dependent dioxygenases, such as histone demethylases and the TET (ten-eleven translocation) protein family, by acting as their competitive inhibitor (Figure 3-2; Dysregulated energetics) (*204*). This results in a higher level of DNA methylation in *IDH*-mutant glioblastomas (*65, 204*). Moreover, the age of patients with G-CIMP tumors tends to be younger at the time of diagnosis (*22*). Interestingly, a subset of patients with *IDH*-mutant

glioblastomas that also carry demethylated *MGMT* promotor are associated with worse clinical outcome compared to those with *IDH*-mutant, *MGMT* promotor methylated tumors (*205*).

Coordination between glioblastoma cells and the tumor microenvironment

Abnormal angiogenesis

Abnormal microvascular proliferation is a key diagnostic feature that distinguishes glioblastoma from lower-grade diffuse gliomas and correlates with disease aggressiveness and tumor recurrence (*206*). Microvascular proliferation in glioblastoma is defined by 1) endothelial cell proliferation and 2) loss of vascular integrity (*207*). VEGF (vascular endothelial growth factor) is one of the key angiogenic factors that is overexpressed in gliomas and glioblastomas compared to normal brain tissues (Figure 3-2; angiogenesis) (*208*). Both VEGF and its receptor, VEGFR, are highly upregulated *in vivo* and correspond to tumor grade (*209*). In part, the upregulation of VEGF was thought to be secondary to tumor necrosis, since HIF-1α induces transcription of *VEGF* in a hypoxic environment (*210*). 2-HG in *IDH*-mutant tumors can also suppress the degradation of HIF-1α by inhibiting its ubiquitination enzyme, resulting in a blockade of HIF-1α proteasomal degradation (*211, 212*). Additionally, VEGF can induce vascular permeability, which leads to vasogenic edema that is frequently observed in glioblastoma both radiologically and histopathologically (*213, 214*).

Bevacizumab (an anti-VEGF antibody) has been approved for clinical use in recurrent glioblastomas therapy (*109, 110*). However, the use of bevacizumab is not widely standardized since it only prolongs tumor progression by a few months and does not improve patient overall survival (*109, 110, 146, 215*). This could be due to an interplay between VEGF and other pro-angiogenic factors in glioblastoma, specifically their potential redundant and/or independent signaling pathways. Examples of these factors/receptors include FGF/FGFR and PDGF/PDGFR (*216*). Additionally, angiogenesis in glioblastoma can be enhanced through several mechanisms including direct and indirect activation of endothelial cells, as well as modulation of the extracellular matrix by production of proteolytic enzymes (*217-219*).

Abnormal vasculatures in glioblastomas are also the niche of glioblastoma cancer stem cells, which appear to preferentially co-localize with endothelial cells (*220*). Active communication between endothelial cells and cancer stem cells cooperatively maintain the niche, particularly via NOTCH signaling (*221, 222*). Specifically, NOTCH ligands are produced by the endothelial cells and can stimulate the proliferation and maintain the self-

renewal capability of cancer stem cells. Moreover, cancer stem cells may contribute to the formation of the vascular structure in the tumor by differentiating into either endothelial cells (*223-226*) or even pericytes that provide structural support for the vasculature (*227*). However, the ability of cancer stem cells to differentiate to both components of the vascular structure is still debatable (*227, 228*).

Glioblastoma and the immune system

A critical advancement in cancer therapeutics in the past decade has been in success of immunotherapy as alternative treatment in many types of cancer (*229, 230*). Tumors of the central nervous system (CNS) including glioblastoma, however, were previously overlooked as candidates for immunotherapy due to the perceived notion of them being in an anatomic immune-privilege site and being poorly immunogenic. However, the discoveries of CNS and tumor immune cell infiltration as well as the CNS lymphatic system has reshaped the understanding of the potential role of the immune system in human brain diseases and malignancies (*231-234*). Particularly, there are routes that facilitate CNS immune surveillance, which readily enables entry of immune cells into the brain parenchyma. Therefore, immunotherapy can and should be further evaluated as an alternative treatment strategy for patients with glioblastoma (Figure 3-2; Inflammation and avoid immune destruction).

Brain resident microglia (referred from now on as “microglia”) are the primary effector cells of the innate immune system in the CNS. Unlike other immune cells, microglia are intrinsic components of the brain and are maintained by local expansion of brain resident cells, rather than by postnatal hematopoietic progenitors (*235-237*). The abundance of microglia has been shown to be inversely correlated with the grade of diffused glioma; lower-grade tumors had a higher microglial signature based on transcriptional analysis (*73*). In contrast, the signature of monocyte-derived macrophage (referred from now on as “macrophage”) was increased in high-grade diffuse gliomas, including glioblastomas, and was associated with a higher endothelial transcriptional signature (*73*). This was not surprising as microvascular proliferation is a pathognomonic feature of glioblastomas. Specifically, leaky microvascular proliferation and disruption of blood-brain barrier can contribute to the infiltration of macrophages and subsequent interaction between the tumor and the immune system (*145*). Glioblastoma cells can release chemoattractants to recruit microglia and macrophages, while these immune

cells can also release factors that regulate glioblastoma cell growth (*237*). Therefore, interference of the interaction between cancer cells and microglia/macrophages is a potential avenue for immunotherapy in glioblastoma.

Lymphocytes are the key immune cell types that can be found in the cerebrospinal fluid (CSF) of healthy individuals (*238*), suggesting that they have immediate access to the brain parenchyma and are involved in immune surveillance (*239*). The presence of $CD8^+$ T cells in glioblastoma has been shown to be associated with prolonged patient survival (*240, 241*). Tumor-infiltrating lymphocytes (TILs) are present in varying abundance in primary glioblastomas, suggesting that there is a spectrum of immunogenicity that can vary between different tumors (*242, 243*). Suppression of the signal by which cancer cells prohibit T cell response is a key mechanism of many immunotherapies that have shown success in solid tumors (*244-247*). The expression of PD1 (program death 1) and presence of lymphocytes in glioblastomas suggested that these therapeutics can also benefit glioblastoma patients (*145*). Thorough characterization of tumor immunogenicity, as well as the presence and diversity of immune cell populations within the tumor microenvironment, can be informative for predicting clinical responses to immunotherapy.

Targeting the aberrant signaling network in glioblastoma

Genetic alterations are key features that have been used for the selection of novel glioblastoma treatment. Most of the targeted therapies had been designed to directly inhibit the activity of RTKs, usually as single agents. Currently, there are only four drugs that are approved by the FDA for use in glioblastoma treatment and bevacizumab is the only approved targeted therapy (the other three being non-targeting alkylating agents temozolomide, carmustine, and lomustine).

Inhibition of receptor tyrosine kinase signaling

Since alterations of *EGFR* and activation of its downstream signaling effectors are commonly observed in glioblastomas, inhibition of EGFR function is thought to provide survival benefit. The hypothesis that genetic alterations can predict clinical responses to RTK targeted therapies is also based on the observations seen in other human malignancies, such as melanoma and lung adenocarcinoma (*108, 248, 249*). Many forms of EGFR

inhibition have been tested as innovative therapies in glioblastomas (Figure 3-3), including drugs that target the tyrosine kinase domain (i.e. erlotinib and gefitinib) (*108, 249-251*), block EGFR extracellular ligand binding domain (i.e. cetuximab) (*144*), or even target a commonly mutated form of EGFR, EGFRvIII (i.e. rindopepimut) (*252, 253*). EGFR tyrosine kinase inhibitors (TKIs), gefitinib and erlotinib, small molecules that were previously successful in the treatment of patients with lung adenocarcinoma with specific *EGFR* mutations (*34, 248*), were among the first drugs tested in clinical trials for glioblastoma patients. Although highly promising based on known genetic alterations, gefitinib and erlotinib provided very limited clinical benefit in a small group of patients (*108*). Additionally, neither gefitinib nor erlotinib significantly diminishes EGFR activity *in vivo* (*147*). Cetuximab also shows very limited activity in glioblastoma patients (*144*).

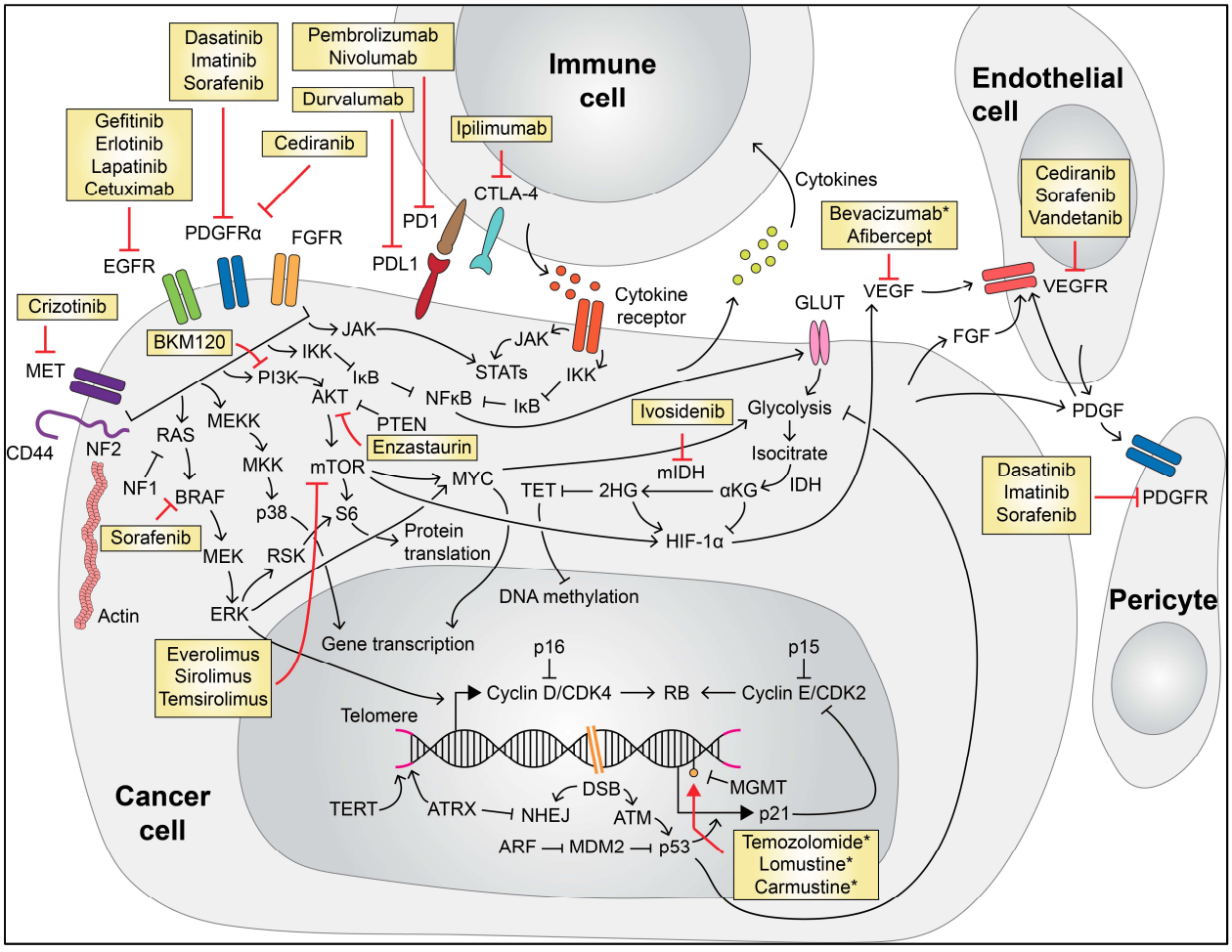


Figure 3-3 Targeting the complex signaling network in glioblastoma cancer cells and the tumor microenvironment. Drugs that have been previously tested as innovative targeted therapies in glioblastoma are shown (yellow boxes). *Drugs that are currently FDA-approved for glioblastoma treatment. For acronyms, see Figure 3-2.

EGFR status alone is ineffective in predicting the clinical response of glioblastoma patients to EGFR inhibition. Further attempts have been made in trying to identify alternative biomarkers that would predict clinical responsiveness to EGFR TKIs. Intact PTEN was shown to be closely associated with clinical response to EGFR TKIs, especially those with detectable EGFRvIII (*108*). However, there were exceptions as some patients with PTEN loss responded to erlotinib or gefitinib, whereas others with intact PTEN did not. Disruption of downstream signaling of EGFR was also proposed as a mechanism that promoted tumor progression without EGFR activity (*254*). Overall, these findings suggested that several signaling molecules are likely to be involved in the resistance to targeted therapies. It is also possible that these mechanisms can continuously adapt in response to external pressure, which can be unique kinetic features of individual tumors.

Inhibitions of non-EGFR RTKs, especially PDGFRα, were thought to be promising alternative therapeutic approaches. However, clinical trials of imatinib treatment in glioblastomas, either as monotherapy or in combination with other chemotherapeutics, did not significantly improve tumor progression or patient overall survival (*255, 256*). Evaluations of bulk tumor lysates have shown that multiple RTKs can be co-activated within the same tumor, enabling post-translational activations of several key effector bottlenecks (*142*). Moreover, cancer cell subsets containing diverse genetic amplification of many RTK-encoding genes can co-exist within the same tumor, and even within the same cells (*44*). This suggests that the diversity of single-cell signaling activation might be responsible for the failures of mono-therapies that specifically target individual RTKs.

Approaches to target molecules downstream to RTKs have also been explored since direct inhibition of surface RTKs showed minimal activity in clinical trials. PI3K and mTOR were some of the main candidates since they are highly activated in glioblastomas. Many of the compounds show very little promise, despite being well-tolerated clinically (*250, 257-259*). Interestingly, cancer signaling activity was shown to be a predicting factor of clinical responsiveness to inhibition of downstream signaling effectors; a subset of patients with recurrent glioblastoma with high levels of ribosomal protein S6 kinase phosphorylation appeared to benefit from treatment with an mTOR inhibitor, temsirolimus (*260*). This further emphasizes that considerations to select innovative therapeutics in glioblastoma should go beyond genomic alterations.

Anti-angiogenic therapies

Based on the classic pathological finding of glioblastomas, anti-angiogenetic therapy was thought to be highly successful as a novel therapeutic approach (Figure 3-3). Bevacizumab is currently approved for use as adjunctive therapy for recurrent glioblastomas; this is based mainly on the radiographic response observed clinically (*261*). However, its use is largely dependent on experts' opinion, on a case-by-case basis, rather than being a standard treatment protocol since bevacizumab shows very minor clinical improvement and has severe adverse effects. Moreover, clinical trials have surprisingly shown that bevacizumab only had minor benefits to progression free survival but not overall survival (*110, 146*). This discordance between tumor progression and patient survival could be attributed to the inefficiency of gadolinium-enhanced MRI to accurately detect early recurrence of the tumors due to the treatment with bevacizumab affecting vascular permeability (*262*).

Other anti-angiogenic therapies such as aflibercept, a decoy receptor for VEGF-A, VEGF-B, and PDGF, also have poor clinical benefit and tolerability (*263*). Even combined inhibition of EGFR and VEGFR by vandertanib does not significantly improve patient overall survival (*264*). The mechanisms that explain why these seemingly logical treatments have failed have yet to be elucidated. Hypothetically, anti-angiogenic therapies may re-route the cancer cell metabolism towards a process that promote tumor progression and do not require nearby blood vessels as the source of oxygen (such as anaerobic glycolysis). Alternative mechanisms that should be further tested include the re-wiring of cancer cell signaling network, or alterations of the homeostasis of the perivascular niche and the resident cancer stem cells, secondary to anti-angiogenic therapy.

Immunotherapy

The multi-step interaction between cancer cells and the immune system involves presentation of antigens from cancer cells, by the antigen-presenting cells, to T cells which are primed against the specific antigen (*265*). T cells then circulate back to the tumor site, recognize the cancer cells expressing the antigen, and attack the cancer cells leading to cancer cell death (*265*). External interference or even augmentation of the immune-cancer interaction are potential mechanistic routes for immunotherapy. The discovery of TILs and antigen-presenting cells has suggested that glioblastoma could be a candidate for many immunotherapeutic strategies, including CAR (chimeric antigen receptor) T-cells (*266, 267*), immune checkpoint inhibitors (*268-270*), as well as dendritic

cell (DC) vaccine therapy (*252, 253*). One of the challenges is to accurately identify patients who would benefit from such therapy.

PDL1 (programmed death-ligand 1) was shown to be highly expressed in the majority of primary and secondary glioblastomas and appeared to be higher in cancer cells compared to the seemingly normal tissue surrounding the tumor (*271*). PDL1, which is expressed by cancer cells, interact with PD1 on tumor-infiltrating T cells to inhibit the anti-tumor immune response upon T cells' recognition of cancer cells. A subset of glioblastomas was shown to have higher PDL1 expression than others, suggesting that there is a spectrum of glioblastoma immunogenicity and that some patients may respond more favorably to checkpoint inhibitors than others (*242, 271*). Blockade of PD1/PDL1 interaction had shown success in many types of cancer, most prominently melanoma, and many anti-PD1 agents are being tested as alternative therapy for glioblastomas. Although many other immune checkpoint inhibitors and vaccines have shown promising results in clinical trials (*145*), thorough characterization of the immune compartment in glioblastomas is needed to dissect the mechanistic connections between immune cells and cancer cells, as well as to explore potential biomarkers to establish standard selection criteria for enrolling patients for immunotherapy.

Diversity of adult glioblastomas

Molecular studies of glioblastomas indicate that there are additional layers of biological complexity and diversity beyond the current standard WHO classification (*3-5, 22, 205*). Indeed, glioblastomas are among the types of human malignancies with the highest genetic mutational burden which were mostly associated with core signaling pathways (*3, 272, 273*). Detailed transcriptional analysis further elucidated the aberrancies of signaling-related genes by classifying adult glioblastomas into four molecular subclasses based on the commonality of specific genetic and transcriptomic alterations: 1) proneural (*PDGFRA* alterations and *IDH* mutation), 2) classical (*EGFR* alterations), 3) mesenchymal (*NF1* deletion), and 4) neural subclasses (predominantly containing features of neural tissues, although likely due to sampling at tumor edges and normal tissue contamination) (*2, 4, 29, 274*). Interestingly, tumors of distinct molecular subclasses were not readily distinguishable by histopathological findings alone (*2*). These findings suggested that there are core biological features that are

significantly distinct between adult glioblastomas, and that regarding them as a unified disease entity might insufficiently capture their critical distinctive features that can guide novel therapeutic avenues.

Although the discovery of inter-tumor molecular diversity of adult glioblastomas provided a significant progress towards individualized medicine, it has not sufficiently captured the outcome diversity in glioblastoma patients. Therefore, transcriptomic molecular profiling has not been clinically implemented despite it being highly informative in revealing the diversity of tumors between patients. It is important to note that the transcriptional molecular classification of glioblastoma was based on the average abnormality of bulk tumors (*2, 4*). Most importantly, bulk-tumor information does not adequately reflect the unique cellular composition within individual tumors. Features of non-cancerous component within a tumor, including tumor-infiltrating immune cells and vascular cells, can immensely mask the significant biological characteristics in cancer cells, especially in rare cell subsets, that drive poor outcome (*29*). Tracking of genetic alterations in individual glioblastomas also revealed the evolutionary dynamics within a tumor, enabling systematic tracing of the cancer cell developmental hierarchy and route of migration (*42*). Diversity of genetic aberrancies was demonstrated in individual cancer cells from adult glioblastomas (*44, 47*). Amplifications of genes encoding commonly altered RTKs in glioblastomas were seen in different cells, and not uncommonly within the same cells (*44*). Interestingly, the diversity of clinically used molecular markers *IDH* mutational status and *MGMT* promotor methylation was also evident within individual tumors (*275*). This latter finding also raised the hypothesis that there is cell-cell variation of *IDH* mutational status and *MGMT* promoter methylation status within an individual tumor. And if so, should the *IDH*-wildtype cells within a tumor that was tested *IDH*-mutant at sample level, or the unmethylated *MGMT* cells within a sample-level methylated tumor, be recognized and quantified as their own entity, since these molecular lesions are mechanistically related to unfavorable responses to standard chemotherapy? Transcriptional analysis of single cancer cells from adult glioblastomas revealed co-existence of cells that harbor features of distinct molecular subclasses within individual tumors, as well as cells that had combined features of multiple subclasses (*5*). Indeed, features of single cells were strongly associated with patient outcome in blood malignancies where clinical tracking of specific cell subsets can inform disease prognostication (*15, 86*).

The presence of multiple cancer cell subsets has been reported in glioblastomas and has been proposed to be a contributing factor towards therapy resistance. However, whether it is the degree of cellular diversity

within an individual glioblastoma, or the presence of a specific cell subsets that drive therapy resistance, has yet to be elucidated. Single-cell analysis revealed that clones derived from individual cells from a single tumor differentially responded to *ex vivo* drug treatments (*30, 48*), suggesting the presence of multiple cancer cell subsets that carry potential contrasting resistance mechanisms at the single cell level. Although there is ample evidence in other types of human cancer that genomic and/or transcriptomic alterations can strongly predict clinical responsiveness to therapies that specifically target the proteins encoded by the altered genes, the same logic does not seem to apply to glioblastoma, as mentioned previously. Taken together, the next critical steps are 1) to pinpoint the cancer cell subset(s) in glioblastoma that drive poor outcome, 2) to characterize the defining features of the unfavorable cell subsets, and 3) to identify novel therapeutic approaches that can eliminate the subsets and improve patient survival.

Single-cell quantification of signaling and proteins: The future of glioblastoma clinical stratification and target discovery

Alterations in cancer cell signaling facilitate the maintenance of the functional hallmarks that sustain malignant capabilities. Signaling states are cumulative responses that are secondary to cell intrinsic aberrancies (i.e. genetic mutations, epigenetic alterations, etc.) or to the changes in the tumor microenvironment (i.e. availability of nutrient or oxygen, presence of growth factors or cytokines, interactions with immune or vascular cells, etc.). Moreover, signaling is a process by which cells integrate internal and external information to govern various cellular processes (*9*). This suggests that mapping of the signaling network in individual cells can illustrate the cellular functional capabilities as well as their global alterations, which make them unique.

Although genomic and transcriptomic data have suggested that signaling alterations are key pathogenesis in adult glioblastomas, very few studies focused on quantifying the actual signaling activity in primary patient samples. Critically, studies that characterize intracellular signaling of single cell from primary patient samples, without first undergoing *ex vivo* culture or subset selection, do not exist. Per-cell mapping of signaling at basal state (*13*), as well as after external potentiation (*12, 14, 15, 80*), have proven to be successful in stratifying the clinical outcome of patients with various hematologic malignancies. Additionally, signaling activity and proteins are actionable cellular targets that can be immediately tested for their therapeutic efficacy.

Therefore, by coupling single cell signaling quantification with clinical outcome assessment in adult glioblastoma, it is possible to pinpoint both the cancer cell subsets that dictate poor survival, and the subset-specific signaling hindrance that may be the mechanism of therapy resistance.

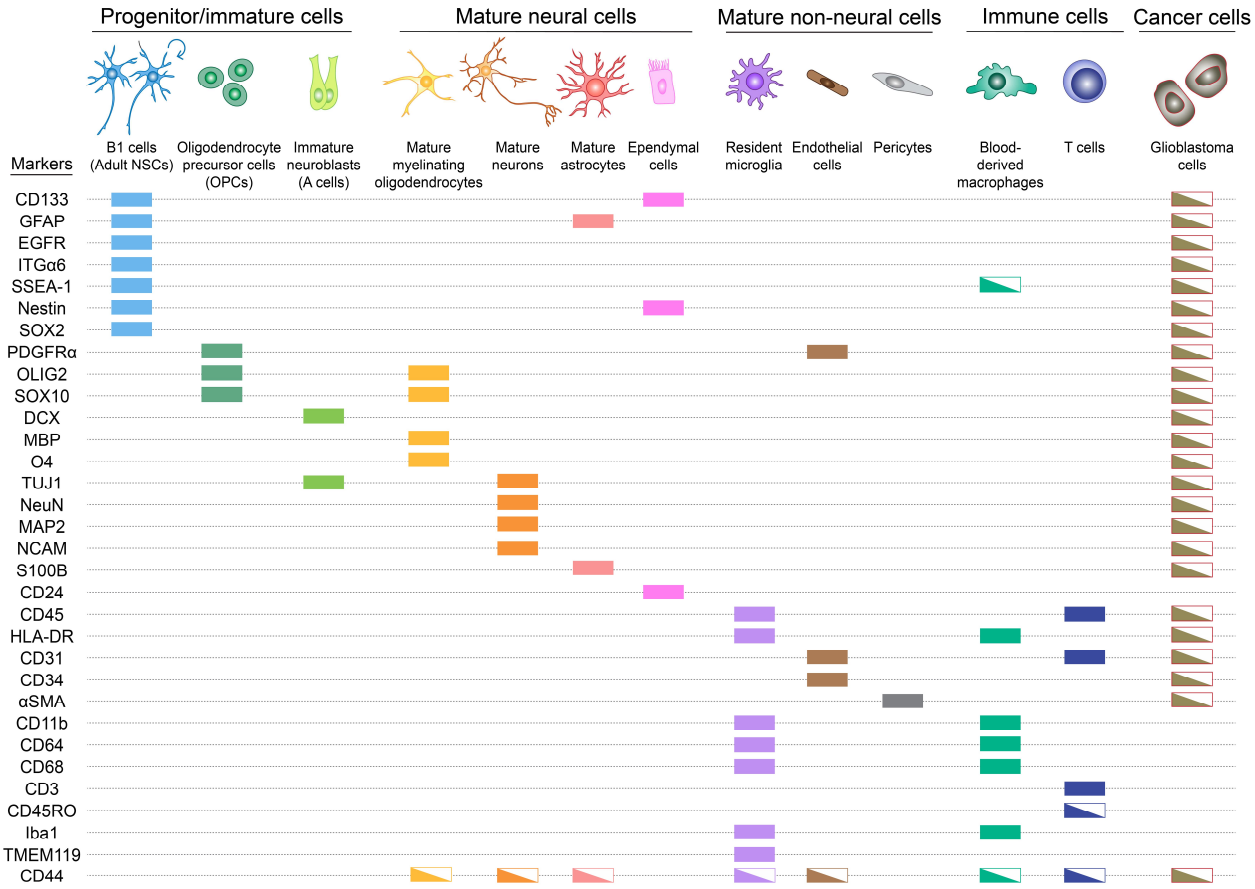


Figure 3-4 Understanding the protein phenotype of cells in a normal brain is a critical for the characterization of single cells in glioblastoma. Many cell types comprise an adult human brain and have rather well-defined cellular protein expression. In adult glioblastomas, the tumor microenvironment is composed of cancer cells, brain resident cells, and peripherally-derived immune cells. Abbreviations: CD133, prominin 1; GFAP, glial fibrillary acidic protein; EGFR, epidermal growth factor receptor; ITGα6, integrin alpha 6; SSEA-1, stage-specific embryonic antigen 1; SOX2, SRY-BO2; PDGFRα, platelet-derived growth factor receptor alpha; OLIG2, oligodendrocyte transcription factor 2; DCX, doublecortin; MBP, myelin basic protein; TUJ1, neuron-specific class III beta-tubulin; NeuN, neuron nuclear antigen; S100B, S100 calcium binding protein B; CD24, heat stable antigen; CD45, protein tyrosine phosphatase, receptor type, C; HLA-DR, human leukocyte antigen – antigen D related; CD31, platelet endothelial cell adhesion molecule; CD34, CD34 antigen; αSMA, alpha smooth muscle actin; CD11b, integrin alpha M chain; CD64, Fc-gamma receptor 1; CD68, scavenger receptor class D member 1; CD3, T-cell co-receptor; CD45RO, a CD45 isoform; Iba1, ionized calcium-binding adapter molecule 1; TMEM119, transmembrane protein 119; CD44, hyaluronic acid receptor. Filled, lineage-defining proteins; Half, proteins that can be expressed in certain cell types).

Single-cell technologies enable accurate distinction between stromal and cancer-specific features (*5, 13*), which is crucial for subsequent selections of innovative therapies. Critically, analysis of adequate number of cells must be a priority to maximize sample representation. Multi-dimensional flow cytometry has long been the standard for the assessments of blood and bone marrow specimens, in both research and clinical settings (*276, 277*). In the traditional, standard fluorescence flow cytometry, single cells are stained with a cocktail of antibodies (each detects a specific protein epitope), each are labeled with distinct fluorophores. Stained cells are then run through the micro-fluidics system of a flow cytometer, in a single-cell stream, and are then subjected to wavelength-specific light excitation. The fluorophores that are conjugated to the antibodies would emit light at different emission spectra, which allows 1) determination of the presence of specific proteins in individual cells, and 2) quantification of the abundance of each protein in each cell in comparison to the other cells in the sample.

Flow cytometry, therefore, fits the requirements for the single-cell risk stratification approach. First, flow cytometry enables quantification of both surface and intracellular proteins, including post-translationally modified signaling effectors (*80, 84, 278*). Second, the quantification is achieved at the single cell level. And third, thousands to millions of single cells can be quantified in a single setting, resulting in a higher representation of cell subsets present in the primary tumor sample compared to other single-cell technologies such as single-cell RNA sequencing or quantitative imaging (*5, 71-73, 279-281*).

The design of antibody panels for flow cytometry analyses requires knowledge of the cellular composition of the sample of interest, as well as either known (or hypotheses) of cellular targets that are to be quantified. For tissue types with long-standing history of single-cell analyses, such as blood and bone marrow, the cell subsets and their immunophenotypes are rather well-characterized (*81, 282*). The clinical single-cell analyses of hematologic malignancies are based on that knowledge, with additional quantification of cancer-specific features that are associated with patient clinical outcome (*12, 15, 80*). To adopt a similar approach for single-cell glioblastoma analysis, it is therefore crucial to understand the protein expression of cell types that are present in a healthy human brain, and to build the glioblastoma antibody panel on top of that for feature discovery (Figure 3-4). With the ultimate goal of using single-cell signaling profile to stratify clinical outcome of glioblastoma patients, signaling effectors would also be included as targets for quantification (Figure 3-2) (*12, 13, 15, 80*).

With the growing knowledge cell phenotype and the continuously increasing targets that should be quantified to thoroughly capture the relevant biological features, the number of cellular targets (and, thus, the number of antibodies in the panel) poses a critical challenge for conventional fluorescence flow cytometry (Figure 3-4). Once excited by light, a given fluorophore that is coupled to an antibody emits a spectrum of wavelengths that can overlap with the emitted light from other fluorophores, which can interfere with data analysis. To minimize spectral overlap and to maximize the accuracy of target quantification, the number of antibodies per panel is unavoidably limited. A standard fluorescence flow cytometry panel includes approximately 8-15 antibodies (occasionally over 20 parameters, at which signal compensation prior to data analysis can be challenging). The number antibodies per panel is even more conservative in clinical settings since accurate signal quantification is critical for patient care.

Mass cytometry had overcome the limitation of conventional fluorescence flow cytometry (*50, 83*). Similar to fluorescence flow cytometry, mass cytometry is a high-throughput, single-cell, antibody-based technology that enables per-cell quantification of protein targets, including assessment of intracellular signaling activity (*83*). However, the antibodies that are used in mass cytometry analysis are conjugated to rare metal isotopes instead of fluorophores. Cells are run through a mass cytometer in a single-cell stream and are then atomized into clouds of ions by an argon plasma. The mass of the ions in the clouds, which would correspond to specific antibodies and cellular targets, are then quantified by inductively-coupled mass spectrometry (ICP-MS) (*83*). The precise time-of-flight (TOF) quantification of the isotopic mass eliminates the spectral overlap which is a common challenge with the use of fluorophores. Therefore, mass cytometry enables accurate quantification of at least 35 parameters at single cell level (*50, 283*). The significantly increased dimensionality of mass cytometry analysis had enabled simultaneous quantification of not only identity proteins (to distinguish stromal cells from cancer cells) (*284, 285*), but also cell viability (*286*), cell cycle (*287*), as well as signaling states in individual cells (*12, 83, 288*).

The increasing complexity of single-cell data demands systematic analysis algorithms. Conventionally, the analysis of single-cell fluorescence flow cytometry data involves sequential interpretations of biaxial plots (*81, 83*). With 35 dimensions of single-cell data, the number of pairwise plots would be close to 600. This traditional analysis is not only cumbersome and impractical, it is also ineffective to thoroughly capture and

potentially discovery the underlying biology of the dataset. Computational algorithms have enabled unbiased analyses and visualization of single-cell data in comprehensible and biologically meaningful ways. Examples for these tools include t-SNE (t-distributed stochastic neighbor embedding) (*289, 290*), SPADE (spanning-tree progression analysis of density-normalized events) (*282*), local linear embedding (LLE) (*291*), and ISOMAP (*292*). Additionally, some algorithms, such as SPADE (*282*), FlowSOM (*293*), DensVM (*294*), and ACCSENSE (*295*), can automatically assign cells into subsets or clusters, based on their phenotypic similarity. Overall, when these tools are used either independently or in combination, they can reveal novel biological findings (such as new biomarkers or rare cell subsets) that otherwise would have been overlooked by conventional analyses.

Conclusion

To date, adult glioblastoma remains incurable. Despite the ample attempts to identify novel therapeutics, the approved treatment options that improves patient outcome are limited. Critically, signaling network alterations appeared to be the key mechanism of glioblastoma pathogenesis. However, there is no clinical stratification that effectively couples single-cell diversity with the identification of immediately targetable signaling activity.

Single-cell protein analysis using fluorescence flow cytometry is already part of the routine clinical diagnostic workflow of blood and bone marrow specimens. Implementation of approaches that can accurately dissect higher data dimensions in individual cells would enable a comprehensive outlook of clinically relevant cell subsets, as well as their unique cytotypes. Mass cytometry is, therefore, an appealing platform since it can precisely and simultaneously quantify high dimensions of identity proteins as well as the clinically significant signaling activity and cellular functions in individual cells. This would reveal patient-specific, trackable, and potentially targetable, biomarkers that can shape the individualized therapeutic approaches. Subsequent disease monitoring could then be effectively tailored to encompass patient-specific features.

Given the degree of cellular diversity and the pertinence of signaling alterations in glioblastoma, the next generation of clinical stratification and therapeutics should incorporate quantitative, single-cell, profiling of the signaling network. Ultimately, single-cell analyses could be integrated into routine clinical testing of glioblastoma for rapid assessment of clinically relevant cell subsets and biomarkers to guide individualized therapy. One of the critical roadblocks is the lack of a standardized, widely applicable, single-cell acquisition protocol that can be

easily implemented in the routine workflow. Additionally, the current clinical data analysis relies heavily on manual expert interpretation. As the increasing data dimensionality starts to become part of routine clinical care, automated computational tools can significantly expedite the lab turnaround time, while maximizing quantitation accuracy and minimizing analytical bias. The key to a successful clinical application is, therefore, the development of a robust automation strategy.

The goal of this dissertation is to tackle the important hurdles towards a clinical stratification that captures the alterations of signaling activities in adult glioblastoma single cells. This dissertation describes the discovery of novel, clinically distinct glioblastoma cells defined by altered signaling and protein identity, while simultaneously providing a frame of work that can be readily implemented in other types of human malignancies. Moving forward, this body of work is a critical starting point for clinical diagnostics, stratification, and disease monitoring, that are continuously evolving towards automation, quantitation, and therapy individualization, based on the diversity of individual cancer cells.

CHAPTER 4

Characterizing Phenotypes and Signaling Networks of Single Human Cells by Mass Cytometry

Authors: Nalin Leelatian, Kirsten E. Diggins, and Jonathan M. Irish

This work is presented as it appears in manuscript form in *Methods in Molecular Biology* 2015 (Leelatian et al., 2015).

License 4263221151542, issued January 6, 2018

Preface

Human tissues comprise of diverse cell populations that can be characterized and distinguished by unique expressions of cell identity proteins. Single-cell analysis approaches enable dissection and deconvolution of tissue cellular diversity. Flow cytometric tools are appealing platforms for in-depth cellular identity characterization since they directly assess per-cell protein expression. With the increasing biological knowledge on features that are associated with cell identity, a higher number of cellular features are needed to be simultaneously quantified to 1) capture the complex tissue microenvironment, and 2) enable discovery of novel cellular characteristics or cell types that may be associated with clinical features. Fluorescence flow cytometry, a single-cell analysis tool that is commonly used for characterizing blood and bone marrow in both research and clinical settings, can simultaneously quantify 8-10 features per cell in a standard setting, and up to 15-20 features per cell with newer instrumentations. Overlapping emission spectra of fluorophores, however, is the key limitation that restricts the number of cellular features that can be simultaneously measured in individual cells. Mass cytometry is a flow cytometry-based technique that utilizes isotope-labelled antibodies, paired with a time-of-flight analysis. This approach significantly minimizes signal overlap, thus, enables simultaneous quantification of 35 or more features per cell. This chapter describes the use of mass cytometry and subsequent computational analyses for the characterization of different cell types found in human peripheral blood mononuclear cells. Additionally, mass cytometry is used here as a tool to dissect the signaling kinetics in a human acute myeloid leukemia cell line. This was accomplished by quantifying the change of phosphorylation of signaling effector

molecules in response to short-term potentiation *ex vivo*. The work presented here describes the tools essential for the overall goal of this dissertation, which is to characterize the cellular phenotype and signaling profiles of cell subsets that are associated with favorable and unfavorable clinical outcomes of adult patients with glioblastoma.

Abstract

Single cell mass cytometry is revolutionizing our ability to quantitatively characterize cellular biomarkers and signaling networks. Mass cytometry experiments routinely measure 25-35 features of each cell in primary human tissue samples. The relative ease with which a novice user can generate a large amount of high quality data and the novelty of the approach have created a need for example protocols, analysis strategies, and datasets. In this chapter, we present detailed protocols for two mass cytometry experiments designed as training tools. The first protocol describes detection of 26 features on the surface of human peripheral blood mononuclear cells. In the second protocol, a mass cytometry signaling network profile measures 25 node states comprised of five key signaling effectors (AKT, ERK1/2, STAT1, STAT5, and p38) quantified under five conditions (Basal, FLT3L, SCF, IL-3, and IFNγ). This chapter compares manual and unsupervised data analysis approaches, including bivariate plots, heatmaps, histogram overlays, SPADE, and viSNE. Data fi les in this chapter have been shared online using Cytobank (http://www. cytobank.org/irishlab/).

Introduction

Computational tools and instrumentation advances have introduced a new era of single cell systems biology research where it is straightforward to comprehensively characterize all cell types, known and unknown, in primary tissues (*296*). Traditional aggregate analysis techniques, such as Western blotting, characterize cellular features with the assumption that the total signal reflects the sum of a homogeneous underlying population of cells. However, even among clonally derived cell lines, biologically meaningful cell-to-cell differences in protein expression and phosphorylation are the rule, not the exception (*11*). The presence of multiple cell subpopulations with distinct expression signatures is common in primary samples taken directly from the body, such as healthy tissue and tumor specimens (*17, 297*). In the case of blood cancers, cancer cell

subsets defined by abnormal signaling are associated with patient clinical outcomes, including overall survival (*15, 80*). Single cell tools such as flow cytometry enable high content single cell measurements of cellular identity and functional response (*10, 84*). Overlap in the emission spectra of conventional fluorescent probes can create experiment design and data analysis challenges that are particularly a drawback in quantitative single cell comparisons (*283*). Moreover, autofluorescence of some tissue types can overlap with fluorophore emission spectra and complicate quantitative analysis.

Mass cytometry employs metal isotope reporters (mass tags) that are not normally found in biological specimens (*83*). Typically, mass tags are coupled to an antibody or other target-specific probe so that their abundance in a cell corresponds to the abundance of a target of interest. Intercalator reagents containing iridium or rhodium are routinely used to mark single cells (*298*). In addition to making cells detectable as an event to the instrument, comparison of event length and uptake of a cell marking intercalator helps distinguish single cells, cell doublets, and other particles (Figure 4-1). Intercalator reagents aim to provide a consistent, strong signal that is minimally variable with experimental conditions. Regular marking of events with an invariant signal helps to identify cell events when parsing mass cytometry data into flow cytometry standard (FCS) format, gating single cells (Figure 4-1), and analyzing cells with computational tools (Figures 2-2 and 2-3).

While mass spectrometry avoids fluorescence associated problems, there are aspects of the technology that can be valuable to monitor and test. Mass cytometry issues include 1) impure isotopic mass tags, 2) spillover between closely spaced spectral channels when signal is very abundant (+1 and −1 spillover), 3) variable oxide formation (primarily +16 spillover), and 4) other less common confounding signals not originating from the cells (e.g., barium in buffers, gadolinium contrast agent from patient magnetic resonance imaging). This chapter will not specifically address these aspects of the technique except to say that they can be minimized by following best practices for instrument use, reagent quality control, and experiment design [(*283, 296*).

A key advantage of mass cytometry is the multiplexed detection of many features of each cell. Typical experiments measure approximately 35 features of every cell (*83, 288, 294, 299*), with 42 being state of the art (*288*). The theoretical limit on the instrument has not been approached and is likely between 100 and 200 features per cell using the current technology. Mass cytometry therefore, allows single-cell deep profiling of cell identity, phenotype, response, and functional outcome. Relative to microscopy, mass cytometry is high content

and high throughput at the single cell level: a typical experiment quantifies 35 features on each of ≥ 100,000 cells from a sample in ~15–20 min. Mass cytometry has numerous applications for characterizing the cellular heterogeneity of healthy and diseased tissues and for tracking changes in populations over time in primary tissue samples (*11*).

Here we present protocols for two mass cytometry experiments: 1) quantifying cell surface biomarkers expressed on healthy human peripheral blood mononuclear cells (PBMCs) and 2) quantifying intracellular signaling network responses in Kasumi-1 cells using phospho-flow (*278, 300*). Data from experiments provided in this chapter are available online (http://www.cytobank.org/irishlab). In addition, computational tools are an integral part of analyzing multidimensional datasets. In this chapter we provide examples of multidimensional data visualization. As data analysis can be daunting in 25-dimensional datasets, this chapter compares analysis of the human PBMC cell surface immunophenotyping dataset by three methods: 1) traditional bivariate gating, heatmaps, and histogram overlays (*301*), 2) Spanning-Tree Progression Analysis of Density-Normalized Events (SPADE (*282*)), and 3) visualization of t-Stochastic Neighbor Embedding (viSNE (*290*)).

Materials

1. Ficoll-Paque solution.
2. 15 mL and 50 mL conical tubes.
3. Cell culture medium: RPMI 1640 containing 10% fetal bovine serum (FBS), 100 U/mL penicillin, and 100 µg/mL streptomycin, stored at 4°C. Heat by immersing in 37°C water bath for 15-20 min.
4. Freezing medium: 12% DMSO and 88% FBS, keep cold on ice.
5. Cryopreservation tubes, 1.8 mL.
6. 12 x 75 mm round-bottom polystyrene cytometry tubes.
7. Water bath set at 37°C.
8. Cell culture incubator set at 37°C with 5 % CO_2.
9. Absolute methanol stored at -20°C or lower.
10. 1× phosphate buffered saline (PBS).
11. Staining medium: 1% bovine serum albumin (BSA) in phosphate buffered saline (PBS).

12. Deionized water.

13. Intercalator: 500 μM iridium. Prepare a 50X working solution (12.5 μM) by diluting with PBS.

14. Cytometry tubes with 35-μm cell strainer caps.

15. Trypan blue, prepared as recommended by manufacturer.

16. Hemocytometer.

17. 16% paraformaldehyde (PFA) aqueous solution.

18. Antibodies: refer to Tables 2-1 and 2-2.

19. Stimuli: refer to Table 4-3.

20. 15-30 mL of human peripheral blood.

21. Kasumi-1 cells (ATCC; CRL-2724).

22. Mass cytometer (CyTOF) (Fluidigm).

23. Data in this manuscript were analyzed using Cytobank (http://www.cytobank.org).

Table 4-1 Antibody panel for cell identity

#	Target	Tag	Clone	μL[§]	Catalog #*
1	CD19	142Nd	HIB19	1	3142001B
2	CD117	143Nd	104D2	1	3143001B
3	CD11b	144Nd	ICRF44	1	3144001B
4	CD4	145Nd	RPA-T4	1	3145001B
5	CD8a	146Nd	RPA-T8	1	3162001B
6	CD20	147Sm	2H7	1	3147001B
7	CD34	148Nd	581	1	3148001B
8	CD61	150Nd	VI-PL2	1	3150001B
9	CD123	151Eu	6H6	1	3151001B
10	CD45RA	153Eu	HI100	1	3153001B
11	CD45	154Sm	HI30	1	3154001B
12	CD10	156Gd	HI10a	1	3156001B
13	CD33	158Gd	WM53	1	3158001B
14	CD11c	159Tb	Bu15	1	3159001B
15	CD14	160Gd	M5E2	1	3160001B
16	CD69	162Dy	FN50	1	3162001B
17	CD15	164Dy	W6D3	1	3164001B
18	CD16	165Ho	3G8	1	3165001B
19	CD44	166Er	BJ18	1	3166001B
20	CD38	167Er	HIT2	1	3167001B
21	CD25	169Tm	2A3	1	3169003B
22	CD3	170Er	UCHT1	1	3170001B
23	IgM	172Yb	MHM-88	1	3172004B
24	HLA-DR	174Yb	L243	1	3174001B
25	CD56	176Yb	CMSSB	1	3176003B

[§]μL of antibody per 100 μL total volume.
*Source was Fluidigm for all antibodies.

Methods

Immunophenotyping of surface markers of cell identity in human PBMCs

Human PBMCs are heterogeneous populations of cells that can be defined and categorized into distinct cell types by different surface marker expression levels. CyTOF allows simultaneous measurement of all the key surface markers of cell identity to enable deep profiling of cell subsets in human PBMCs. The protocol outlined below describes the method for live-cell surface staining of human PBMCs to characterize expression of 25 different cell surface markers.

Preparation and freezing of PBMCs

1. Collect 15-30 mL of human peripheral blood from each donor in sterile tubes with heparin or EDTA anticoagulant.
2. Transfer the blood to a 50 mL conical tube and add an equal amount of 1X PBS into the same tube and pipet to mix.
3. Add Ficoll-Paque solution into two new 50 mL conical tubes at 12.5 mL per tube (see Note 1).
4. Slowly overlay 20 mL of diluted blood sample onto the Ficoll-Paque solution in each tube (see Note 2).
5. Slowly add PBS into each tube to increase the volume in each tube to 50 mL.
6. Centrifuge at 400 *x g* for 30 min at room temperature. Set the deceleration on the centrifuge to the lowest speed to avoid mixing the PBMCs with the rest of the solution.
7. Check for a white ring of PBMCs (buffy coat) right above the red blood cell pellet. Gently aspirate and discard the majority of the plasma above the ring. Avoid disturbing the buffy coat.
8. Gently pipette the buffy coat using a wide bore pipet tip and transfer the buffy coat into a new 15 mL conical tube. Combine buffy coat from all the tubes of the same blood donor.
9. Add PBS to the 15 mL conical tube to increase the volume in each tube to 15 mL.
10. Centrifuge at 300 *x g* for 10 min at room temperature. Set the deceleration on the centrifuge to high brake speed.
11. Aspirate to discard the supernatant.
12. Resuspend cell pellet in 2 mL of PBS. Count cells with trypan blue (see Note 3).

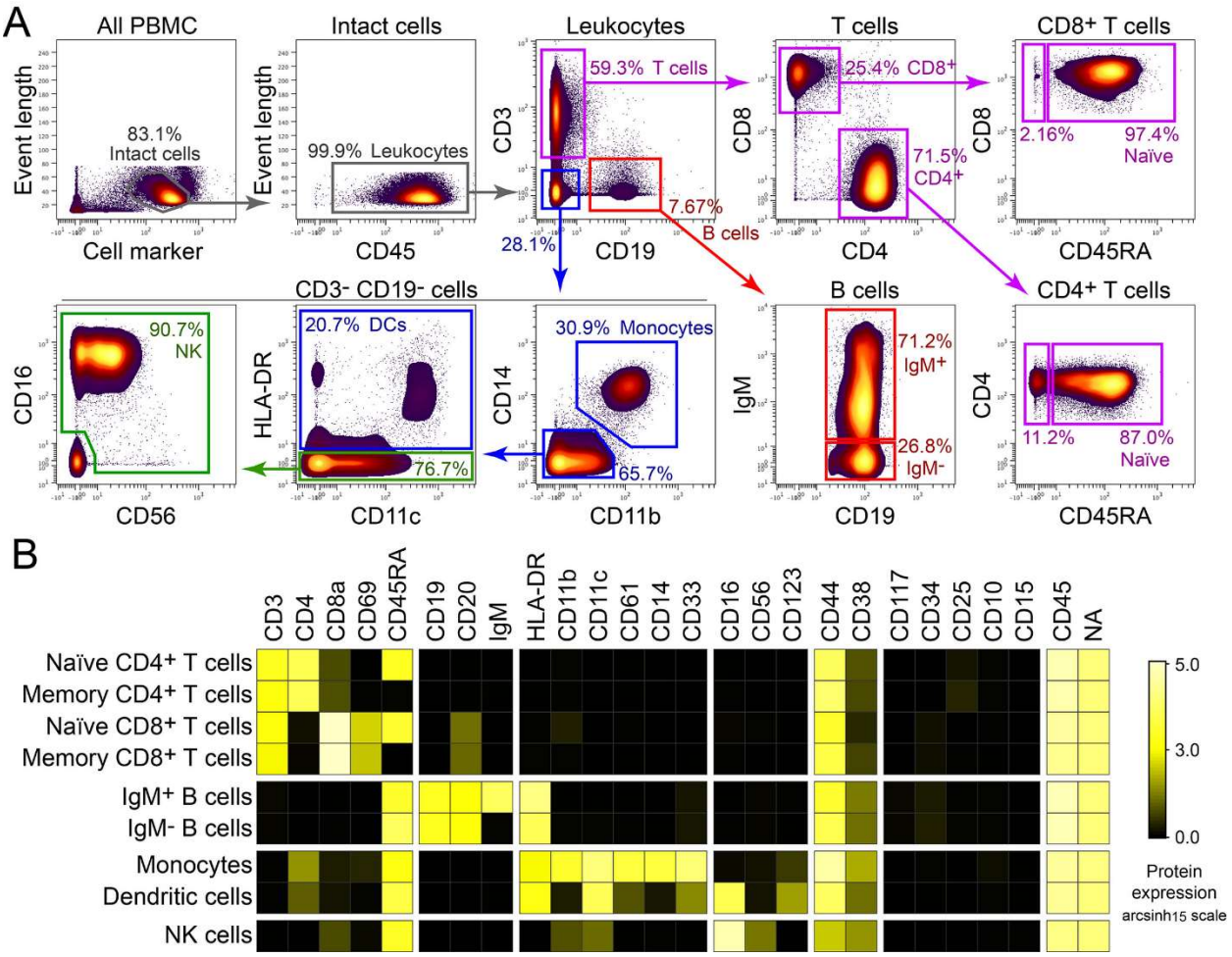


Figure 4-1 Phenotyping human PBMC subsets with traditional bivariate gating and heatmap analysis. (A) Bivariate plots compare features measured on healthy human PBMCs by mass cytometry. Heat corresponds to proportional cell abundance in a plot region for the population indicated above the plot. Intact cells (grey gate) were defined using event length and an iridium-based cell marker (Ir-191 intercalator). Among the intact cells, leukocytes were defined as CD45+ events. CD3 and CD19 were then used to identify T cells and B cells, respectively. Subsets of T cells and B cells were identified using additional markers (CD4, CD8, CD45RA, and IgM). The non-T non-B cells ($CD45^+$ $CD3^-$ $CD19^-$) were gated as monocytes, dendritic cells, and natural killer (NK) cell using CD14, CD11b, CD11c, HLA-DR, CD16, and CD56. (B) A heatmap compares expression of 27 measured features on the same cells populations shown in (A) using a log-like $arcsinh_{15}$ scale. The heat corresponds to the $arcsinh_{15}$ fold difference in median expression for a given marker compared to the table minimum.

13. Add 10 mL of PBS to each tube and centrifuge at 300 *x g* for 10 min at room temperature again. Repeat the wash as needed until the supernatant in relatively clear.
14. Aspirate to discard the supernatant and avoid disturbing the cell pellet.
15. Resuspend the cells in freezing medium to yield a concentration of 10-15 × 10^6 cells/mL.
16. Aliquot cell suspension into cryopreservation tubes at 1 mL per tube.

17. Freeze the cells slowly at the rate of -1°C/min in -80°C freezer (see Note 4). Transfer the cryopreservation tubes into liquid nitrogen the next day for long-term storage.

Live-cell surface staining of PBMCs

1. Remove a cryopreservation tube with PBMCs from liquid nitrogen. Immediately immerse the cryopreservation tube in 37°C water bath. Remove the cryopreservation tube form the water bath once the cell suspension is completely thawed. This step should not take longer than 1-2 min.
2. Add 10 mL of warm cell culture medium into a 15 mL conical tube. Transfer thawed PBMCs form the cryopreservation tube into the same 15 mL conical tube. Pipet to mix cells with the cell culture medium.
3. Pellet the PBMCs at 300 *x g* for 5 min at room temperature. Aspirate to discard the supernatant.
4. Add 5 mL of warm cell culture medium and pipet to mix the PBMCs with the medium. Count the cells using trypan blue (see Note 3) and adjust the volume to yield a concentration of 1-2 x 10^6 live cells/mL. Pipet the cells to get a single cell suspension.
5. Label two cytometry tubes as “Intercalator only” and “Surface panel.”
6. Aliquot the PBMC suspension into cytometry tubes at 1 mL per tube. Put the tubes with cell suspension into a 37°C incubator with 5% CO_2. Rest cells for 15 min before staining.
7. Take the cytometry tubes from the incubator. Pellet the cells at 300 *x g* for 5 min at room temperature.
8. Check for cell pellets and discard supernatant by decanting (see Note 5).
9. For the tube labeled “Intercalator only,” vigorously vortex the cell pellets in void volume and add 1 mL of ice-cold methanol (stored at -20°C or lower) to permeabilize the cells. Pipet or briefly vortex to mix (see Note 6) and store the tube at -20°C for at least 10 min (see Note 7).
10. For the tube labeled “Surface panel,” resuspend the cell pellets with 75 µL of staining medium (see Note 8).
11. Transfer 75 µL of cell suspension from the tube labeled “Surface panel” to a new cytometry tube. Add antibodies from Surface Marker Panel (Table 4-1) to the tube to a final volume of 100 µL. Pipet or briefly vortex the cells with the antibodies to mix. Incubate the cells with the antibodies at room temperature for 30 min.

12. Add 1 mL of 1X PBS and pipet or briefly vortex to mix. Pellet the cells at 300 *x g* for 5 min at room temperature.
13. Check for cell pellets and discard supernatant by decanting (see Note 5). Repeat steps 12 and 13 once.
14. Vigorously vortex the cell pellet in void volume and add 1 mL of ice-cold methanol (stored at -20°C or lower) to permeabilize the cells, pipet or briefly vortex to mix (see Note 6) and store the tube at -20°C with the tube labeled “Intercalator only” for at least 10 min.

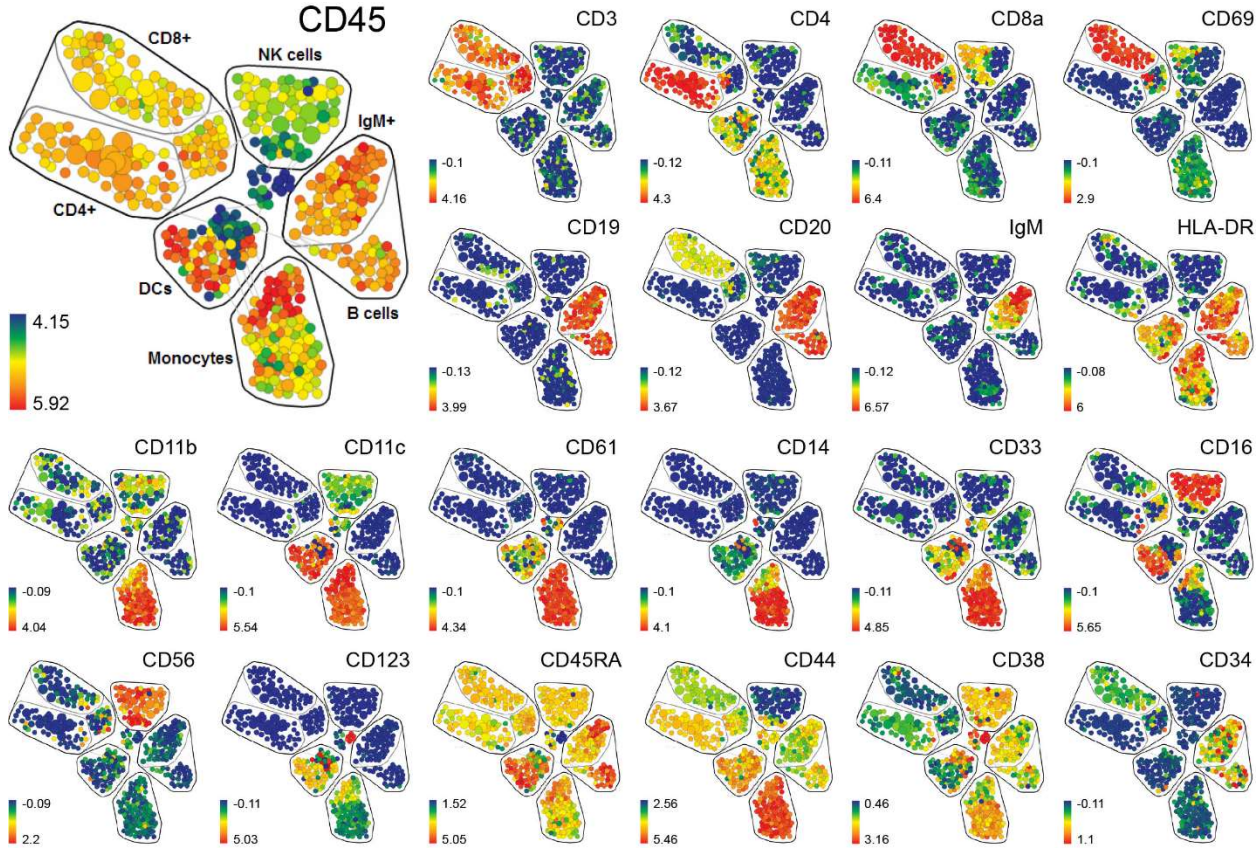


Figure 4-2 SPADE clusters PBMCs into populations based on similar marker expression. SPADE plots show clustered populations of healthy human PBMCs (circles) connected using a minimum spanning tree. Each circle represents a population of cells with a similar phenotype for the 21 markers shown. The size of each circle is proportional to the number of cells in that population. The heat color for each circle corresponds to the median expression of the indicated marker on the cells within that circle. Heat corresponds to the arcsinh_{15} fold difference in median expression for a given marker. Note the scale min and max differ for each marker. Black outlines termed “bubbles” and associated population labels derive from manual interpretation of cellular identity based on marker expression.

Cell rehydration and DNA intercalation

1. Take the tubes from -20°C.

2. Add 1 mL of staining media to each tube, pipet or briefly vortex to mix. Pellet the cells at 800 *x g* for 5 min at room temperature.
3. Check for cell pellets and discard supernatant by decanting (see Note 5). Repeat steps 2 and 3 once.
4. Add 200 µL of 1X PBS to each cytometry tube, pipet or briefly vortex to mix.
5. Add 4 µL of 50X Intercalator to each tube. Pipet or briefly vortex to mix the cells. Incubate cells with intercalator for 30 min at room temperature.
6. Add 1 mL of 1X PBS to each tube, pipet or briefly vortex to mix. Pellet the cells at 800 *x g* for 5 min at room temperature.
7. Check for cell pellets and discard supernatant by decanting (see Note 5). Repeat steps 6 and 7 once.
8. Add 1 mL of deionized water to each tube, pipet or briefly vortex to mix.
9. Strain the cells through cytometry tubes with 35-µm cell strainer cap. Cells are now ready for CyTOF.

Phospho-flow signaling in Kasumi-1 Human AML Cell Line

A phospho-flow signaling experiment can characterize the cellular heterogeneity of intracellular signaling responses to various stimuli among cells within a given cell population. Here, we are using Kasumi-1, a human AML cell line, to demonstrate the concept. This experiment includes addition of various stimuli (Fms-related tyrosine kinase 3 ligand (FLT3L), stem cell factor (SCF), interleukin- 3 (IL-3), and interferon-γ (IFNγ)) to the cells to stimulate signaling, followed by fixation and immunostaining with antibodies specific for phosphorylated proteins (Figure 4-4).

Cell culture

1. Culture Kasumi-1 cells according to the manufacturer's recommendation.
2. The day before the experiment, passage cells and resuspend the cells at 0.5–1.0 x 10^6 cells/mL in the recommended cell culture media. Keep the cells in a 37°C incubator with 5% CO_2.

Stimulation of cells

1. Label 6 cytometry tubes as "Intercalator only," "Unstimulated," "FLT3L," "SCF," "IL-3," and "IFNγ."

2. Take Kasumi-1 cells from the incubator and aliquot 1 mL of cell suspension into each cytometry tube.
3. Prepare the stimuli while the cells are resting in the incubator. Make a 50X solutions of each stimulus by diluting them in PBS (Table 4-3).
4. Add 20 µL of the corresponding 50X stimuli to each cytometry tubes. For tubes labeled “Intercalator only” and “Unstimulated,” do not add any stimulus. Place all the tubes back in the incubator and incubate for 15 min (see Notes 9 and 10).

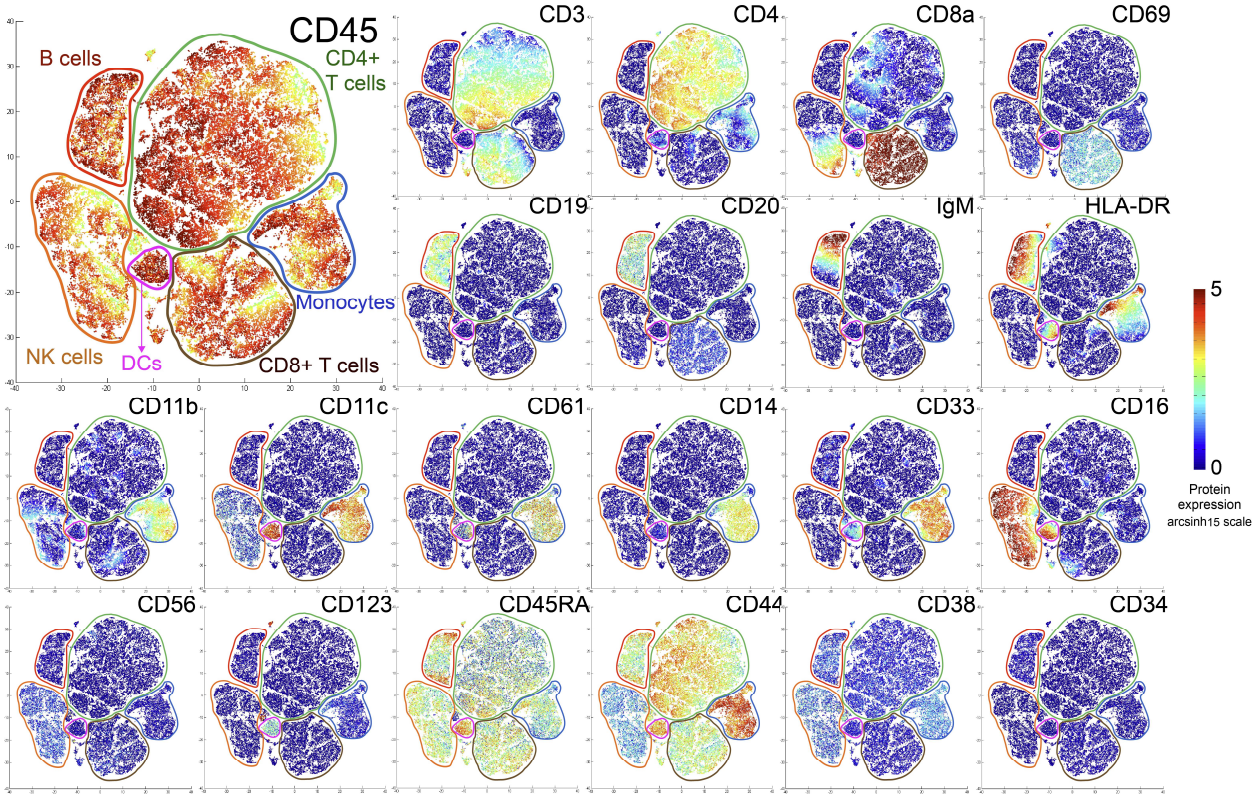


Figure 4-3 viSNE arranges cells in a 2D map representing phenotypic similarity. viSNE maps show healthy human PBMCs arranged according to phenotypic similarity for the 21 displayed markers measured by mass cytometry. The axes are unitless dimensions that reflect phenotypic differences. The distance between any two cells on the map corresponds to how similar or different the cells are from each other in high-dimensional space. Heat corresponds to the arcsinh_{15} fold difference in median expression for a given marker. This allows for a global single-cell view of every parameter in every cell. Cell populations can then be identified through various techniques, including automated clustering or manual analysis of well-characterized markers.

Fixation and permeabilization of cells

1. At the end of the stimulation timepoint, add 100 µL of 16% PFA to each cytometry tube and vortex the tube to mix (see Note 11). Allow cells to fi x for at least 5 min at room temperature.

2. Wash the cells by adding 1 mL of PBS to each tube and pipet or vortex to mix. Pellet the cell suspension at 800 *x g* for 5 min at room temperature. Check for cell pellets and discard supernatant by decanting (see Note 5).
3. Vigorously vortex the cell pellets in void volume to resuspend the cells.
4. Add 1 mL of ice-cold methanol to each tube (stored at -20°C or lower) to permeabilize the cells. Pipet or briefly vortex to mix (see Note 6) and store the tube at -20°C for at least 10 min

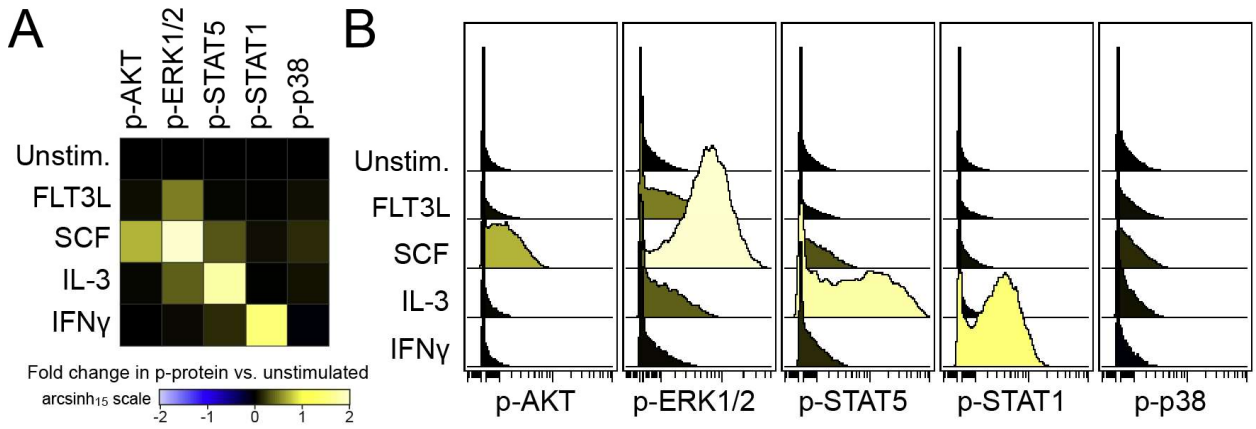


Figure 4-4 Mass cytometry phospho-flow analysis of AML cell signaling responses. (A) A heatmap compares phospho-protein abundance in Kasumi-1 AML cells following 15 minutes of stimulation by FLT3L, SCF, IL-3 or IFNγ. Each row in the heatmap or histogram overlay corresponds to a stimulation condition and each column corresponds to a phospho-protein (p-AKT, p-ERK1/2, p-STAT5, p-STAT1, or p-p38). Heat corresponds to the $arcsinh_{15}$ fold difference in median expression for a given marker compared to the unstimulated condition (Unstim). (B) The same data as in (A) are shown in histogram overlay format. Histogram overlays illustrate the distribution of marker expression within a population and highlight heterogeneity (e.g. p-STAT5 response to IL-3).

Intracellular phospho-protein staining and DNA intercalation of cells

1. Take the tubes from -20°C.
2. Add 1 mL of staining media to each tube, pipet or briefly vortex to mix. Pellet the cells at 800 *x g* for 5 min at room temperature.
3. Check for cell pellets and discard supernatant by decanting (see Note 5). Repeat steps 2 and 3 once.
4. Add 200 µL of 1X PBS to the tube labeled “Intercalator only.” Pipet or briefly vortex to mix and set the tube aside.
5. For the remaining tubes, resuspend the cell pellets with 95 µL of staining media per tube (see Note 8).

6. Transfer 95 µL of cell suspension to new cytometry tubes. Add intracellular antibodies from the signaling panel (Table 4-2) to each tube. Pipet or briefly vortex the cells with the antibodies to mix. Incubate the cells with the antibodies at room temperature for 30 min.
7. Add 1 mL of 1X PBS to each tube and pipet or briefly vortex to mix. Pellet the cells at 800 *x g* for 5 min at room temperature.
8. Check for cell pellets and discard supernatant by decanting (see Note 5). Repeat steps 7 and 8 once.
9. Add 200 µL of 1X PBS to all the tubes (except the tube labeled “Intercalator only”) and pipet or briefly vortex to mix.
10. Add 4 µL of 50X intercalator to all the tubes including the tube labeled “Intercalator only.” Pipet or briefly vortex to mix. Incubate the cells with intercalator for 30 min at room temperature.
11. Add 1 mL of 1X PBS to each tube, pipet or briefly vortex to mix. Pellet the cells at 800 *x g* for 5 min at room temperature.
12. Check for cell pellets and discard supernatant by decanting (see Note 5). Repeat steps 11 and 12 once.
13. Add 1 mL of deionized water to each tube, pipet or briefly vortex to mix.
14. Strain the cells from each tube through a cytometry tube with 35-µm cell strainer cap. Cells are now ready for CyTOF.

Table 4-2 Antibody panel for cell signaling

#	Target	Tag	Clone	µL[§]	Catalog #*
1	p-STAT5	150Nd	47	1	3150005A
2	p-AKT	152Sm	D9E	1	3152005A
3	p-STAT1	153Eu	58D6	1	3153003A
4	p-p38	156Gd	D3F9	1	3156002A
5	p-ERK1/2	167Er	D13.14.4E	1	3167005A

[§]µL of antibody per 100 µL total volume.
*Source was Fluidigm for all antibodies.

Table 4-3 Stimulation conditions

Stimulus	Time (min)	50X Dose [§]
FLT3L	15	10 µg/mL
SCF	15	25 µg/mL
IL-3	15	10 µg/mL
IFNγ	15	10 µg/mL

[§] Add 20 µL of 50X to 1 mL cells to yield 1X

High-dimensional data analysis

CyTOF simultaneously measures 30+ features per cell, providing a global view of biological events in individual cells. However, such high dimensional data is difficult to analyze effectively with conventional bivariate data visualization and serial manual gating alone. Computational tools help users visualize relationships in high dimensional space and separate rare cells into populations based on distinct, multi-feature phenotypes. As data analysis can be daunting in 25-dimensional datasets, this chapter compares analysis of the human PBMC cell surface immunophenotyping dataset by three methods: 1) traditional bivariate gating (see Note 12), heatmaps, and histogram overlays (using Cytobank (*301*), Figure 4-1, see Note 13) Spanning-Tree Progression Analysis of Density- Normalized Events (SPADE (*282*), Figure 4-2, see Note 14) visualization of t-Stochastic Neighbor Embedding (viSNE (*290*), Figure 4-3, see Note 15).

Notes

1. Avoid spilling Ficoll-Paque solution on the side of the 50 mL conical tubes.
2. Gently overlay blood and PBS mixture on top of the Ficoll-Paque solution. This can be done by adjusting the release speed of the serological pipet aid to minimum. Do not mix the Ficoll-Paque solution and the diluted blood sample since this would minimize PBMCs isolation.
3. To count live cells, mix 10 µL of trypan blue working solution with 10 µL cell suspension and count cells under a microscope using a hemocytometer.
4. To maximally preserve cell viability, cells should be frozen down at a slow rate close to −1°C/min. This can generally be achieved by placing cryopreservation vials in plastic containers that are immersed in isopropanol prior to transfer to -80°C. Commercially available products can also be used. Transfer viably cryopreserved cells from -80°C to liquid nitrogen within 48 h.
5. To avoid losing the cell pellet, decant only once and try to discard as much supernatant as possible. Avoid re-decanting after inverting the cytometry tubes to an upright position since this can loosen the cell pellet or detach it from the bottom of the tube.
6. Cell pellets may clump when methanol is added if they are not resuspended well. To prevent cell clumping, vigorously vortex the pellets in void volume prior to adding ice-cold methanol.

7. Following permeabilization ice cold methanol for 10 min, samples can be maintained long term at -20°C or -80°C for months without additional apparent changes to target epitopes.
8. The goal of this step is to resuspend each cell pellet to yield the total volume indicated in the protocol and to avoid losing cells in the following steps. Therefore, less staining media than what is indicated in the protocol might be needed depending on the void volume in the tube.
9. Stimuli should be added as soon as the cells are taken out of the incubator to avoid cooling the cells below 37°C.
10. In general, 15-min stimulations work well for many signaling responses, including the cytokine stimulations used here. However, if this protocol should be adapted for other stimulations, it is crucial to find optimal time points and concentrations for each signaling response.
11. Adding 100 µL of 16% PFA to each tube would yield a final concentration of approximately 1.6% PFA. Cells must be in single-cell suspension prior to adding PFA to avoid fixing cells in clumps.
12. Bivariate analysis: employs pairwise comparisons of markers to select, or gate, cell populations. Identification of cell subsets has traditionally relied on sequential bivariate gating. Here, we use the PBMC surface immunophenotyping data to demonstrate the concept (Figure 4-1) before using the high dimensional population identification tools SPADE and viSNE. In this example, established cell surface markers were used for sequential manual gating of well-characterized human blood mononuclear cell populations, including B cells, T cells, monocytes, dendritic cells, and natural killer (NK) cells.
13. Heatmaps and histogram overlays: are ways to visualize a statistic, such as median expression or fold change, across a large number of populations or experimental variables. Unlike bivariate analysis and viSNE, these tools do not provide a single- cell view of the data. Instead, these tools summarize a key population feature using color. Histogram overlays illustrate the distribution of a single marker for a population and shade the graphic based on the statistic, whereas heatmaps generally just show the color. To demonstrate these data visualization tools, we have used the Kasumi-1 phospho-flow dataset (Figure 4-4). Kasumi-1 phospho-flow experiment data are available online (www.cytobank.org/irishlab).
14. SPADE: With high content single cell experiments measuring more than 25 features per cell, sequential bivariate gating can overlook populations of cells with low or unexpected patterns of marker expression.

When applied to flow data, SPADE clusters cells based on phenotypic similarities in high-dimensional space. SPADE is a powerful complement to manual analysis approaches. After the initial clustering step, SPADE creates a minimum-spanning tree that illustrates the relationships between cell populations in two dimensions. Each node of the tree is a cluster of cells, and nodes can be colored to reflect the expression intensity of each marker (*299*). Here, we subjected the human PBMC immunophenotyping dataset to SPADE analysis to create a minimum-spanning tree (Figure 4-2). Known cell subsets in human PBMCs were identified as "bubbles" (Figure 4-2) through manual review of biomarker expression on different populations on the tree.

15. viSNE: is a computational tool that projects cells onto a two dimensional map such that the distances between cells in 2D reflect the distance between them in high-dimensional space (*294*). Thus, cells that are close together on a viSNE map are phenotypically similar for the markers used to create the map. Users then identify and characterize populations of cells based on the groups formed in the viSNE map. We performed a viSNE analysis of the human PBMCs immunophenotyping dataset (Figure 4-3). As with SPADE, expression intensity of each parameter can be visualized as heat intensity on the viSNE map, enabling identification of known cell subsets in human PBMCs. Importantly, viSNE allowed a global single-cell view of the data. viSNE enabled identification of rare cell subsets and cell populations without supervision or prior knowledge of the expected population distribution.

Acknowledgements

The authors thank P.B. Ferrell for use of Kasumi-1 mass cytometry data. This work was supported by the NIH/NCI R00 CA143231, NIH/NCI R25 CA136440 (K.E.D.), the Vanderbilt International Scholars Program (N.L.), and Vanderbilt-Ingram Cancer Center (VICC NIH/NCI P50 CA68485) pilot grants including a Young Ambassador award.

Conflict of interest disclosure: J.M.I. declares a competing financial interest (cofounder and board member of Cytobank Inc.).

CHAPTER 5

Single Cell Analysis of Human Tissues and Solid Tumors with Mass Cytometry

Authors: Nalin Leelatian*, Deon B. Doxie*, Allison R. Greenplate, Bret C. Mobley, Jonathan M. Lehman, Justine Sinnaeve, Rondi M. Kauffman, Jay A. Werkhaven, Akshitkumar M. Mistry, Kyle D. Weaver, Reid C. Thompson, Pierre P. Massion, Mary A. Hooks, Mark C. Kelly, Lola B. Chambless, Rebecca A. Ihrie, and Jonathan M. Irish

This work is presented as it appears in manuscript form in *Cytometry B Clinical Cytometry* 2017 (Leelatian, Doxie et al., 2017).

License 4263230224263, issued January 6, 2018

Preface

Single-cell mass cytometry is a high-throughput analytical platform that enables simultaneous quantification of 35 or more features per individual cells, in hundreds of thousands of cells in a single setting. This technology was initially introduced as a new-generation flow cytometry that enhances the power for single-cell feature characterization in human bone marrow and blood samples. The applications of mass cytometry, as well as fluorescence flow cytometry, in human solid tissues and tumors had been very limited since cell suspensions are required for the analyses. Prior to the development of the protocol described in this chapter, methods for obtaining single cells from solid tissues and tumors varied largely between different studies, even among similar tissue types. The goal of this development was to establish a tissue dissociation protocol that 1) maximizes cell viability, 2) preserves the cellular diversity of the original tissues, 3) obtains cells suitable for multi-platform flow cytometry analyses, and 4) can be applied in multiple human solid tissue and tumor types. Human glioblastoma was among the key human tissue types that were tested here. Therefore, the final protocol is a fundamental building block of the overall goal of this dissertation. Experimental parameters that could impact the key objectives mentioned above were systematically tested on multiple tissue types; these included 1) the use of mechanical dissociation (fine mincing), 2) types and combinations of tissue digestion enzymes, 3) duration of enzymatic dissociation, and 4) preservation of cellular diversity as assessed by multiple analytical approaches

including fluorescence flow cytometry, mass cytometry, and immunohistochemical stains. This systematic comparison led to the development of a universal tissue dissociation protocol which is applicable for obtaining viable single cells from human glioblastoma, human melanoma, human tonsils, and small cell lung cancer patient-derived xenografts, while maintaining the cellular diversity of the original tissue. Additionally, the cells derived from this protocol are suitable for both fluorescence flow and mass cytometry analyses, and can be further evaluated for use in other single-cell analysis platforms such as single-cell RNA-sequencing. This protocol was subsequently applied for the collection of viable single cells from adult glioblastoma surgical resections for cellular profiling, which is further elaborated in Chapter 6.

Abstract

Background: Mass cytometry measures 36 or more markers per cell and is an appealing platform for comprehensive phenotyping of cells in human tissue and tumor biopsies. While tissue disaggregation and fluorescence cytometry protocols were pioneered decades ago, it is not known whether established protocols will be effective for mass cytometry and maintain cancer and stromal cell diversity.

Methods: Tissue preparation techniques were systematically compared for gliomas and melanomas, patient derived xenografts of small cell lung cancer, and tonsil tissue as a control. Enzymes assessed included DNase, HyQTase, TrypLE, collagenase (Col) II, Col IV, Col V, and Col XI. Fluorescence and mass cytometry were used to track cell subset abundance following different enzyme combinations and treatment times.

Results: Mechanical disaggregation paired with enzymatic dissociation by Col II, Col IV, Col V, or Col XI plus DNase for 1 hour produced the highest yield of viable cells per gram of tissue. Longer dissociation times led to increasing cell death and disproportionate loss of cell subsets. Key markers for establishing cell identity included CD45, CD3, CD4, CD8, CD19, CD64, HLA-DR, CD11c, CD56, CD44, GFAP, S100B, SOX2, nestin, vimentin, cytokeratin, and CD31. Mass and fluorescence cytometry identified comparable frequencies of cancer cell subsets, leukocytes, and endothelial cells in glioma ($R = 0.97$), and tonsil ($R = 0.98$).

Conclusions: This investigation establishes standard procedures for preparing viable single cell suspensions that preserve the cellular diversity of human tissue microenvironments.

Introduction

In preparing single cell suspensions of healthy and malignant tissue, a common goal is to preserve viability while maintaining cellular diversity and preserving rare subsets. Multidimensional cytometry is well suited to this challenge because it can simultaneously characterize known cell types and reveal novel cell subsets (*11, 296*). Mass cytometry uses antibodies to quantify features of individual cells in primary tissues (*81, 302*) and has been applied to characterize cell subsets in human bone marrow, blood, and germinal center tissues as well as diverse murine tissues (*12, 83, 290*). However, mass cytometry remains relatively untested in the context of solid tumors. Fluorescence flow cytometry and fluorescence activated cell sorting (FACS) have been used to prospectively isolate functionally distinct cell subsets and suggest that mass cytometry analysis could help to further characterize solid tumors (*11*). A key goal of this study was to evaluate the suitability of different cell preparation techniques for mass cytometry and to develop standard procedures and quality controls that do not require measuring light scatter. An additional goal was to use the multidimensionality of mass cytometry to characterize preservation of cellular diversity under different solid tumor cell preparation techniques.

In this study, mechanical and enzymatic dissociation protocols were systematically tested on multiple types of fresh human solid tumors and tissues to develop an efficient, reliable method for dissociation and single-cell analysis by mass cytometry. Human tonsils and lymphoma tumors reliably dissociate with mechanical force alone and we have previously established protocols for their study by fluorescence cytometry (*15, 303*) and mass cytometry (*82*). Preparation techniques for tissue samples derived from intraoperative resections of gliomas (grades II-IV), melanomas, and patient derived xenografts (PDX) of small cell lung cancer (SCLC) were compared. As a control, the same techniques were applied to human tonsillar tissue. The abundance of different cell types, such as leukocytes, endothelial cells, epithelial cells, fibroblasts, and cancer cell subsets, was tested under these conditions. Established protein markers for expected cell types in tissues tested in this study were used in fluorescence cytometry (Table S5-1) and mass cytometry (Table S5-2, Table S5-4). The common markers were selected so that both rare and abundant cell types could be compared between mass and fluorescence cytometry. The additional markers in the mass cytometry panel provided a more comprehensive analysis of cell diversity.

Six enzymes for cell separation were selected to compare in solid tumor preparation protocols for mass cytometry analysis: HyQTase, TrypLE, collagenase (Col) II, Col IV, Col V, and Col XI. Enzyme choice was based in part on prior use in several solid tumor types and preparation of single cell suspensions containing cancer cell and immune subsets for FACS (*304-309*). DNase was also tested to determine its ability to enhance live cell yield from dissociation. Dissociation kinetics for enzyme combinations in distinct tissue types were also characterized. Finally, specific enzymes and dissociation duration times were selected based on optimal viable cell yield and representation of expected cell populations.

Materials and Methods

Tissue Sample Collection – All samples were obtained with patient consent, with Vanderbilt institutional review board (IRB) approval, in accordance with the Declaration of Helsinki, and were de-identified. Gliomas were intraoperative specimens from WHO grade II, III, or IV tumors (IRB #131870), collected in sterile normal saline. Melanomas were cutaneous and lymph node resections (IRB #030220), collected in MEM (Corning/Mediatech, Corning, NY) with 10% FBS + 1X Pen/Strep (GE Healthcare, Pittsburgh, PA). Small-cell lung cancer (SCLC) patient derived xenograft (PDX) samples were obtained as a gift from the Rudin laboratory (LX-22, (*310*)) and propagated solely as patient-derived xenografts in female athymic nude mice (HSD:Athymic Nude-Foxn1$^{nu/nu}$) obtained from Envigo with Vanderbilt institutional animal care and use committee (IACUC) approval. SCLC PDX were collected in RPMI 1640 (Corning/Mediatech, Corning, NY) plus 10% FBS+ 1X Pen/Strep. Glioma, melanoma, and SCLC PDX samples were transported at room temperature without delay to the laboratory and processing began within 30 minutes of collection from patients. Human tonsillar tissue was obtained from routine tonsillectomies (IRB #121328), collected in RPMI 1640 (Corning/Mediatech, Corning, NY) plus 10% FBS + 1X Pen/Strep, transported on ice, and processed within 4 hours of collection.

Mechanical and Enzymatic Dissociation – Sequential dissociation steps are described in detail in the main text. “Coarse mincing” indicates no additional mechanical dissociation of tissues (i.e. tissues were left as obtained intraoperatively). “Fine mincing” indicates additional mechanical dissociation using scalpels. Conventional mechanical dissociation of tonsils included fine mincing and immediate filtration of tissue through a 70 μm cell

strainer without additional enzymatic dissociation, as previously established (*15, 82, 303*). Dissociation enzymes were obtained from Sigma Aldrich (Darmstadt, Germany) (collagenase II, IV, V, and XI), ThermoFisher (Waltham, MA) (TrypLE-Express), and GE Healthcare (PA) (HyQTase). Collagenases were used at 1 mg/mL. HyQTase and TrypLE-Express were used at 1X according to the manufacturer's recommendations. DNase I (Sigma Aldrich) was used at a final concentration of 0.25 mg/mL. For conditions involving collagenases and no enzyme, cells were resuspended in recommended media for specific tissue types prior to adding indicated enzymes (gliomas, DMEM/F12 + Glutamax, (Gibco/Life Technologies, MA) with a defined hormone and salt mix (*311*) and 50 µg/mL gentamicin; melanomas, MEM with 10% FBS + 1X Pen/Strep; Tonsils and SCLC PDXs, RPMI 1640 + 10% FBS + 1X Pen/Strep. For dissociation conditions with HyQTase or TrypLE, tissues were dissociated in working concentrations of enzymes (with or without DNase), without addition of cell culture media, according to the manufacturer's recommendations. Enzymatic dissociations were performed in a 37°C incubator with 5% CO_2, with constant rocking on a nutating platform mixer at 18 rpm. Cells were then strained with 70 µm and 40 µm cell strainers prior to further analysis.

Quantification of cell viability – Cell suspensions obtained from different dissociation protocols were resuspended in corresponding cell culture media at volumes proportional to initial tissue weight (1 mL per 100 mg of tissue). Viable cells were quantified using Trypan Blue staining, normalized to the initial tissue weight, and reported as millions of live cells per gram of tissue.

Statistical testing – Enzyme conditions were compared as groups (horizontal lines) using a Student's t-test. The relationship between cell subset abundance measured by fluorescence or mass cytometry was compared using Pearson's correlation R and Spearman's rank correlation ρ (rho).

Cell line and cell culture – Jurkat cells were obtained from Utpal Dave at Vanderbilt, and were grown in RPMI 1640 + 10% FBS + 1X Pen/Strep as recommended. MeWo cells were obtained from Kimberly Dahlman and Jeffery Sosman with permission of Antoni Ribas (UCLA) and were grown in MEM + 10% FBS + 1X Pen/Strep, as recommended.

Flow cytometry - Cell suspensions were evenly divided for parallel phenotyping with fluorescence and mass cytometry according to the protocols below. Conditions were identical between mass and fluorescence cytometry with the exception of an additional staining step including saponin for mass cytometry analyses of glioma and melanoma that include SOX2. This type of saponin step has been established to have no significant impact on subsequent mass cytometry staining (*312*).

Fluorescence flow cytometry – For fluorescence cytometry, live surface staining was performed for surface marker detection (Supplemental Table S5-1). After washing with PBS and pelleting twice (at 200 *x g* for 5 min each time), cells were fixed with 1.6% paraformaldehyde (Electron Microscopy Services, Fort Washington, PA) for 10 min at room temperature, washed with PBS (HyClone Laboratories, Logan, UT), pelleted at 800 *x g,* and permeabilized with 100% ice-cold methanol (Fisher Scientific, Waltham, MA) at -20°C overnight following established protocols (*15, 278*). Cells were washed twice with cell staining media composed of PBS plus 1% BSA (Fisher Scientific, Waltham, MA) and pelleted at 800 *x g*. For each comparison, cells were stained in 100 µL staining media for 30 minutes at room temperature. All antibodies are listed in Supplemental Tables. Note that some antibodies that detect cell surface antigens (CD45-BV786, CD44-PE, and CD31-PE-Cy7) were used after fixation and methanol permeabilization due to concerns for stabilization of fluorochromes after methanol exposure. After staining, cells were washed twice with PBS, pelleted at 800 *x g*, and resuspended in PBS for analysis on a 5-laser LSRII (BD Biosciences, San Jose, CA) at the Vanderbilt Flow Cytometry Shared Resource.

Mass cytometry – Solid tissue cells obtained from the same dissociation conditions as those analyzed by fluorescence flow cytometry were stained live for cell surface markers, fixed, permeabilized, and washed as for fluorescence flow cytometry above and in concordance with established mass cytometry protocols (*81*). Permeabilization with 0.02% Saponin (Millipore, Darmstadt, Germany) in PBS was also included before methanol permeabilization of gliomas and melanomas as part of an optimized multi-step protocol that included detecting SOX2, which was not included in the fluorescence panel. Metal-tagged antibodies were used to stain cells in 100 µL cell staining media for 30 minutes at room temperature (Supplemental Table S5-4). After staining, cells were washed once with PBS, once with deionized water, pelleted at 800 *x g*, and resuspended in deionized

water containing normalization beads (Fluidigm). Standard bead-based normalization was used as previously described (*313*). Cells were collected on a CyTOF 1.0 at the Vanderbilt Flow Cytometry Shared Resource. Original data were normalized with MATLAB normalization software prior to further analysis using Cytobank (*301*) and established mass cytometry analysis methods (*314*). viSNE analysis was performed using 60,000 cCasp3^{-}HH3^{+} cells per sample. For glioma G-LC-15, the following markers were used for viSNE analysis: CD31, CD64, CD45RO, S100B, CD45, PDGFRα, SOX2, CD24, CD44, CD3, GFAP, αSMA, HLA-DR, and CD56. For tonsil T02-23, the following markers were used for viSNE analysis: CD4, IgD, CD16, CD45RO, CD45RA, CD45, CD27, CD86, CD33, CD11c, CD14, CD19, CD38, CD8, CD3, IgM, HLA-DR, and CD56. Samples of the same tissue type dissociated with different types of collagenase were analyzed simultaneously by viSNE.

Histone H3 testing – Healthy peripheral blood mononuclear cells (PBMCs) were used as controls in testing histone H3 as a nucleated cell marker for multiple flow cytometry platforms. PBMCs were stained live for detection of cell surface markers (Supplemental Table S5-2). After being washed twice with PBS, cells were then fixed with 1.6% paraformaldehyde and permeabilized with 100% ice-cold methanol for intracellular staining. Stained PBMCs were then evenly divided and half of the cells were stained with iridium at a final concentration of 0.25 μM in PBS for 15 minutes at room temperature. Cells were then washed once with PBS, once with deionized water, pelleted at 800 *x g*, and resuspended in deionized water containing normalization beads. Cells were collected as described above.

Results

Tissue dissociation with collagenase and DNase improved live cell yield

A matrix of dissociation conditions was tested to identify optimal protocols for multiple solid tumor types and tonsil controls (Figure 5-1, Figure S5-1, and Figure S5-2). The mechanical dissociation protocol (*see Materials and Methods*) was first compared to fine mincing of tonsil tissue followed by a 2-hour enzymatic dissociation with combinations of collagenase and DNase. For tonsils, a combination of fine mincing, collagenase, and DNase resulted in superior live cell yield per gram of tissue compared to conventional

dissociation methods (Figure S5-1, $p < 0.05$). Additionally, fine mincing of tonsils did not adversely affect cell viability (Figure S5-2) when compared to coarse mincing (left as obtained intraoperatively).

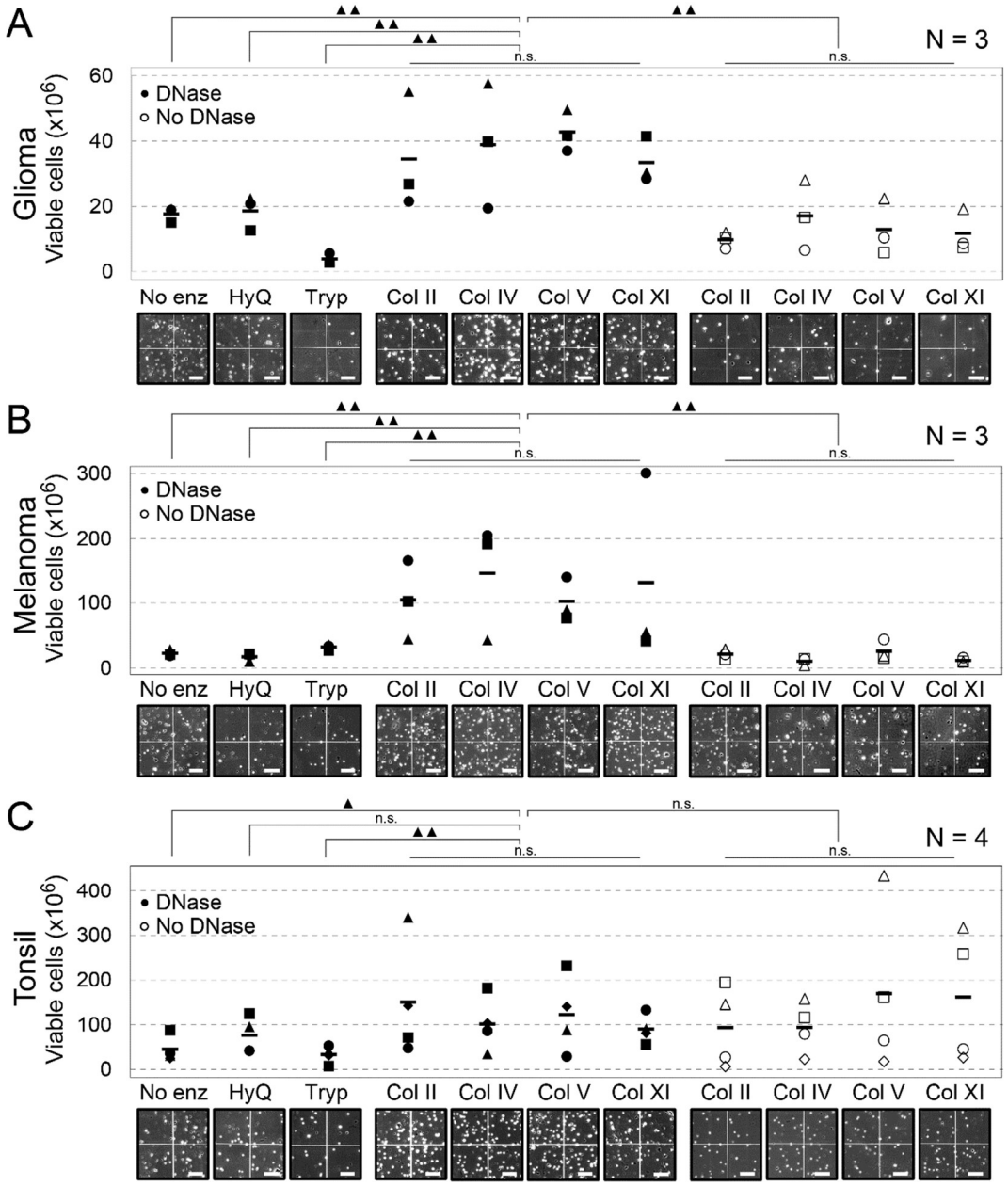


Figure 5-1 Collagenase plus DNase treatment provides better yield of live cells from three human tissues than no enzyme, TrypLE, HyQTase, or collagenase treatment alone. Graphs show millions of viable cells per gram yielded by different tissue preparation conditions following fine mincing for (A) gliomas, (B) melanomas, and (C) tonsil tissue. In addition to DNase (closed symbols), preparation enzymes tested included no additional enzyme (No enz), recombinant trypsin TrypLE (Tryp), HyQTase (HyQ), and collagenase (Col) II, IV, V, or XI. Average live cell yield is indicated for each condition by the thick horizontal line. Individual tissues or tumors are represented by different symbols. Representative trypan blue stained images are depicted under each condition. Scale bars = 100 µm. Symbols denote not significant (n.s.), $p < 0.05$ (♦), or $p < 0.01$ (♦♦). N indicates number of separate individual sample donors tested under each condition for each tissue type.

Since freshly resected tissues and tumors frequently differ in size, fine mincing was selected as an initial mechanical dissociation step for all tissue types. To determine the optimal enzymes for disaggregation of human gliomas, seven different enzymatic conditions were tested for their ability to yield live, single cells (Figure 5-1A, N = 3). Intraoperative samples of gliomas were finely minced and incubated with a cocktail of DNase plus one enzyme (either HyQTase, TrypLE, Col II, Col IV, Col V, or Col XI) or DNase alone for 2 hours at 37°C, with continuous rocking. Increased live cell yield per gram of tissue was seen in conditions containing collagenase and DNase as compared to other conditions ($p < 0.01$). Additionally, DNase plus collagenase improved live cell yield for glioma compared to collagenase alone ($p < 0.01$). No significant differences were observed in live cell yield per gram of glioma tissue between conditions using different types of collagenases plus DNase. High-resolution images of trypan blue stains are shown in Figure S5-3.

The same matrix of conditions was tested on intraoperative samples of human melanomas (Figure 5-1B, N = 3). As with glioma, no significant difference in live cell yield was observed between different types of collagenases, and viable cell yields were highest in conditions containing collagenases and DNase ($p < 0.01$). In freshly resected tonsils (Figure 5-1C, N = 4), collagenases with DNase gave a higher live cell yield than either DNase alone ($p < 0.05$) or TrypLE plus DNase ($p < 0.01$). However, collagenases with DNase did not significantly differ from HyQTase with DNase, and addition of DNase did not result in higher or lower live cell yield, in tonsil dissociation.

Enzymatic dissociation with collagenase and DNase for 1-2 hours provided superior live cell yields

While incubation in enzyme solutions enhanced tissue disaggregation (Figure 5-1 and Figure S5-1), excessive incubation might adversely affect cell viability. A dissociation time course was performed on intraoperative glioma specimens to determine the optimal time point for highest live single cell yield (Figure 5-2A). Gliomas were finely minced and incubated in collagenases plus DNase for 30 minutes, 1 hour, 2 hours, 4 hours, or 6 hours (Figure 5-2A, N = 3). Live cell yield per gram of glioma tissue significantly decreased after 4 hours of enzymatic dissociation with Col II, Col V, or Col XI plus DNase compared to earlier time points (Col II and Col XI, $p < 0.001$; Col V, $p < 0.05$), whereas it significantly decreased after 6 hours of dissociation with Col IV plus DNase (Figure 5-2A, $p < 0.001$).

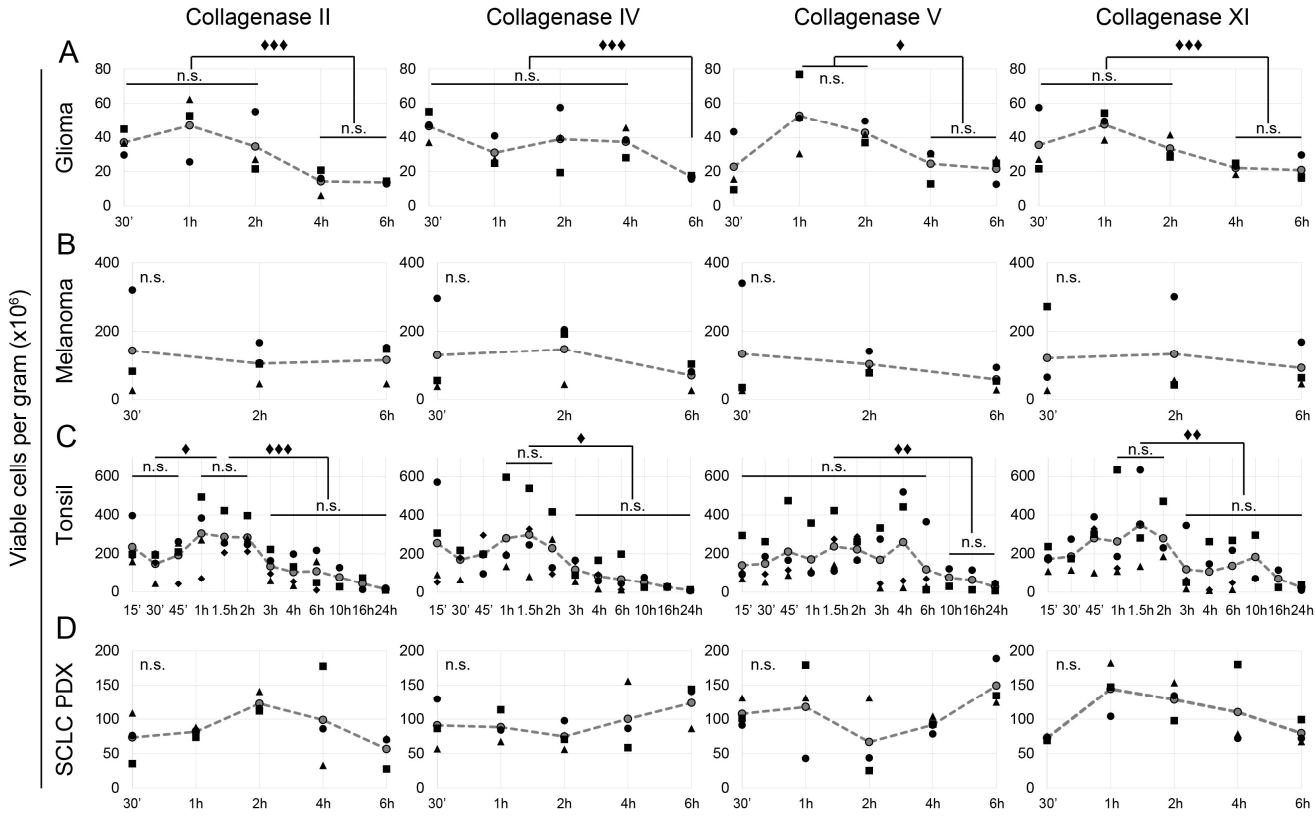


Figure 5-2 Collagenase and DNase treatment for 1 or 2 hours provided better overall live cell yield than other times. (A) Gliomas (N = 3) were finely minced and treated for varying times with DNase and either Col II, Col IV, Col V, or Col XI. Yield of live single cells (x10^6) per gram was quantified from Trypan blue images after 30 minutes (’), 1 hour (h), 2h, 4h, and 6h (filled symbols). Individual tissues or tumors are represented by different symbols. Grey circles mark average yield and are connected with dashed lines to indicate dissociation kinetics. Dissociation kinetics were similarly assessed for (B) melanomas (N = 3), (C) tonsil tissue (N = 4, except for 10h, 16h, 24h where N = 2), and (D) SCLC PDX tumors (N = 3) (D). Symbols denote not significant (n.s.), $p < 0.05$ (♦), $p < 0.01$ (♦♦), or $p < 0.001$ (♦♦♦).

Dissociation kinetics of tonsils were also characterized for time points ranging from 15 minutes to 24 hours (Figure 5-2C). Finely minced tonsils dissociated with Col II plus DNase for 1-2 hours gave higher live cell yield when compared to earlier time points ($p < 0.05$) as well as later time points ($p < 0.001$). Similarly, viable cell yield decreased significantly after 1-2 hours when tonsils were dissociated with either Col IV or Col XI plus DNase (IV, $p < 0.05$; IX, $p < 0.01$). Live cell yield from the combination of Col V and DNase also decreased after 6 hours ($p < 0.01$). Live cell yield from intraoperative melanoma specimens and SCLC patient-derived xenografts (PDXs) did not significantly decrease after 6 hours of dissociation, regardless of the type of collagenase (Figure 5-2B and 3-2D).

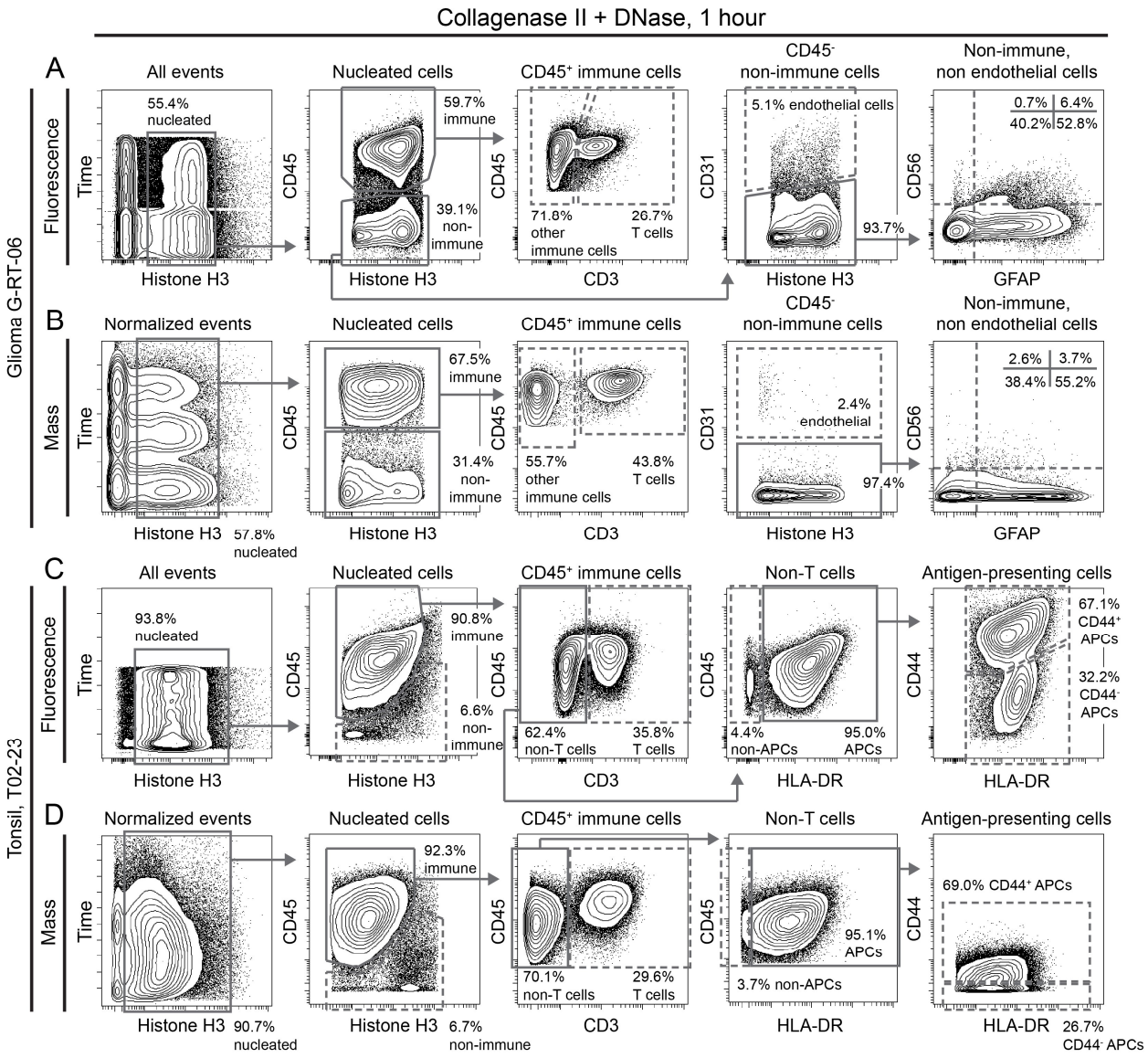


Figure 5-3 Frequency of cell types in glioma, and tonsil tissue quantified by fluorescence and mass cytometry. Biaxial plots show gating for established cell types in human tumors and tissues prepared using Col II plus DNase for 1 hour. Nucleated cells (HH3$^+$) were identified. Immune cells (CD45$^+$), T cells (CD45$^+$ CD3$^+$), APCs (CD45$^+$ CD3$^-$ HLA-DR$^+$), endothelial cells (CD31$^+$ CD45$^-$), and non-immune non-endothelial cells (CD45$^-$ CD31$^-$) were also found. (A) In fluorescence cytometry analysis of glioma from an individual patient (G-RT-06), CD56 (NCAM) and GFAP expression are shown for CD45$^-$ CD31$^-$ cells. (B) A similar gating scheme was applied to mass cytometry data from G-RT-06. In tonsil tissue from donor T02-23, CD44 and HLA-DR are shown for CD45$^+$ CD3$^-$ HLA-DR$^+$ cells, for both fluorescence (C) and mass cytometry analysis (D). Frequency of terminal populations (dashed gates) was compared between fluorescence and mass cytometry in Table 5-1.

Testing histone H3 as a nucleated cell marker compatible with mass and fluorescence cytometry

An anti-Histone H3 (HH3) monoclonal antibody was next tested as a potential marker of nucleated cells that would function equivalently in fluorescence and mass cytometry. Jurkat T leukemia cells gated as intact

cells were 98.9% positive for HH3 in fluorescence cytometry (Figure S5-4A). Similarly, when Jurkat cells were gated first as $HH3^+$, they were observed to be >99.8% intact cells when gated using light scatter in fluorescence cytometry (Figure S5-4B). Peripheral blood mononuclear cells (PBMCs) were used to further test HH3 because PBMC have well-studied cell subsets that have been extensively characterized by both fluorescence and mass cytometry (*81, 300, 315*). PBMCs from a healthy donor were stained with a panel of 16 mass-tagged antibodies (Table S5-2). Frequencies of known cell subsets identified by biaxial gating were closely correlated in the same mass cytometry dataset gated using HH3 or established iridium-based gating (Figure S5-5, Pearson correlation R = 1.00, Spearman rank of subset abundance rho (ρ) = 1.00, Table S5-3), supporting the use of HH3 as nucleated cell marker across multiple flow cytometry platforms.

Assessment of cell subset diversity in solid tumor following collagenase and DNase treatment

Two- to seven-dimensional fluorescence flow cytometry has been used extensively to characterize presence and abundance of cell subsets in patient-derived tissues. Glioma cell subsets consistent with those documented in prior studies were present after a 1-hour dissociation with DNase plus Col II using fluorescence flow cytometry (Figure 5-3A, Col II). In glioma sample G-RT-06, 55.4% of all events were identifiable as intact nucleated cells based on HH3 staining. $CD45^+$ immune cells comprised 59.7% of live intact cells, which included $CD3^+$ T cells (26.7%) as well as other immune cell types (71.8%). Presence of immune cell subsets was confirmed with immunohistochemistry (IHC) staining of formalin-fixed paraffin-embedded (FFPE) sections of the same sample (Figure S5-6). Additionally, $CD31^+$ endothelial cells were detected (5.1% of non-immune cells), as were cell subsets that differentially expressed CD56 (NCAM) and GFAP. The abundance of nucleated cells and other known cell subsets was similar between different collagenase types (Figure S5-7).

To determine if cells derived from dissociations using collagenase and DNase were suitable for mass cytometry analysis, cells obtained from intraoperative glioma resections (G-RT-06) were stained with 16 isotope-labelled antibodies (Table S5-4). Histone H3 was used to identify intact nucleated cells. A biaxial analysis sequence similar to that used for fluorescence flow cytometry analysis was used for comparison of subset abundance identified by these two cytometry platforms (Figure 5-3B). A strong correlation of cell subset abundance between the two methods was observed and quantified (Table 5-1; Pearson's R = 0.97, Spearman's

rank ρ = 0.93). Similar comparisons were performed in tonsils (Figure 5-3D). Strong correlations of subset abundance between the two different cytometry platforms was also observed in tonsil (Table 5-1; Pearson's R = 0.98, Spearman's rank ρ = 0.90).

Subsets of immune cells in tonsils were also identified by fluorescence flow cytometry, including $CD3^+$ T cells, $CD44^+$ antigen-presenting cells (APCs), $CD44^-$ APCs, and additional immune and non-immune cell types, as expected (Figure 5-3C, Col II). Abundance of tonsil cell subsets was similar between dissociations using different collagenase types (Figure S5-8). Single cells obtained from resected melanomas (MP-04) and a melanoma cell line, MeWo, were analyzed by fluorescence flow cytometry and were observed to have intrinsic auto-fluorescence on some channels, whereas glioma and tonsil samples studied here showed no auto-fluorescence (Figure S5-9). Mass cytometry was next used to study melanoma tumors (Figure S5-10). $CD45^+$ immune subsets, including $CD45^+HLA\text{-}DR^+$ antigen-presenting cells, $CD45^+CD3^+$ T cells ($CD8^+$ and $CD8^-$, and $CD45RO^+$ memory and $CD45RO^-$ non-memory), as well as $CD31^+$ endothelial cells were identified in melanoma. Additionally, among the non-immune, non-endothelial cells, other cell subsets were identifiable by nestin, SOX2, CD44, HLA-ABC, vimentin, and cytokeratin.

Table 5-1 Mass and fluorescence cytometry detect comparable frequencies of cell types in glioma, melanoma, and tonsil tissue

Table 1 - Mass and fluorescence cytometry detect comparable frequencies of cell types in glioma, melanoma, and tonsil tissue

Tissue	Subsets	Percent†		R*	Rank		ρ§
		MC	FC		MC	FC	
Glioma G-RT-06	$CD56^+GFAP^-$	0.7	0.1		1	1	
	Endothelial cells	1.0	0.8		2	3	
	$CD56^+GFAP^+$	1.1	0.2		3	2	
	$CD56^-GFAP^-$	11.4	6.1	0.97	4	4	0.93
	$CD56^-GFAP^+$	18.1	15.4		5	6	
	T cells	27.8	15.1		6	5	
	Other immune cells	39.3	30.9		7	7	
Tonsil T02-23	Non-APCs	2.1	2.5		1	2	
	Non-immune	5.2	2.3		2	1	
	$CD44^-$ APCs	16.2	17.7	0.98	3	3	0.90
	T cells	27.1	33.8		4	4	
	$CD44^+$ APCs	40.3	39.1		5	5	

† Frequency of terminal populations (dashed gates in Figure 3, percent) was measured by mass cytometry (MC) or fluorescence cytometry (FC)
* Pearson's correlation coefficient R
§ Spearman's ranked ρ

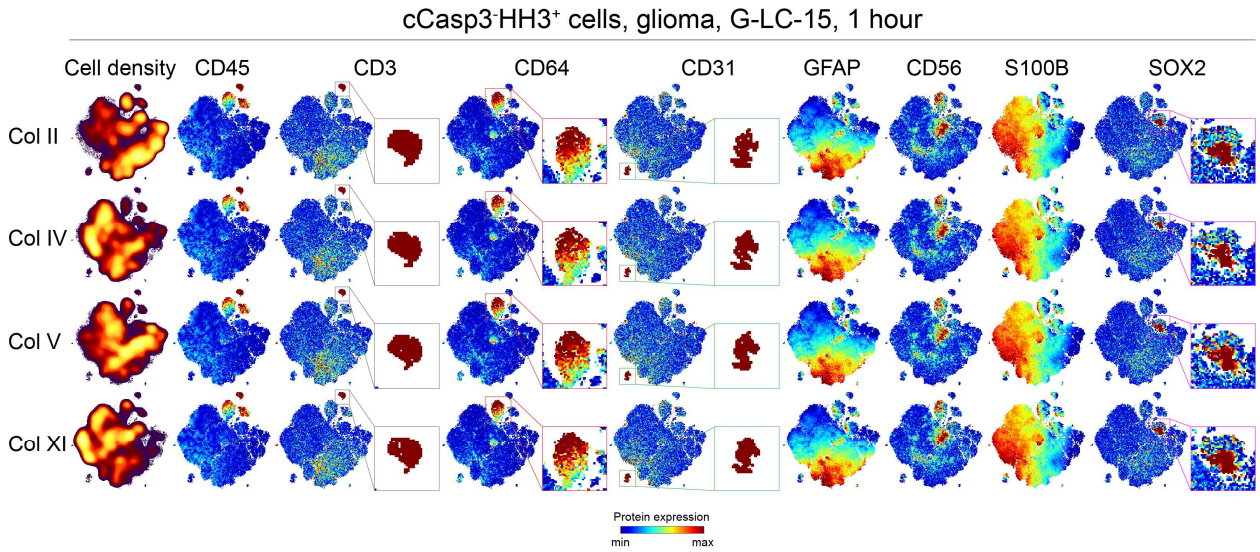


Figure 5-4 Treatment of a glioma with different collagenases yielded comparable cell subset frequencies. viSNE plots show non-apoptotic nucleated cells (cCasp3$^-$HH3$^+$) from glioma G-LC-15 obtained following 1-hour treatment with DNase plus either Col II, VI, V, or XI. Heat plots indicate cell density (first column) or expression of 8 proteins indicating cell type (CD45, CD3, CD64, CD31, GFAP, CD56, S100B, and SOX2). viSNE mapping was run together. Color-coded inserts next to the complete map highlight cell subsets (grey = CD45$^+$CD3$^+$ T cell, 0.9 ± 0.1%; red = CD45$^+$CD64$^+$ microglia, 3.9 ± 1.0%; green = CD45$^-$CD31$^+$ endothelial cells, 0.7 ± 0.2%; fuchsia = SOX2$^+$ stem-like cells, 1.2 ± 0.5%).

To characterize the effects of different types of collagenase on the presence of cell subsets, mass cytometry analysis of cells derived from glioma dissociation at one hour with DNase plus either Col II, Col IV, Col V, or Col XI was performed (Figure 5-4). This time point was selected based on its highest live cell yield across multiple tissue types, shown above. viSNE analysis (*290*) was used to compare cell subsets in the different dissociation conditions. Known cell subsets in gliomas were present in all conditions, including CD45$^+$ immune cells (CD3$^+$ T cells, and CD64$^+$ microglia), CD45$^-$CD31$^+$ endothelial cells, GFAP$^+$ glial cells, S100B$^+$ astrocyte-like cells, and SOX2$^+$ stem-like cells. Established cell subsets were also observed in tonsil specimens dissociated for one hour in all types of collagenase (Figure S5-11). As expected, the majority of cells were CD45$^+$ immune cells. Additionally, known immune subsets, including CD3$^+$CD4$^+$ helper T cells, CD3$^+$CD8$^+$ cytotoxic T cells, CD19$^+$IgD$^+$ naïve B cells, and CD19$^+$CD27$^+$ memory B cells, were identified. These findings suggest that both mass cytometry and fluorescence cytometry identify key cell subsets in glioma and tonsil dissociated with collagenase plus DNase.

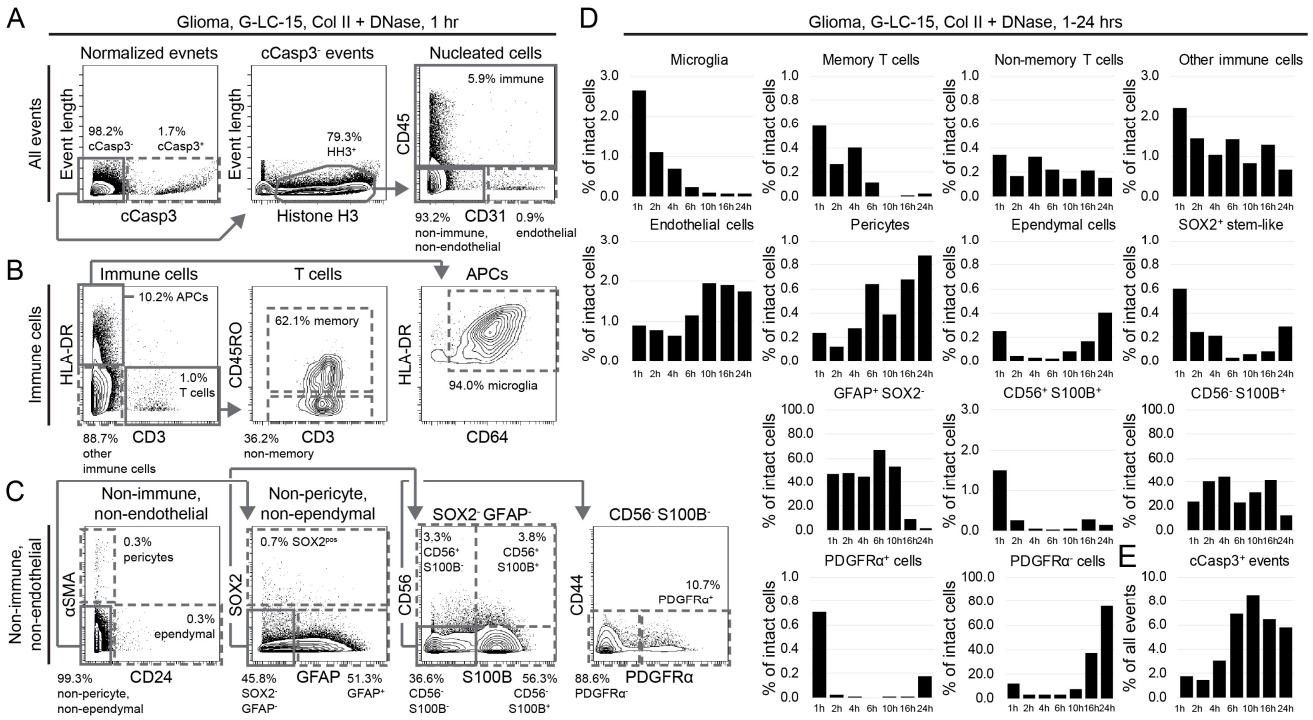


Figure 5-5 Enzymatic treatment times longer than one hour differentially impact glioma tumor cell subsets. Biaxial plots and bar graphs quantify cell subsets measured in mass cytometry analysis of glioma G-LC-15 after varying treatment times with collagenase II and DNase. (A) Gating for apoptotic cells (cCasp3$^+$) and live immune cells (cCasp3$^-$CD45$^+$), endothelial cells (cCasp3$^-$CD31$^+$), and non-immune, non-endothelial cells (cCasp3$^-$CD45$^-$CD31$^-$). (B) Subsets of glioma tumor-infiltrating immune cells were identified, including microglia (HLA-DR$^+$CD64$^+$), CD45RO$^+$ and CD45RO$^-$ subsets of CD3$^+$ T cells, and other immune cells. (C) Pericytes (CD45$^-$CD31$^-$αSMA$^+$), ependymal cells (CD45$^-$CD31$^-$CD24$^+$), SOX2$^+$ stem-like cells (CD45$^-$CD31$^-$SOX2$^+$), GFAP$^+$ cells (CD45$^-$CD31$^-$GFAP$^+$), and astrocyte-like cells (CD45$^-$CD31$^-$S100B$^+$) were quantified as subsets of G-LC-15. (D) Gating for cell types as in (A-C) was applied to mass cytometry analysis of cells from G-LC-15 treated with collagenase II plus DNase for 1, 2, 4, 6, 10, 16, or 24 hours. (E) Percentage of apoptotic cells as in (A) was measured for each dissociation time, as in (D).

Longer dissociation times led to disproportionate cell death and loss of cellular diversity

To determine if the abundance of cell subsets changed over time with enzymatic dissociation, time course dissociations of glioma sample, G-LC-15 (Figure 5-5), and tonsil sample, T02-23 (Figure S5-13), with DNase plus Col II were performed. Cell subsets were identified using sequential biaxial analysis and given the indicated labels following expert review. Apoptotic cells, defined by high cCasp3 signal, were excluded from subsequent cell subset quantification (Figure 5-5A). Within the population of HH3$^+$ nucleated cells, marker analysis identified CD45$^+$ immune cells and CD31$^+$ endothelial cells. Known subsets of immune cells were present within the CD45$^+$ population, including microglia (HLA-DR$^+$CD64$^+$), memory T cells (CD3$^+$CD45RO$^+$), and non-memory T cells

($CD3^+CD45RO^-$) (Figure 5-5B). Within the $CD45^-CD31^-$ population, pericytes (αSMA^+) and ependymal cells ($CD24^+$) were seen, as well as rare $SOX2^+$ stem-like cells, $GFAP^+$ glial cells, $PDGFR\alpha^+$ cells, and $S100B^+$ astrocyte-like cells (Figure 5-5C). Quantification of these cell subsets was performed in samples obtained from different dissociation durations to characterize maintenance and enrichment of cell subsets over time (Figure 5-5D). Among immune cells, a decrease in microglia (after 1 hour) and memory T cells (after 4 hours) was noted, whereas the proportion of non-memory T cells appeared to remain constant over the full range of times tested. $SOX2^+$ stem-like cells were most abundant after 1 hour of dissociation and decreased thereafter. Even though the proportion of $SOX2^+$ stem-like cells increased at 24 hours after dissociation, the overall decrease in viable cells after 4-6 hours of glioma dissociation (Figure 5-2A) suggested an overall loss in total viable stem-like cells at later time points. Additionally, the abundance of $GFAP^+$ glial-like cells (known to be present in most gliomas, Figure S5-12) remained constant during the initial 10 hours of dissociation and showed a decrease after 16 hours. This suggested that longer dissociation depletes key cell subsets in glioma. Most of the nucleated, non-apoptotic cells that remained after 24 hours of dissociation lacked expression of the key cell identity markers used in this study. Moreover, the abundance of $cCasp3^+$ apoptotic events also increased over time (Figure 5-5E).

A similar time course strategy was applied to tonsil specimen dissociation (Figure S5-13A). A decrease in the abundance of most immune cell subsets was observed at all time points greater than 1 hour of dissociation with Col II plus DNase (Figure S5-13B). This decrease affected all T cell subsets, plasma cells/blasts, germinal center B cells, class-switched memory B cells, and unswitched memory B cells. Notably, abundance of naïve B cells remained constant during the initial 6 hours of dissociation and only decreased after 10 hours. $CD27^-IgD^-$ B cells increased in abundance at time points extending to 6 hours, followed by a decrease at 10 hours. Dendritic cells were the only immune cell subsets that continued to increase in abundance at 24 hours of dissociation. As expected, longer dissociation times likewise led to an increase in apoptotic cells (Figure S5-13C).

Discussion

A common protocol of collagenase II plus DNase for 1 hour was identified as effective for preparing viable and mass cytometry compatible single cell suspensions of all tested human solid tumors and healthy tissues.

Multiple types and combinations of enzymes and dissociation kinetics were compared in freshly resected patient-derived tissues and patient-derived xenografts. Unexpectedly, collagenase also resulted in greater viable cell yield from tonsils when compared to the conventional dissociation method (Figure 5-1 and Figure S5-1), indicating that the protocol for preparation of tonsil and lymphoma tumors could be further refined. DNase clearly improved live cell yield from gliomas and melanomas and is strongly recommended for tissues where there may be ongoing cell death. Even though DNase was not observed to improve tonsil dissociation, DNase also did not adversely affect tonsil cell viability. Live cell yield from glioma dissociation began to decrease after 4-6 hours. However, live cell yields from melanoma and SCLC PDX were constant throughout the dissociation duration tested (6 hours) for all types of collagenase. In contrast, live cell yield from tonsils was maximal during the initial 2 hours of dissociation, except for collagenase V, which significantly decreased only after 6 hours.

Critically, dissociation of tissue using combined collagenase and DNase preserved cellular diversity, as seen by mass cytometry and standard fluorescence flow cytometry (Figure 5-3 and Figure 5-4). At one hour after dissociation, known cell subsets were present as expected in each of the tested tissue types. These included immune cells in tonsil, infiltrating immune cells in glioma and melanoma, and tissue-specific cell subsets, such as cancer cell subsets, endothelial cells, glial cells, pericytes, and stem-like cells in gliomas. A difference in abundance of T cells observed between fluorescence and mass cytometry was determined to be due to use of different anti-CD3 antibody clones, as has been previously reported (*315, 316*). While immune cells and $GFAP^+$ cells in glioma were confirmed with IHC stains and observed to be in relatively close agreement between IHC and flow cytometry, small tissue sections and sections that do not sample all tumor regions may over- or under-represent cell subsets or overlook rare cells. The quantitative analysis of a large number of whole cells by multidimensional flow cytometry (10^5 to 10^7) provides a strong complement to the location information provided by imaging cytometry (*50*).

Longer dissociation times led to increased cell death and disproportionate depletion of cell subsets in both gliomas and healthy tonsil. Additionally, the abundance of glial/astrocyte-like cells, as well as rare stem-like cells in glioma, decreased over time. Even though the proportions of some cell subsets increased at later time points (endothelial cells, pericytes, $SOX2^+$ stem-like cells in gliomas, and dendritic cells and $CD27^-IgD^-$ B cells in tonsils), the significant increase in cell death over a long period of dissociation would result in an overall

decrease in total yield of those cell types. Comparison of the results from gliomas, melanomas, SCLC xenografts, and tonsil tissue indicates that different tissues may be sensitive to prolonged enzymatic digestion. Dissociation conditions should be evaluated closely and carefully matched to tissue type and study goals. However, based on the results here, no more than 1 hour of dissociation is recommended unless the protocol is being optimized for a specific purpose. In future single-cell-level studies of other complex solid tissues, it will be critical to identify conditions that efficiently generate single-cell suspensions while preserving rare subpopulations of interest. Additionally, cell viability stains such as Cisplatin can be included in future mass cytometry experiments that aim to test cell functions like signaling, proliferation, viability, or cytokine production (*286*).

Acknowledgements

The authors thank Charlie Rudin for the LX22 PDX line, Utpal Dave for Jurkat T leukemia cells, and Kimberly Dahlman and Jeffery Sosman for MeWo cells.

Funding sources: Study and researchers were supported by NIH/NCI R00 CA143231 (J.M.I.), R25 GM062459 (D.B.D.), T32 CA009592 (D.B.D.), F31 CA199993 (A.R.G.), F31 HD007502 (J.S.), the Vanderbilt-Ingram Cancer Center (VICC, P30 CA68485), the Vanderbilt International Scholars Program (N.L.), a Vanderbilt University Discovery Grant (J.M.I. and N.L.), a VICC Provocative Question award (M.C.K. and J.M.I.), R01 NS096238 (R.A.I) and VICC Ambassadors awards (J.M.I. and R.A.I.).

Conflict-of-interest disclosure: J.M.I. is co-founder and board member and Cytobank Inc. and received research support from Incyte Corp.

Supplemental data

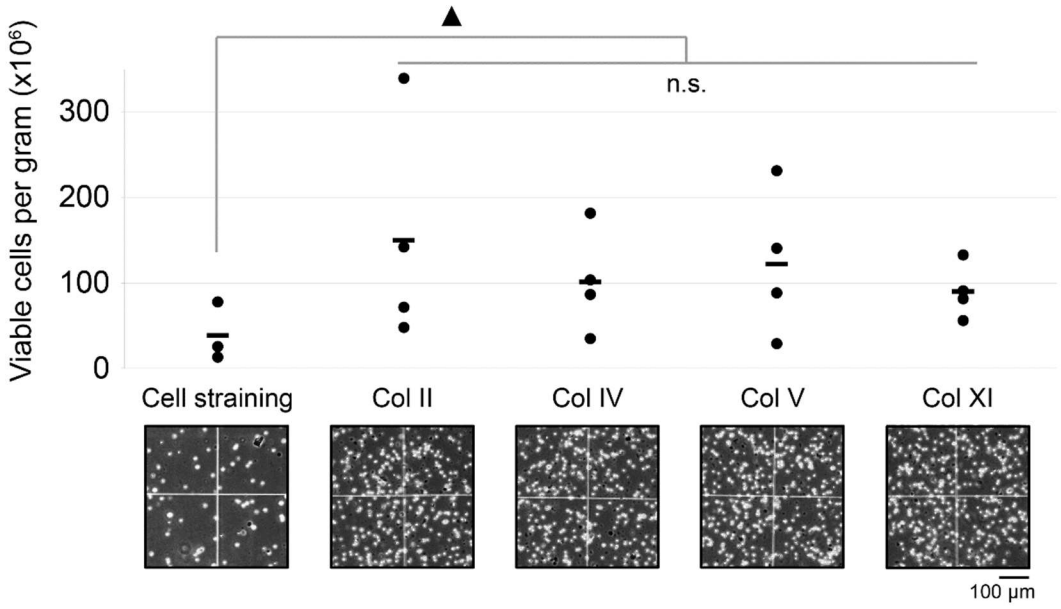


Figure S5-1 Tonsil dissociation with fine mincing and enzymes gave higher live cell yield compared to conventional dissociation method. Enzymatic dissociations of tonsils by fine mincing and incubation with collagenases and DNase (2 hours) were compared to traditional mechanical dissociation (see *Materials and Methods*). Viable cells ($x10^6$) per gram of tissue were quantified. Average live cell yield of each condition are shown as horizontal lines. Scale bars = 100 µm. (n.s. = not significant; Col = collagenase; no enz = no enzyme; HyQ = HyQTase; Tryp = TrypLE). (Cell straining, N = 3; Col, N = 4). (▲ $p < 0.05$)

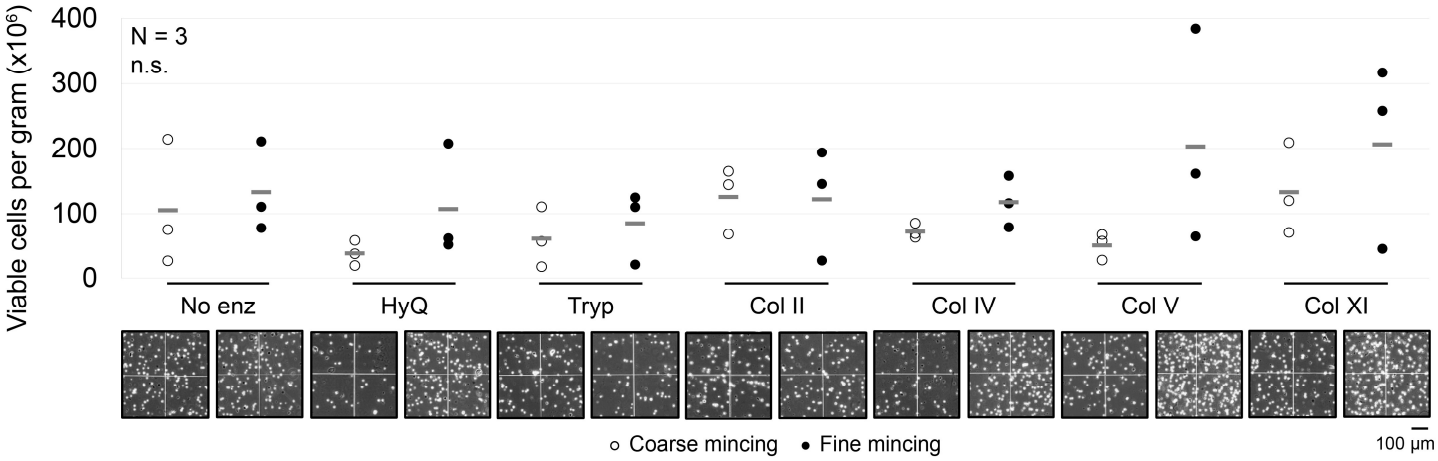


Figure S5-2 Fine mincing did not adversely affect live cell yield from tonsil dissociation. Live cell yield (x10^6) per gram of tonsils obtained by coarse (open circles) and fine (filled circles) mincing of tonsils were compared after a 2-hour incubation in different enzyme combinations. All conditions contained DNase. Average live cell yield of each condition are shown as horizontal lines. Representative Trypan Blue stained images of each conditions are shown. Scale bars = 100 µm. (n.s. = not significant; Col = collagenase; no enz = no enzyme; HyQ = HyQTase; Tryp = TrypLE).

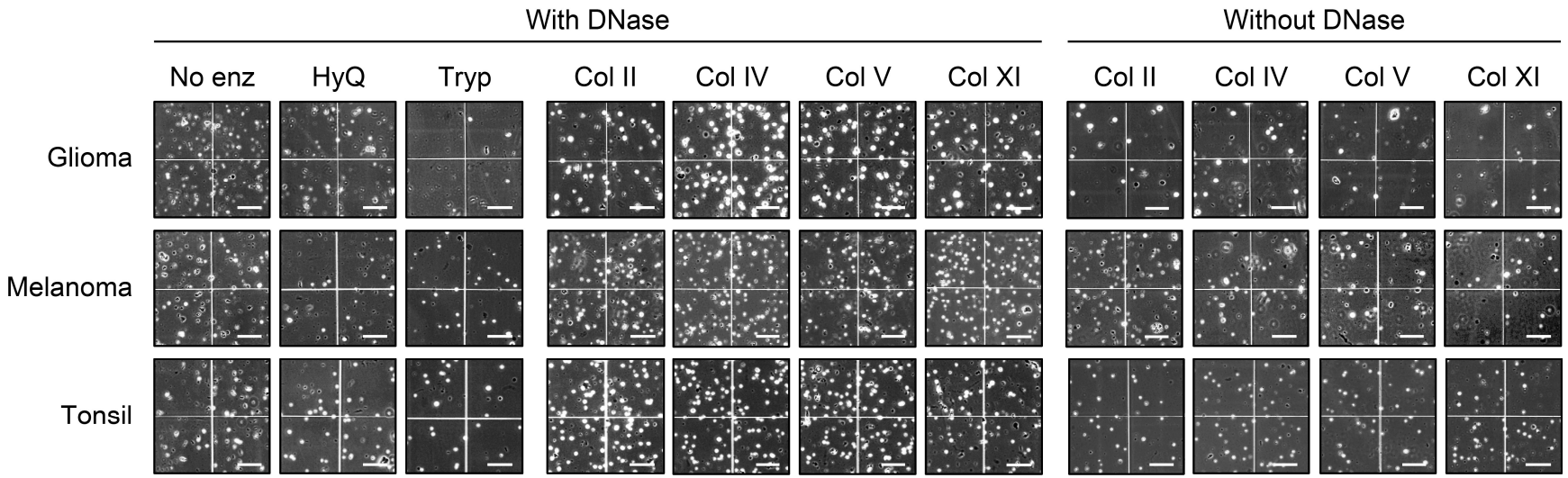


Figure S5-3 Trypan Blue staining allowed quantification of live cell yield. Higher resolution of Trypan Blue stains shown in Figure 1. No enz = no enzyme; HyQ = HyQTase; Tryp = TrypLE; Col = collagenase. Scale bars = 100 µm.

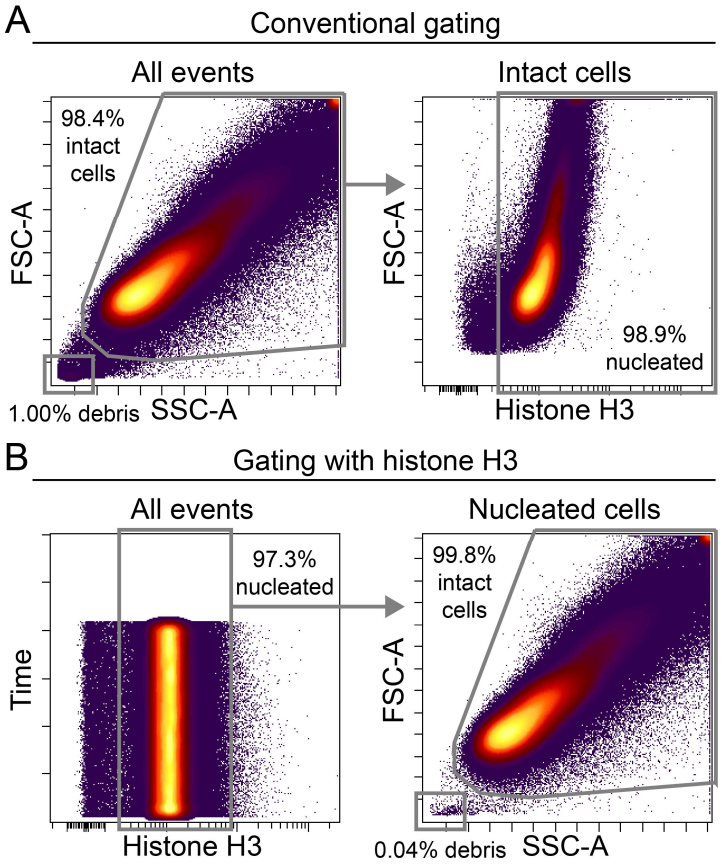


Figure S5-4 Histone H3 effectively identifies intact Jurkat cells via fluorescence flow cytometry. (A) Intact Jurkat T cells (98.4%) were identified by conventional biaxial analysis using SSC-A (x-axis) and FSC-A (y-axis). 98.9% of intact Jurkat cells were $HH3^{+}$ (nucleated). (B) Sequential gating starting with HH3 identified 97.3% nucleated events, 99.8% of which were defined as intact cells based on FSC-A and SSC-A biaxial analysis.

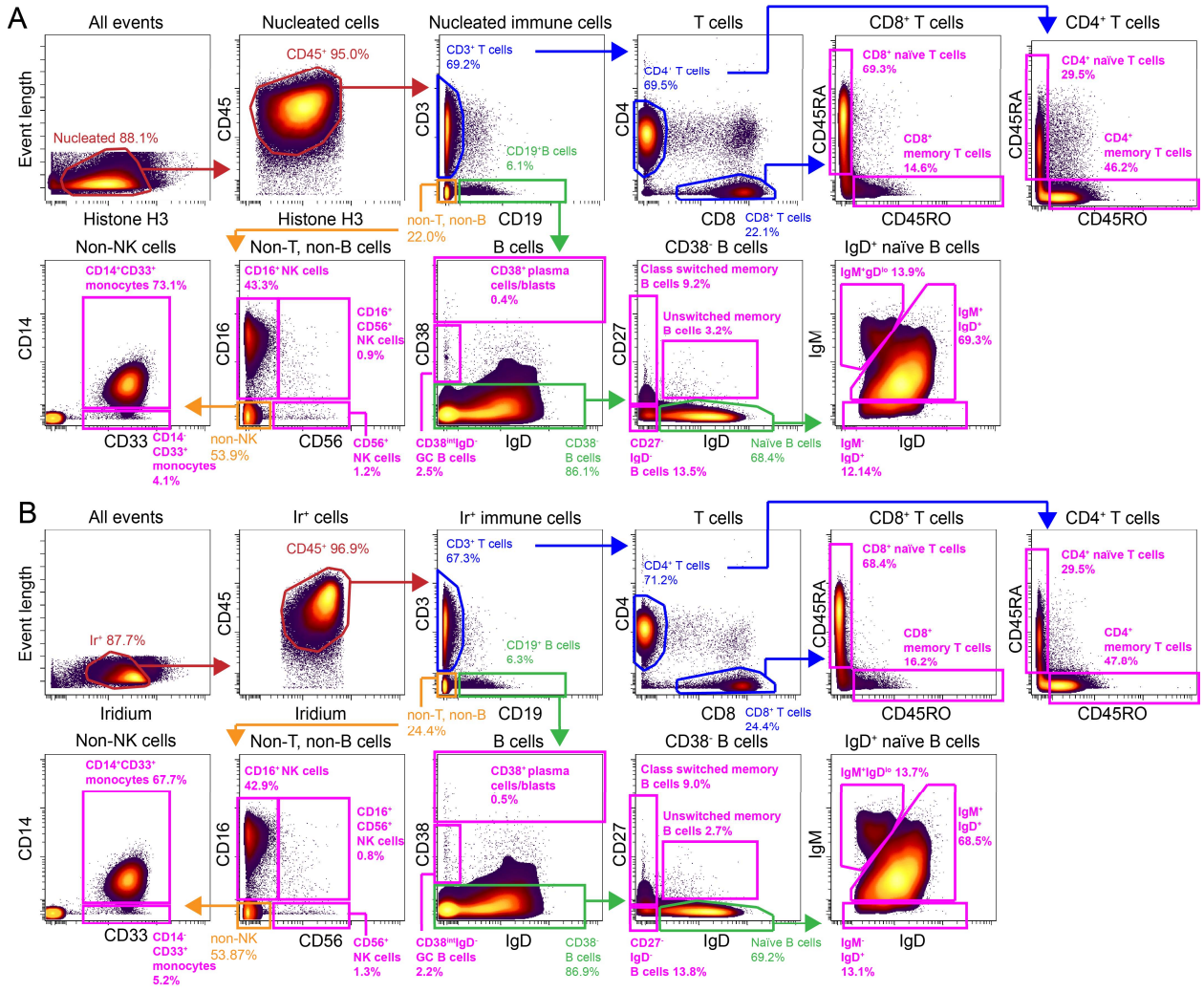


Figure S5-5 Histone H3 is an antibody-based nucleated cell marker for mass cytometry. Live surface stained healthy human PBMCs were stained intracellularly with HH3 antibody and iridium. Either HH3 (A) or iridium (Ir) (B) was used for the initial intact cell gates. $CD45^+$ events were identified. Sequential biaxial gating was used to identify known cell subsets from either $HH3^+CD45^+$ or Ir^+CD45^+ events. Gating subsets of $CD45^+$ immune cells is the same for both $HH3^+$ nucleated cells and Ir^+ intact cells. Pearson analysis and Spearman rank comparing abundance of terminal cell subsets (fuchsia) gated using either HH3 or iridium as intact cell marker are shown in Table S5.

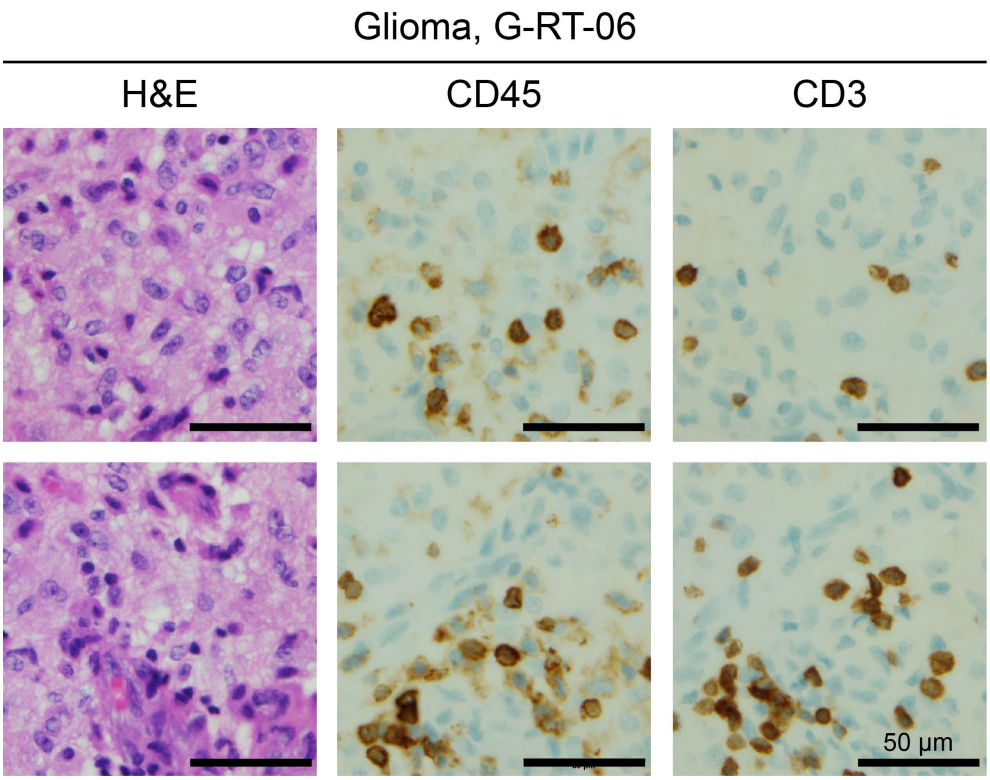


Figure S5-6 Glioma infiltrating immune cells were identified by immunohistochemistry. Hematoxylin and eosin (H&E) stains of FFPE sections of glioma, G-RT-06, are shown. IHC stains with CD45 and CD3 antibodies of the same sample are also depicted. Scale bars = 50 µm

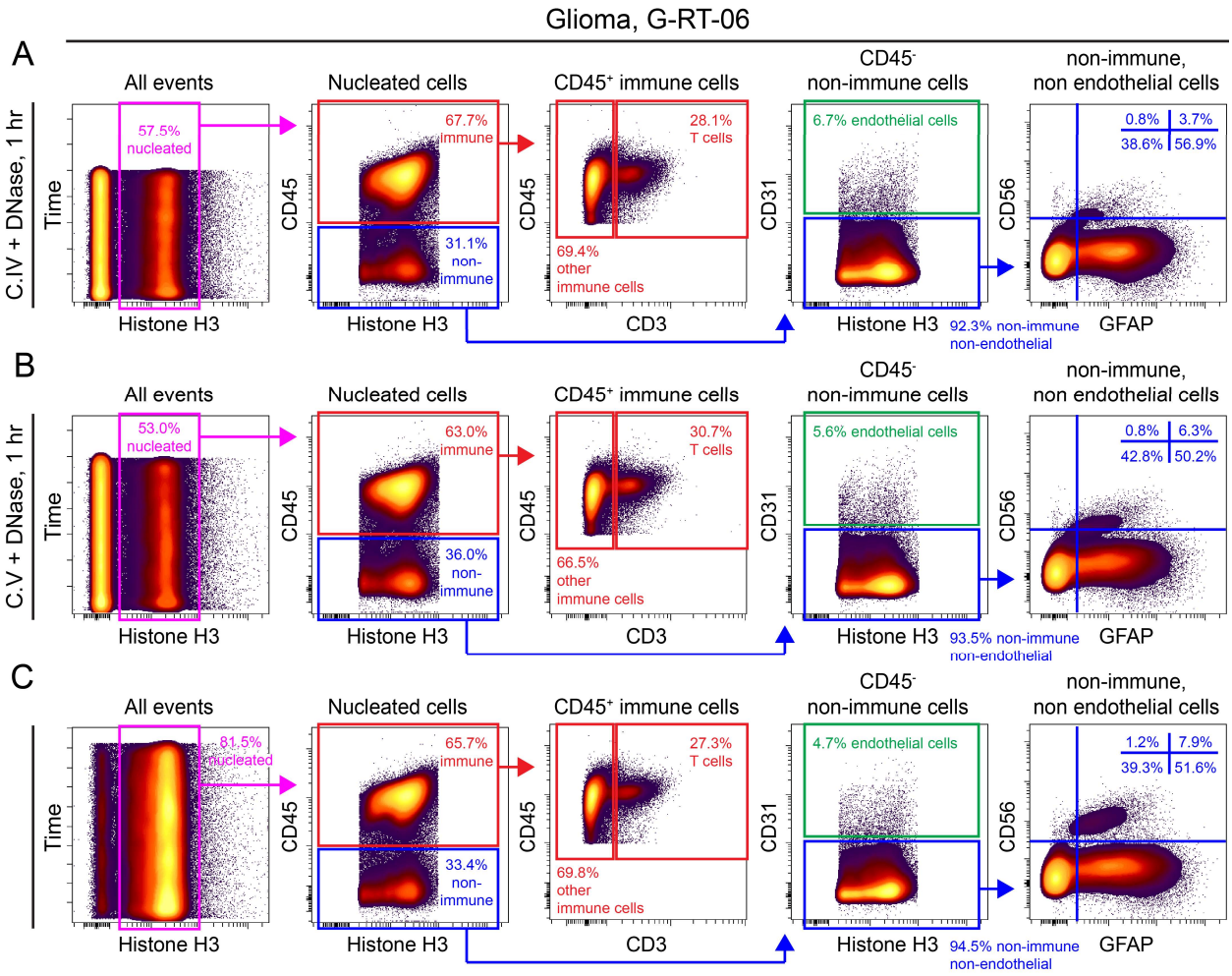


Figure S5-7 Known cell subsets in glioma were identified by fluorescence flow cytometry after dissociation with either collagenase IV, V, or XI. Sequential biaxial gating of glioma G-RT-06 after 1-hour dissociation with DNase plus either (A) Col IV, (B) Col V, or (C) Col XI is shown. The gates shown are the same used in DNase plus Col II dissociation of the same glioma shown in Figure 5A. Abundance of subsets are shown as percentages.

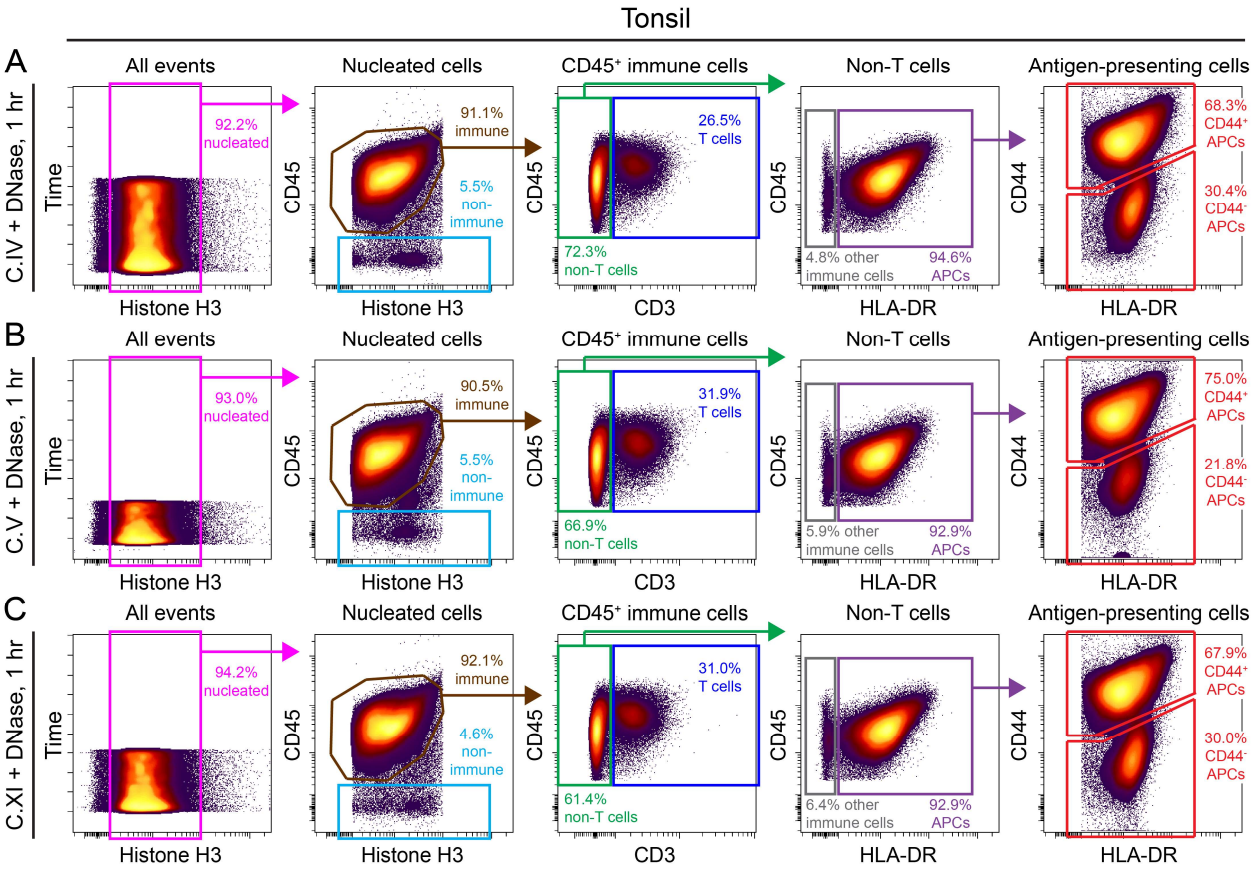


Figure S5-8 Known cell subsets in tonsil were identified by fluorescence flow cytometry after dissociation with either collagenase IV, V, or XI. Biaxial gating of patient-derived tonsil after 1-hour dissociation with DNase plus either (A) Col IV, (B) Col V, or (C) Col XI, is shown. Gating scheme is similar to that used to identify cell subsets of the same tonsil sample after DNase plus Col II dissociation shown in Figure 5C. Abundance of cell subsets are shown as percentages.

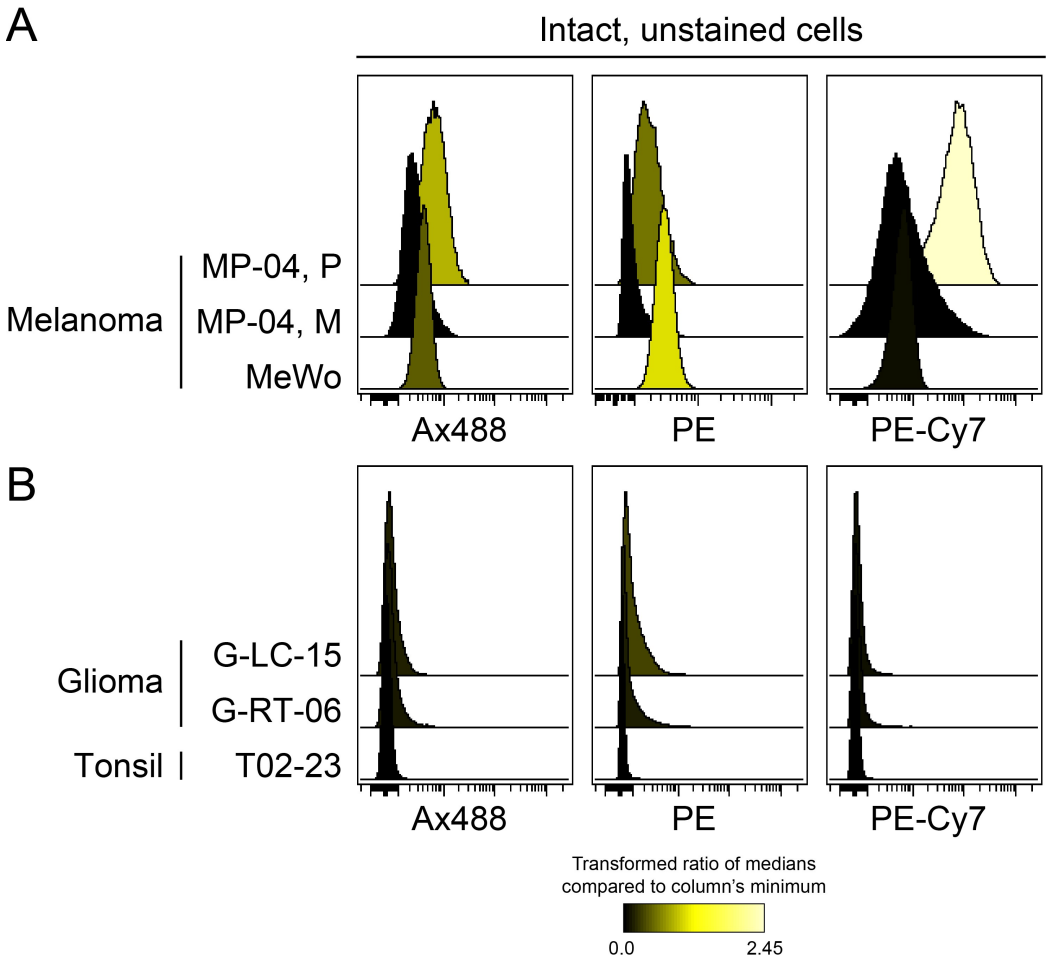


Figure S5-9 Unstained melanoma cells showed variable auto-fluorescence signal. (A) Unstained intact cells from primary (P) and metastatic (M) sites of melanoma MP-04, as well as MeWo melanoma cell line, and (B) two patient-derived glioma samples (G-LC-15, and G-RT-06) and one tonsil (T02-23), were measured for their auto-fluorescence signal on Ax488, PE, and PE-Cy7 channels. Histograms display transformed ratio of medians of signal intensity compared with the minimal signal of each column (channel).

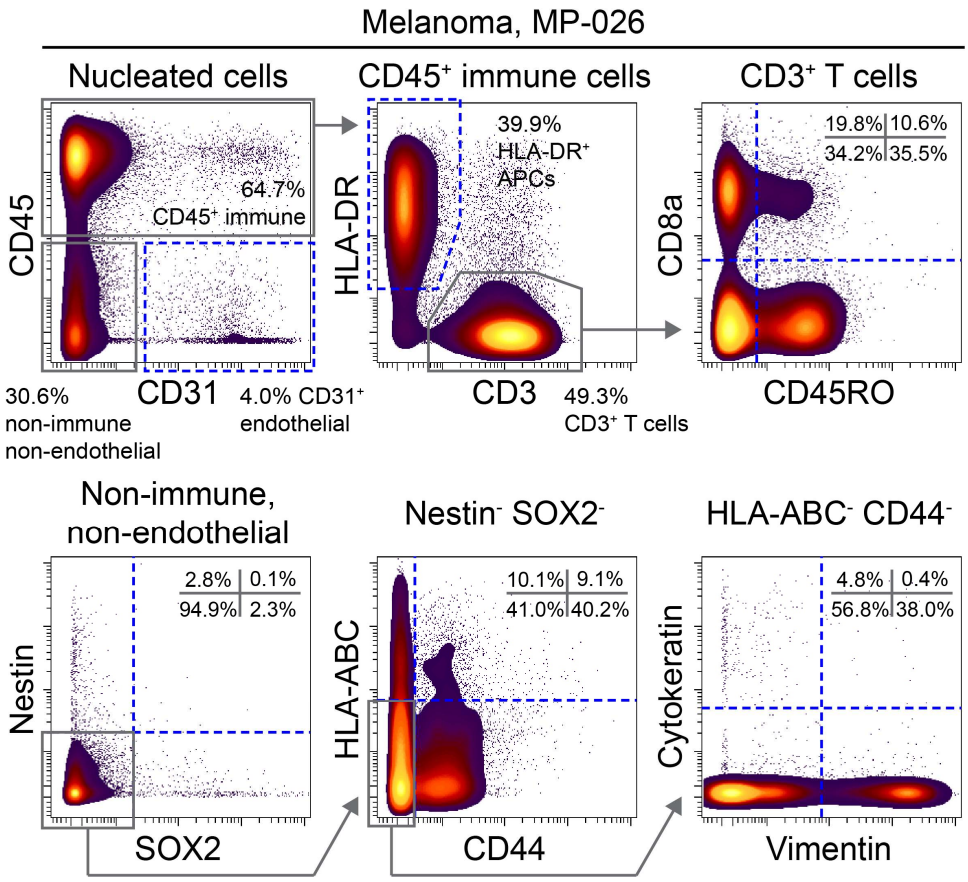


Figure S5-10 Cell subsets in melanoma can be characterized by mass cytometry. HH3⁺ nucleated cells from melanoma sample MP-026, identified by mass cytometry, were characterized for cell subsets by biaxial analysis. Immune cell subsets, endothelial cells, and non-immune, non-endothelial cell subsets were identified using 12 cell identity markers (CD45, CD31, HLA-DR, CD3, CD8a, CD45RO, Nestin, SOX2, HLA-ABC, Cytokeratin, and Vimentin).

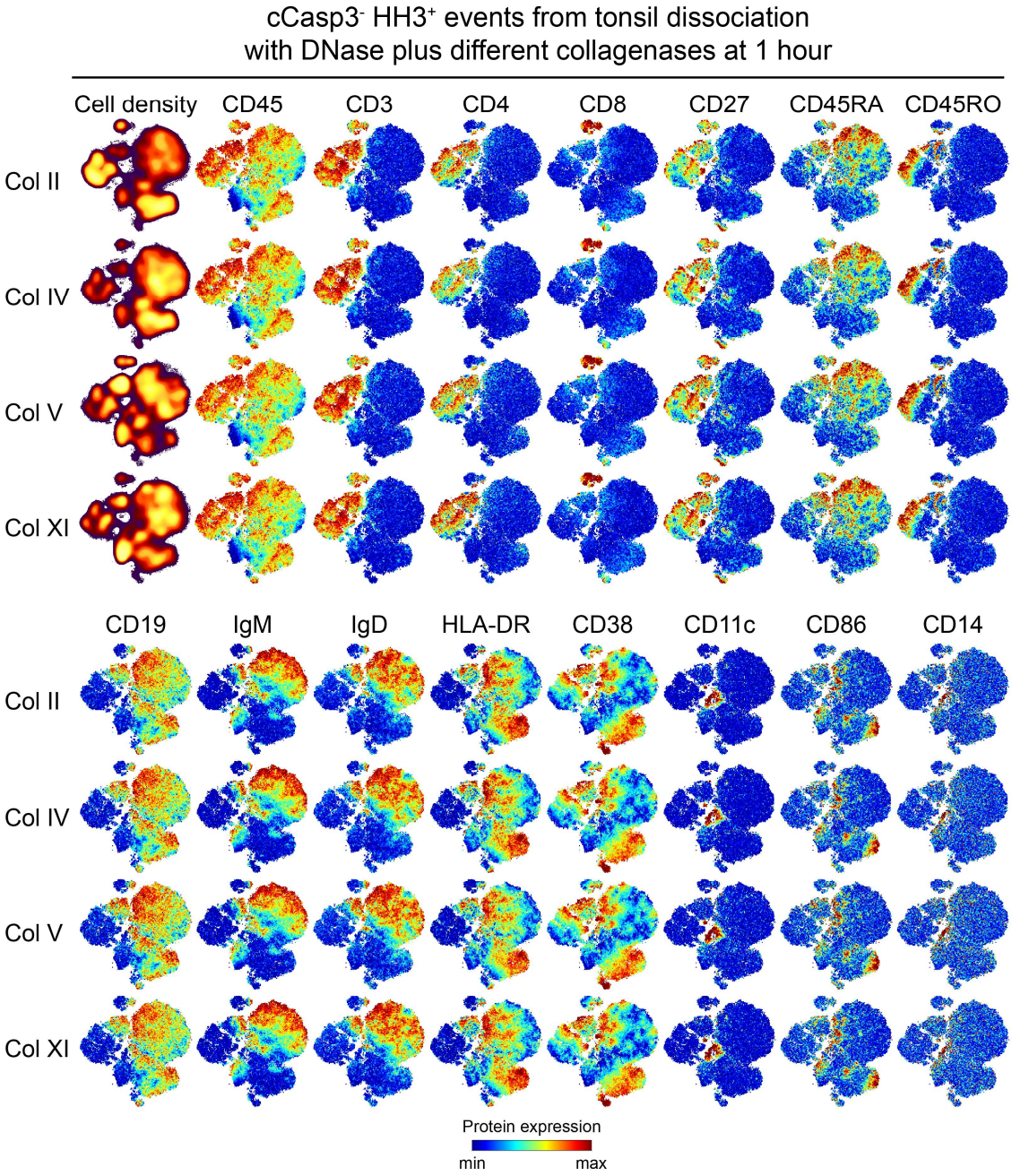


Figure S5-11 Presence and abundance of tonsil cell subsets were comparable after 1-hour dissociation with different types of collagenases plus DNase. Patient-derived tonsils were dissociated for 1 hour with DNase plus either Col II, IV, V, or XI. Nucleated ($HH3^+$) $cCasp3^-$ events were mapped simultaneously by viSNE. Contour plots of different dissociation conditions are shown to illustrate cell density (first column, top row). Heat plots show expression of 15 cell identity markers (CD45, CD3, CD4, CD8, CD27, CD45RA, CD45RO, CD19, IgM, IgD, HLA-DR, CD38, CD11c, CD86, and CD14).

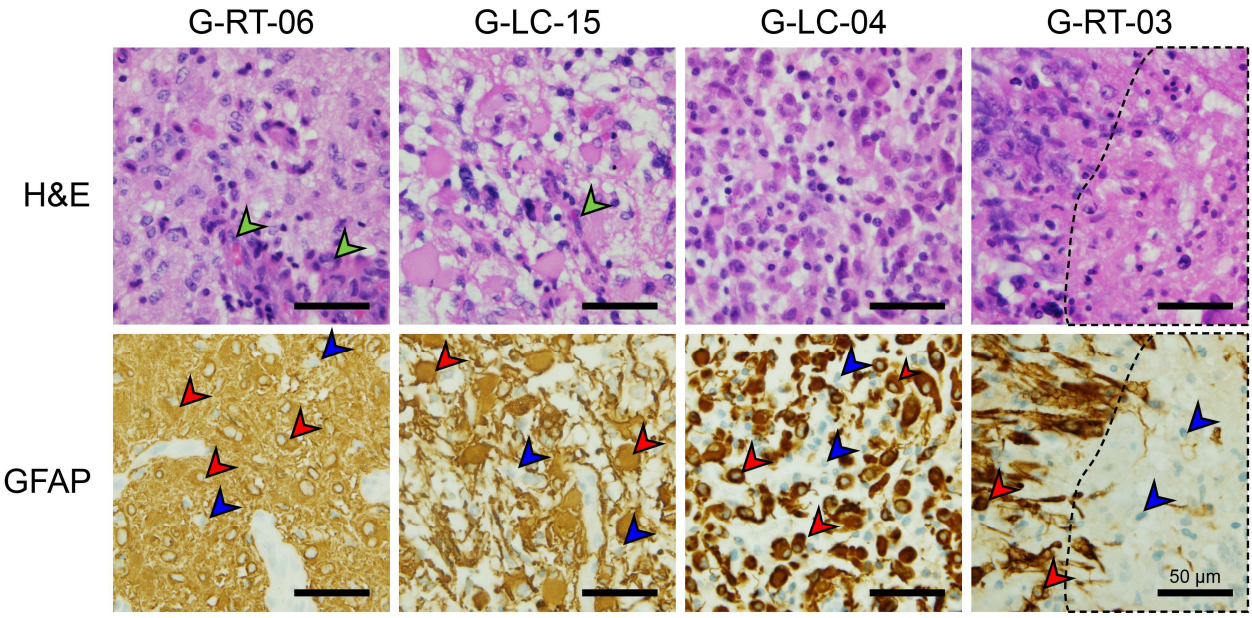


Figure S5-12 GFAP$^+$ cell subsets are present in gliomas. Hematoxylin and eosin (H&E) stains of 4 gliomas are shown (top row). Green arrowheads depict blood vessels or vascular proliferation, and dashed borders show area of necrosis. GFAP staining (bottom row) of the same tumors illustrates GFAP$^+$ (red arrowheads) and GFAP$^-$ (blue arrowheads) cells. Scale bars = 50 µm.

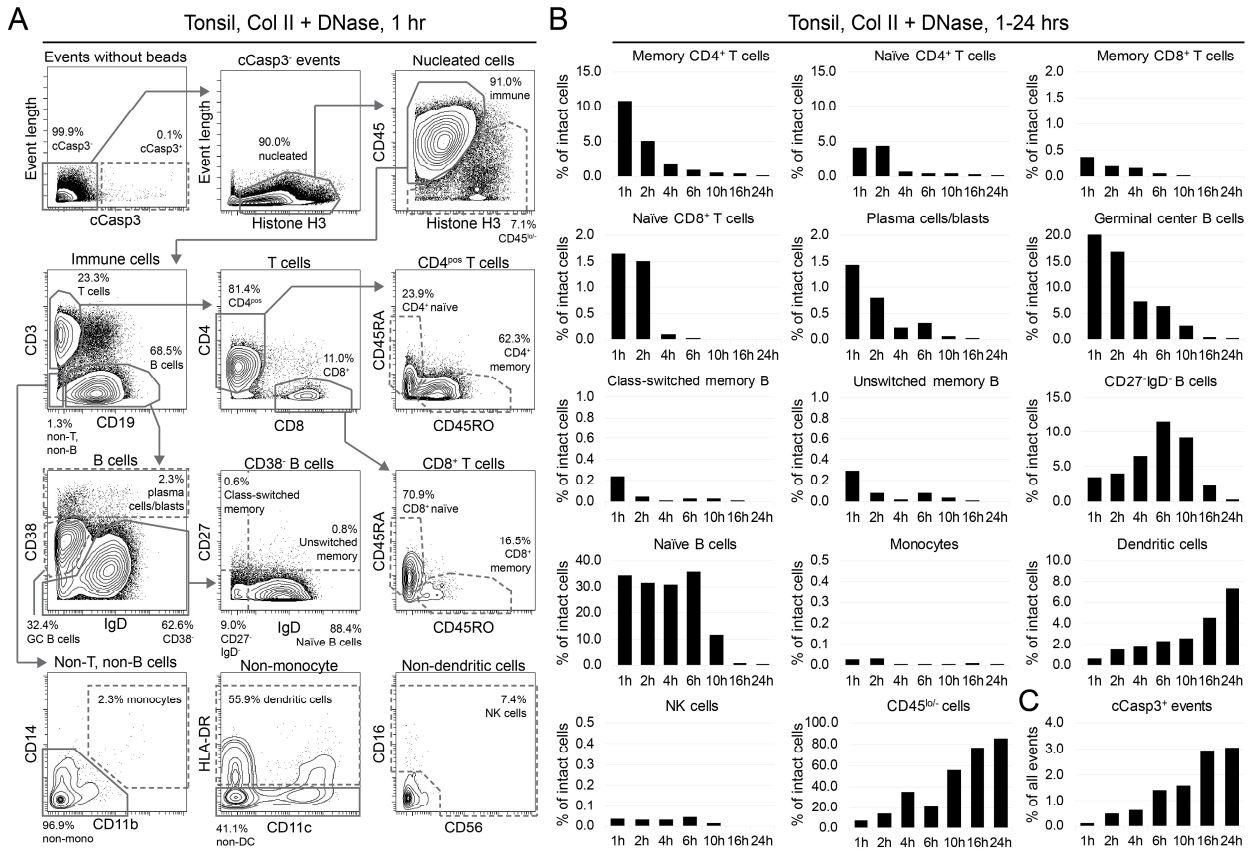


Figure S5-13 Disproportionate depletion and selection of immune cell subsets was observed in tonsil samples with collagenase II dissociation over time. (A) Biaxial gating was used to identify apoptotic (cCasp3$^+$) events, as well as intact nucleated immune cell subsets, in tonsils after 1-hour dissociation with Col II plus DNase. Terminal gates are outlined as dashed gates. (B) A similar gating scheme as in (A) was applied to cells obtained from different duration after dissociation with Col II plus DNase (x-axis) of the same tonsil sample. Abundance of terminal cell subsets were quantified as percentages compared to non-apoptotic nucleated (cCasp3$^-$HH3$^+$) cells (y-axis). (C) Abundance of apoptotic cells from different time points after dissociation is shown as percentage compared to all events.

Table S5-1 Fluorescence antibodies

Target	Dye	Clone	Company	Staining		Panels	
				Live	MeOH	Glio	Ton
CD3	BUV395	UCHT1	BD	✓		+	+
Histone H3	Pacific Blue	D1H2	CST		✓	+	+
CD45	BV786	HI30	BD		✓	+	+
CD19	Ax488	HIB19	BioLegend	✓			+
CD56	Ax488	HCD56	BioLegend	✓		+	
Nestin	Ax594	10C2	BioLegend		✓		
CD44	PE	BJ18	BioLegend		✓		+
CD31	PE-Cy7	WM59	BD		✓	+	
HLA-DR	Ax647	L243	BioLegend	✓			+
GFAP	Ax647	1B4	BD		✓	+	
MCAM	APC	SHM-57	BioLegend		✓		
Dead	Ax700	-	Invitrogen	✓		+	+

Table S5-2 Mass cytometry antibody panel for healthy PBMCs

Target	Mass	Clone	Company	Staining	
				Live	MeOH
CD4	145	RPA-T4	Fluidigm	✓	
IgD	146	IA6-2	Fluidigm	✓	
CD16	148	3G8	Fluidigm	✓	
CD45RO	149	UCHL1	Fluidigm	✓	
CD45RA	153	HI100	Fluidigm	✓	
CD45	154	HI30	Fluidigm	✓	
CD27	155	L128	Fluidigm	✓	
CD33	158	WM53	Fluidigm	✓	
CD14	160	M5E2	Fluidigm	✓	
CD19*	161	HIB19	BioLegend	✓	
CD38	167	HIT2	Fluidigm	✓	
CD8	168	SK1	Fluidigm	✓	
CD3	170	SP34-2	Fluidigm	✓	
IgM	172	MHM-88	Fluidigm	✓	
CD56*	175	HCD56	BioLegend	✓	
Histone H3	176	D1H2	Fluidigm		✓
Iridium	191	-	Fluidigm		✓

* Custom conjugated

Table S5-3 Pearson analysis and Spearman rank comparing histone H3 and Iridium as intact cell markers

PBMC subsets	Percent		Pear	Rank		Spear
	HH3	Ir	(R)	HH3	Ir	(ρ)
Plasma cells/blasts	0.02	0.03		1	1	
Germinal center B cells	0.2	0.1		2	2	
Unclass-switched memory B cells	0.2	0.2		3	3	
$CD16^{pos}CD56^{pos}$ NK cells	0.2	0.2		4	4	
$CD56^{pos}$ NK cells	0.3	0.3		5	5	
$IgM^{neg}IgD^{pos}$ naïve B cells	0.4	0.5		6	7	
Class-switched memory B cells	0.5	0.5		7	6	
$CD33^{pos}CD14^{neg}$ monocytes	0.5	0.7		8	9	
$IgM^{pos}IgD^{lo}$ naïve B cells	0.5	0.5	1.00	9	8	1.00
$CD27^{neg}IgD^{neg}$ B cells	0.7	0.8		10	10	
Memory $CD8^{pos}$ T cells	2.2	2.6		11	11	
$IgM^{pos}IgD^{pos}$ naïve B cells	2.4	2.6		12	12	
$CD14^{pos}CD33^{pos}$ monocytes	8.7	8.9		13	13	
$CD16^{pos}$ NK cells	9.5	10.5		14	14	
Naïve $CD8^{pos}$ T cells	10.5	11.4		15	15	
Naïve $CD4^{pos}$ T cells	14.1	14.1		16	16	
Memory $CD4^{pos}$ T cells	21.9	22.9		17	17	

HH3 = Histone H3; Ir = iridium; Pear = Pearson analysis; Spear = Spearman rank; ρ = Spearman rho
Percent = % of each subset compared to either $HH3^{pos}CD45^{pos}$ cells, or $Ir^{pos}CD45^{pos}$ cells

Table S5-4 Mass cytometry antibody panels for dissociated solid tissues and tumors

Target	Mass	Clone	Company	Staining			Panels		
				Live	Sap	MeOH	Glio	Mel	Ton
HLA-ABC	141	W3-32	Fluidigm	✓				+	
cCasp3	142	D3E9	Fluidigm			✓	+	+	+
CD11b	144	ICRF44	Fluidigm	✓					+
CD31	145	WM59	Fluidigm			✓	+	+	
CD4	145	RPA-T4	Fluidigm	✓					+
CD64	146	10.1	Fluidigm	✓			+		
IgD	146	IA6-2	Fluidigm	✓					+
CD8a	146	RPA-T8	Fluidigm	✓				+	
CD16	148	3G8	Fluidigm	✓					+
CD45RO	149	UCHL1	Fluidigm	✓			+	+	+
S100B*	153	19-S100B	BD			✓	+		
CD45RA	153	HI100	Fluidigm	✓					+
CD45	154	HI30	Fluidigm			✓	+	+	+
CD27	155	L128	Fluidigm	✓					+
CD86	156	IT2.2	Fluidigm	✓					+
Vimentin	156	RV202	Fluidigm			✓		+	
CD33	158	WM53	Fluidigm	✓					+
CD11c	159	Bu15	Fluidigm	✓					+
CD14	160	M5E2	Fluidigm	✓					+
PDGFRα*	161	16A1	BioLegend	✓			+		
CD19*	161	HIB19	BioLegend	✓					+
Cytokeratin	161	C-11	Fluidigm			✓		+	
SOX2*	163	O30-678	BD		✓		+	+	
CD24	166	ML5	Fluidigm	✓			+		
CD38	167	HIT2	Fluidigm	✓					+
Nestin*	168	10C2	BioLegend			✓		+	
CD8	168	SK1	Fluidigm	✓					+
CD44*	169	BJ18	BioLegend			✓	+	+	
CD3	170	SP34-2	Fluidigm	✓			+	+	+
GFAP*	171	1B4	BD			✓	+		
IgM	172	MHM-88	Fluidigm	✓					+
αSMA*	173	ab54723	Abcam			✓	+		
HLA-DR	174	L243	Fluidigm	✓			+	+	+
CD56*	175	HCD56	BioLegend	✓			+		+
Histone H3	176	D1H2	Fluidigm			✓	+	+	+

* Custom conjugated
Sap = 0.02% saponin; Glio = glioma; Mel = melanoma; Ton = tonsil

CHAPTER 6

Single Cell Signaling Profiles Reveal Clinically and Biologically Distinct Glioblastoma Cells

Authors: Nalin Leelatian, Akshitkumar M. Mistry, Justine Sinnaeve, Kirsten E. Diggins, Allison R. Greenplate, Jocelyn S. Gandelman, Daniel Liu, Kyle D. Weaver, Reid C. Thompson, Lola B. Chambless, Bret C. Mobley, Rebecca A. Ihrie*, and Jonathan M. Irish*

* Denotes equal contribution

This work is presented in the format that is being prepared for submission, as of February 2018.

Preface

Despite the immense effort to identify meaningful clinical stratifications and therapeutics based on molecular alterations, adult glioblastomas still carry a high fatality rate. Previous glioblastoma studies had two critical limitations. First, although there was ample evidence that signaling alterations were the hallmarks of glioblastoma pathogenesis, most of the studies used genomics and transcriptomics information to infer signaling rather than directly quantifying the ongoing signaling activity. Second, the studies that quantified signaling activity either did not dissect the signaling in single cells, despite the increasing awareness of cellular diversity in glioblastoma, or presumed that the data observed in animal models would reflect signaling in patient tumors.

The work presented in this chapter aims to tackle both limitations. By coupling single-cell mass cytometry signaling quantification with a risk stratification approach, two novel glioblastoma cell subsets were uncovered. These cells were defined by unique signaling enrichments, abnormal developmental phenotype, and a significant capacity to stratify patient clinical outcomes. Additionally, a relationship between brain resident microglia, tumor infiltrating immune cells, and the clinically distinct glioblastoma cells was discovered, indicating a tight association between cancer cells and the tumor microenvironment. Single-cell mass cytometry was critical to revealing these findings since the discovery relied on the ability to pinpoint specific signaling mechanism in rare cell subsets (occasionally 1 in 10,000). This explains why the findings here have not been identified in previous

studies that utilized other technical platforms. Importantly, the proteomic profiles of the stratifying cell subsets suggest immediately testable cellular targets for innovative therapeutics.

Abstract: While genomic and transcriptomic profiles have revealed extensive cellular diversity in adult glioblastoma, such approaches have not yet stratified patient outcomes. Single cell mass cytometry was used here to systematically reveal glioblastoma cells based on per-cell measurements of 42 proteins and phospho-proteins that govern neural and immune cell development and function. Risk stratification analysis pinpointed two distinct glioblastoma populations that were closely associated with better or worse clinical outcomes. The defining biological features of glioblastoma negative prognostic cells included potentiation of essential survival signaling effectors, and inflammatory signaling was observed within positive prognostic cells. Prognostic signaling in glioblastoma cells was closely related to microglial activation phenotype and leukocyte infiltration. Once revealed, seven cellular features were sufficient to identify prognostic glioblastoma cells in a clinically-compatible workflow.

Main Text

Glioblastoma is the most common primary malignant brain tumor in adults (*317*). These aggressive tumors are composed of vascular cells, immune cells, neural-lineage cells, and cancer cells with diverse transcriptional, genomic, and protein expression profiles (*5, 48, 284*). As in other cancers, altered signaling in the biochemical networks that govern growth, survival, and interaction with the microenvironment is thought to confer hallmark capabilities to glioblastoma cells, including survival, proliferation, invasion, and immune evasion (*7*). However, the extreme diversity of cell types and molecular lesions within and between glioblastomas has obscured the identity of biological events that drive poor outcomes. There remains an urgent need to better understand how the cellular composition of glioblastomas and per-cell signaling in different tumor cell types relate to tumor progression and patient survival.

Genomic and transcriptomic profiles have shown that the inter- and intra-tumor cellular diversity in glioblastoma is largely driven by alterations in receptor tyrosine kinases (RTKs) (*2, 3*). Previous studies have either measured signaling states in bulk primary tumors (*140, 142, 318-320*) or characterized genomic and

transcriptomic profiles in small numbers of single cells (*5, 30, 44, 47, 48*). However, stratification of patient clinical outcomes has not resulted from these approaches. Very little is known about the activation states of signaling effector proteins in primary glioblastoma cells. Mass cytometry simultaneously measures key phospho-protein signaling nodes while also tracking proteins controlling cell lineage and identity (*83*), and we recently reported proof-of-concept studies of human glioblastoma using mass cytometry (*285, 321, 322*). Mass cytometry studies also leverage single cell analysis tools like viSNE (*290*), FlowSOM (*323*), and MEM (*321*) to reveal and characterize phenotypically distinct cells. Prior studies with phospho-specific flow cytometry have revealed cell subsets defined by signaling that are closely linked to clinical outcomes (*10, 12, 13, 15, 80*). This study investigated associations between patient outcomes, intracellular signaling, and aberrant expression of developmental programs in glioblastoma cells.

Results

Distinguishing glioblastoma, immune, and vascular cells in tumors

Millions of single cells from freshly dissociated, surgically resected *IDH*-wildtype glioblastomas were comprehensively analyzed by single-cell mass cytometry. Cells from 28 patient samples were analyzed using a 32-dimensional antibody panel (Figure 6-1A, Table S6-1, and Table S6-2, signaling & proteins). For this cohort, the median progression-free survival (PFS) and the median overall survival (OS) after diagnosis were 6.3 and 13 months, respectively (Figure S6-1). These survival times were similar to those observed in larger population of patients undergoing standard therapy (*26, 178*). A comprehensive cytomic view of each tumor was generated using viSNE analysis (*290*). As a representative, patient LC26 is shown (Figure 6-1B and Figure 6-1C). This analysis confirmed the presence of multiple cell types within a tumor (*5, 48, 284*). Expression of canonical identity proteins was used to characterize non-glioblastoma cells, including $CD45^{+}HLA\text{-}DR^{+}$ antigen-presenting cells (APCs), $CD45^{+}HLA\text{-}DR^{-}$ non-APC immune cells, and $CD45^{-}CD31^{+}$ endothelial cells (Figure 6-1B and Figure 6-1C). Differential expression of surface and intracellular proteins was observed between the key cell types. In addition to the canonical proteins, other proteins were also highly expressed in non-glioblastoma cells, in accordance to previous reports, including CD34, PDGFRα, and ITGα6 in endothelial cells (*324-326*), and CD44 in immune cells (*327, 328*).

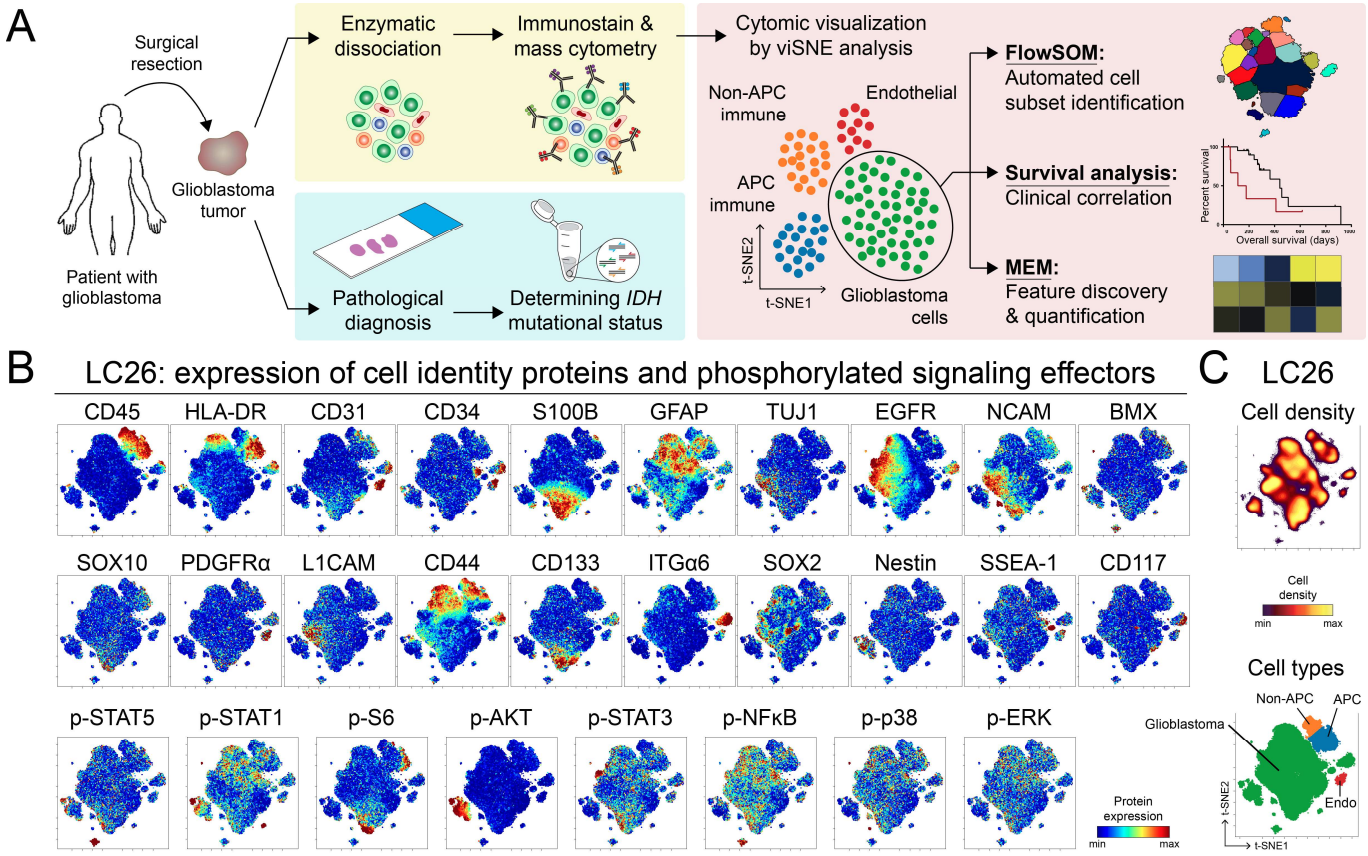


Figure 6-1 Single-cell quantification of phospho-protein signaling and identity proteins in glioblastoma. (**A**) A diagram of the study depicts isolation of viable cells from glioblastoma tumors, pathological confirmation of *IDH*-wildtype status, immunostaining for mass cytometry analysis, viSNE dimensionality reduction, and expert identification of four main cell types, as in Figure 6-1B. Next, glioblastoma cells from each patient were computationally isolated and combined for new analyses using FlowSOM, MEM, statistical tests for clinical associations, as in Figure 6-2. (**B**) viSNE plots from one patient's glioblastoma tumor (LC26) show individual cells colored according to expression of 20 identity proteins and 8 phosphorylated signaling effectors. Heat indicates protein or phospho-protein expression per cell and position indicates phenotypic similarity. (**C**) Plots show cell density and cell identity for the same cells from tumor LC26 showing in Figure 6-1B. Protein features including CD45, CD31, HLA-DR, and others shown in (B) were used to identify cells as antigen presenting cells (APC, blue), other immune cells (non-APC, orange), endothelial cells (Endo, red), and glioblastoma cells (green).

Biologically distinct glioblastoma cells within an individual tumor

Intra-tumor single-cell diversity among $CD45^{-}CD31^{-}$ glioblastoma cells was apparent in the t-SNE plot as multiple, well-separated islands (Figure 6-1B and Figure 6-1C). In addition, identity-protein expression ranged across multiple orders of magnitude and contrasted between the cell-type islands, indicating glioblastoma cell subsets were distinguishable by differences in per-cell protein expression. This analysis also revealed unexpected diversity of developmental phenotypes in subsets of glioblastoma cells, including lineage aberrancy

(such as co-expression of astrocytic S100B and stem-like CD133, or neuronal TUJ1 and stem-like SOX2) as well as co-expression of molecular subclass markers (such as mesenchymal CD44 and classical EGFR) (*2*). Moreover, the phosphorylation states of eight key signaling effectors were quantified, revealing intra-tumor signaling diversity of glioblastoma single cells (Figure 6-1B, p-STAT5-Y694, p-AKT-S473, p-STAT3-Y705, p-S6-S235/S236, p-STAT1-Y701, p-NFκB(p65)-S529, p-ERK1/2-T202/Y204, and p-p38-T180/Y182). For example, simultaneous phosphorylation of S6, STAT5, and STAT3 was observed in a subset of glioblastoma cells that expressed S100B, but not in cells that expressed EGFR, GFAP, or CD44 (Figure 6-1B). Thus, the per-patient approach revealed unexpected combinations of cellular identity proteins and unique signaling features within cells from an individual glioblastoma tumor.

Systematic analysis of glioblastoma cells across tumors

By focusing initially on each patient, the analysis was able to best isolate the major cell lineages within the tumor and discern immune and endothelial populations. This optimally identified each patient's glioblastoma cells in a context-specific manner. From each patient, at least 4,710 glioblastoma cells were identified and then pooled for a new comparison across the patient cohort. This is to determine whether cell subsets were common or unique to individual tumors. A total of 131,880 glioblastoma cells were identified from 28 patients and combined for simultaneous, in-depth profiling based on cellular identity and signaling features (Figure 6-2A). Automated clustering by FlowSOM (*293*) revealed 43 phenotypically distinct glioblastoma cell subsets (Figure S6-2), the majority of which were observed in multiple patients' tumors Figure S6-3).

Clinically distinct glioblastoma cells

It may be the case that phenotypically distinct cell subsets are associated with differential clinical outcomes, as has been observed in blood cancers (*15*). To assess risks associated with newly revealed glioblastoma cell subsets, a computational workflow was designed to compare associations between cell subset abundance and patient overall survival (Figure 6-2B and Figure S6-2, see Supplementary Methods). Using this automated pipeline, glioblastoma cell subsets with contrasting and statistically significant associations with patient survival were revealed and characterized (Figure 6-2B). Glioblastoma Negative Prognostic (GNP) cells

(red; subsets 29, 32, 33, and 37) were defined as cell subsets with associated hazard ratios (HR) of death >1 ($p<0.05$) (Figure 6-2C), whereas Glioblastoma Positive Prognostic (GPP) cells (green; subsets 1, 2, 13, and 22) were defined as cell subsets with associated HR of death <1 ($p<0.05$) (Figure 6-2D). The remaining cell subsets were not significantly deterministic of patient survival. Assessment of the presence and abundance of GNP and GPP cells in the entire patient cohort illustrate the extensive variation in abundance of clinically significant glioblastoma cell subsets across patients (Figure 6-2E, Figure S6-4A, and Figure S6-4B). Interestingly, there was no linear correlation between GNP and GPP cell abundance despite their contrasting connections with patient survival (Figure S6-4C).

To determine if the cumulative abundance of the multiple individual cell subsets comprising the GNP or GPP populations was predictive of clinical outcome, cumulative GNP and GPP cells were tested with Kaplan-Meier survival analysis using the Wilcoxon test (Figure S6-5). Patients whose tumors had over 5% GNP cells (in the glioblastoma cell fraction) had significantly shortened OS (Figure S6-5A, $p=0.0022$) and PFS (Figure S6-5B, $p=0.02$). In contrast, presence of GPP cells over 5% was associated with a significantly greater OS (Figure S6-5C, $p=0.0063$) and PFS (Figure S6-5D, $p=0.015$). The preliminary statistical analysis indicated that the abundance of GNP and GPP glioblastoma cell subsets was closely associated with patient survival. To determine if the effect of cell subset abundance was a) continuous, and b) independent of other widely recognized predictors of survival (age (*112, 329*), *MGMT* promoter methylation status (*107, 330*), and treatment including extent of surgical resection (*113, 114*), therapy with temozolomide (*1*), and radiation (*331, 332*)), a multivariate Cox proportional-hazard model was developed by incorporating the above-listed survival predictors with GNP or GPP cell abundance. Multivariate examination of GNP cell abundance as a continuous variable, keeping the other predictors constant, indicated that each additional 1% increase of GNP cells was associated with an approximately 8% increase in the mortality rate compared to baseline (HR 1.08 [95% CI 1.02-1.14]; $p=0.008$). In a similar analysis of GPP cells, each additional increase in 1% of GPP cells was associated with an approximately 19% decrease in the mortality rate (HR 0.81 [0.67-0.98]; $p=0.03$) and approximately 5% decrease in time to tumor progression or recurrence compared to baseline (HR 0.95 [0.91-0.99]; $p=0.03$). Strikingly, when GNP and GPP were assessed simultaneously, abundance of GNP cells was the primary predictor of OS (HR 1.06 [1.00-1.12]; $p=0.05$), while abundance of GPP cells was the primary predictor of PFS (HR 0.94 [0.90-0.99];

p=0.024). Thus, the abundance of GNP and GPP cell subsets were each associated with distinct patient outcomes, independent of the standard prognosticators of glioblastoma patient survival, and each other.

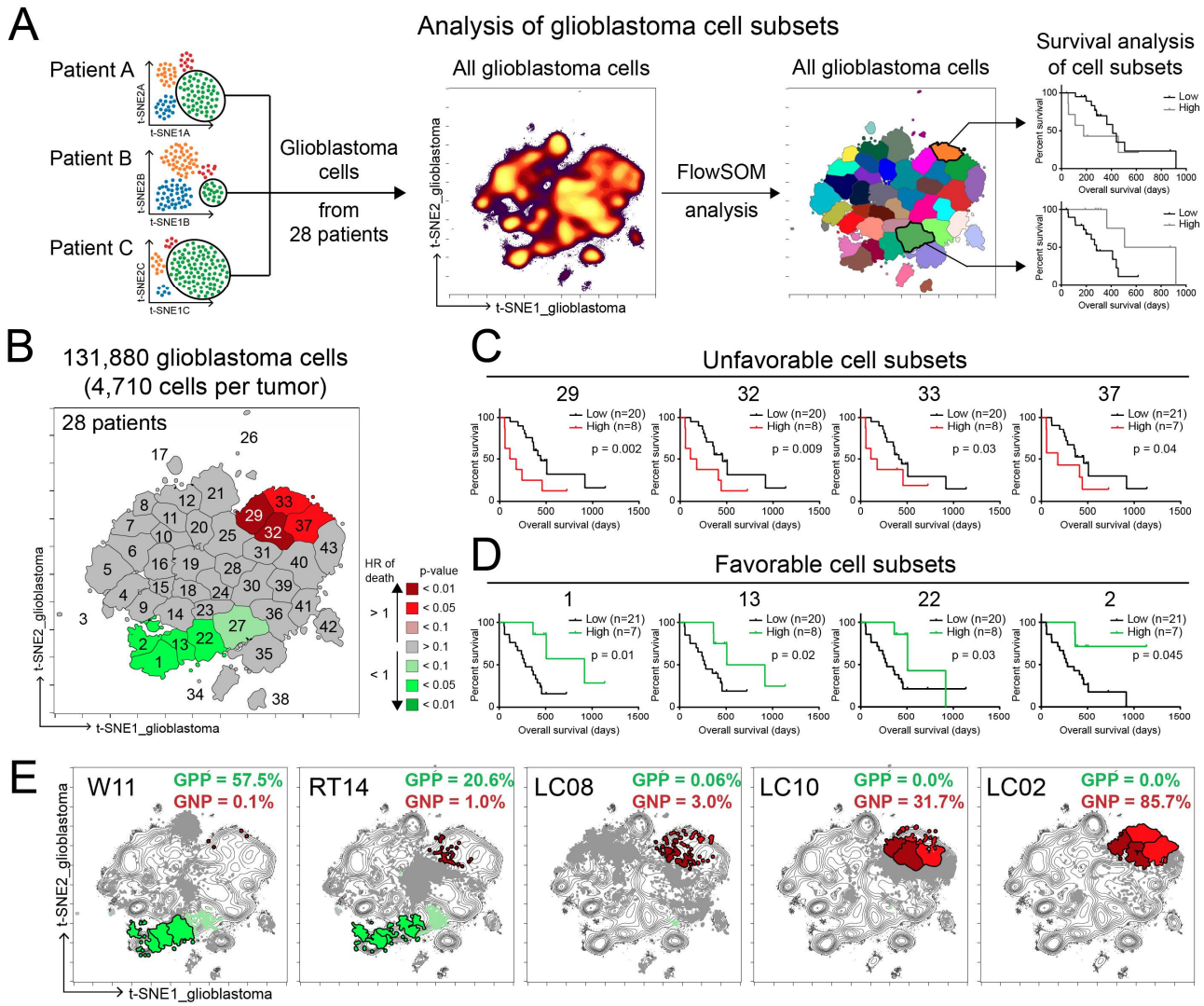


Figure 6-2 Risk stratification reveals clinically distinct Glioblastoma Negative Prognostic cells and Glioblastoma Positive Prognostic cells. (**A**) A diagram depicting the automated identification of glioblastoma cell subsets that linked to clinical outcome. Glioblastoma cells identified from 28 patients were computationally pooled for a new combined viSNE analysis. FlowSOM was used to automatically reveal glioblastoma cell subsets which were systematically assessed for their association with patient overall survival. (**B**) Forty-three glioblastoma cell subsets were identified across 28 patients. For each subset, overall survival was compared between patients with high vs low cell abundance (see Supplementary Methods). Cell subsets are color-coded based on hazard ratio (HR) of death (HR>1, red; HR<1, green) and p-values. (**C**) Overall survival analyses of abundance of Glioblastoma Negative Prognostic (GNP) cells (subsets 29, 32, 33, and 37) and (**D**) Glioblastoma Positive Prognostic (GPP) cells (subsets 1, 13, 22, and 2) are shown. (**E**) Abundance of GNP (red) and GPP (green) cells in 5 patients (W11, RT14, LC08, LC10, and LC02) are displayed. The non-prognostic cell subsets are shown in gray. The underlying contour depicts all glioblastoma cells from 28 tumors.

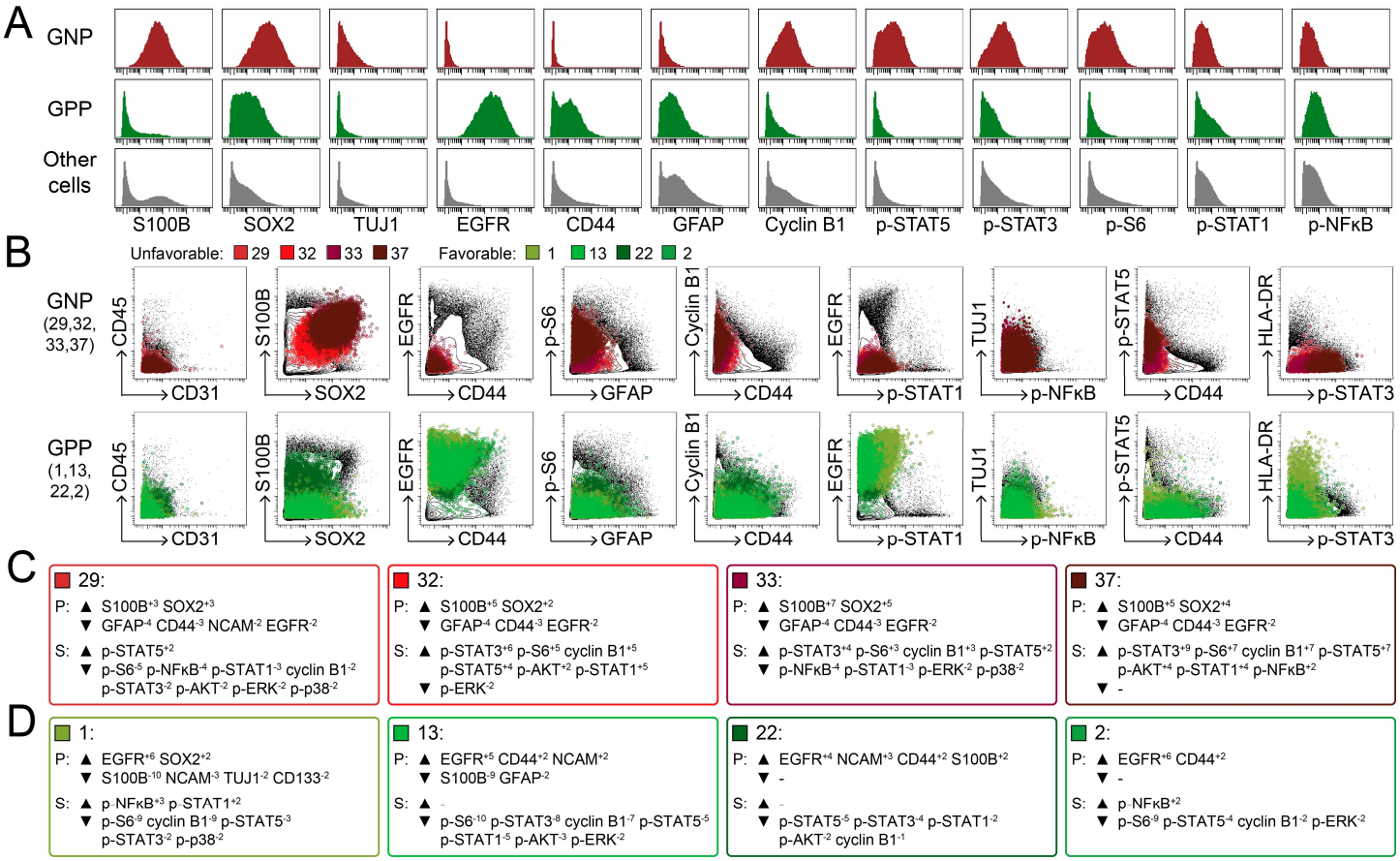


Figure 6-3 Clinically distinct glioblastoma cells are identified by abnormal developmental phenotype and altered signaling profiles. (**A**) Histogram plots of GNP (red), GPP (green), and other glioblastoma cells (gray) are shown to illustrate the differential expression of developmental proteins and phosphorylated signaling effectors. (**B**) Individual cell subsets that comprised GNP cells and GPP cells were mapped over biaxial plots to illustrate their shared and distinctive features. The underlying contours include glioblastoma, immune, and endothelial cells from 28 tumors. (**C**) MEM quantified the enrichment of identity proteins (P) and phosphorylated signaling effectors (S) in cell subsets comprising GNP population. (**D**) MEM analysis was also performed in cell subsets comprising GPP population.

Exceptional biological features of clinically distinct glioblastoma cells

To identify the signature of each prognostic glioblastoma cell subset, a systematic manual comparison between GNP, GPP, and other glioblastoma cells was performed (Figure 6-3A and Figure S6-6). GNP cells aberrantly co-expressed mature neural-lineage proteins (astrocytic S100B, and neuronal TUJ1) and stem-like SOX2. Additionally, GNP cells showed abnormal proliferative activity (M-phase marker, cyclin B1) as well as phosphorylation of RTK signaling effectors promoting cell survival, growth, and replicative capacity (p-STAT5, p-S6, p-STAT3). In contrast, GPP cells co-expressed higher levels of both EGFR and CD44, had lower signaling effector phosphorylation levels, and were less proliferative as measured by cyclin B1. To exclude the possibility

that GNP or GPP subsets contained immune or endothelial cells, these populations were compared on biaxial plots for the respective definitional markers CD45 and CD31 and were shown to exclusively contain non-immune, non-endothelial cells (Figure 6-3B, first column). Additional biaxial analyses of defining and distinguishing features of GNP and GPP cells were performed (Figure 6-3B).

Marker Enrichment Modeling (MEM) was used to objectively quantify the signature of GNP and GPP cells (Figure S6-7 and Figure S6-8) (*321*). MEM automatically derives a label for cell types based on the magnitude and variance of the features in a given subset relative to a reference population, which was used here to calculate enrichment of total proteins (P) and phospho-protein signaling effectors (S) in GNP cells (Figure 6-3C) and GPP cells (Figure 6-3D). The reference point for the MEM quantification of a given subset was all other glioblastoma cells (Figure S6-7; see Supplementary Methods for MEM interpretation). MEM provided an unbiased assessment of signature features and confirmed aberrant enrichment of lineage proteins (▲S100B$^{+5\pm1.6}$, SOX2$^{+3.5\pm1.3}$) and phosphorylated STAT5 (▲p-STAT5$^{+4\pm2.6}$) in GNP cells. Specifically lacking from clinically unfavorable GNP cells was EGFR, GFAP, and CD44 (▼EGFR$^{-2\pm0}$, GFAP$^{-2\pm0}$, CD44$^{-2\pm0}$). In contrast, total EGFR and CD44 were positively enriched in GPP cells (▲EGFR$^{+5.5\pm1}$, CD44$^{+2\pm1}$) and proliferation (▼cyclin B1$^{-4.5\pm3.8}$) and pro-survival phospho-proteins (▼p-S6$^{-9\pm4.7}$, p-STAT5$^{-4.5\pm1}$, p-STAT3$^{-3\pm3}$) were consistently lacking in GPP cells (Figure 6-3C and Figure 6-3D).

Towards tracking clinically distinct glioblastoma cells in the clinic

For clinical application of the high-dimensional analysis that was used to discover GNP and GPP cells, it would be useful if these distinct populations could be estimated or identified using traditional, lower-dimensional approaches. Traditional biaxial cytometry gating was previously shown to closely match findings by mass cytometry in glioblastoma (*284*) and could be implemented using the widely-available fluorescence flow cytometry platforms. A conventional biaxial gating scheme, using only the top 7 cellular features identified by MEM scoring, successfully re-discovered GNP and GPP cells without the use of high-dimensional tools such as t-SNE/viSNE (Figure S6-9A). Populations of GNP and GPP cells identified by traditional gating were then mapped back onto the t-SNE analysis to determine whether they comprised distinct islands or were mixed. Cells gated using the traditional biaxial approach occupied the same, distinct t-SNE phenotypic islands as those

identified by computational tools Figure S6-9B and Figure 6-2B). MEM analysis showed that the cells identified by traditional gating were statistically similar in identity, protein expression, and phosphorylation as those cells identified by the automated cell subset discovery approach (Figure S6-9C, Figure 6-3C, and Figure 6-3D). The similarity of the manually and computationally identified cells was also apparent in histogram plots (Figure S6-9D). It was crucial to couple high-dimensional single-cell mass cytometry with an automated, risk-stratification approach, to discover and quantify the features of clinically significant glioblastoma cell subsets, in an unbiased manner. However, once revealed, GNP and GPP cells could be reliably identified by traditional analysis approach compatible with clinical flow cytometric profiling.

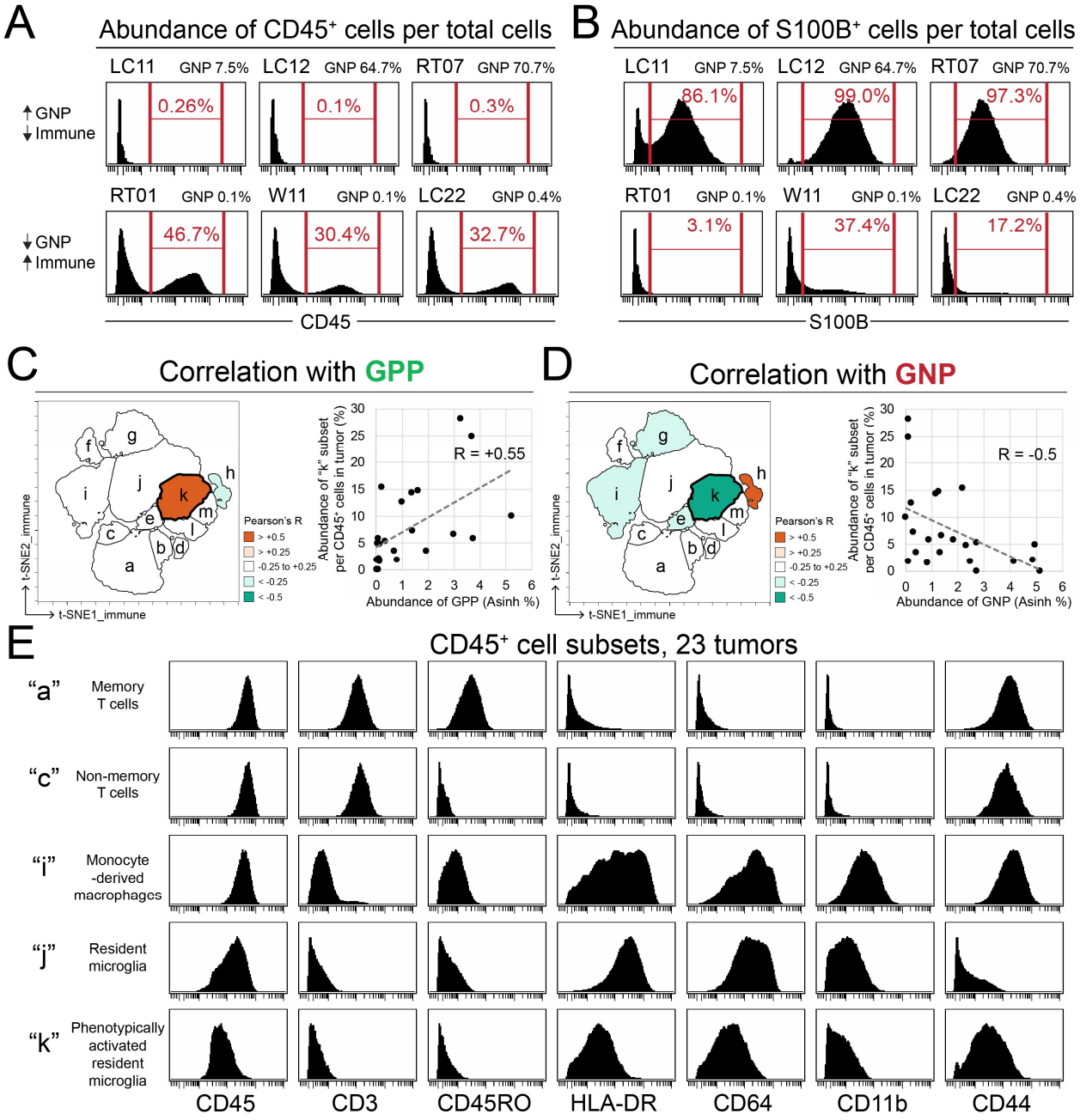


Figure 6-4 Clinically distinct glioblastoma cells were closely related to immune infiltration and microglial activation phenotype. (**A**) Histogram plots showing abundance of CD45$^+$ immune cells in GNP-high, immune-low tumors (top row: LC11, LC12, RT07), and GNP-low, immune-high tumors (bottom row: RT01, W11, LC22). (**B**) An analysis of S100B$^+$ cells of the same tumors is shown. (**C**) CD45$^+$ cells from 23 tumors were combined for a viSNE analysis. Protein features shown in Fig. S11 were used to identify 13 immune cell subsets. Each subset was colored based on its correlation with GPP cell abundance, as determined by Pearson correlation coefficient R (R > +0.5, orange; R = 0, white; and R < -0.5, teal). The correlation between subset "k" and GPP cells is shown (x-axis, GPP abundance; y-axis, subset "k" abundance) (**D**) A similar correlation analysis was performed between immune cell subsets and GNP cells. The correlation between abundance of subset "k" and GNP cells is also depicted. (**E**) Histogram plots show the expression of key immune markers in subsets "a" (memory T cells), "c" (non-memory T cells), "i" (monocyte-derived macrophages), and "j" (resident microglia), and "k" (phenotypically activated resident microglia).

Immune microenvironment associations with clinically relevant glioblastoma cells

Infiltrating immune cells comprise a large proportion of non-cancer cells in glioblastomas. Recent studies have revealed a high degree of variation in total immune abundance and relative proportions of immune cell subsets across different glioblastoma tumors (*73*). In this dataset, the abundance of tumor-infiltrating CD45$^+$ cells (Figure 6-4A) and S100B$^+$ cells (Figure 6-4B) was quantified and compared between tumors, revealing an inverse correlation between the abundance of immune cells and clinically unfavorable GNP cells (Figure S6-10A). This correlation was not observed, however, with the clinically favorable GPP cells (Figure S6-10B).

To further dissect the abundance and phenotype of immune cell subsets, single cells from 23 patients were characterized with antibodies against immune-cell surface proteins and t-SNE analysis on the CD45$^+$ cells (Figure 6-4C and Table S6-2). Nine immune-related surface features were used to generate the immune-specific viSNE map (Figure S6-11). Consistent with previous reports, the intratumor heterogeneity among CD45$^+$ cells suggested the presence of diverse immune cell subsets. Using both the cellular distribution on the viSNE map and per-cell protein expression profiles, 13 subsets of CD45$^+$ cell were identified by expert gating on t-SNE axes (Figure 6-4C and Figure 6-4D, left). The abundance of each CD45$^+$ cell subset was compared to abundance of GPP (Fig.4C, right) and GNP cells (Figure 6-4D, right), using Pearson correlation coefficient R. Of the 13 subsets, subset "k" was the only CD45$^+$ cell subset that showed a contrasting correlation with GPP cells (Figure 6-4C, right; R = +0.55) and GNP cells (Figure 6-4D, right; R = -0.5). Based on histogram analysis, the cellular phenotype of subset "k" was CD45^{int}CD11b^{int}HLA-DR^{int}, which resembled resident microglia (Figure 6-4E, bottom row) (*333, 334*). Additionally, subset "k" also expressed CD44, a feature that was previously shown to indicate

microglial activation (Figure 6-4E, right-most column) (*335, 336*). This analysis also confirmed the phenotype of other immune cell subsets present in glioblastomas, including memory T cells ("a" $CD3^{+}CD45RO^{+}$), non-memory T cells ("c" $CD3^{+}CD45RO^{-}$), monocyte-derived macrophages ("i" $CD45^{hi}CD11b^{hi}$HLA-DR^{hi}), and non-activated resident microglia ("j" $CD45^{int}CD11b^{int}$HLA-$DR^{int}CD44^{-}$) (Figure 6-4E).

Discussion

Value of single-cell signaling profiles

By coupling single-cell quantification of identity proteins and phosphorylated signaling effectors, this study uncovered novel glioblastoma cells that differentially predicted clinical survival with robust statistical significance. Single-cell mass cytometry was critical to revealing this signature in two ways. First, assessment of a large number of cells per tumor enabled this study to pursue statistically powerful approaches towards the identification of rare and novel cell subsets. Second, per-cell quantification of phosphorylation-specific signaling effectors was key to the discovery of novel cell subsets, which revealed potential mechanisms of tumor progression as well as interactions between distinct cell types in the tumor microenvironment. The GNP signature was defined by abnormal neural-lineage features and simultaneous high basal phosphorylation of many signaling effectors downstream of RTKs, in particular STAT5, which is known to be associated with cell growth and survival. In contrast, the GPP signature was defined by EGFR and CD44 enrichment, lack of proliferation, and specific lack of STAT5 phosphorylation. This result agrees with single-cell transcript analysis that revealed CD44 expression is widespread and heterogeneous within tumors (*5*). Thus, tracking signaling in subpopulations of tumor cells and aligning this information with developmental markers proved critical to identifying prognostic populations.

Cellular heterogeneity and homogeneity in glioblastoma

Previous studies of blood cancers have reported contrasting results regarding the role of intra-tumor signaling heterogeneity. Studies of leukemia have clearly demonstrated that signaling heterogeneity is associated with stem-ness and poor clinical outcomes (*12, 80*). In contrast, in lymphoma, emergence and dominance of the tumor by a subset with abnormal suppression of signaling revealed signaling homogeneity in the worst outcome cases (*15*). Here, homogeneity in either GNP or GPP cells was associated with the most extreme ends of good and bad clinical outcome and heterogeneity was of intermediate outcome. For example,

GNP cells comprised more than 70% of poor outcome samples LC02 and RT07 and GPP cells comprised more than 90% of LC06, a good outcome case (Figure 6-2, Figure 6-S4, and Table S6-1). This difference is likely a result of the analysis strategy that focused on defining a cell population based on degree of risk stratification. Yet, it remains striking that GNP and GPP cells are consistent in phenotype across tumors and were found at greater than 1% abundance in >50% of the glioblastomas studied here (Figure S6-3). The abundance of GNP cells in post-therapy glioblastomas should now be measured to determine whether these cells further increase in abundance following treatment resistance.

Immune cells, microglia, and immunotherapy implications

Identification of negative prognostic signaling events and their associated cell types resulted in a clinically applicable detection method for characterizing human tumors (Figure S6-9) and revealed a close relationship between signaling in cancerous cells and microenvironmental changes. Prior studies of low-grade, *IDH*-mutant glioma suggested that vascularity of higher-grade tumors is accompanied by an increase in cells with a monocyte-derived macrophage transcriptional signature (*73*). The results here confirm that glioblastomas can contain a significant population of macrophages. In particular, tumors composed of GPP cells were also more likely to contain $CD45^{+}$ hematopoietic cells, including $CD45^{hi}CD11b^{hi}CD64^{+}HLA\text{-}DR^{+}$ macrophages (Figure 6-4). Separately, the abundance of $CD44^{hi}$ microglia ($CD45^{int}CD11b^{int}CD64^{+}HLA\text{-}DR^{+}$) was directly associated with GPP cell abundance and inversely associated with GNP cell abundance. Microglial CD44 expression denotes activation (*335, 336*). These results are consistent with the idea that prevalence and activation of tissue-resident microglia may be critical in eliciting anti-tumor immunity. It is not clear from these studies whether an altered immune microenvironment precedes development of an aggressive glioblastoma or whether the more aggressive tumors suppress anti-tumor immunity.

Further supporting the importance of the glioblastoma immune microenvironment is that clinically favorable glioblastoma cells showed enrichment of p-STAT1 and p-NFκB (Figure 6-3), transcription factors that activate antigen presentation, inflammation, and anti-tumor immunity (*337-342*). In melanoma, STAT-driven expression of HLA-DR, an MHC class II molecule, is associated with improved responses to checkpoint inhibitor immunotherapy (*343*). Here, unfavorable GNP cells were associated with diminished tumor-infiltrating immune

cell abundance (Figure 6-4 and Figure S6-8). While active proliferation marked by cyclin B expression was a hallmark of GNP cells (Figure 6-4), it is unlikely that the GNP cell can out-proliferate the cytotoxic immune cells and dominate the tumor microenvironment without some mechanism for suppression of the cytotoxic $CD8^+$ T cells observed within most tumors (Figure 6-3 and Figure 6-4). GPP cells with inflammatory STAT1 and NFκB signaling were not correlated with recruitment of total $CD45^+$ leukocytes (Figure S6-10) or a blood-derived immune cell subset (Figure 6-4 and Figure S6-11), but GPP were associated with activated, resident $CD44^+$ microglia (Figure 6-4 and Figure S6-11). These findings suggest that immunotherapy is likely to be most productive in tumors containing GPP cells and that additional research is needed to understand whether GNP cells directly suppress microglia or immigrant leukocytes.

Relationship between glioblastoma cell signaling and outcome

GNP cells were closely associated with the unfavorable clinical outcome, while presence of GPP cells primarily predicted delayed time to progression or recurrence. Although the divergent connections of GNP and GPP cells with patient outcome were apparent, correlation between these two novel cell subsets was not observed (Figure S6-4C). These findings indicated that GPP and GNP cells might have a complex mechanistic relationship, which would be valuable to explore in future studies. The discovery of GNP cells, abnormal both in neural-lineage phenotype and basal signaling activity, has the potential to lead to new therapeutic approaches for glioblastoma. The negatively prognostic STAT5 phosphorylation in GNP cells, for example, should be explored as a potential therapeutic target and a biomarker of therapy response. Critically, the identification of signaling features that were specifically enriched in clinically distinct glioblastoma cells led to a development of a lower-dimensional pipeline which can be immediately adopted for clinical stratification.

Acknowledgements: We thank the Irish and Ihrie labs at Vanderbilt University for helpful discussions, and Dr. Chris Wright for his invaluable suggestions during the preparation of this manuscript. **Funding:** NIH/NCI R00 CA143231 (J.M.I.), the Vanderbilt-Ingram Cancer Center (VICC, P30 CA68485), the Vanderbilt International Scholars Program (N.L.), a Vanderbilt University Discovery Grant (J.M.I. and N.L.), Alpha Omega Alpha Postgraduate Award (A.M.M), Society of Neurological Surgeons/RUNN Award (A.M.M), T32 HD007502 (J.S.), F31 CA199993 (A.R.G.), R25 CA136440-04 (K.E.D.), the Vanderbilt Medical Scholars Program (J.S.G.), a VICC

Provocative Question award (J.M.I.), R01 NS096238 (R.A.I), DOD W81XWH-16-1-0171 (R.A.I.), the Michael David Greene Brain Cancer Fund (R.A.I), VICC Ambassadors awards (J.M.I. and R.A.I.), and the Southeastern Brain Tumor Foundation (J.M.I. and R.A.I.). **Author contributions:** N.L., R.A.I., and J.M.I. designed the study. N.L. and J.S. performed experimental work. N.L A.M.M., J.S., R.A.I., and J.M.I., performed data analysis, developed figures, and wrote the manuscript. A.M.M. compiled patient data and performed survival analysis. K.E.D. and A.R.G. developed R scripts data analysis and visualization. J.S.G. and D.L. implemented FlowSOM R script. K.D.W., R.C.T., and L.B.C. provided clinical samples. B.C.M. provided pathological diagnosis and tumor molecular status. R.A.I. and J.M.I. provided financial support. All authors contributed in reviewing the manuscript. **Competing interests:** J.M.I. is a co-founder and a board member of Cytobank Inc. and received research support from Incyte Corp and Janssen. **Data and materials availability:** Data files are available online at https://www.cytobank.org/irishlab/.

Supplemental Material

Materials and Methods

Patient samples

Surgical resection specimens of 28 *IDH*-wildtype glioblastomas collected from the Vanderbilt University Medical Center between 2014 and 2016 were processed into single cell suspensions following an established protocol (Table S6-1) (*285*). Only samples that were confirmed to be *IDH*-wildtype glioblastomas by standard pathological diagnosis were used. All samples were collected with patient informed consent in compliance with Vanderbilt Institutional Review Board (IRB #131870), and in accordance with the declaration of Helsinki.

Patient characteristics and collection of clinical data

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 classified as either gross total or subtotal resection by both a neurosurgeon and a neuroradiologist independently. Gross total resection was defined as consensual agreement 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 (*1*), 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 October 2017, noting time to first, definitive radiographic progression or recurrence of glioblastoma consensually 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.

Mass cytometry analysis

Cells derived from patient samples were prepared as previously described (*285, 312*). Rhodium viability stain and cleaved caspase 3 antibody were included to exclude non-viable and apoptotic cells, respectively. 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 harsh permeabilization with ice-cold methanol (Table S6-2) (*285*). Detection of total histone H3 was used to identify intact, nucleated cells (*284*). Cells were resuspended in deionized water containing standard normalization beads (Fluidigm) (*313*), and collected on a CyTOF 1.0 at Vanderbilt University. A 32-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 (*313*), and were arcsinh transformed (cofactor 5), prior to analysis on a web-based Cytobank software (*301*). Mass cytometry data are publicly available at http://vanderbilt.cytobank.org/.

Cell subset identification by FlowSOM

Cell subset identification was performed in R, using the previously published FlowSOM R package (*293*). t-SNE values (t-SNE1_glioblastoma and t-SNE2_glioblastoma; Fig. 2B) from viSNE analysis of $CD45^{-}CD31^{-}$ glioblastoma cells from 28 patients were used as parameters for cell subset clustering. The number of cell subsets assigned to the algorithm ranged from 5 to 45. The optimal number of clusters was defined as 43, since it was the fewest number of clusters that effectively predicted patient clinical outcome while minimizing intra-cluster signal variance (Figure S6-2).

MEM quantification

MEM analysis was performed in R, using the previously published R package (*321*). In short, MEM captured and quantified cell subset-specific feature enrichment by scaling the magnitude (median) differences between samples, depending on the spread (interquartile variance, IQR) of the data. These values were then computed in comparison to the remaining cells in a given dataset. MEM values can be interpreted as either being positively enriched (▲, positive values) or negatively enriched (▼, negative values). The variation of a given cellular feature across GNP or GPP cell subsets were quantified as ± standard deviations (SD).

Survival and statistical analysis

Time from surgical resection to death (overall survival, OS) and time from surgical resection to the initial radiographic progression/recurrence or death before radiographic assessment (progression free survival, PFS)

were plotted and analyzed using GraphPad Prism version 7.00 for Windows (GraphPad Software, San Diego, CA, USA) using right-censored Kaplan-Meier curves. 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 were compared using the Gehan-Breslow-Wilcoxon method (i.e. Wilcoxon test), reporting a hazard ratio between the survival curves.

The abundance of glioblastoma cell subsets and their clinical significance was assessed using outcome-guided analysis (Figure 6-2 and Figure S6-2). Patients were divided into Low and High groups, based on the distribution (interquartile variance) of the abundance of a given cell subset across the cohort. A Wilconxon test was then used to assess the p-value and the hazard ratio (HR) of death. Glioblastoma cell subsets that were significantly ($p < 0.05$) associated with early death ($HR > 1$) were termed Glioblastoma Negative Prognostic (GNP) cells, whereas the subsets that were significantly ($p < 0.05$) associated with prolonged survival ($HR < 1$)) were termed Glioblastoma Positive Prognostic (GPP) cells. Abundance of GNP and GPP cells is each tumor was quantified as percentages per total glioblastoma cells.

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, glioblastoma *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 3.4 (R foundation for Statistical Computing, Vienna, Austria), utilizing the rms library (version 4.5). 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.

Table S6-1 – Patient characteristics

Sample ID	Gender	Age	TMZ	RT	EOR	MGMT meth	PFS (days)	OS (days)	Status
K01	Female	59	Yes	Yes	STR	No	176	364	Dead
LC02	Male	67	No	No	STR	No	53	53	Dead
LC03	Male	60	No	No	STR	No	57	57	Dead
LC04	Male	65	Yes	Yes	GTR	No	263	918	Dead
LC06	Male	41	Yes	Yes	GTR	No	958	1136	Alive
LC08	Female	55	Yes	Yes	STR	No	210	441	Dead
LC09	Male	68	Yes	Yes	STR	Yes	66	110	Dead
LC10	Male	78	No	Yes	STR	No	117	178	Dead
LC11	Female	76	Yes	Yes	STR	Yes	111	411	Dead
LC12	Male	40	Yes	Yes	GTR	Yes	606	724	Alive
LC18	Male	69	Yes	Yes	STR	No	133	240	Dead
LC21	Female	69	Yes	Yes	GTR	Yes	267	267	Dead
LC22	Male	64	Yes	Yes	STR	No	263	366	Dead
LC25	Female	60	Yes	Yes	GTR	No	185	391	Alive
LC26	Male	71	Yes	Yes	STR	Yes	363	386	Alive
LC27	Male	62	Yes	Yes	STR	No	79	252	Dead
RT01	Female	69	Yes	Yes	STR	No	162	198	Dead
RT07	Female	70	Yes	Yes	STR	No	49	57	Dead
RT10	Male	56	No	No	STR	Yes	22	113	Dead
RT14	Female	55	Yes	Yes	GTR	Yes	472	483	Alive
RT15	Female	80	No	Yes	STR	No	98	353	Dead
W02	Male	75	Yes	Yes	STR	Yes	446	507	Dead
W04	Male	60	Yes	Yes	STR	No	220	456	Dead
W05	Male	60	Yes	Yes	STR	No	241	282	Dead
W11	Male	69	Yes	Yes	GTR	No	491	504	Alive
W12	Male	50	Yes	Yes	STR	Yes	190	505	Alive
W14	Male	47	Yes	Yes	STR	Yes	125	452	Alive
W15	Male	50	Yes	Yes	STR	Yes	400	441	Alive

TMZ = temozolomide
RT = radiation

EOR = extent of resection

STR = subtotal resection

GTR = gross total resection

MGMT meth = MGMT promotor methylation

Table S6-2 – Mass cytometry antibody panels

Target	Mass	Clone	Signaling & proteins		Immune	Stain		
			Panel	viSNE	Panel	Live	Sap	MeOH
Rhodium	103	-	●		●	✓		
Cyclin B1	139	GNS-1	●					✓
TUJ1	141	TUJ1	●	■				✓
cCasp3	142	5A1E	●					✓
CD117	143	104D2	●			✓		
S100B	144	19/S100B	●	■				✓
CD64	144	10.1			●	✓		
CD31	145	WM59	●	■*		✓		
CD34	148	581	●	■		✓		
CD45RO	149	UCHL1			●	✓		
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	159	GoH3	●	■		✓		
CD133	160	AC133	●	■		✓		
PDGFRα	161	16A1	●	■		✓		
SOX2	163	O30-678	●	■			✓	
SSEA-1	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	●	■	●	✓		
NCAM/CD56	175	HCD56			●	✓		
Histone H3	176	D1H2	●		●			✓

● = included in the panel

■ = included for generation of viSNE map

* Excluded from viSNE analysis of only glioblastoma cells

Live = live surface stain

Sap = 0.02% saponin stain

MeOH = stain after ice-cold methanol permeabilization

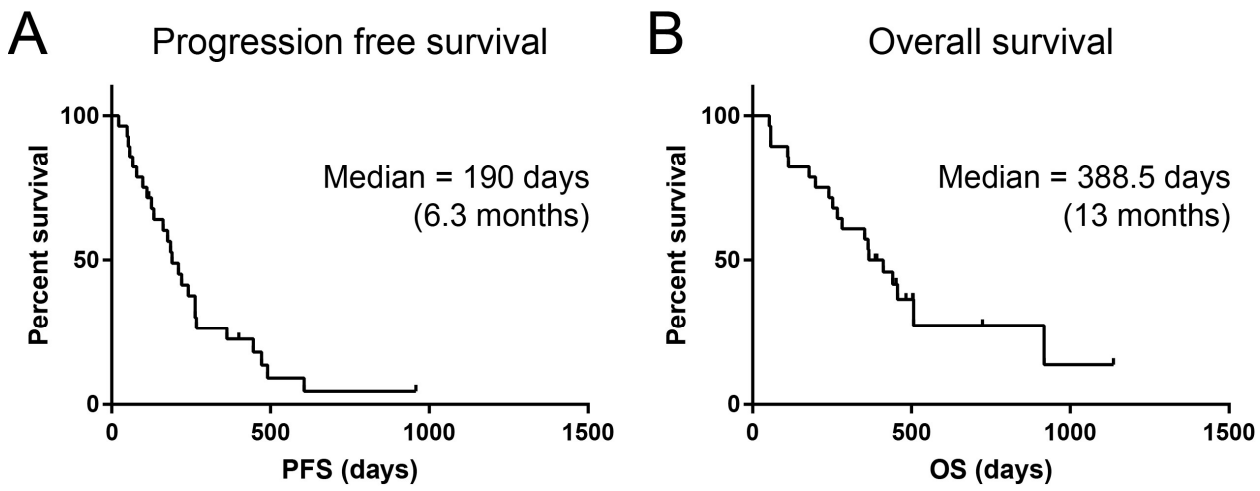


Figure S6-1 Survival of 28 adults with IDH-wildtype glioblastomas. (**A**) Progression-free survival (PFS) and (**B**) overall survival (OS) of 28 patients with *IDH*-wildtype glioblastomas are depicted in Kaplan-Meier survival plots. Censored values are indicated by tick marks on the curves. The median PFS was 190 days (6.3 months), and the median OS was 388.5 days (13 months).

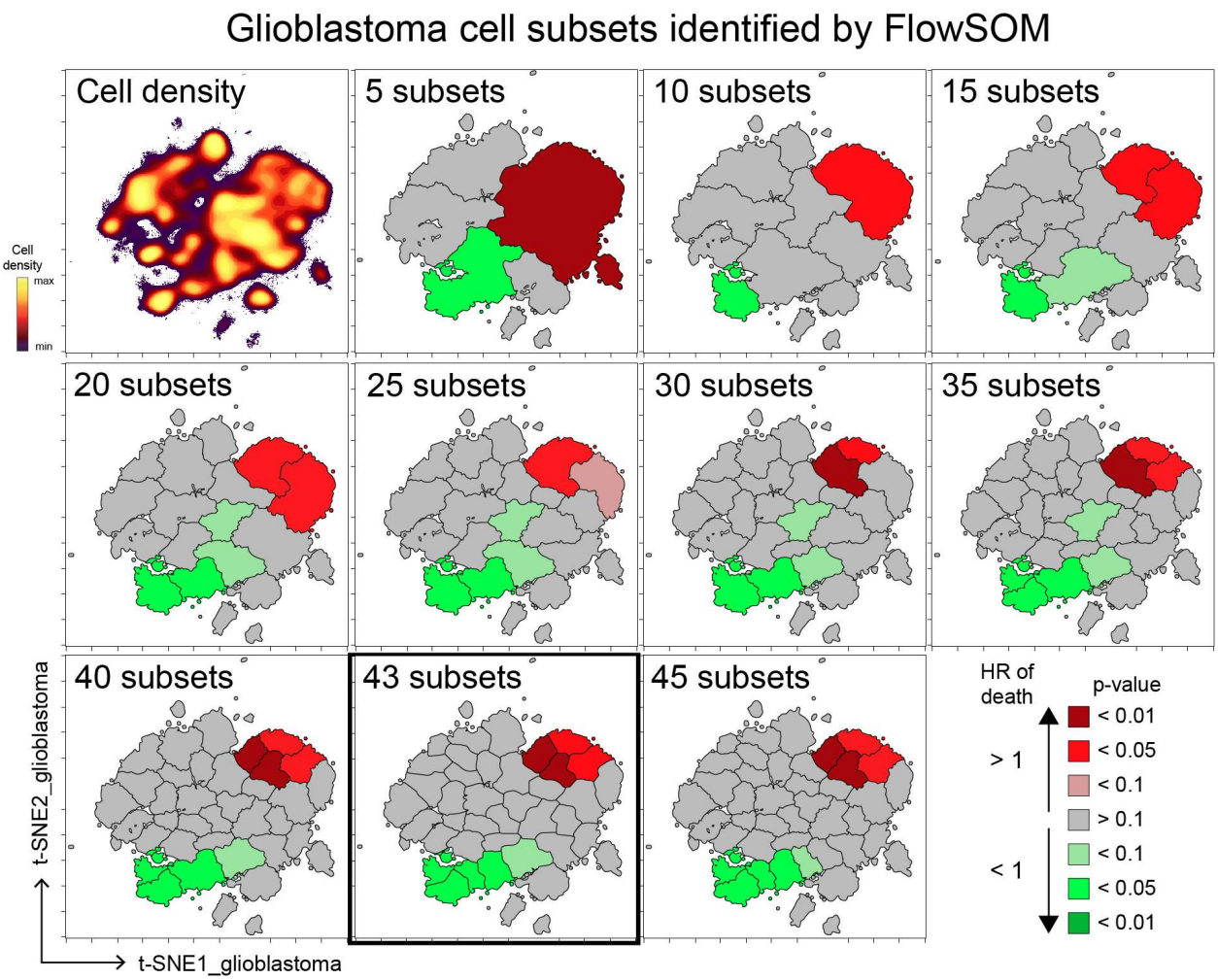


Figure S6-2 Identification of glioblastoma cell subsets that significantly stratify patient overall survival. Glioblastoma cells from 28 tumors were combined for a new viSNE analysis. The top-left map shows cell density. Automated FlowSOM clustering was used to identify glioblastoma cell subsets, ranging from 5 to 45 subsets. The overall survival of patients with high vs low abundance of each cell subset was assessed using the Wilcoxon test. Cell subsets were are color-coded based on hazard ratio (HR) of death (HR>1, red; HR<1, green) and p-values. The optimal cell subset number, 43, is marked with thick borders (see Figure 6-2). For cell FlowSOM and survival analyses, see Supplementary Methods.

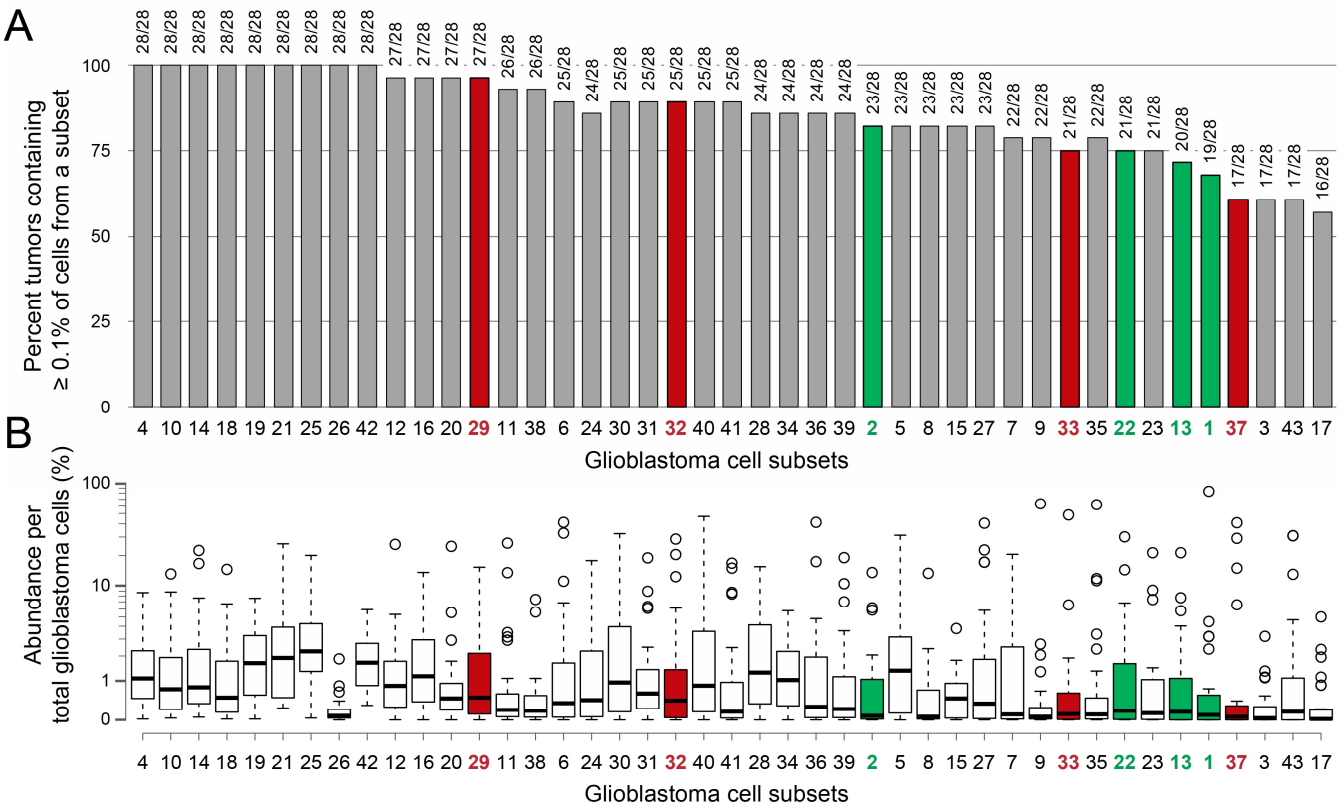


Figure S6-3 Glioblastoma cell subsets were present across patients at varying abundance. (**A**) Percent of tumors (y-axis) whose tumors were composed of at least 0.1% of a given glioblastoma cell subset (x-axis) is depicted. Glioblastoma cell subsets are arranged from being present in 100% of tumors (left) to being present in fewest tumors (right). (**B**) Abundance of glioblastoma cell subsets (%, y-axis) per tumor is shown. Each data point (open circle) represents one tumor. Horizontal lines depict the median abundance of cell subsets across all tumors. Boxes and error bars represent the 25th to 75th percentile and the standard deviation, respectively. Cell subsets that constitute GNP cells are labeled in red, and those that constitute GPP cells are labeled in green.

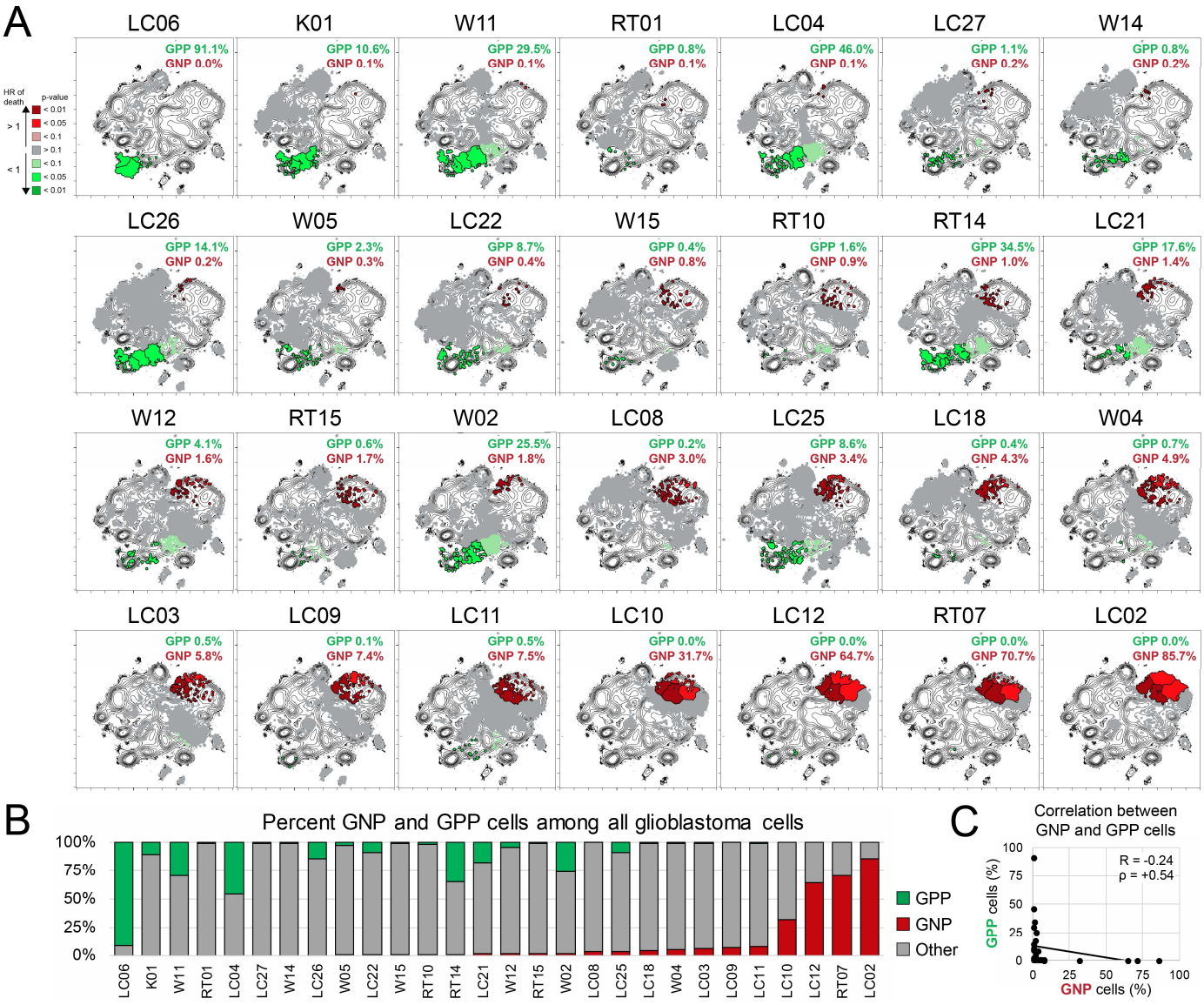


Figure S6-4 Abundance of GNP and GPP cells were not significantly correlated to each other. (**A**) Glioblastoma cells in each tumor are displayed over a cell density contour of the viSNE analysis of all the glioblastoma cells from 28 tumors. Abundance of GPP (green) and GNP (red) cells in each tumor is shown as percentages. Tumors are arranged from that with the lowest abundance of GNP cells (LC06; 0.0%) to the tumor with the most abundant GNP cells (LC02; 85.7%). Glioblastoma cells that were non-predictive of clinical outcome are shown in gray. (**B**) Stacked bar graph illustrates the abundance of GPP cells (green), GNP cells (red), and other glioblastoma cells (gray) per total glioblastoma cells (y-axis) in each patient (x-axis). (**C**) Correlation between the abundance of GNP (x-axis) and GPP cells (y-axis) was assessed by Pearson's correlation coefficient (R = -0.24) and Spearman's rank (ρ = +0.54) tests. Each data point represents one tumor.

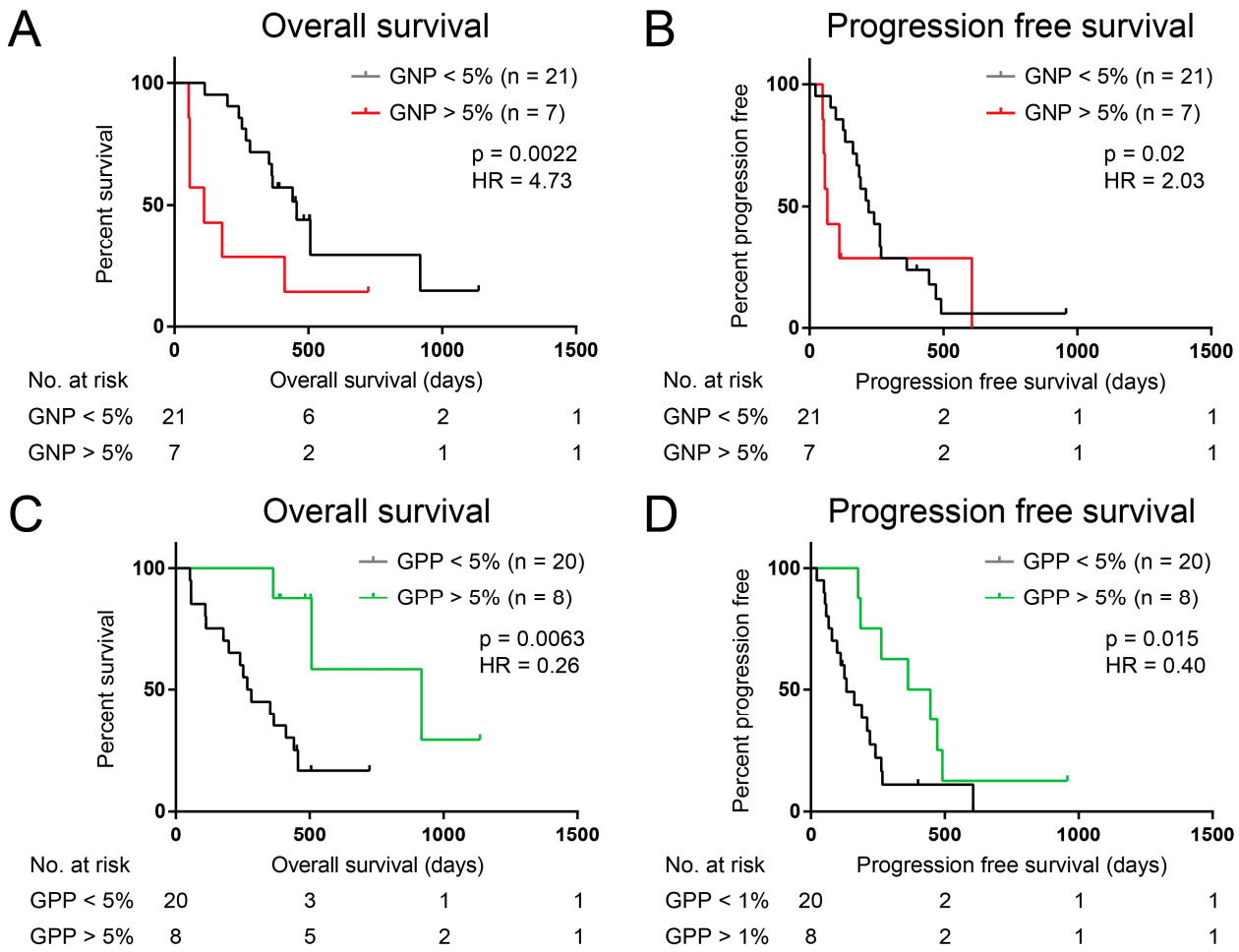


Figure S6-5 Glioblastoma Negative Prognostic cells and Glioblastoma Positive Prognostic cells contrastingly linked to patient clinical outcomes. The associations of GNP cell abundance with overall survival (**A**) and progression free survival (**B**) was assessed using Wilcoxon tests. The p-values and hazard ratios are depicted in the corresponding Kaplan-Meier plots. Similar analyses of GPP cell abundance were performed for their associations with overall survival (**C**) and progression free survival (**D**).

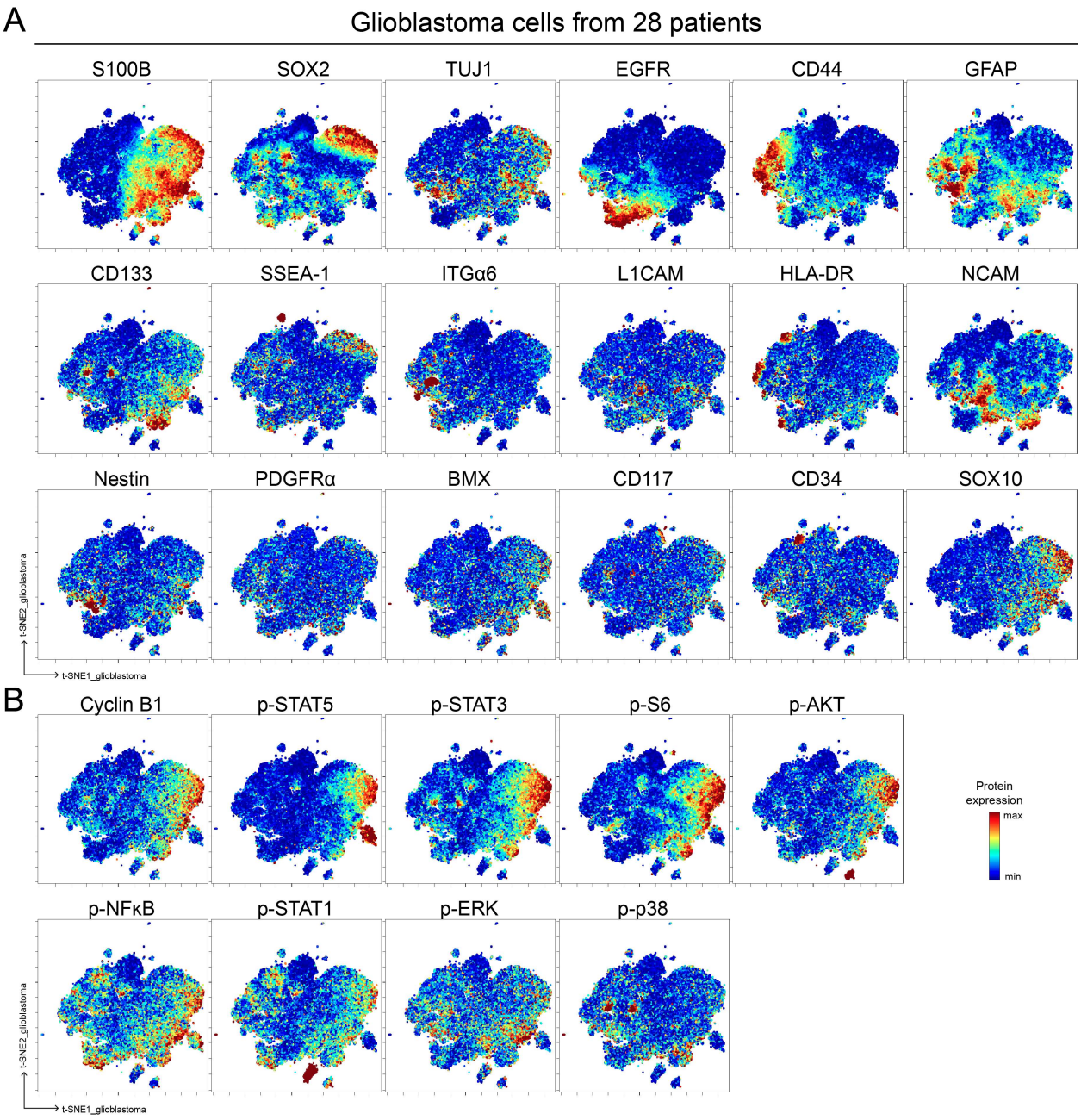


Figure S6-6 Expression of identity proteins and phosphorylation-specific signaling effectors in individual glioblastoma cells. viSNE analysis was performed on glioblastoma cells combined from 28 tumors. (**A**) Per-cell expression levels of 18 identity proteins, (**B**) 8 phosphorylated signaling effectors and proliferation marker cyclin B1 are depicted.

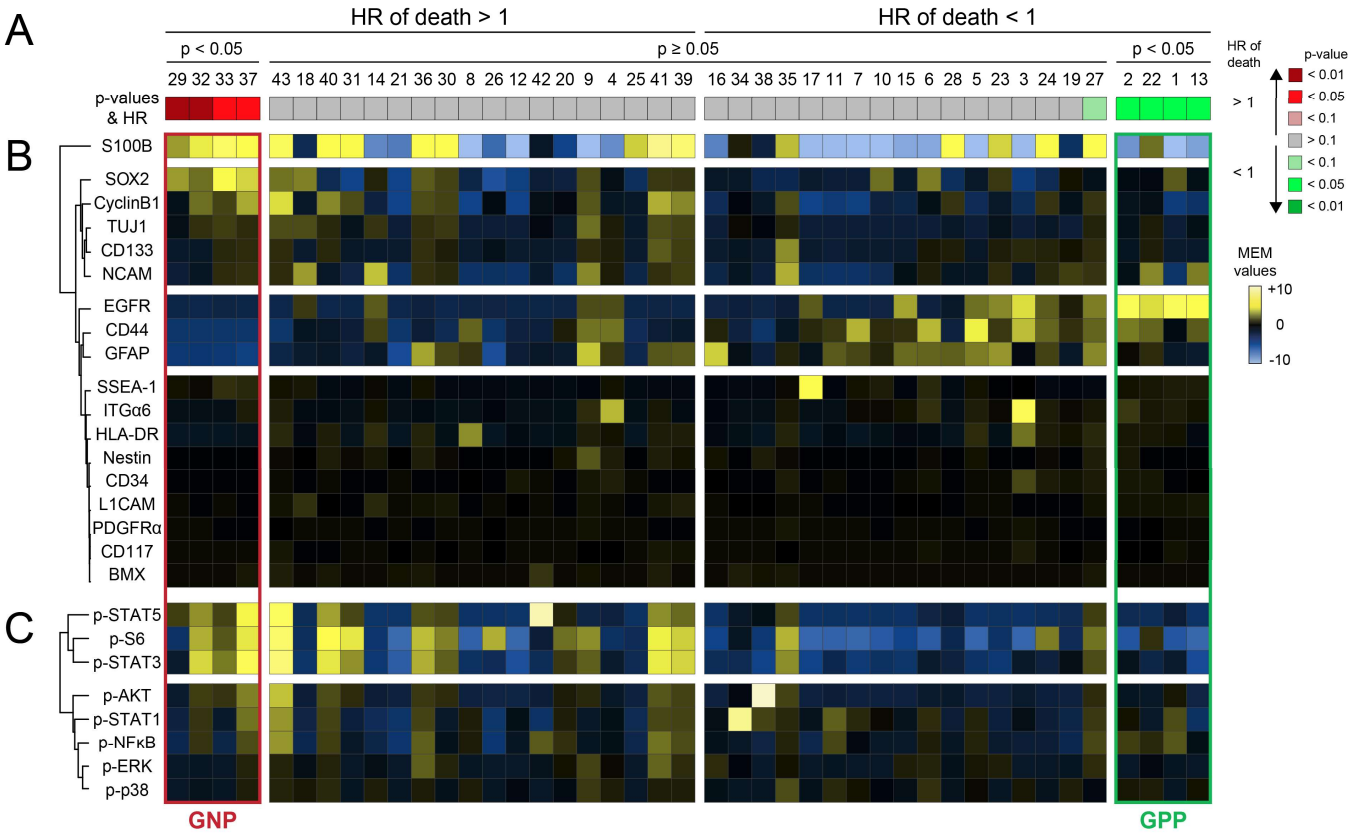


Figure S6-7 Glioblastoma cell subsets showed differential enrichment of identity proteins and phosphorylated signaling effectors. (**A**) Forty-three glioblastoma cell subsets that were automatically identified by FlowSOM are arranged according to their associations with overall survival (HR>1, red, left; HR<1, green right) and statistical significance (p-values). (**B**) Heatmap represents the identity-protein MEM values of glioblastoma cell subsets (columns). GNP cells are labeled in red, which GPP cells are labeled in green. Hierarchical clustering of identity proteins was performed based on their MEM values and is depicted on the left of the heatmap. (**C**) MEM values and hierarchical clustering of phosphorylation-specific signaling effectors of each glioblastoma cell subset were also analyzed. HR = hazard ratio of death.

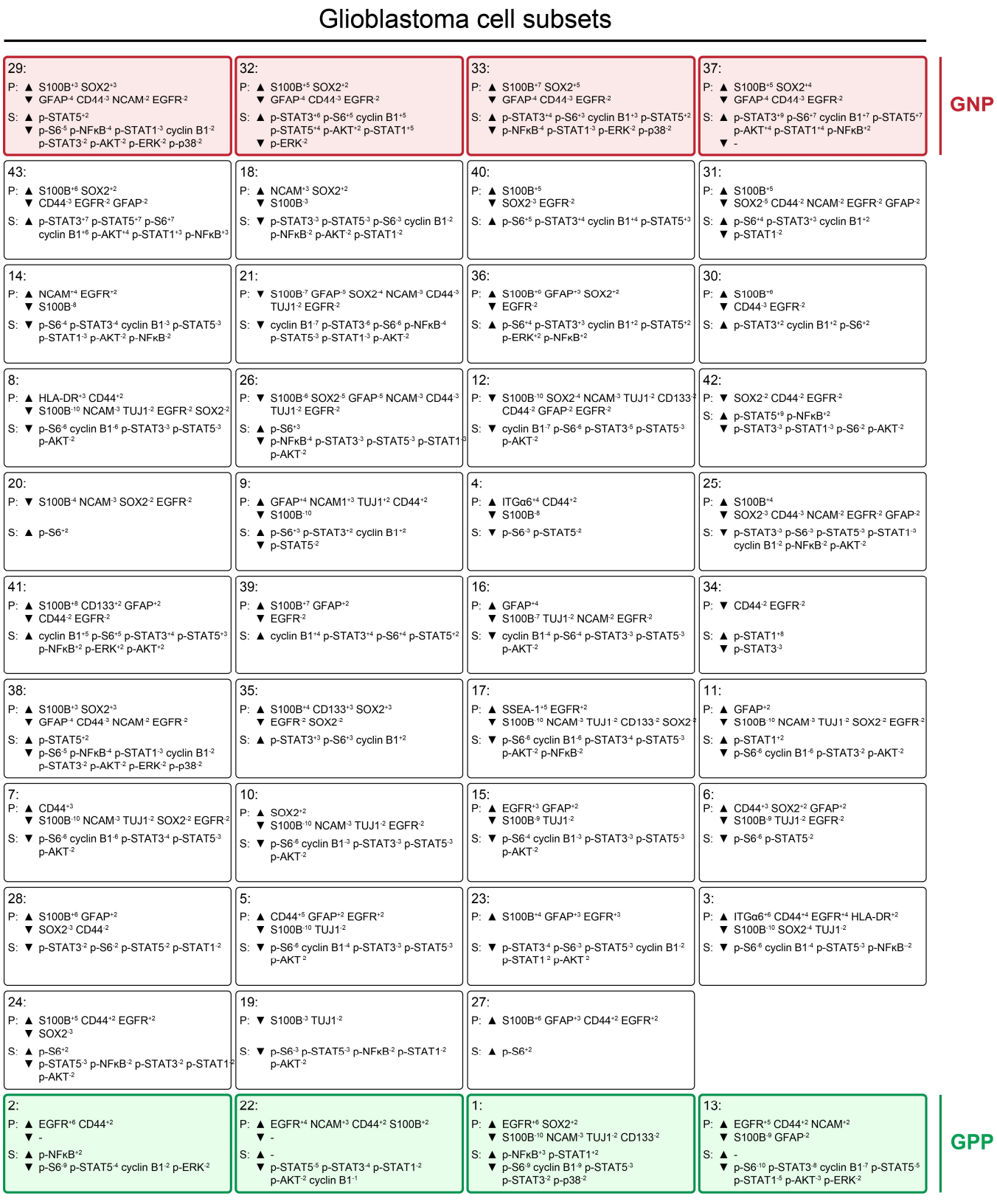


Figure S6-8 Quantitative MEM labels of the enriched identity proteins and signaling features of glioblastoma cell subsets. Enrichment of identity proteins (P) and phosphorylated signaling effectors (S) of glioblastoma cell subsets was quantified using MEM. GNP and GPP cells are labeled in red and green, respectively. See also Figure 6-3 and Figure S6-7.

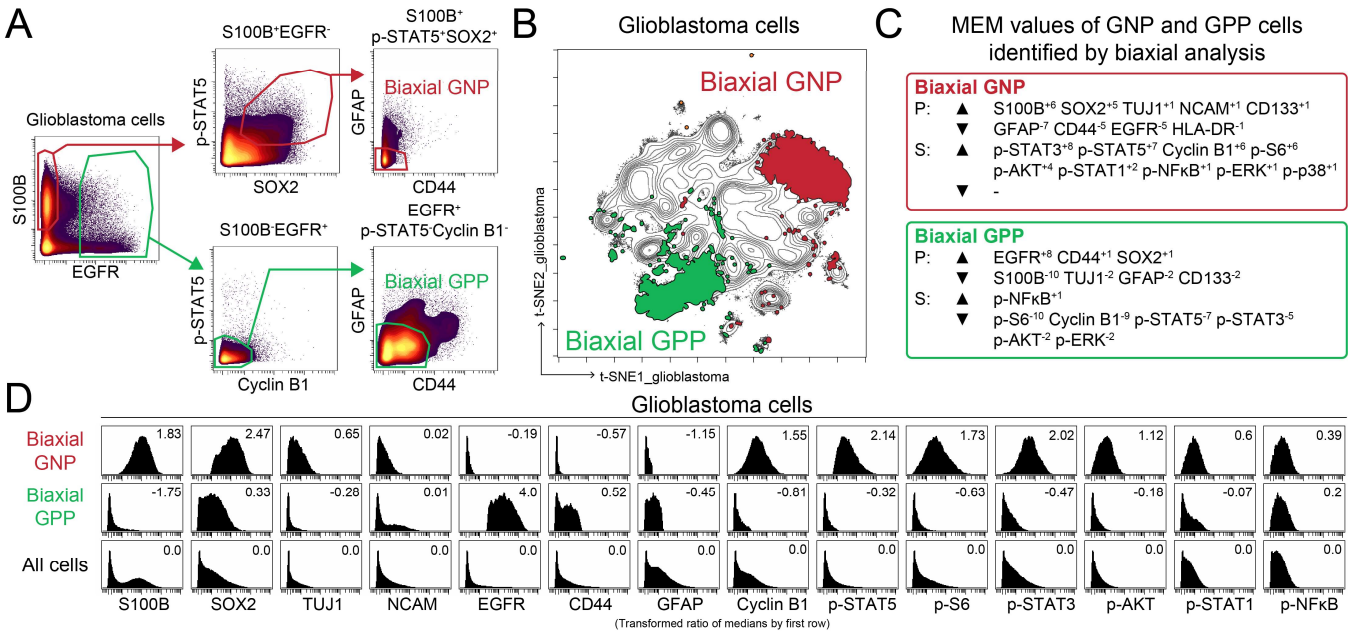


Figure S6-9 A standard low-dimensional biaxial analysis workflow effectively identifies clinically distinct glioblastoma cell subsets. (**A**) Biaxial plots demonstrating a sequential gating scheme compatible to standard clinical flow cytometry workflow. Biaxial GNP and biaxial GPP cells were identified using red and green gates, respectively. (**B**) Biaxial GNP (red) and biaxial GPP (green) cells are overlaid over a contour of glioblastoma cells from 28 tumors subjected to a common viSNE analysis as shown in Figure 6-2. (**C**) MEM analysis was used to quantify enriched identity proteins (P) and phosphorylated signaling effectors (S) of biaxial GNP and biaxial GPP cells. (**D**) Histogram analysis depicts the expression of key identity proteins and phosphorylated signaling effectors of biaxial GNP (red) and biaxial GPP cells (green). This was compared to all glioblastoma cells (black, bottom row). The transformed ratio of medians compared to all glioblastoma cells are also specified.

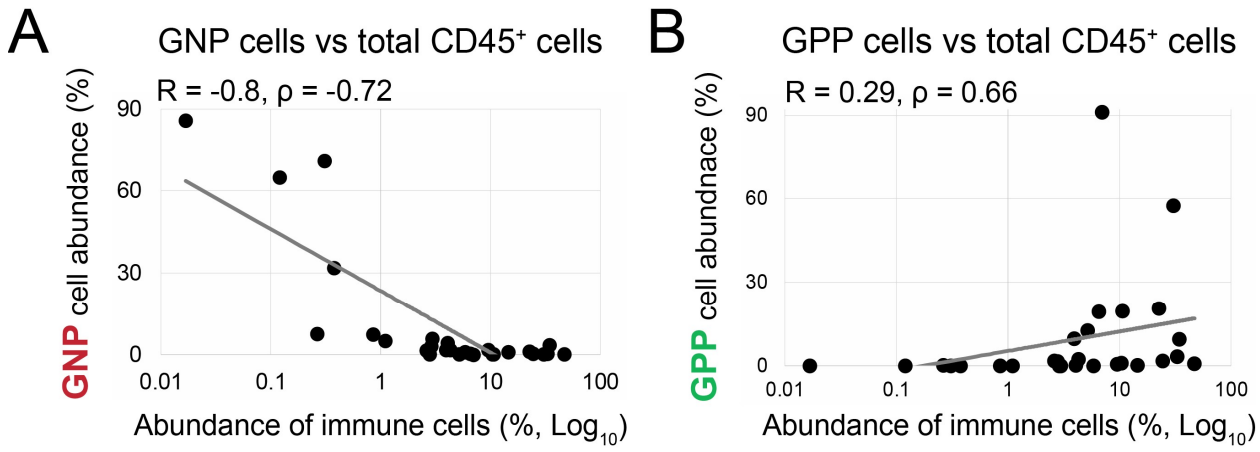


Figure S6-10 Abundance of CD45+ immune cells inversely correlated with GNP cells but was not associated with GPP cells. (**A**) The correlation between the abundance of total CD45⁺ immune cells (x-axis, % per total cell in tumor) and GNP cells (y-axis, % per glioblastoma cells) was assessed using Pearson's correction coefficient (R) and Spearman's rank (ρ). (**B**) A similar correlation between GPP cells and immune cells was also evaluated as depicted.

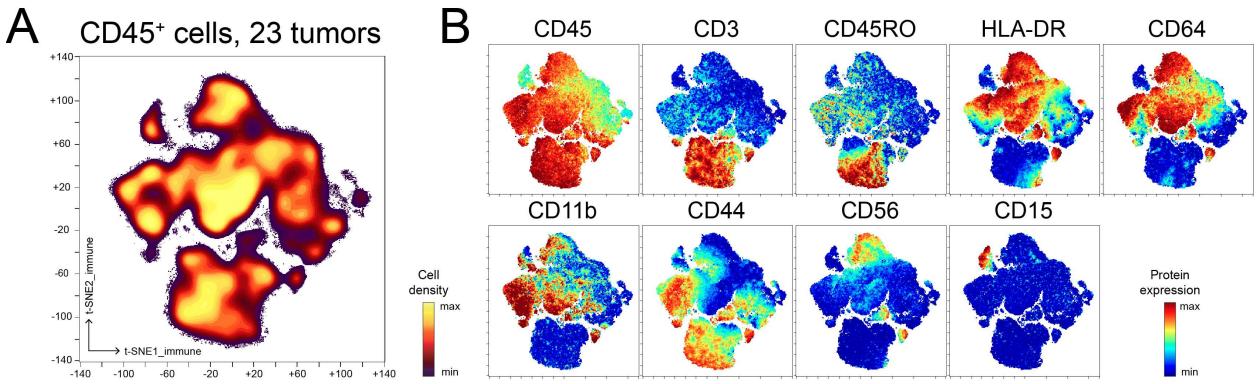


Figure S6-11 viSNE analysis revealed cellular diversity among CD45+ immune cells. (**A**) CD45+ cells from 23 glioblastomas were subjected to an immune-specific viSNE analysis. Cell density is depicted. (**B**) Per-cell expression of 9 surface immune identity proteins that were used to generate the viSNE map are shown.

CHAPTER 7

Summary and Future Directions

Summary of dissertation

Cellular diversity is a core component of human malignancies. A thorough understanding of the cell subsets and their mechanistic hindrance that drives tumor progression, therapy resistance, and eventual recurrence, is essential for disease stratification and the development of novel therapeutics. Advanced technological approaches and computational tools have made systematic, unbiased, single-cell analyses attainable. Integration of single-cell assessments with pertinent automated analytical approaches, which are tailored for specific systems and objectives, is critical for the discovery of features that drive poor outcome.

This dissertation presents a modular workflow that combines single-cell proteomics quantification with outcome-guided feature discovery of human adult glioblastoma, a rapidly fatal brain tumor (Chapter 6). This automated workflow was systematically built upon tools that were successfully established for quantitative single-cell analysis of hematologic malignancies (Chapter 4). This pipeline also couples a new technological development that enables derivation of viable single cells while preserving cellular diversity, a critical hurdle in solid tumor single-cell analysis (Chapter 5 and Appendix A), with an advanced automated feature quantification (Appendix E). This has led to the discovery of novel glioblastoma cancer cell subsets, and their subset-specific signaling bottlenecks, that strongly and quantitatively predicted tumor progression and patient survival. The novel signaling hallmarks in the stratifying glioblastoma cell subsets can be used as predictive biomarkers and should be immediately tested as targets for innovative therapies. Moreover, this automated pipeline can be readily evaluated and modified for applications in other solid malignancies.

Measurements and displays of single-cell proteomic features

Chapter 4 presents the basic single-cell mass cytometry protocol and the key analytical tools. Using a well-characterized, cytometry-ready human PBMCs as an example, this study demonstrated the power of mass cytometry for per-cell quantification of proteomics features that reflect cell identity. Dimensionality-reduction tools, such as SPADE (*282*) and viSNE (*290*), efficiently displayed the global phenotypic connectivity of single cells in PBMCs. The eventual subset delineation still heavily relied on manual identification by means of gating.

This step can be difficult in tissues that are less well-characterized or those that display erratic phenotypic aberrancies, such as cancers. Although there have been many developments of automated cell subset identification algorithms, identifying the approach that can efficiently and automatically derive clinically meaningful cell subsets is especially challenging.

The critical obstacle is to pinpoint the definitions of the cell subsets. Cells in human PBMCs carry distinct protein phenotypes that correlate well with their physiologic functions. For example, $CD38^+$ plasma cells produce antibodies, whereas $CD8^+$ T cells execute cytotoxic functions. In human malignancies, however, the significance of cell subset definition is generally shifted towards the identification of cells that govern therapy resistance or tumor recurrence (*14, 15, 86*).

Previous work in hematologic malignancies demonstrated that single-cell signaling activities that occur in response to external potentiation were closely associated to patient outcome (*14, 15, 80*). Even though surface protein identity and signaling activation showed strong correlations in healthy blood cells, signaling profiles better defined leukemic cells that predicted poor survival (*12, 14, 15*). These studies are examples of alternative approaches to define cell subsets beyond identity protein expression. A single-cell *ex vivo* signaling potentiation protocol, which was the basis for these stratifications, is also described in detail in Chapter 4. Using a human leukemia cell line as an example, post-translational modifications of key signaling effectors were quantified. This protocol efficiently revealed distinct cellular mechanistic adaptations to various environmental cues.

Solid tissue and tumor dissociation: Overcoming the roadblock towards single-cell analysis

Single-cell characterization in healthy blood and hematologic malignancies have made significant advancement during the past few decades. As samples are readily in suspension, they require minimal preparation for various single-cell analytical platforms, including all flow cytometry. To broadly apply these valuable single-cell tools to enable similar in-depth analyses in solid tissues and tumors, a combined mechanical and enzymatic dissociation protocol was developed, as described in Chapter 5 (see Appendix A for a complementary step-by-step protocol).

This protocol is a major advancement for solid tumor single-cell biology. Not only can it be applied in many types of solid tissues and tumors (including glioblastomas), viable cells that are derived using this

standardized universal protocol are suitable for quantitative analysis on multiple flow cytometry platforms (fluorescence and mass). Moreover, the protocol maximizes cell viability and preserves the diversity of cells within the original tissues and tumors, making it ideal for the discovery of features even in a rare cell population.

This advancement had bridged the critical gap in the characterization of cell subsets and the degree of cellular diversity within individual glioblastomas. Even though cellular diversity has been observed in previous glioblastoma studies, the conclusions were usually drawn from a small number of single cells and were rarely based on samples that were directly derived from patients without prior *ex vivo* cell expansion or xenograft generation (*5, 44, 48, 73*). In contrast, the development described here enabled 1) high-dimensional single-cell characterization in a large number of cells (millions of cells per sample), and 2) quantification of cellular proteomics features, which can also include post-translational modifications that infer signaling activities (Chapter 6).

Outcome stratification and the discovery of therapeutic target candidates in adult primary glioblastomas

The discovery of novel glioblastoma cell subsets presented in Chapter 6 resulted from an analysis that was built upon 1) single-cell protein quantification (Chapter 4), 2) the development of a universal tumor dissociation workflow (Chapter 5), and 3) a system-tailored risk stratification approach. In contrast to the molecular alterations in glioblastomas that were identified using bulk-tumor analyses (*2, 3, 22*), the single-cell approach precisely distinguished stromal features (immune and vascular) from glioblastoma-specific characteristics. Importantly, a clustering algorithm, resulting in automated cell subset identification, was incorporated into the workflow enabling an unbiased discovery of features that may be the mechanisms of poor outcomes (*293*).

Critically, the overall objective was to identify glioblastoma cells that were closely associated with extreme patient outcomes (either favorable or unfavorable), as well as their biological defining features. The novel clinical stratifying glioblastoma cell subsets were characterized by two core features. First, single cells in each subset had minimal variations of protein identity expression and signaling effector phosphorylation, indicating that the glioblastoma cell subsets were relatively uniform and that the biological characteristics were distinct between subsets. Second, these cell subsets significantly predicted tumor progression and patient overall survival.

Comprehensive analysis further revealed that GPP cells were the main predictor of prolonged time to tumor progression, while GNP cells were closely associated with shorter overall survival. This is the first discovery in glioblastoma that there are cancer cells with contrasting clinical influences co-existing within individual tumors. Interestingly, the abundance of these two populations did not appear to be strongly inversely correlated. This suggested that although it is possible that the GPP and GNP cells may mechanistically interact and oppose each other, there are other factors that independently influence the presence and the balance of these cells, such as presence of immune cell subsets and their differential activation.

Importantly, the association of the cell subsets and patient outcome was the critical defining feature of glioblastoma cell subsets and led to the discovery of clinically distinct populations. Moreover, single-cell mass cytometry quantified simultaneous activation states of multiple signaling effectors, which are considered to be core converging points of multiple pathways that are commonly activated in glioblastoma. Therefore, these signaling effectors are representations of the global signaling activity that could be secondary to activation of different surface receptors. Furthermore, the signaling states that define the glioblastoma prognostic cells should be tested as biomarkers that predict responsiveness to targeted therapies (*260*), and should be evaluated as indicators of specific, individualized combinatorial therapy.

Evidence of lower-grade gliomas suggested a tight connection between tumor grades and the component of the tumor-infiltrating immune cells, indicating a close relationship between cancer cells and the cell in the tumor microenvironment (*73*). The work presented in Chapter 6 demonstrates that the abundance of total immune cells in glioblastoma varied significantly between tumors, ranging from less than 0.1% to more than 50% of total cells. Further investigation revealed that the abundance of the brain resident microglia that were phenotypically activated significantly contributed to the differences observed. Specifically, unfavorable tumors lacked the phenotypically activated resident microglia, which were highly abundant in the more clinically favorable glioblastomas. The signaling profile of a subset of GPP cells suggested that glioblastoma-infiltrating immune cells might be directly interacting with cancer cells to elicit an anti-tumor response. This discovery is especially valuable for future studies that aim to characterize the efficacy of immunotherapy, or even how targeting cancer cells impact the homeostasis within the tumor microenvironment.

Future directions

Bridging cutting-edge single-cell discoveries and clinical applications

The modular workflow and biological discovery described in this dissertation has significant potential for clinical applications. With the increasing knowledge of biological features that predict outcome and suggest targeted therapies, high-dimensional analyses of single cells can become the future of standard clinical laboratory practice. Considering the extent of cancer cell diversity even within a single tumor, the use of high-dimensional single-cell technologies, such as mass cytometry, as an initial laboratory approach could become a routine. The goal would be to map the cytomic characteristics and to profile the signaling regulators of individual cancer cells, in order to identify key features to track throughout the course of therapy, which may potentially indicate cellular targets for individualized treatment.

As demonstrated in Chapter 6, high-dimensional single-cell mass cytometry and risk stratification were critical to the identification of GPP and GNP cells in an unbiased fashion. Furthermore, MEM accurately quantified and revealed the most critical characteristics of the stratifying cell subsets. By using clinical significance as the defining feature of cell subsets, the clinically stratifying glioblastoma cells were revealed. This demonstrates that cancer cell subsets can, and probably should, be defined beyond the expression of proteins or other cellular features. Critically, signaling profiles of cancer cells should strongly be considered part of the routine clinical quantification since they were consistently associated with clinical outcome across many types of human malignancies. Here, signaling profiles showed very promising implications for clinical diagnostics, as well as for the development of stratification approaches that also suggest immediately testable therapeutic targets. Moving forward, assessments for GNP and GPP cells should also be evaluated as part of future clinical trials to evaluate the randomization of different therapeutic arms.

Novel surgical approaches to assist surgery in real-time

The highly infiltrative nature of glioblastomas is a major hurdle for surgical treatments which aim to physically remove the bulk of the tumor. Grossly, it is extremely challenging to distinguish the tumor from the surrounding brain that is often infiltrated by cancer cells. This almost always lead to an incomplete resection that can only be identified microscopically, despite the lack of signal enhancement per post-operative radiological

evaluations. Many advanced surgical approaches have been tested to maximize the extent of safe surgical resection of glioblastomas to maximally remove cancer cells, while preserving critical brain functions. Gamma knife has been shown to successfully identify the infiltrative "margins" of *IDH*-mutant tumors (*119*). This approach is based upon real-time mass spectrometry analysis of vaporized surgical waste to identify 2-HG, a cancer-specific biomarker in *IDH*-mutant tumors. However, its benefit so far had been limited to 5-10% of adult glioblastoma patients whose tumors can produce 2-HG.

The discovery of proteomic features of GNP cells described in this dissertation has the potential to broaden the clinical applicability of Gamma knife to benefit patients with *IDH*-wildtype glioblastomas. The key proteomics defining characteristics of GNP cells can be readily tested for mass spectrometry analysis, which will aid the surgical removal of grossly undisguisable, clinically unfavorable, glioblastoma cells. The next step is to determine if maximal removal of GNP features would improve patient outcome, which might be dependent upon whether these cells are responsible for tumor recurrence and/or patient death (see below under single-cell characterization of recurrent glioblastomas).

A practical approach towards individualized disease monitoring

The overarching goal of the work in this dissertation is to create a diagnostic platform towards individualized clinical care for glioblastoma patients. It is possible that, in the near future, disease monitoring and therapeutic regimens would be tailored based on the unique profiles of single cells that comprise individual glioblastomas. High-dimensional quantitative feature discovery based on studies of specific patient cohorts can suggest "disease-specific" features to be initially evaluated. For example, the high-dimensional characterization of GNP and GPP cells in Chapter 6 suggests that there are key molecules (i.e. p-STAT5, p-STAT1, p-NFκB, S100B, SOX2, EGFR, CD44, and cyclin B1) that should be quantified for the clinical stratification of adult glioblastoma patients. Importantly, the clinically distinct GNP and GPP cells were composed of different cell subsets that have very similar, but not identical, signaling and protein characteristics. For example, p-STAT3 and p-S6 were enriched in 3 of the 4 cell subsets that constitute the GNP population, while p-NFκB is enriched in only 2 of the 4 subsets that make up the GPP population. The components of GNP and GPP populations also can differ slightly between tumors. Yet, all of these subsets significantly dictate distinct clinical outcomes. Thus,

the assessments of the subset-defining features may need to be tailored in a patient-specific manner to enable effective disease monitoring over the course of therapy.

Targeting the signaling regulators that drive poor outcome

A key consideration towards using signaling profiles as surrogates for individualized therapy is whether they could be used as targets for therapy. For over a decade, signaling profiles have been successfully shown to stratify outcome in many types of hematologic malignancies (*12-15, 80*). However, profiles of downstream signaling effectors have not been used to create unique therapeutic approaches for individual patients.

One of the critical missing links between the discovery of the stratifying signaling characteristics and successful clinical implementation is the confirmation that the signaling profiles of unfavorable cancer cells are also present *in vivo*, which would validate that the signaling states are viable therapeutic targets. Imaging technologies might not be able to quantify features of millions of cells like flow cytometry. However, they can partially address this concern. Especially with the invention of techniques such as imaging mass cytometry (IMC) (*279*) and multiplexed ion beam imaging (MIBI) (*280*), the per-cell dimensionality that can be obtained and quantified from imaging approaches has significantly increased from conventional immunohistochemistry (IHC) and immunofluorescence (IF) analyses. These approaches can assist with the identification of unfavorable cancer cells, defined by unique high-dimensional intracellular signaling activity, in intact (i.e. undissociated) patient samples. Alternatively, a simplified low-dimensional imaging workflow, using IHC or IF stains, can be immediately derived and tested using the critical cellular features discovered by mass cytometry. This would generate a clinically applicable platform that utilizes samples that are readily available in clinical laboratories.

Single-cell high-throughput drug screening can aid the identification of compounds, or combinations of compounds, that can effectively eliminate the unfavorably signaling states or the cells that harbor such states (*344, 345*). Specifically, cell barcoding technologies (either fluorescence (Appendix B and D, (*344, 345*)) or metal (*346*)) are promising approaches for rapid identification of individualized treatment regimens. These techniques enable simultaneous and accurate comparison of hundreds of therapeutic conditions. Patient samples that contain diverse populations of cancer cells can be treated with different therapeutic compounds *ex vivo* to track the loss of the unfavorable cancer cells, or the loss of specific adverse signaling events. Importantly, single-cell

approaches can determine the specificity of treatments on cancer cells compared to other stromal cells in the tumor microenvironment (*347*). Additionally, the unfavorable cancer cells can be physically isolated from the sample (i.e. by cell sorting, based on their unique phenotype) to be treated with compounds *ex vivo*; the cell subset-specific changes in phenotype and signaling states can then be tracked overtime. These approaches distinguish between the change in the intrinsic cellular features (such as signaling network re-wiring) from the loss or death of cell subsets. Eventually, the drug combinations that show the most promising results can be further evaluated in *in vivo* models, as well as in clinical trials.

Traditionally, clinical trials are designed to test two or more arms of treatments on specific patient populations. The extensive inter-tumor diversity, however, must be taken into consideration for the selection of therapeutic compounds for the development of clinical trials in this new era. The idea of “N-of-1” trials and experimentations have been introduced based on the notion that no two tumors are alike and, therefore, an exploratory approach for drug selection should be adopted (*348*). This is because different drugs or combinations of drugs might be needed to achieve a similar end goal in different patients, due to the differential sensitivity of cancer cells to treatments. Critically, specific desirable outcomes, such as a decrease in a signaling state, need to be clearly defined. Using the unfavorable signaling state in GNP cells as an example, the end goal for the N-of-1 compound evaluation could be a decrease in glioblastoma cell STAT5 phosphorylation level. Notably, the N-of-1 screening process could include not only drugs that are known to directly target STAT5 or its immediate upstream molecule, such as JAK, but also other novel compounds to initially identify drugs that exhibit STAT5 inhibitory effect regardless of their specific mechanisms (i.e. using STAT5 phosphorylation as a surrogate for drug efficacy as oppose to only targeting STAT5).

While eventually the promising therapeutic compounds must be tested in patients, the initial step is to evaluate the efficacy and specificity of the drugs in an *ex vivo* setting. At least in the field of glioblastoma, neurosphere culture assay is commonly used as a mean to maintain viable cancer cells *ex vivo* for functional characterization or for therapy evaluation (Figure 7-1A) (*30, 45, 349*). Brain- or glioblastoma-derived cells are maintained as 3-dimensional spheres in a non-adherent serum-free culture condition that is supplemented with specific growth factors (EGF and basic FGF (bFGF)) (*349, 350*). However, neurosphere culture selects for specific cell populations and cannot maintain the cellular diversity seen *in vivo* (Figure 7-1B). Particularly, the

standard neurosphere culture enriches for stem-like cells while depleting immune cells and cells with mature neural phenotype (Figure 7-1B). Ideally, the platform for N-of-1 compound screening (either the cell culture conditions or animal models) should mimic the tumor microenvironment to enable accurate prediction of the drug efficacy *in vivo* in order to make this individualized screening approach applicable in clinical settings.

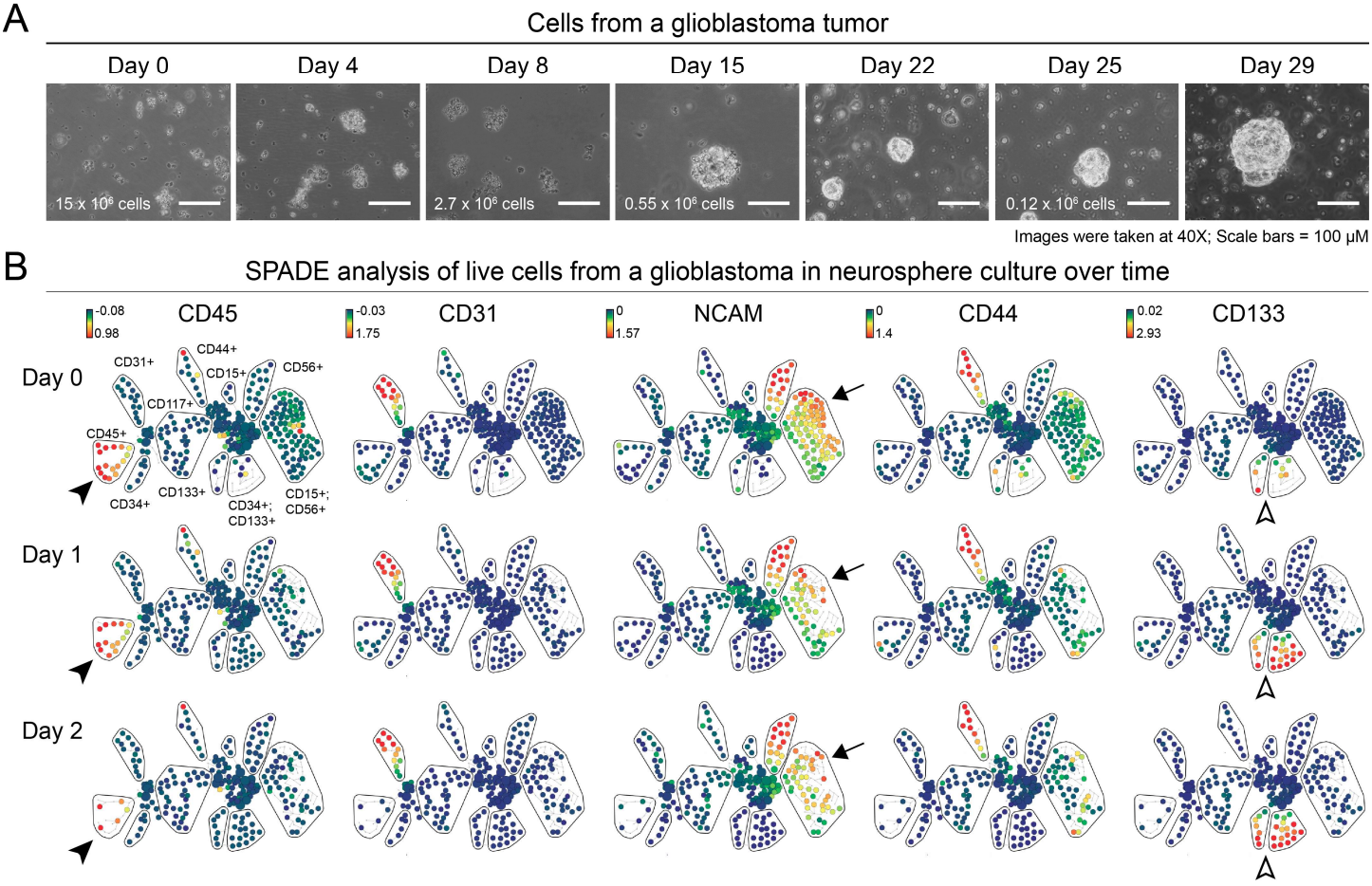


Figure 7-1 Neurosphere culture sustains the viability of some cell types derived from adult glioblastoma but does not maintain cellular diversity. (A) Cells from a freshly dissociated glioblastoma were subjected to a standard neurosphere culture protocol (serum-free, with EGF and bFGF supplement) (*350*). Representative images of the cells and formed neurospheres from days 0, 4, 8, 15, 22, 25, and 29 are shown. Images were taken at 40X (scale bars = 100 µm). (B) Cells from days 0, 1, and 2 (rows) were collected and stained with a panel of fluorescently tagged antibodies (columns, showing 5 proteins quantified) and analyzed with a fluorescence flow cytometer. SPADE analysis (*282*) was performed on live, intact cells. Each node represents a cell subset with similar protein expression. Heat represents the degree of expression of the indicated surface proteins. Filled arrow heads illustrate $CD45^{+}$ immune cells. Arrows illustrate $CD45^{-}NCAM^{+}$ neural-like cells. And open arrow heads illustrate $CD45^{-}CD133^{+}$ stem-like cells. Acronyms: EGF, epidermal growth factor; bFGF = basic fibroblast growth factor; NCAM, neural cell adhesion molecule.

Elucidating the cancer cell signaling mechanisms with signaling potentiation

Co-activation of surface receptors and their convergence via downstream molecules suggested that quantification of post-translational modifications of signaling effectors better recapitulates the signaling activity than the assessment of the presence of receptors alone (*142*). However, the "static" signaling characteristics of clinically distinct glioblastoma cells described in Chapter 6 of this dissertation only reflect the baseline activity of the cells. Further experimentations are needed to elucidate their actual signaling mechanisms, specifically to understand their potential adaptive mechanisms under different environmental stimulations. Previous work in healthy blood and hematologic malignancies revealed that signaling potentiation of cells *ex vivo* further elaborated the understanding of the disease by two folds. First, combination of baseline and potentiated signaling better correlated with significant clinical features, including responses to chemotherapy as well as clinically-relevant cytogenetics, compared to basal signaling alone (*80*). Second, tracking of signaling responses overtime revealed mechanistic interactions between different cell types that would have been overlooked by static analysis (*83*). It would be especially valuable to map the glioblastoma cell subset-specific signaling network, especially of the GNP cells, by dissecting the "potential" signaling routes and how they communicate with other cells in the tumor microenvironment under specific circumstances.

Signaling potentiation of human glioblastoma cell lines illustrated how signaling kinetics of different cell types can be elucidated using a high-throughput flow cytometry assay (Figure 7-2). Mass cytometry was used to profile the expression of several identity proteins, which revealed distinct phenotypes between four different human glioblastoma cell lines (Figures 7-2A and S7-1). MET and PDGFRα expressions were shown as examples (Figure 7-2B). Notably, LN-229 and LN-18 cells both showed expression of MET on the cell surface (the antibody used in the assay is specific to the extracellular domain of MET). When both cell lines were stimulated with HGF, the sole ligand of MET, LN-229 showed a much higher level of AKT phosphorylation compared to basal, unstimulated state (Figure 7-2C). Here, the level of phosphorylation of a signaling effector, under a specific potentiation condition, is defined as a signaling node state. When coupled with fluorescence cell barcoding to enhance the analysis throughput (see Appendix B), a total of 1,845 signaling node states could be simultaneously compared across four different glioblastoma cell lines (Figure 7-2D).

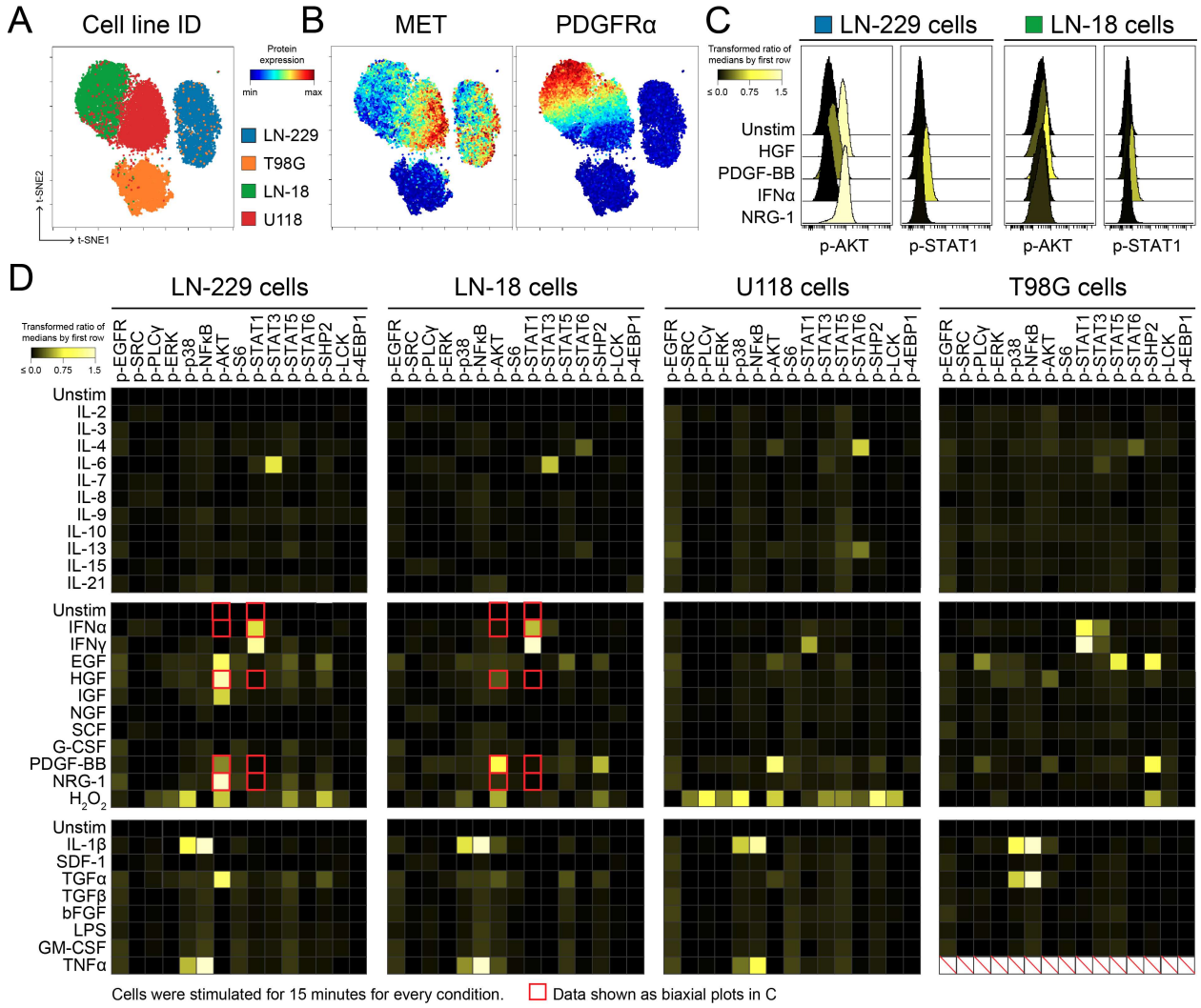


Figure 7-2 Ex vivo potentiation revealed diverse signaling responses between human glioblastoma cell lines. (A) viSNE analysis of four human glioblastoma cell lines (blue, LN-229; orange, T98G; green, LN-18; red, U118). The expression of identity proteins used to generate the viSNE map is depicted in Figure S7-1. (B) Here, per-cell expression of MET and PDGFRα is shown. (C) Histogram analysis of p-AKT and p-STAT1 (x-axes) of LN-229 (2 left columns) and LN-18 (2 right columns) under different stimulation conditions. Color of histograms reflect the median values of transformed ratio compared to their corresponding unstimulated condition (unstim, first rows). (D) Heatmap illustrating 465 signaling node states of LN-229, LN-18, and U118 cells, and 450 signaling node states of T98G cells. Color reflects the median values of transformed ratio compared to the unstimulated state. Note: TNFα stimulation data was not available for T98G cells.

The kinetics of signaling secondary to external potentiation can be similarly dissected in different cancer cell subsets in patient samples. Signaling potentiation can also be coupled with signaling inhibition or drug treatment to map how the signal re-wires if a certain point in the network is impeded. This is especially valuable for identifying potential drug resistance mechanisms. The selection of growth factors and ligands for signal potentiation should also be based on those that are known, or are hypothesized, to be present in the tumor

microenvironment of human glioblastomas. This should include factors in the cerebrospinal fluid (CSF) and the neural stem cell niche. These experimentations may reveal how a given cancer cell subset might differentially behave if it were present in a different anatomic location within the brain.

Dissecting the mechanism of glioblastoma recurrence with single-cell biology

The identification of GNP and GPP cells described in Chapter 6 is the first time single-cell analysis has been successfully used to stratify clinical outcomes of adults with *IDH*-wildtype glioblastoma. This discovery has significant roles for clinical implementation as disease stratification and for selection of novel therapeutic compounds. Evidence for the significance of the signaling states of GNP cells as targets for therapy is the demonstration that GNP cells are present in recurrent tumors, especially at higher abundance. This would indicate that GNP cells could be the mechanism underlying therapy resistance. Most of the studies in adult glioblastomas, including previous single-cell genomic and transcriptomic studies as well as the work presented in this dissertation, were conducted in primary tumors that were collected prior to therapy (*5, 73*). Given what little is known about the cellular diversity of glioblastomas at recurrence, single-cell analysis of cell subsets and their signaling profiles of paired primary-recurrent tumors is now needed.

Since the analysis of millions of single cells using flow cytometry-based approaches gives a better representation of the cellular diversity within the tumor, a method to collect patient samples that would enable subsequent derivation of viable single cells is key. As demonstrated throughout this dissertation, primary tumors can be easily collected intra-operatively for single-cell preparation (Chapters 5 and 6). This is not always available in the case of recurrent glioblastomas where tumor resection is not routinely performed, due to the extent of the tumors and the patients' quality of life. A large resource of recurrent glioblastomas could potentially be collected after death, during autopsy. However, since the tissues and cells can rapidly deteriorate, a rapid autopsy protocol could be especially valuable for studies that would largely benefit from high-quality viable cells (*351*). Rapid autopsy is generally performed within 1-3 hours of after death to minimize post-mortem demise and enables collection of large quantity of samples from different sites for systematic comparison of their cellular constituents (*351*). This process has shown significant benefits in many studies and could be readily adopted for studying the signaling profiles of individual cancer cells in recurrent glioblastomas.

Cancer-immune interaction

The relationship between the abundance of the novel glioblastoma stratifying cell subsets and the brain resident microglia that were phenotypically activated suggested a mechanistic relationship between the two cell components. The abundant GNP cells in the clinically unfavorably tumors were actively proliferating, which could inevitably prevent the infiltration of immune cells within the tumor microenvironment, although it is unlikely the only explanation without mechanistic evidence of immune cell suppression. Alternatively, GNP cells could potentially be less immunogenic than GPP cells, leading to overall inactive and scarce immune cells within the tumor microenvironment. Animal models can be used to compare the immunogenicity between GNP and GPP cells (*352-354*). Subsequently, certain immunogenic chemotherapies can be tested to determine if the immunogenicity of GNP cells can be enhanced to elicit a proper anti-tumor immune response (*355*).

Additionally, the presence of activated microglia may directly alter the signaling activity in GPP cells and/or induce an anti-tumor signaling phenotype. It has been previously demonstrated that different resident brain cells, including microglia, can secrete microvesicles that contain factors and cytokines, enabling them to interact with neighboring cells (*356*). IL-1β, a known activator of the canonical NFκB signaling pathway (Figure 7-2D), is one of the many factors that can be secreted by microglia (*357, 358*). Notably, enrichment of p-NFκB is one of the defining features of some GPP cells described in Chapter 6. Whether or not activated microglia can directly suppress the unfavorable signaling effector STAT5 needs to be evaluated. Overall, these observations suggest that there could be an ongoing, variable, immune surveillance and activation in adult glioblastomas. Further dissection of the mechanisms can provide alternative therapeutic avenues in addition to directly targeting the cancer cells.

Conclusion

A thorough characterization of signaling alterations in individual cancer cells is key to tackle the fatality of adult glioblastomas. This dissertation provides a novel modular workflow that dissects the diversity of the critical hallmarks of glioblastoma pathogenesis. This was accomplished by 1) the development of a protocol for the derivation of viable single cells from glioblastoma patient samples, 2) the use of a system-tailored single-cell mass cytometry analysis of proteins and signaling alterations, and 3) the coupling of a patient outcome

assessment with cell subset characterization for the discovery of clinically distinct glioblastoma cells. Overall, this workflow has the potential to become a prototype towards the individualization of glioblastoma patient care (Figure 7-3).

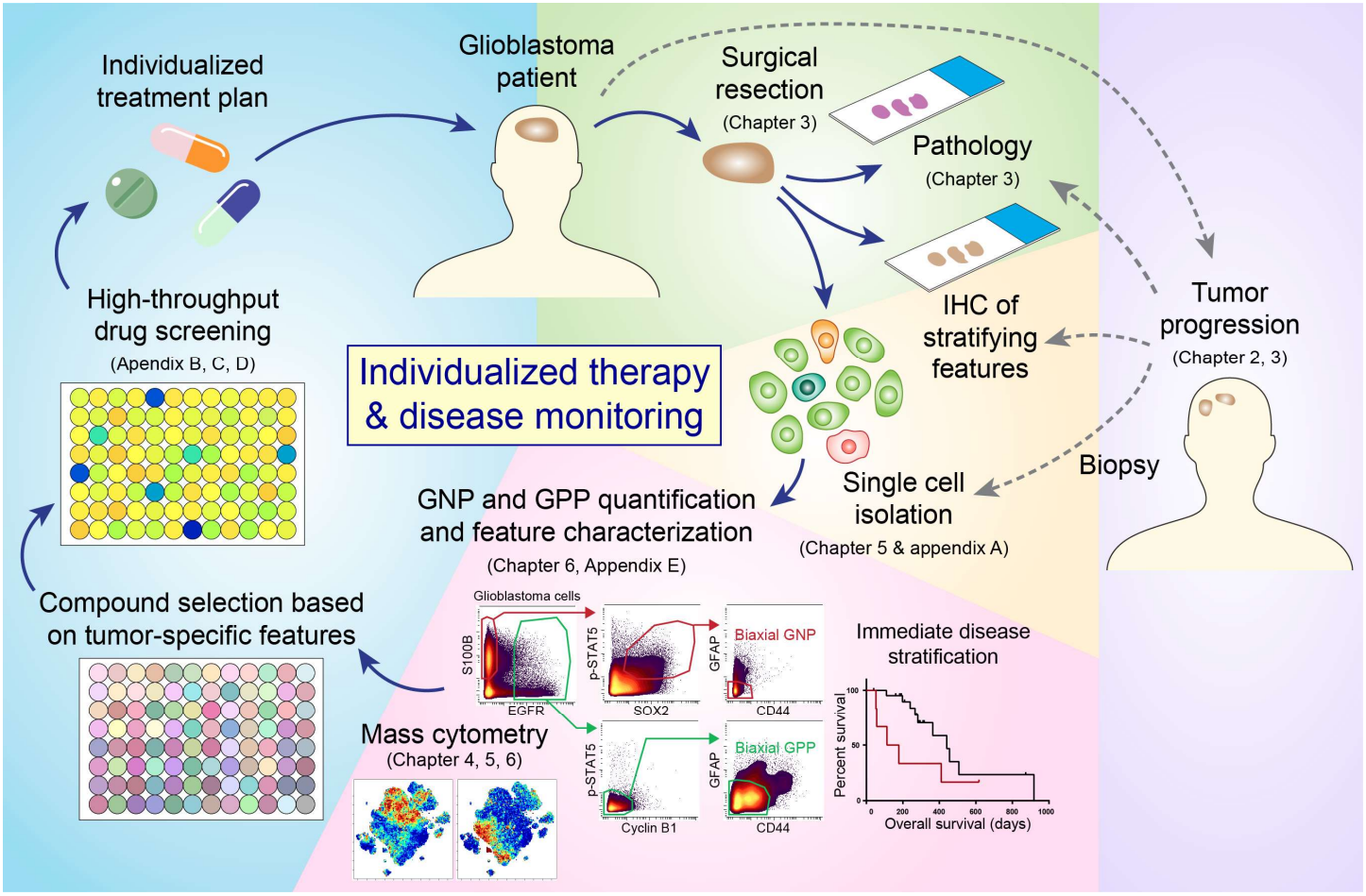


Figure 7-3 Towards individualized disease monitoring and clinical care in adult glioblastomas. In addition to the standard pathological diagnosis (green), single-cell isolation (orange) and quantification of protein identity and signaling profiles in individual glioblastoma cells (pink) has the potential to be part of the standard clinical care. The rapid turn-around time of standard flow cytometry and the quantitative clinical prediction of GNP and GPP cells enable immediate disease stratification. High-dimensional single-cell mass cytometry can be standardized for additional in-depth feature quantification, which can benefit subsequent selection of therapeutic compounds to be tested by high-throughput drug screening approaches. With a drug screening platform that accurately predicts responsiveness *in vivo*, this immediately leads to a development of an individualized treatment strategy. In cases of tumor progression (purple), the same analytical pipeline can be repeated to identify a new optimal therapeutic approach.

The discovery of clinically distinct glioblastoma cells, their protein and signaling phenotypes, as well as their correlation with the tumor microenvironment, suggests numerous diagnostic and therapeutic avenues that should be promptly investigated. The single-cell stratification scheme can be immediately tested for clinical utility

since the abundance of both GNP and GPP cells can significantly predict patient outcome and tumor progression with a large effect size, independent of other known clinical predictors. Importantly, a lower-dimensional diagnostic approach that mimics the standard analytical workflow of hematologic specimens can easily be adopted for clinical use. Critically, signaling activity that is definitional of the GNP cells should be tested as novel therapeutic targets, or as surrogates for drug activity. Moreover, the single-cell signaling platform described here can serve as the foundation for dissecting potential drug resistance mechanisms, such as signaling network rewiring, as novel compounds are being investigated. Additionally, the mechanisms underlying the cancer-immune interaction should also be comprehensively explored as an alternative approach to targeting glioblastoma via modifications of the tumor microenvironment. Technical approaches that are valuable for high-throughput drug discovery are also described in the Appendices of this dissertation.

A critical limitation of single-cell mass cytometry is the lack of intra-tumor spatial information. Quantitative imaging to dissect the regional preference of the prognostic glioblastoma cells can further reveal the mechanisms responsible for therapy resistance, such as contact with the ventricular subventricular zone or the perivascular niche. Additionally, identification of physical contact between GPP cells and microglia may further advocate that the anti-tumor signaling profile of GPP cells is a response to immune cell activation. Moreover, it would be valuable to determine if GNP and GPP cells often co-localize within the same region of the tumor or if they exist in distinct regions, which could indicate the presence of specialized tumor microenvironment that may be primed by different cancer cell types. Regardless, single-cell snapshot proteomics described in this dissertation presents an immediate approach for clinical stratification that simultaneously identify future therapeutic targets.

As high-throughput single-cell feature quantification becomes a routine in research and clinical settings, experimental and analytical automation are mandatory. This is to maximize the efficiency and accuracy of the system and to eliminate bias. The discovery of GNP and GPP cells relied on a systematic and automated feature discovery pipeline which was tailored to fit a specific system, to address a specific question. Importantly, this workflow can be a valuable addition to the characterization and clinical stratification of other types of solid malignancies. Signaling features and the clinical significance of cancer cells should now be included as cell subset defining features, incorporated as part of routine disease stratification, and used as evidence for the selection of individualized therapeutic regimens.

Supplement data

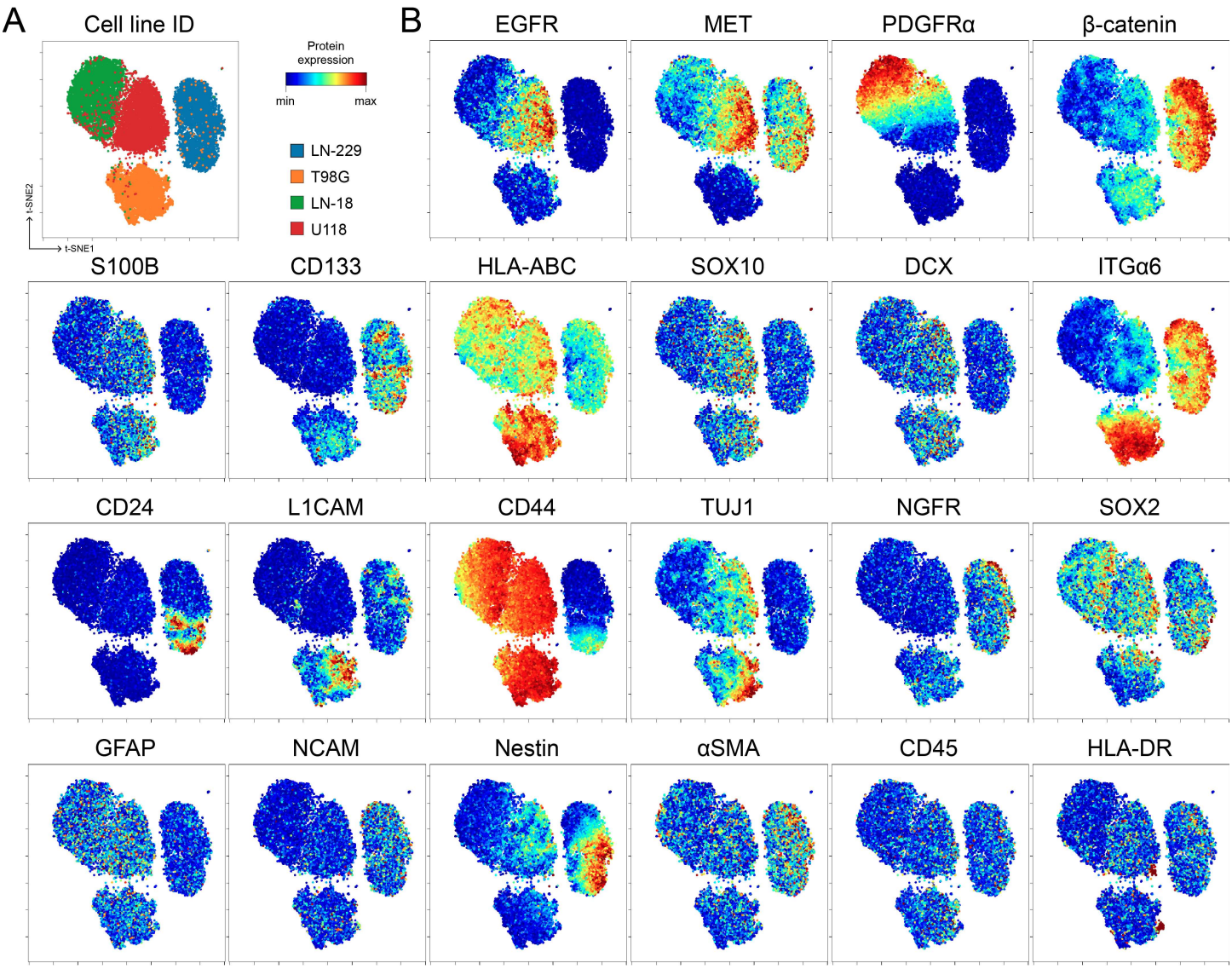


Figure S7-1 Protein expression in human glioblastoma cell lines. (A) Single-cell mass cytometry data of four human glioblastoma cell lines were subjected to a viSNE analysis (LN-299, blue; T98G, orange; LN-18, green; U118, red). (B) Twenty-two surface and intracellular identify proteins that were quantified and were used to generate the viSNE map are shown.

APPENDIX A

Preparing Viable Single Cells from Human Tissues and Tumors for Cytomic Analysis

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

*Corresponding authors

This work is presented as it appears in manuscript form in *Current Protocols in Molecular Biology* (Leelatian et al., 2017).

Preface

One of the major challenges in understanding the complex relationship of cells in solid tissues and tumors is the lack of a standard protocol that enables quantification of multiple cellular features of numerous single cells that are representative of the original tissues. As described in previous chapters, mass cytometry is an appealing platform for characterizing single cells in solid tumors. However, viable single-cell suspension is required for this approach. This chapter describes a step-by-step protocol for solid tissue and tumor dissociation to derive single cells for mass cytometry analysis. Its development and optimization are described in detail in Chapter 5. Description of high-dimensional data analysis algorithm and display of data are presented here. Moreover, a practical approach for troubleshooting when implementing the protocol in a new tissue type is discussed.

Abstract

Mass cytometry is a single-cell biology technique that samples > 500 cells per second, measures > 35 features per cell, and is sensitive across a dynamic range of > 10^4 relative intensity units per feature. This combination of technical assets has powered a series of recent cytomic studies where investigators used mass cytometry to measure protein and phospho-protein expression in millions of cells, characterize rare 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 unit establishes a primer for single-cell analysis of solid tumors and tissues, and has been tested with mass cytometry. 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 (*359*) 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 (*10, 11*). 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. Mass cytometry 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 (*83, 296, 359-361*). 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 (*284*).

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 (*11, 359, 360*). Our group and others have implemented this technology in studies of donor and patient cells that are obtained as a suspension, such as blood and bone marrow (*81, 83, 282, 315, 362-365*) or that can be disaggregated from lymphoid structures by mechanical force alone (*13, 15, 82, 303, 366*). Clinical diagnoses of blood malignancies use fluorescence flow cytometry characterization of cell surface marker expression, as well as cell subset quantification (*19, 367-369*). Additionally, flow cytometry has been used clinically to identify minimal residual disease and to detect disease progression in leukemia (*290, 370, 371*). Fluorescence flow cytometry has also been applied to studies of solid tissues and tumors for research purposes (*45, 305-309*).

In addition to their ability to characterize cell surface markers, flow cytometry technologies allow simultaneous detection and quantification of intracellular targets in individual cells (*10, 80, 84*). Commercially

available fluorescence flow cytometers generally measure 8 to 12 targets per cell using target-specific antibodies conjugated to individual fluorophores (*283*). The number of targets is limited due to the overlap of emission spectra of different fluorophores. 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 (*50, 83, 302*). 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 (*315, 372, 373*). Additionally, multiple cellular targets of interest can be measured simultaneously, and the numbers are greater than those routinely measured in current fluorescence-based cytometry (*50, 359-361*). 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 (*294, 296, 374*). 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 (*374*) (*12, 360, 375, 376*).

Mass cytometry–based characterization of human bone marrow (*83*), blood (*299*), and tonsil (*82*) cell subsets has been accomplished in prior studies and described in protocols (*81*). However, mass cytometry has just recently been developed and applied in solid tissues and organs (*321*). 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 (*377-379*). 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) (*284*). 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 mass cytometry analysis of cells from human tonsils, gliomas, and melanomas (Basic Protocol 2), and a section in the Commentary describing computational analysis of multi-dimensional data obtained from mass cytometry based on established approaches (*282, 290, 314*).

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) (*284*). 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

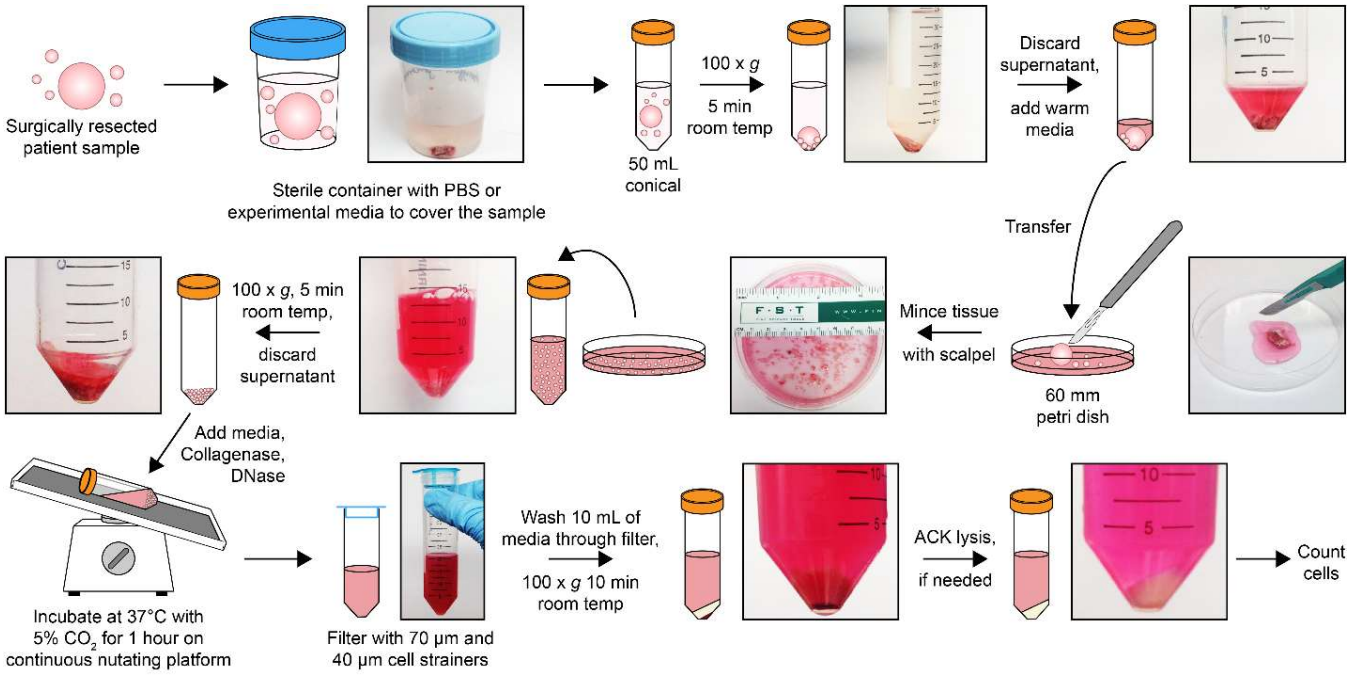


Figure A-1 Step-by-step illustration of tissue dissociation protocol. Surgically resected patient samples were transported in PBS, normal saline, or experimental medium at room temperature. Mechanical dissociation was followed by 1-hr enzymatic dissociation using collagenase II and DNase I (see text).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

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 (*278, 380*), 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 (*15, 82, 303*). Conversely, multiple growth factor supplements are added to the neurosphere culture medium to ensure growth of human glioma cells (*381, 382*).
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 (*284*).
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.

Remove RBC and quantify viable cells

1. 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.
2. 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.
3. 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, aliquotted 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.

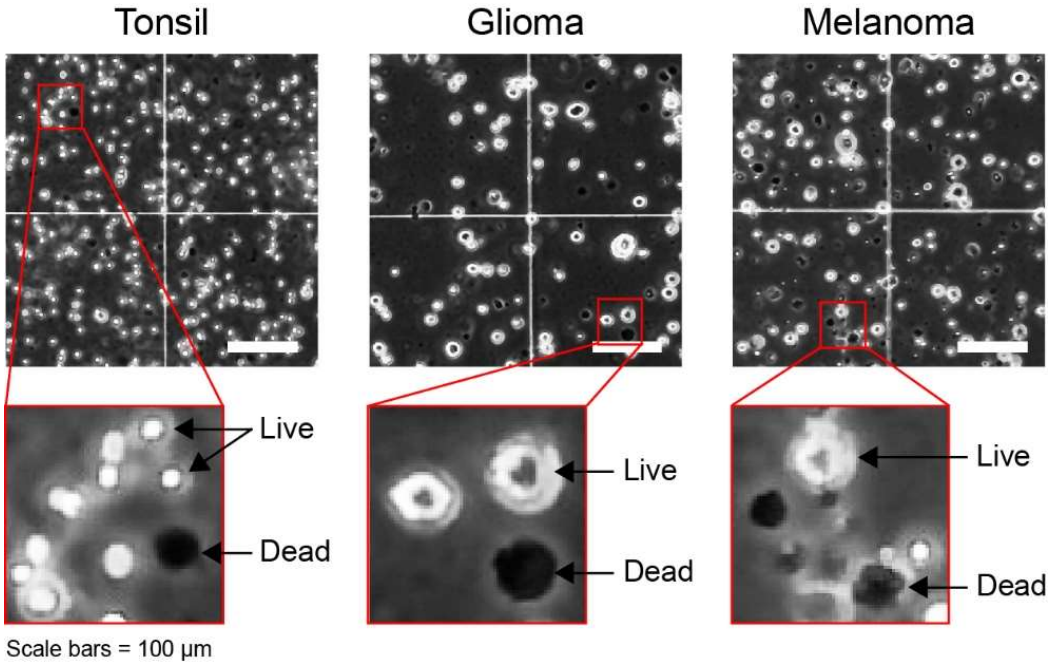


Figure A-2 Trypan blue stain for viable cell quantification. Trypan blue stain was used to quantify cell viability after mechanical and enzymatic dissociation. Representative images of dissociated human tissues including tonsil, glioma, and melanoma are shown. Red boxes show higher resolution of live (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.

Basic protocol 2: Preparation of cells for mass cytometry

This section describes a protocol for immunostaining of single-cell suspensions derived from human tissues and tumors. Tonsils, glioma tumors, and melanoma tumors are used 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 (*81, 84, 276, 383*).

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 (*278, 380, 383, 384*). 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 (*278*). Examples of permeabilization reagents include saponin (Perm 1), methanol (Perm 2), ethanol, and Triton X-100, among many others. For the protocol described here, saponin permeabilization (Perm 1; for SOX2 antibody staining) is used prior to methanol permeabilization (Perm 2; for staining of the remaining intracellular targets) in human glioma and melanoma.

Materials

1. Perm 1: room-temperature 0.02% (w/v) saponin (Calbiochem, cat. no. 558255) in PBS
2. Perm 2: ice-cold 100% methanol (Fisher Scientific, cat. no. A412-4) kept at −20°C until immediately prior to adding to cells
3. Dissociated single cells (Basic protocol 1)
4. Experimental media:
    a. For glioma: DMEM/F12 plus GlutaMax (Gibco/Life Technologies, cat. no. 10565018) with a defined hormone and salt mix (*311*) 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: 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)
5. 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)
6. Staining medium: 1% (w/v) bovine serum albumin (BSA; Fisher Scientific, cat. no. BP9703100) in PBS
7. Live Stain reagent mix: A combined solution of all relevant antibodies (antibody list in Table A-1)
8. Phosphate-buffered saline (PBS; Corning/Mediatech, cat. no. 21040CV)
9. 16% paraformaldehyde (PFA; Electron Microscopy Sciences, cat. no. 15710)
10. Saponin Stain reagent mix: A combined solution of all relevant antibodies (antibody list in Table A-1)
11. Methanol Stain reagent mix: A combined solution of all relevant antibodies (antibody list in Table A-1)
12. 1X Four Elements Calibration Beads (Fluidigm, cat. no. 201078)
13. 15-ml conical tubes (Corning Falcon, cat. no. 430055)
14. Benchtop centrifuge with swing-out rotor (Sorvall model ST 16; Thermo Scientific)
15. 5-ml round-bottom FACS tubes without cap (Corning Falcon, cat. no. 352052)

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	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		✓	✓	✓		
	IgD	IA6-2	Fluidigm	1:200	0.5	✓			✓		
146Nd	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	✓			✓		
	CD19	HIB19	100	1:100	1	✓			✓		
161Dy	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		✓				✓
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; Sap = saponin; MeOH = post-methanol
Fluidigm = use antibodies provided by Fluidigm

16. Rotor adapters with round buckets that accommodate 5 ml FACS tubes (Thermo Fisher Scientifc, cat. no. 75003680)
17. 5-ml round-bottom FACS tubes with filter caps (Corning Falcon, cat. no. 352235)

Prepare antibody

1. Example 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 of

three example staining steps: live cell staining (Live), staining in 0.02% saponin (Saponin Stain), and staining after methanol treatment (Methanol Stain).

a. 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 (*81, 278, 380*).
b. This protocol does not use metal barcoding, but that technique can be useful in addressing potential batch effects from staining and collecting data at different times (*312, 346*).

Stain viable cells to detect extracellular targets

2. If preparing cells from cryopreservation, thaw cryovial in a 37ºC water bath for 1 to 2 min (until just completely thawed). If preparing cells that were freshly dissociated, skip to step 5.
3. Transfer cells from cryopreservation tubes to a 15-ml conical tube.
4. Resuspend cells in 10 ml of warm experimental medium. 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.
5. Pellet cell suspension 5 min at 100 x g, room temperature, and discard supernatant.
6. Resuspend cells in 1 ml of staining medium and transfer cell suspension to a 5-ml FACS tube (without filter cap).
7. Pellet cell suspension 5 min at 100 x g, room temperature, and discard supernatant by briskly decanting.
    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.
8. Resuspend cell pellet in staining medium to achieve the transfer volume.
    a. The “transfer volume” is the volume transferred in step 9, below, and is calculated by subtracting the summed volume of staining antibodies (the “antibody volume”) from the total volume in which

staining will occur in step 10. 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 10. 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 transfer volume of 70 µl for step 9. For steps 8 to 10, the volume of antibodies varies and is specified in Table A-1 for tonsil, glioma, and melanoma. The total volume in step 10 is 100 µl (including cells in transfer volume and all antibodies).

9. Transfer cell suspension in staining medium to a new FACS tube.
10. Add Live Stain reagent mix (Table A-1) to the same FACS tube. Briefly vortex to mix cells with antibodies. Total staining volume (cells in staining medium plus antibodies) should be exactly 100 µl. 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.
11. Leave cells at room temperature for 30 min.
12. Add 1 ml of staining medium to the FACS tube and pellet cells 5 min at 100 x g, room temperature, then decant to discard supernatant (see step 7).
13. Repeat step 12 once.
14. Add 1 ml of PBS to FACS tube and resuspend the cells by gentle vortexing until there are no visible cell clumps.
15. Add 100 µl of 16% PFA to FACS tube, for a final concentration of 1.6%, and vortex to mix. This step is for cell fixation prior to further intracellular staining steps.
16. Leave cells at room temperature for 10 min.
17. Pellet cells 5 min at 800 x g, room temperature, and decant to discard supernatant (see step 7).
    Cells are now ready for intracellular staining.

Stain to detect intracellular targets

18. Determine the optimal permeabilization conditions required for each intracellular target. If a subset of intracellular antigens requires permeabilization with saponin, as is the case with antibodies used to stain

glioma and melanoma, include steps 19 to 27. 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 28.

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 (*385*). The isotope-labeled antibodies can then be used for immunostaining by following the protocol described below.

19. Resuspend cells in 1 ml of Perm 1, pellet 5 min at 800 x g, room temperature, and decant to discard supernatant (see step 7).
20. Repeat step 19 once.
21. Resuspend cell pellet in appropriate volume of Perm 1 (see step 8).
22. Transfer appropriate volume of cell suspension in Perm 1 to a new FACS tube (see step 8).
23. Add Saponin Stain (Table A-1) reagent mix to the same FACS tube. Briefly vortex to mix cells with antibodies.
    a. Total staining volume (cells in Perm 1 plus antibodies) should be exactly 100 µl. Note that the Saponin Stain mix, in addition to the prior Perm 1, must also contain saponin.
24. Leave cells at room temperature for 30 min.
25. Add 1 ml of Perm 1 to the FACS tube, pellet cells 5 min at 800 x g, room temperature, and decant to discard supernatant (see step 7).
26. Repeat step 25 once.
27. 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 7).
28. Vigorously vortex cell pellet to resuspend cells in void volume.
29. Add 1 ml of Perm 2 to FACS tube. Vortex to thoroughly resuspend cells with Perm 2.
    a. The final concentration of ice-cold methanol after the Perm 2 is added to cells should be > 95%.
30. Keep cells at −20°C for at least 20 min.

a. Alternatively, cells can be stored in Perm 2 at −80°C for days or weeks before proceeding, if necessary.

31. Remove FACS tubes from −20°C and add 1 ml of PBS to each tube.
32. Pellet cells 5 min at 800 x g, room temperature, then decant to discard supernatant (see step 7).
33. Add 1 ml of staining medium to FACS tubes. Pellet cells 5 min at 800 x g, room temperature, then decant to discard supernatant (see step 7).
34. Repeat step 33.
35. Resuspend cell pellet in appropriate volume of staining medium (see step 8).
36. Transfer appropriate volume of cell suspension in staining medium to a new FACS tube (see step 8).
37. Add Methanol Stain reagent mix to the same FACS tube. Briefly vortex to mix cells with antibodies.
    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.
38. Incubate cells at room temperature for 30 min.
39. 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 (see step 7).
40. Repeat step 39 once.
41. 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 7).
42. 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 7).
43. 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.
44. Filter cells using FACS tubes with filter caps.

    Cells are now ready for mass cytometry analysis.

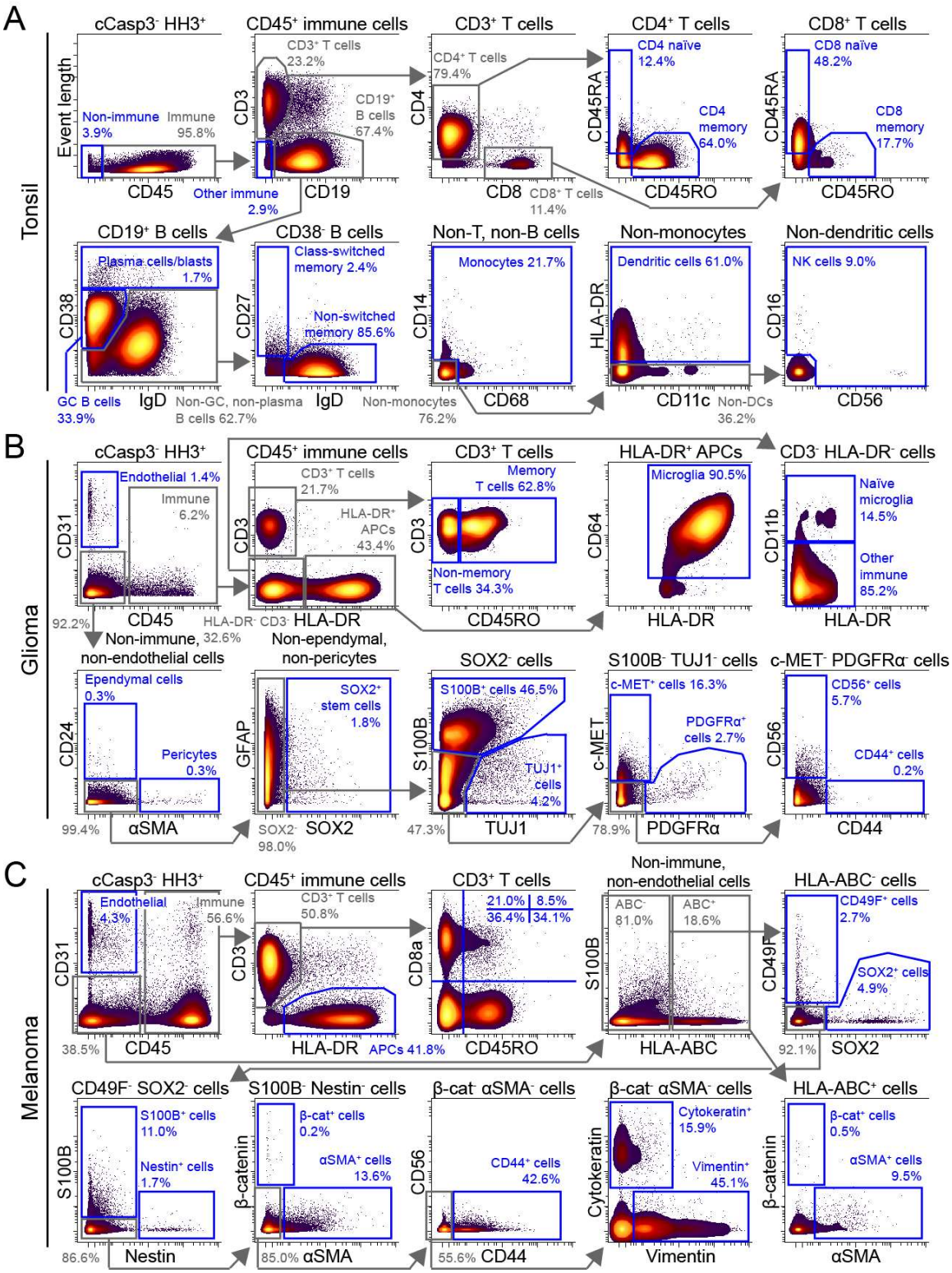


Figure A-3 Biaxial analysis of cells derived from human tissue and tumors. Biaxial plots of non-apoptotic (cCasp3$^-$), nucleated (HH3$^+$) cells from mass cytometry analysis of (A) tonsil, (B) glioma, and (C) melanoma are shown. Intermediate gates are shown in gray, and terminal gates are shown in blue. Cell types or protein identity of cells in each gate are indicated. The percentages of cells in gates are also specified.

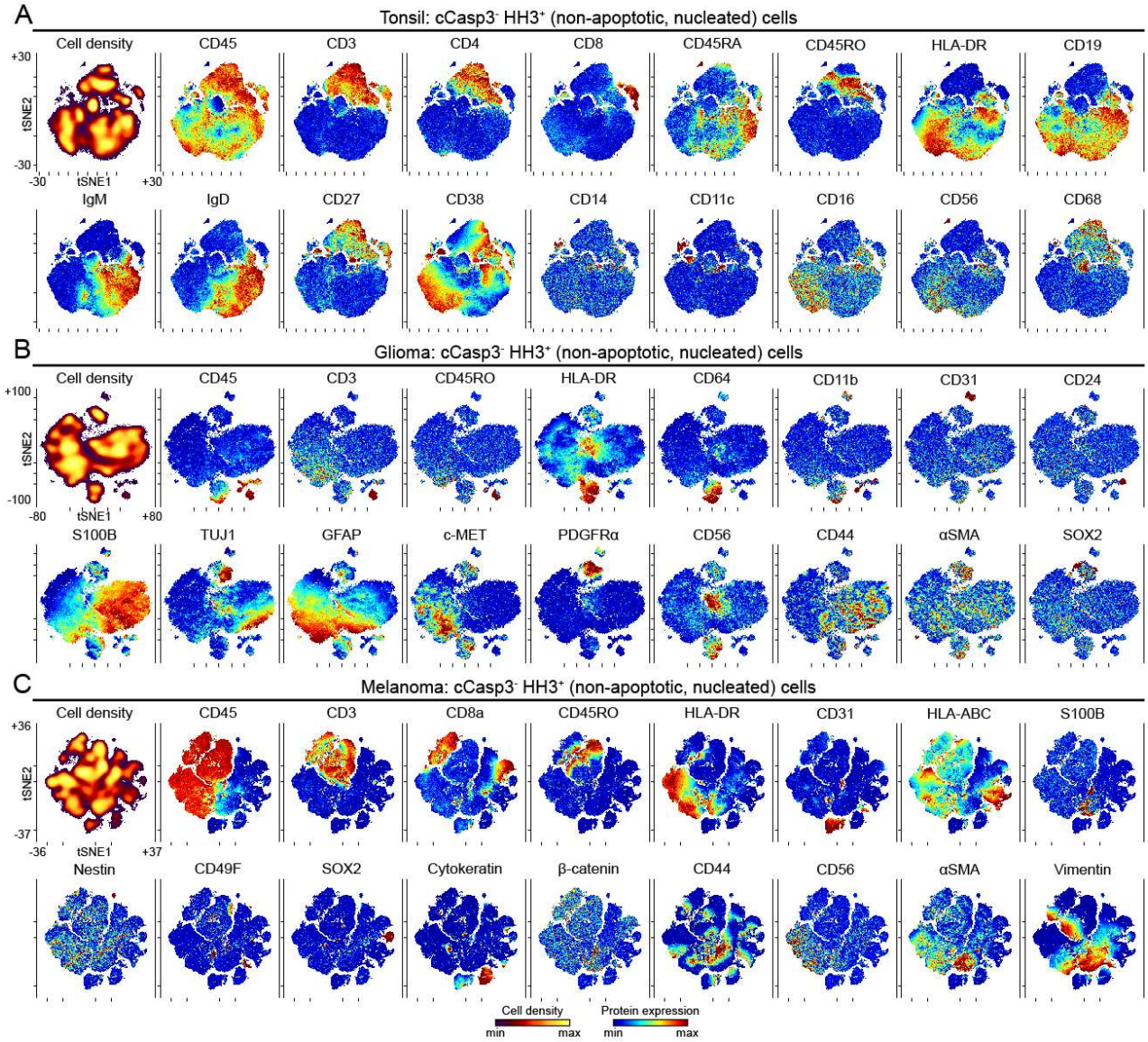


Figure A-4 High-dimensional analysis of mass cytometry data using viSNE. Non-apoptotic (cCasp3^{-}), nucleated (HH3^{+}) cells from mass cytometry analysis of (A) tonsil (plots of 106,568 cells), (B) glioma (plots of 65,834 cells), and (C) melanoma (plots of 94,810 cells), are shown. The first plot of 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 (*298, 385-389*). 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 (*83, 302, 390*). However, as flow cytometry requires individual cells in suspension, mass cytometry's application 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 (*288, 362, 364, 370, 374, 391, 392*). 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, mass cytometry 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 (*284, 321*). 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 (*284*). 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 (*284*). Additionally, inclusion of DNase I in the dissociation protocol described here significantly improves viable cell yield from multiple tissues (*284*). 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 (*284*). 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 (*278*). Parameters that need optimization (*278*) 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 (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, 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 (*312*). 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). 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 (*393*). 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 used anti-histone H3 antibody staining for detection of intact nucleated cells (*284*). 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 (*284*). 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 (*284*)).

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.

Signal normalization using bead standards

During each mass cytometry analysis, the detection sensitivity of the mass cytometer can vary between individual samples. Additionally, signal can vary between different mass cytometry experiments. Specifically, the signal intensity of a given cellular target, which is known to be consistent, can vary

between different samples for different mass cytometry analyses. To allow accurate comparison between samples, bead standards are used (*313*). Known metal isotopes with standard signal intensity are embedded within polystyrene beads. These beads are mixed and analyzed simultaneously with each sample to allow monitoring of signaling variation during data acquisition. The change of the isotope signal of the beads proportionately correlates with the variation of each sample due to detection variability. Therefore, the variation of signal between samples can be adjusted to allow direct comparison of the signal between samples. Normalization is performed using publicly available MATLAB normalization software prior to further data analysis. For more details, refer to (*313*).

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 (*81*). In a typical cytometry analysis workflow (*314, 394*), cells are filtered or assigned to populations based on expression profiles of cellular targets in a process called gating (Figure A-3). 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 (*83, 363, 394, 395*). Figure A-3 shows examples of sequential biaxial gating of cells derived from mass cytometry analysis of healthy tonsil (Figure A-3A), a patient glioma (Figure A-3B), and a patient melanoma (Figure A-3C; (*81, 284*)). 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 (*316, 396*). However, concepts of cell identity are still under active discussion in established single-cell fields like immunology (*395*). 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 (*290, 295, 296, 397*), reference knowledge bases of established cell identities (*301, 398*), annotated repositories of single-cell data, and quantitative labels of cell type (*321*).

Heat plots and viSNE analysis

A new generation of flow cytometry and single-cell analysis tools compress multiple dimensions of information into rich two-dimensional views. Examples include mountain plots (*15*), viSNE and related views of t-SNE axes (*290, 295*), SPADE plots (*282*), and many other tools that have the potential for machine learning of cell identity (for more details, see (*81, 314, 321, 394, 397*). Here, we demonstrate two-dimensional viSNE plots (*290*). viSNE analysis places cells on a two-dimensional map that reflects how individual cells are similar to or different from each other when every measured cellular target is simultaneously taken into account. Cell density and expression of individual markers are displayed for a wide range of features using the same t-SNE axes. Data collected from the same patients as shown in biaxial analyses in Figure A-3 were analyzed by viSNE and are shown in Figure A-4 (tonsil, Figure A-4A; glioma, Figure A-4B; and melanoma, Figure A-4C). 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; (*301*)).

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 (*278*). Once all cells are permeabilized by ice-cold methanol, an additional 1 hr is required for intracellular immunostaining with isotope-labeled antibodies.

Acknowledgements

The authors thank Dr. Jay A. Werkhaven for providing tonsil specimens (Vanderbilt IRB #121328); Dr. Lola B. Chambless, Dr. Kyle D. Weaver, and Dr. Reid C. Thompson for providing glioma specimens (Vanderbilt IRB #131870); and Dr. Mary A. Hooks and Dr. Rondi M. Kauffmann for providing melanoma specimens (Vanderbilt IRB #030220).

Funding sources

Study and researchers were supported by NIH/NCI R00 CA143231 (J.M.I.), R25 GM062459 (D.B.D.), T32 CA009592 (D.B.D.), F31 CA199993 (A.R.G.), T32 HD007502 (J.S.), the Vanderbilt-Ingram Cancer Center (VICC, P30 CA68485), the Vanderbilt International Scholars Program (N.L.), a Vanderbilt University Discovery Grant (J.M.I. and N.L.), a VICC Provocative Question award (J.M.I.), R01 NS096238, DOD W81XWH-16-1-0171, the Michael David Greene Brain Cancer Fund (R.A.I), and VICC Ambassadors awards (J.M.I. and R.A.I.).

Conflicts of interest

The authors declare no conflicts of interest.

APPENDIX B

Fluorescence Cell Barcoding

Author: Nalin Leelatian

Introduction

High-content fluorescence flow cytometry enables quantification of multiple cellular targets in millions of individual cells (*276, 364, 399*). This technology has been extensively incorporated in both research and clinical settings for characterizing cellular phenotype, identifying novel cell subsets, monitoring therapy responsiveness, as well as dissecting mechanisms by which cells resist therapy (*14, 15, 80*). Practically, 5-10 parameters can be simultaneously quantified in every cell. This limit can be further extended to 15-25 parameters, depending on the settings of a specific flow cytometer, antibody panel design, and the capability for pre-analytical signal compensation (*283*).

Due to its quantitative capacity at the single cell level, fluorescence flow cytometry is an appealing platform for high-throughput drug screening assays as well as for the dissection of mechanistic responses of cells to external stimuli and inhibition (*12, 15, 83, 300*). The increasing scale of screening assays creates new challenges for flow cytometry approaches, especially the need to precisely capture the subtle differences in signal across multiple samples (*14, 15, 80, 84*). The time and cost of large-scale flow cytometry platforms also increase exponentially. Therefore, there is an urgent need to minimize these technical challenges, while maintaining the accuracy of data interpretation.

Fluorescence cell barcoding (FCB) increases the throughput of flow cytometry, minimizes inter-sample signal variation, and significantly reduces reagent cost (*344, 400*). In short, multiple samples (such as cells treated with different inhibitors or signaling stimuli) are labeled with different combinations of fluorophores, which have different excitation and/or emission wavelengths, at different concentrations. Thus, each sample would have a unique fluorescence “barcode”. After barcoding, all the samples are than physically combined for simultaneous antibody staining and data acquisition, according to a traditional flow cytometry workflow. Unique barcodes enable computational separation of the samples during subsequent analysis.

A previously published protocol effectively combined 96 samples using three different fluorophores, namely Pacific Blue, Alexa 488, and Alexa 700 (*344*). Notably, these fluorophores have different excitation wavelengths (i.e. they are maximally excited by three different lasers on the cytometer). One of the key challenges in FCB assays is that, when designing a fluorescence panel, some of the channels must be dedicated for FCB, leaving fewer channels for biological measurements. For example, if a 10-color panel is to be used, three channels would be dedicated for FCB, leaving seven additional channels for feature quantification. To maximize the capacity of flow cytometry, an immediate objective should be to optimize the FCB protocol so that fewer fluorophores can be used to barcode at least the same number of samples.

Additionally, the use of FCB fluorophores that have different excitation wavelengths can place a challenge on subsequent data analysis. Specifically, if the antibody that detects a cellular target of interested is conjugated to a fluorophore that has the same excitation wavelengths and a nearby emission spectra as one of the FCB dyes, there is potential for signal spillover between the FCB channels and the channel dedicated for quantifying biological features (*401*). Therefore, the excitation and emission spectra of the FCB dye combinations must also be taken in to consideration when designing and optimizing a FCB workflow.

Here, an FCB protocol that utilizes only two fluorophores, Pacific Blue and Pacific Orange, is described. These fluorophores have different emission spectra but are maximally excited at the same excitation wavelength (~405 nm). The advantage of this protocol is that one laser is fully dedicated for FCB, leaving 2-4 lasers, and up to 15-20 channels, for quantification of biological features. Another advantage is that if the FCB scale is to be expanded, additional dyes that can be maximally excited by the same laser can easily be incorporated. The protocol described here includes details on how to optimize the dye concentrations to combine 12 to 96 samples for simultaneous analysis. The approaches to address signal compensation for accurate data interpretation are also discussed.

Fluorescence cell barcoding protocol

Here, a step-by-step 48-plex (8X Pacific Blue and 6X Pacific Orange) FCB protocol is described. Specifically, this protocol is to be used for barcoding cells that have been fixed and permeabilized with 100% ice-cold methanol for at least 20 minutes (*278, 312*). Fluorophore concentrations for 48-plex FCB are specified

in Table SB-1 (see Table SB-2 for 96-plax FCB). Other FCB conditions, such as live or post-saponin barcoding, should be optimized separately and are not detailed in this protocol, but are discussed briefly later in this chapter.

Reagents

1. Pacific Blue succinimidyl ester.
2. Pacific Orange succinimidyl ester.
3. V-bottom 96-well plate with lid.
4. 12-channel pipet.
5. 200 µL pipet tips.
6. Dimethyl sulfoxide (DMSO).
7. 16% paraformaldehyde (PFA) aqueous solution.
8. Absolute methanol stored at -20°C or lower.
9. 1X phosphate buffered saline (PBS).
10. Staining medium: 1% bovine serum albumin (BSA) in PBS.
11. 12 x 75 mm round-bottom polystyrene cytometry tubes.
12. Parafilm.
13. Aluminum foil

Preparing a 48-plex cell barcoding plate

1. Prepare different concentrations of Pacific Blue and Pacific Orange dyes according to Table SB-1. Dilute dyes in DMSO.
2. Transfer 10 µL of Pacific Blue and 10 µL of Pacific Orange to each well, based on the layout depicted in Figure B-1. For example, for well B4 (row B, column 4), mix 10 µL of Pacific Blue level 2 with 10 µL of Pacific Orange level 4.
3. Cover plate with a lid and wrap the side of the plate with parafilm.
4. Use aluminum foil to cover plate to project the fluorophores from light.
5. Transfer plate to -80 °C for long-term storage, if needed.

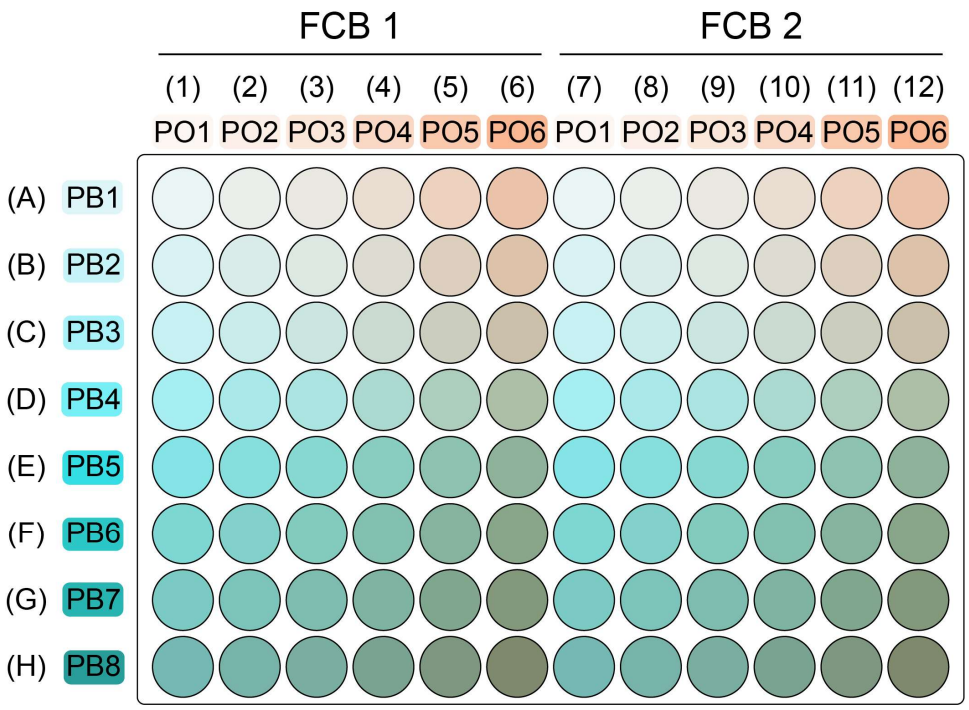


Figure B-1 Plate layout for 48-plex fluorescence cell barcoding

Eight levels of Pacific Blue (PB) and six levels of Pacific Orange (PO) can be combined to generate a 48-plex FCB platform (see Table SB-1). A 96-well plate can hold two sets of 48-plx FCB (FCB1 and FCB2).

Cell fixation and permeabilization

6. Perform the biological experiment, if needed, in a 96-well plate. At this point, cells can be in either PBS or the original experimental solution.
    a. For each barcode set, there should be 48 conditions in total (half of the plate). If there are more than 48 conditions, the total set of barcodes should be increased accordingly.
    b. At least one of the 48 conditions should be an untreated/unstimulated condition. This would serve as a biological internal negative control.
    c. If the experiment was not performed in a 96-well plate, cells must be resuspended in the appropriate volume indicated in Step 7.
7. Transfer 200 µL of cells from each condition of the original 96-well plate to a new, V-bottom 96-well plate, using a multi-channel pipet.
    a. See previously established protocol for cell viability staining, if desired (Ax700 viability stain, (*81*)).
8. Add 20 µL of 16% PFA to each well. Mix cells with PFA by pipetting using a multi-channel pipet.
9. Leave the cells for 10 minutes at room temperature.

10. Centrifuge the 96-well plate at 800 *x g* for 5 min at room temperature.
11. Decant the plate to remove supernatant.
    a. Flip the plate in one, swift motion. Only invert the plate once. Do not re-invert the plate since it would cause the cell pellets to detach from the bottom of the plate.
12. With the lid on, briefly vortex the plate to mix the cell pellets with the void volume.
13. Use a multi-channel pipet to add 200 µL of 100% ice-cold methanol to each well. Pipet to mix.
14. Put the lid back on and wrap the plate with parafilm.
15. Store the plate at -20˚C for at least 20 min or overnight.
    a. If the cells are to be stored more than a week, transfer cells to -80˚C for long-term storage.

Fluorescence cell barcoding

16. Transfer barcode plate to room temperature, if previously kept at -80˚C. Keep the plate in the dark and wait for dye to completely thaw (about 10-15 minutes).
17. Transfer cell plate to room temperature.
18. Centrifuge cell plate at 800 *x g* for 5 min at room temperature.
19. Decant the plate to remove supernatant (see Step 11 Note).
20. With the lid on, briefly vortex the plate to mix the cell pellets with the void volume.
21. Use a multi-channel pipet to add 200 µL of room temperature PBS to each well. Pipet to mix.
22. Repeat Steps 18-21 once.
23. Transfer 180 µL of cells from each well to the barcode plate. Pipet to mix.
    a. Carefully document the sample that goes into each barcode well.
    b. Be careful not to expose cells to any reagents that contain FBS or BSA prior to the barcode step since the excess amine groups from these reagents can impact the incorporation of dyes into cells (see Discussion).
    c. The remaining cells in the original cell plate can be combined and set aside as “Unstained” control. Keep the cells in a 5 mL FACS tube at 4˚C.
24. Leave cells in the dark at room temperature for 30 minutes

25. Use a multi-channel pipet to add 20 µL of 1% BSA in PBS (staining media) to each well. Pipet to mix.
26. Centrifuge cell plate at 800 *x g* for 5 min at room temperature.
27. Decant the plate to remove supernatant (see Step 11 Note).
28. With the lid on, briefly vortex the plate to mix the cell pellets with the void volume.
29. Use a multi-channel pipet to add 200 µL of staining media to each well. Pipet to mix.
30. Repeat Steps 26-28 once.

Creating Pacific Blue and Pacific Orange control tubes

31. Add 200 µL of staining media to the following wells: PB1PO1 (A1), PB8PO1 (H1), and PB1PO6 (A6) (see Figure B-1). Pipet to mix.
32. Label 2 FACS tubes as “Pacific Blue compensation” and “Pacific Orange compensation”
33. Transfer 10 µL of cell suspension from A1 and H1 to “Pacific Blue compensation” tube. Add 200 µL of PBS to the tube. Keep the tube at 4˚C.
34. Transfer 10 µL of cell suspension from A1 and A6 to “Pacific Orange compensation” tube. Add 200 µL of PBS to the tube. Keep the tube at 4˚C.

Combining cells for barcode-only control and for subsequent immunostains

35. Use the remaining volume in the plate to combine all the cells from 48 wells into a new FACS tube labeled “Barcode-only”.
36. Add 200 µL of staining media to a few more wells. Repeat Step 35 twice to completely remove cells from the barcode plate.
37. Centrifuge cell FACS tube at 800 *x g* for 5 min at room temperature.
38. Decant the tube to remove supernatant (see Step 11 Note).
39. Briefly vortex the plate to mix the cell pellet with the void volume.
40. Add 1 mL of staining media. Briefly vortex to mix.
41. Repeat Steps 37-40 twice.
42. Centrifuge cell FACS tube at 800 *x g* for 5 min at room temperature.

43. Decant the tube to remove supernatant (see Step 11 Note).
44. Add the appropriate amount of staining volume for subsequent immunostaining, if needed (see (*81*) for a detailed immunostaining protocol). If no immunostaining is needed, skip to Step 46.
45. Transfer cell suspension into a new FACS tube for immunostaining.
46. Add 1mL of PBS to the remaining cells in the “Barcode-only” tube.
47. Centrifuge cell FACS tube at 800 *x g* for 5 min at room temperature.
48. Decant the tube to remove supernatant (see Step 11 Note).
49. Briefly vortex the plate to mix the cell pellet with the void volume.
50. Repeat Steps 46-49 once.
51. Add 200 µL of PBS to the “Barcode-only” tube and keep the tube at 4˚C until ready for flow cytometry.
    d. If a viability stain, such as Ax700, was used in Step 7, only the $Ax700^-$ live cells should be included for signal compensation (see “De-barcoding and signal compensation” under Discussion). This is accomplished by computationally gating for $Ax700^-$ cells after sample collection, prior to data analysis.

Discussion

Cell permeabilization reagents can impact fluorescence cell barcoding

Since the multi-dimensionality nature of fluorescence flow cytometry enables quantification of multiple parameters in single cells, this technology is often used for simultaneous characterization of both surface and intracellular features (*14, 15, 312, 345, 402*). This usually requires a multi-step staining protocol, which include a surface staining step (usually performed on live cells), followed by one or more intracellular staining steps (*312*). The optimization of FCB for a given experiment depends largely on the step by which sample combination is needed (Figure B-2). Any immunostaining prior to FCB must be performed on each sample separately. Generally, there are two types of FCB; live FCB, and post-permeabilization FCB (*400*).

The succinimidyl esters of the fluorophores used in this FCB protocol react to amine sources (i.e. proteins) in the cells (*403*). The available binding sites per cell, therefore, directly influence the level of fluorophore that would be incorporated. Therefore, more dyes are more likely to be incorporated into larger cells, and less dyes into smaller cells. This factor needs to be taken into consideration during experiment design and

data analysis, especially during the de-barcoding steps (see De-barcoding and signal compensation). Importantly, it is necessary to separately test the individual fluorophores to optimize the concentrations needed for different barcoding protocols, since different cell barcoding conditions (pre- vs post-permeabilization) would directly impact the degree of dye incorporation per cell. In general, live FCB requires a higher concentration of fluorophore to yield the same fluorescence intensity compared to post-permeabilization FCB, due to fewer available proteins and amine groups on the cell surfaces.

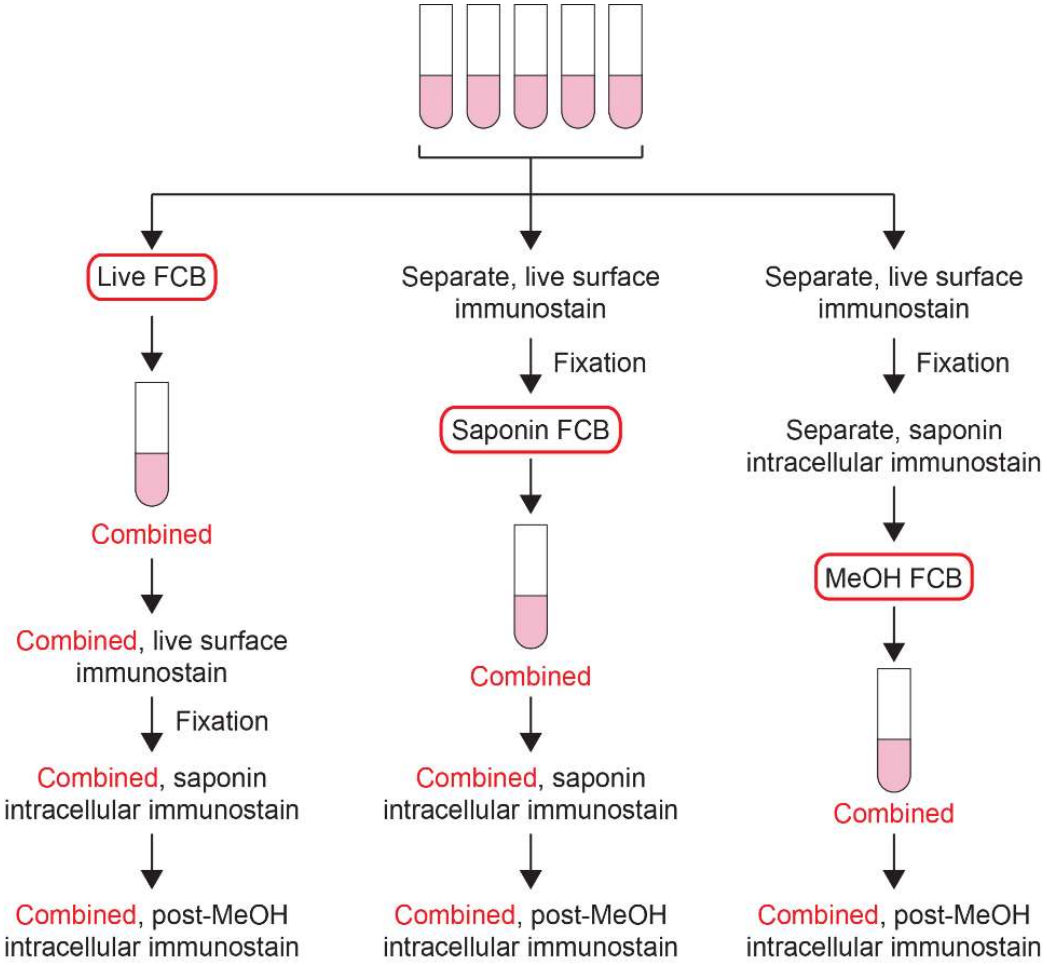


Figure B-2 Fluorescence cell barcoding immediately precedes sample pooling

Samples can be properly combined for simultaneous immunostaining and analysis only after FCB. (Left) If the protocol is optimized for live FCB, samples can be combined after FCB for live, surface immunostaining. (Middle and Right) If post-permeabilization FCB is to be used, samples must undergo surface immunostaining separately (if needed) prior to FCB and sample pooling.

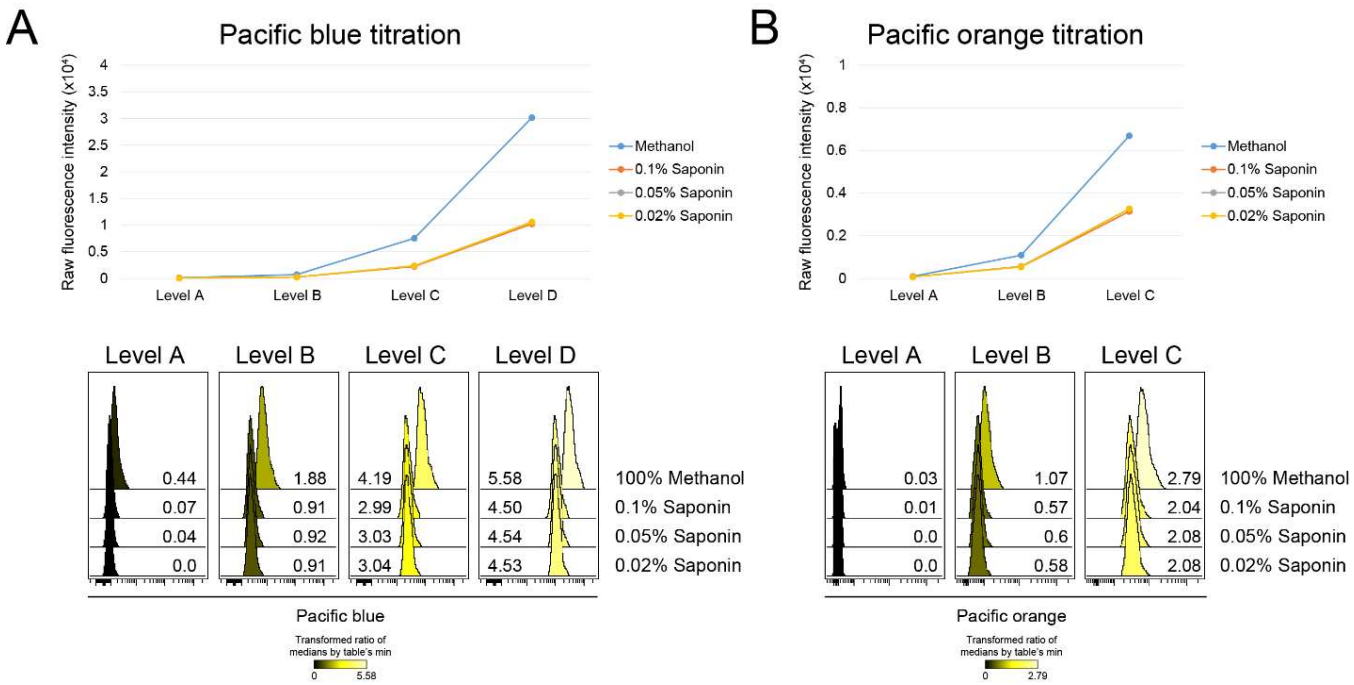


Figure B-3 Permeabilization reagents impact bindings of FCB dye to cells

(A, top) Four levels of Pacific blue were used to barcode cells that were permeabilized with 100% ice-cold methanol (blue), 0.1% saponin (orange), 0.05% saponin (gray), and 0.02% saponin (yellow). Y-axis represents raw fluorescence intensity (x 10^4). (A, bottom) histogram of cells that were permeabilized with different reagents (rows) and barcoded with different concentrations of Pacific Blue (columns), are shown. Values adjacent to histograms indicate transformed ratios of values by table's minimum. (B) Similar analysis was performed using three levels of Pacific Orange.

The type of cell permeabilization reagent also impacts FCB. For example, when the same concentrations of Pacific Blue and Pacific Orange were used to barcode cells that were permeabilized with either 100% ice-cold methanol (see FCB protocol) or different concentrations of saponin, cells that were permeabilized with methanol showed higher fluorescence signals, whereas saponin permeabilization resulted in lower fluorescence signals regardless of its concentration (Figure B-3).

Correlations between concentrations of fluorophores and signal intensity

Precise titration of fluorophore is key to a successful FCB experiment. Specifically, when a new dye is to be used, it is essential to characterize the relationship between dye concentrations and fluorescence intensity. As mentioned previously, the method for cell barcoding directly impacts the dye incorporation and would alter the relationship between dye concentrations and fluorescence intensity. The example detailed here utilized cells that were permeabilized with 100% ice-cold methanol for at least 20 minutes (*278*).

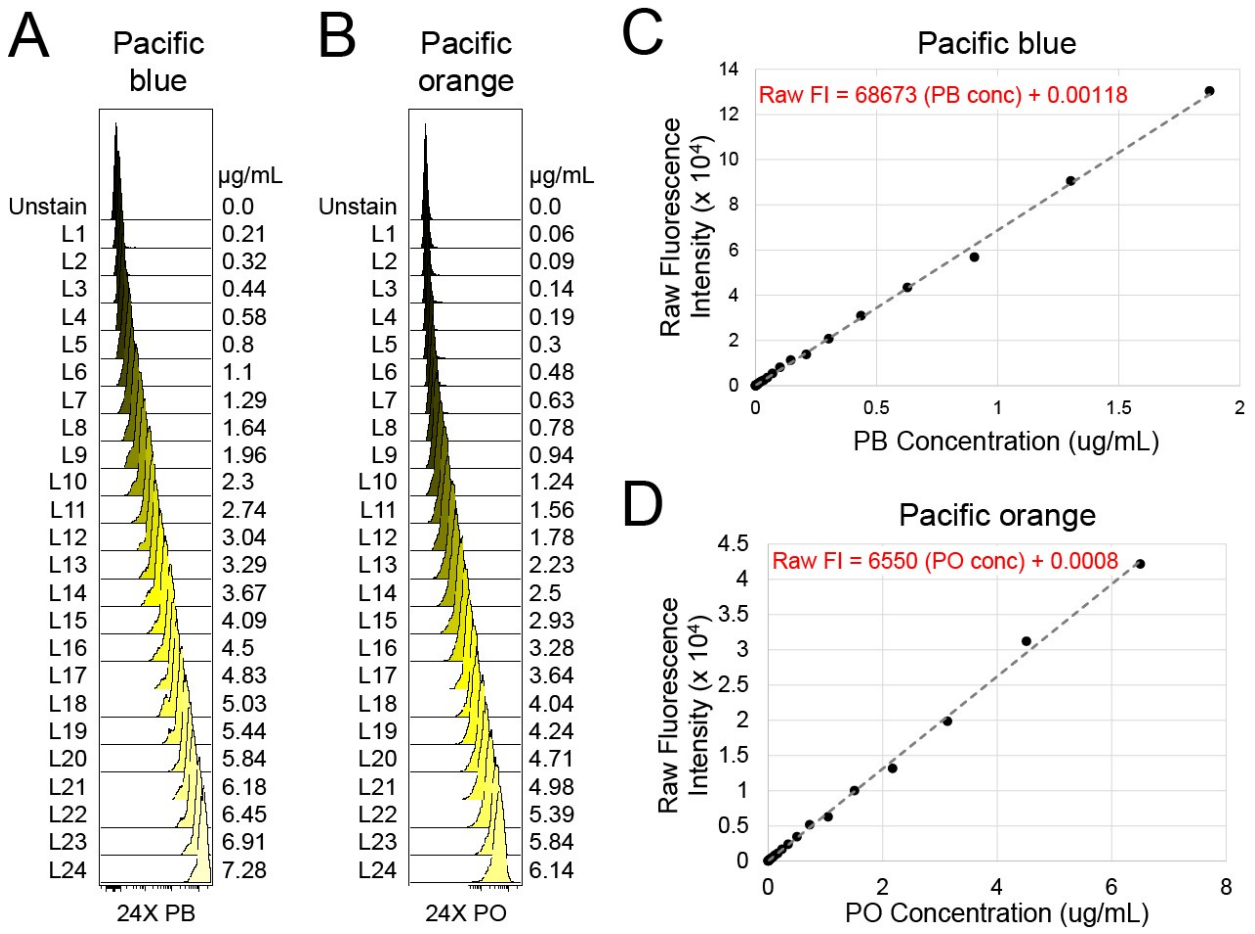


Figure B-4 Pac Blue and Pac Orange titrations and equations

(A) Raji cells stained with 24 levels of Pacific blue (PB) ranging from 0.21 to 7.28 µg/mL were analyzed and histograms are illustrated. (B) Similar titration was performed with Pacific orange (PO) using SKO cells, ranging from 0.06 to 6.14 µg/mL. Unstained cells were included as controls (top rows) (C-D) Relationship between dye concentration and raw fluorescence intensity are shown, and the linear equations are specified for each fluorophore. X-axis represents concentration of fluorophore (µg/mL), and y-axis represents raw fluorescence intensity (x 10^4).

As the concentration of fluorophores increases, the raw fluorescence intensity also increases (Figure B-4). Notably, Pacific Blue was used at a lower concentration than Pacific Orange to achieve similar fluorescence intensity (Figure B-4C and Figure B-4D). The relationships between dye concentrations and fluorescence intensity (FI) could be illustrated as linear correlations that passed closely to the (0,0) coordinated (Figures B-4C and Figure B-4D), as reflected in the dye-intensity equations (1) and (2):

(1) Raw FI = 68,673(PB conc) + 0.00118

(2) Raw FI = 6,550(PO conc) + 0.0008

Note: FI, fluorescence intensity; conc, concentration (µg/mL), PB, Pacific Blue; PO, Pacific Orange

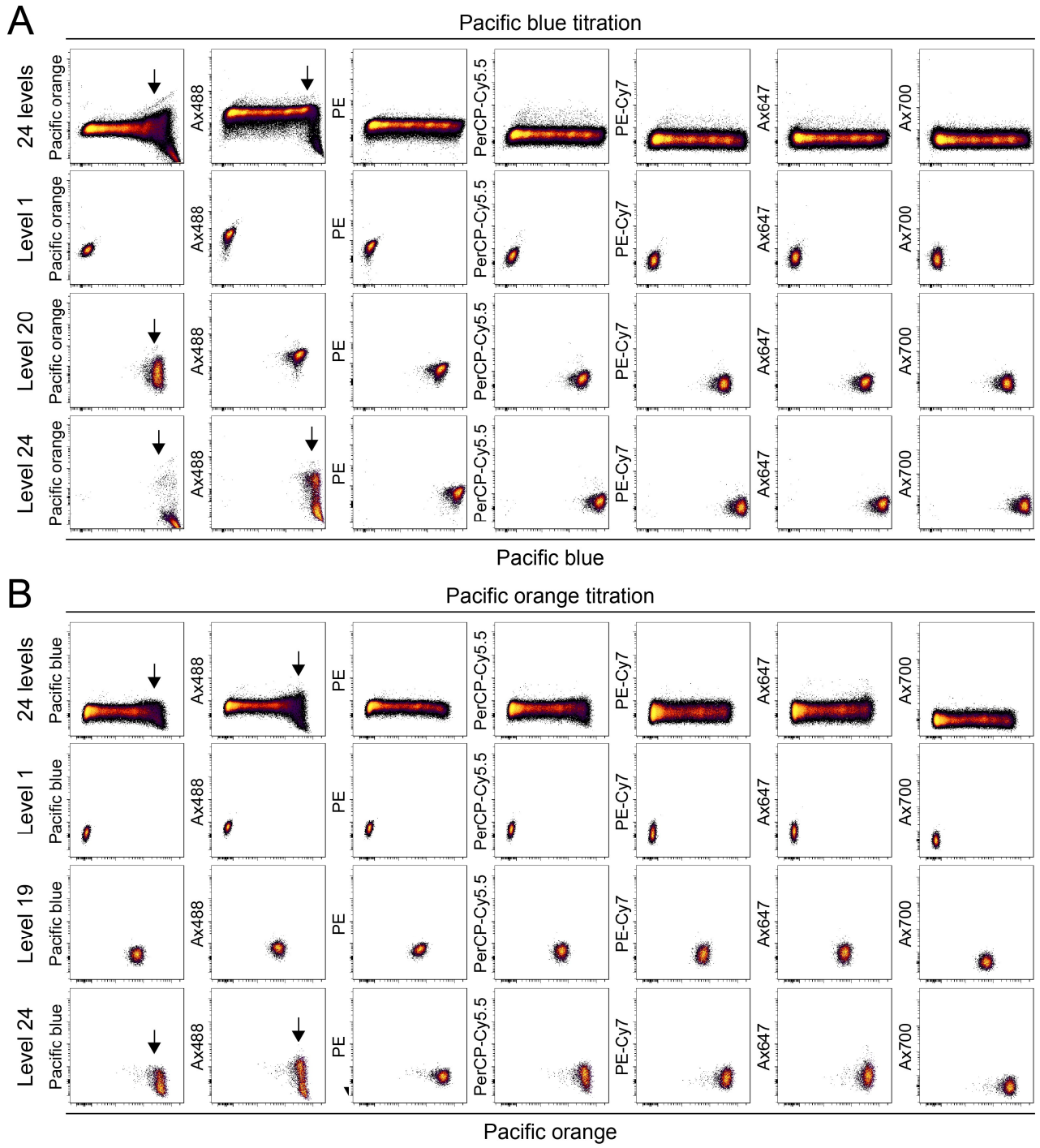


Figure B-5 Higher concentrations of fluorophores lead to significant spillover and spreading error

(A) Assessment of signal spill-over from 24 levels of Pacific blue into other channels after compensation is illustrated. X-axis represents an increasing concentration of Pacific blue from left to right, and y-axis represents recipient channels (Pacific Orange, Ax488, PE, PerCP-Cy5.5, PE-Cy7, Ax647, and Ax700). Sample that contained Raji cells stained 24 levels of Pacific blue is shown in the top row. Individual Pacific Blue levels (levels 1, 20, and 24) are also shown separately. (B) Similar assessment was performed in SKO cells stained with 24 levels of Pacific orange. X-axis represents an increasing concentration of Pacific orange from left to right, and y-axis represents recipient channels. Individual levels that are shown included levels 1, 19, and 24. Arrows demonstrate signal spill-over and spreading errors. Also refer to Figure B-3.

Once the dye-intensity equations are computed, the number of samples for a given FCB experiment can be tailored as needed (see Table SB-1 and Table SB-2). The lower limit of the FCB range is usually the lowest concentration that can be easily distinguished from unstained, non-barcoded samples (L1; Figures B-4A and Figure B-4B). This concentration can be slightly adjusted for ease of subsequent de-barcoding (i.e. to easily separate barcoded samples from cells that were not properly barcoded). The major concern for determining the upper limit of each dye is its potential spill-over into other channels, especially those with similar emission spectra. It is critical that the spill-over from the highest concentration must be easily compensated and have no spreading error in other channels, since these can lead to false positive or false negative signals (Figure B-5). After the upper and lower limits are determined, the concentrations of fluorophores for individual FCB level are then calculated from the equations. FIs that are in between the upper and lower limits are pre-determined, with the goal to equally separate each level on a log or asinh scale. Each FI would be the desired median FI for a given FCB level. Note that log or asinh scales are standard for flow cytometry data visualization, which equally emphasize the fold-change differences at lower and higher FIs. The fewer the number of samples to be barcoded, the easier the FCB optimization and subsequent computational de-barcoding (sample gating and isolation) since there would be more separation between each sample (Figure B-6).

De-barcoding and signal compensation

The purpose of sample de-barcoding is to computationally separate pooled data into individual samples for subsequent analysis. Since each sample is stained with both Pacific Blue and Pacific Orange at specific concentrations, this step requires proper signal compensation between the two FCB channels (Figure B-7A). The practical strategy of FCB compensation is to align the median signal of same level of each dye (e.g. all PB1 or all PO1) in a straight line, on either the y-axis (here, Pacific Orange) or the x-axis (here, Pacific Blue), depending on the data display. As seen in the “before compensation” plot, for a given level of Pacific orange (row, y-axis)), as the signal of Pacific Blue increases (column, x-axis), the signal of Pacific Orange also increases. However, based on the FCB scheme, Pacific Orange signals on any given row should remain constant. Therefore, Pacific Blue is the “source” of the false positive signal seen in Pacific Orange, which is the “detector” or “signal recipient”. Similar spill-over from Pacific Orange into Pacific Blue is also seen, but now the source and

recipient are now switched. Notably, the degree of PO-into-PB spill-over is of a lesser degree than the PB-into-PO spill-over. This is expected as signal spill-over is usually more prominent if the detecting channel has a longer emission wavelength.

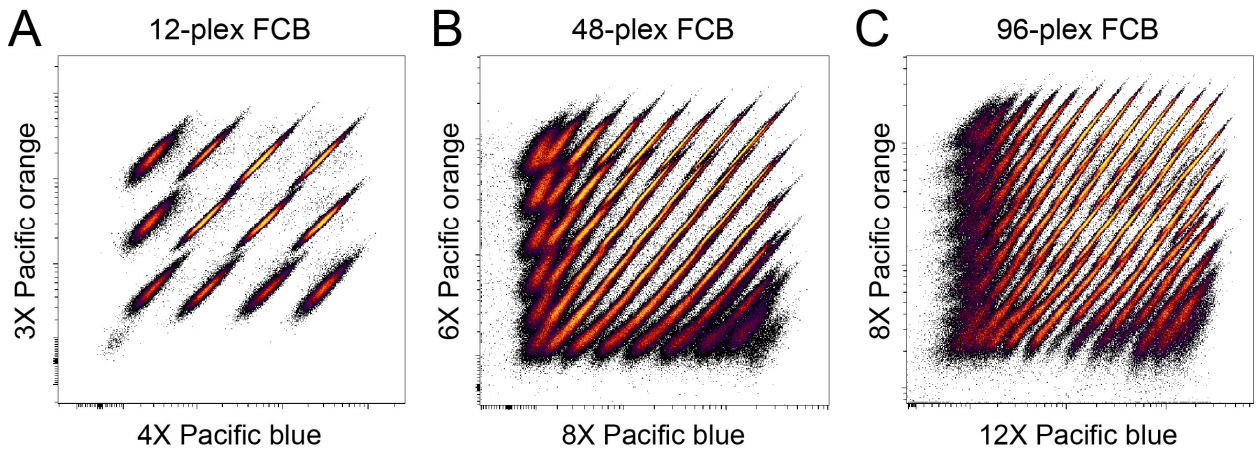


Figure B-6 Complexity of FCB can be augmented to increase experimental throughput
(A) A 12-plex FCB is shown on a biaxial single-cell plot. The scheme of the FCB includes 4 levels of Pacific Blue and 3 levels of Pacific Orange. (B) Augmented 48-plex (8 levels of Pacific Blue and 6 levels of Pacific Orange), and (C) 96-plex (12 levels of Pacific Blue and 8 levels of Pacific Orange) are also shown.

After the FCB channels are properly compensated, individual samples can be isolated by sequential biaxial gating. (Figure B-7B). It is important to computationally separate the sample in a stepwise manner, rather than gating directly on the PB-PO biaxial plot. As mentioned previously, larger cells usually have more proteins and amine groups to react with barcoding dyes. Even in seemingly homogenous samples such as cell lines, the size of individual cells can vary. And therefore, per-cell FI of the barcoding dyes would vary based on cell size (forward-scatter, FSC-A, x-axis; Figure B-7B, top). This variation can be seen on the PB-PO biaxial plot. For example, in the 12-plex FCB, the larger cells within each of the 12 populations would be on the top right part of the population, whereas smaller cells would be on the bottom left (Figure B-6A). These “tails” can significantly overlap with other nearby populations as the number of barcoded samples increases (Figures B-6B and B-6C), making direct gating on the PB-PO biaxial plots challenging. A sequential gating scheme starting with 1) a biaxial plot of cell size (FSC-A) vs one of the barcoding dyes (here, Pacific Orange; Figure B-7B, top), followed by 2) a secondary PB-PO biaxial plot, is performed to avoid incorrect assignment of cells to the wrong sample.

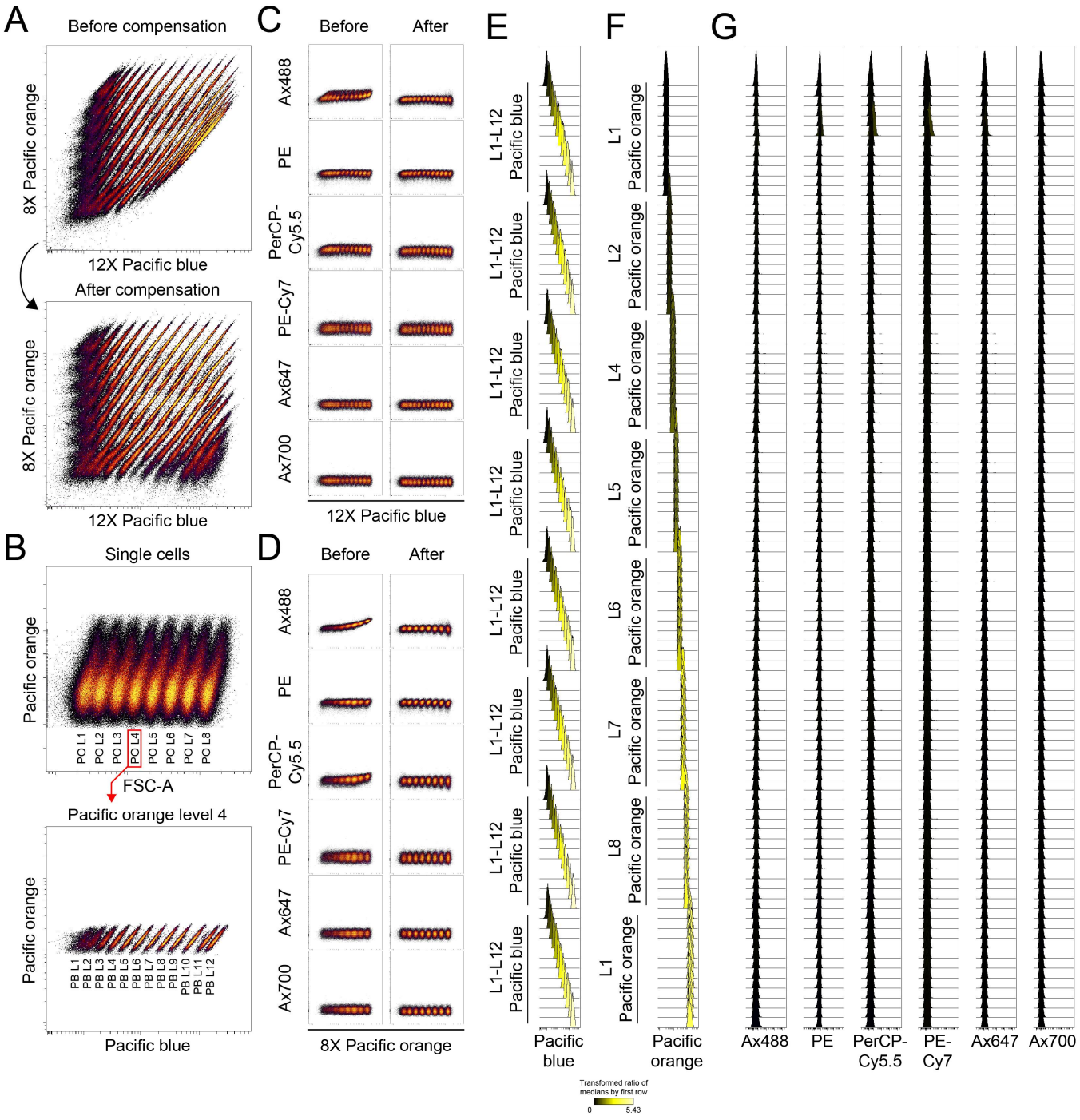


Figure B-7 De-barcoding and signal compensation are crucial for data analysis

(A) A 96-plex FCB of single SKO cells before (top) and after (bottom) signal compensation is illustrated. (B) After Pacific Blue and Pacific Orange compensation, pooled data are computationally separated into individual samples by biaxial gating. The first step is to separate one of the FCB channels (in this case, Pacific Orange), into different levels. And within each level (Pacific Orange L4 is shown as an example), the second FCB dye (Pacific Blue) is used to identify individual samples. (C) Spill-over signals from Pacific Blue, and (D) Pacific Orange, into other recipient channels (y-axes) before (left) and after (right) appropriate compensation are illustrated. Appropriate de-barcoding and compensation resulted in sample separation (E and F), and lack of signal spill-over into recipient channels (G) are shown.

After FCB channels are decompensated and individual samples are isolated, any signal spill-over from FBC channels into other detectors should be corrected (Figure B-7C and B-7D). This can happen even with an ideal antibody panel design, due to the nature of fluorophore excitation and emission spectra. The practical approach of this step is the same as FCB signal compensation; for any given concentration of Pacific Blue and Pacific Orange, there should be no signal spill-over into any other recipient channels. For example, before compensation, as the signal of Pacific Orange increased, the signal detected as Ax488 also increased (Figure B-7D, top left). But since neither Ax488 dye nor Ax488-conjugated antibody was used, there should be no Ax488 signal in any of the samples. A compensation value must, therefore, be added to remove the false positive Ax488 signal that resulted from Pacific Orange spill-over (Figures B-7D, top right, and Figure B-8). After appropriate signal compensation, there were no change in signal in any of the other recipient channels, even at the highest levels of Pacific Blue and Pacific Orange (Figure B-7E, Figure B-7F, and Figure B-7G). This ensures that in an actual experiment with additional fluorescence-tagged antibodies, any signal detected in those recipient channels would be a reflection of true biological changes.

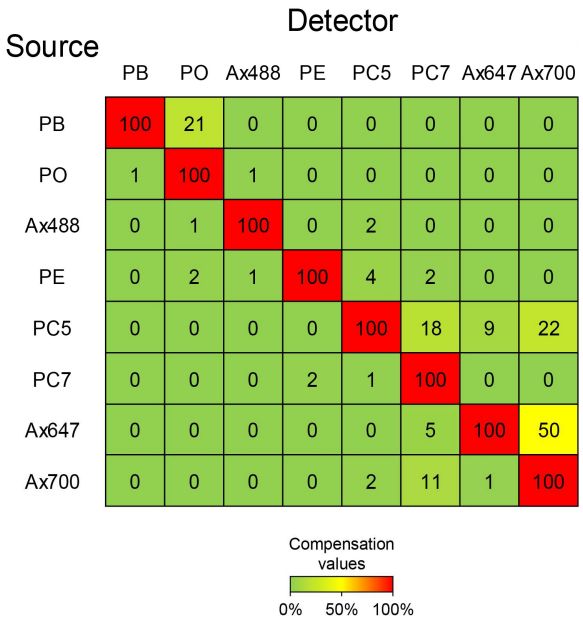


Figure B-8 Matrix depicting compensation values in an example FCB panel

Compensation matrix of 96-plax FCB data shown in Figure B-6. Numbers represent compensation values, ranging from 0% to 100%. For the corresponding source (row) and detector (column), the compensation values are always 100%. Other numbers reflect the percent of false positive signal from other given channels. For example, there is a 21% false positive signal that was detected as Pacific Orange that was in fact a spill-over from Pacific Blue. PB = Pacific Blue, PO = Pacific Orange, PC5 = PerCP-Cy5.5, and PC7 = PE-Cy7.

During FCB optimization, it is critical to compare an unstained sample with samples that are fully barcoded without additional antibodies (barcode-only sample). This ensures that the observed signals are due to FCB alone. However, it is important to assess the signal spill-over form the FCB channels into other channels that might be used for biological quantification in subsequent experiments. And most importantly, both unstained and barcode-only samples should always be collected as controls (see FCB protocol), even while performing actual biological experiments, to ensure appropriate signal compensation for accurate data interpretation.

Conclusions

Fluorescence cell barcoding is a technical approach that significantly enhances the throughput of conventional fluorescence flow cytometry. FCB increases the accuracy of data analysis, especially for inter-sample comparison, and reduces reagent costs. Notably, FCB is especially suitable for experimentation that requires accurate comparisons between multiple samples or conditions. The FCB protocol discussed here can be readily coupled with a liquid handling system to enable high-throughput assay automation. The high flexibility of the FCB protocol allows users to easily tailor the dye combinations to accommodate different experiment purposes. Even though 12-, 48-, and 96-plex FCB schemes were discussed here, additional fluorophores can be titrated and added to the protocol if needed. As mentioned previously, fewer biological readouts would fit into a panel that uses more fluorophores for FCB. Thus, a 384-plex or even a 1536-plex FCB format might be more suitable for experimentations where fewer biological readouts are required (e.g. a drug screening assay where only apoptosis or cell death are the end-points of interest). A challenge in FCB analysis is the signal compensation and sample de-barcoding. A step-wise analysis approach described here enables accurate assignment of cells to a given barcoded sample. This can be easily developed into a computational de-barcode algorithm that, when coupled with robotic liquid handling, would result in a fully automated experimental and analytical pipeline.

Table SB-1 Concentrations of Pacific Blue and Pacific Orange for 48-plex barcoding

Pacific Orange	Concentrations for 1:20 dilution (μg/mL)*
L1	0.218
L2	0.542
L3	1.352
L4	3.371
L5	8.404
L6	20.95

Pacific Blue	Concentrations for 1:20 dilution (μg/mL)*
L1	0.038
L2	0.080
L3	0.172
L4	0.367
L5	0.785
L6	1.679
L7	3.589
L8	7.672

* Add 10 μL of the dyes to each well, according to the outline in Figure B-1. The total volume of dye per well will be 20 μL.

Table SB-2 Concentrations of Pacific Blue and Pacific Orange for 96-plex barcoding

Pacific Orange	Concentrations for 1:20 dilution (µg/mL)
L1	0.370
L2	0.659
L3	1.172
L4	2.087
L5	3.715
L6	6.612
L7	11.770
L8	20.95

Pacific Blue	Concentrations for 1:20 dilution (µg/mL)
L1	0.031
L2	0.051
L3	0.085
L4	0.140
L5	0.230
L6	0.380
L7	0.627
L8	1.035
L9	1.708
L10	2.818
L11	4.650
L12	7.672

APPENDIX C

Using Fluorescence Flow Cytometry to Quantify Cellular Uptake of Compounds and Track Signaling Responses

Authors: Katherine M. Chong, Nalin Leelatian, Sean M. DeGuire, Asa A. Brockman, David C. Earl, Rebecca A. Ihrie, Jonathan M. Irish, Brian O. Bachmann, and Gary A. Sulikowski

Data presented in this chapter was published in *Journal of Antibiotics (Tokyo)* 2016 (Chong, Leelatian et al. 2017), titled “The use of fluorescently-tagged apoptolidins in cellular uptake and response studies”.

Preface

The study described here presents another application of fluorescence flow cytometry to track cellular drug uptake. Here, two forms of apoptolidin, a drug that was shown to preferentially induce cell death in transformed glial cells compared to non-transformed cells, were conjugated to fluorescent cyanine-3 (Cy3). This allows direct tracking of their differential uptake in a panel of human cells in conjunction with the quantification of signaling alterations and functional changes secondary to drug treatments. This approach has significant implications in the identification of novel therapeutic compounds that specifically target cancer cells, or cancer cell subsets, compared to normal cells. Additionally, the high-dimensionality nature of flow cytometry enables simultaneous dissection of the mechanistic alterations, such as decreased phosphorylation of a specific signaling effector, in cancer cells that were targeted by the drugs. This is especially valuable for pre-clinical studies that aim to characterize drug specificity and cytotoxicity mechanisms. Additionally, the drug conjugation reaction can potentially be adapted to enable labeling with metal isotopes for a high-dimensional single-cell characterization by mass cytometry.

Abstract

The apoptolidins are glycomacrolide microbial metabolites reported to be selectively cytotoxic against tumor cells. Using fluorescently tagged active derivatives we demonstrate selective uptake of these four tagged glycomacrolides in cancer cells over healthy human blood cells. We also demonstrate the utility of these five fluorescently tagged glycomacrolides in fluorescent flow cytometry to monitor cellular uptake of the six glycomacrolides and cellular response.

Introduction

Apoptolidin A, isolated from an actinomycete (FU 40), was reported to selectively induce cell death in E1A and E1A/E1B19K transformed rat glia cells, while normal rat glia and 3Y1 fibroblast cells were found to be non-responsive to apoptolidin A (*404*). The pathway of cell death was proposed to be apoptosis based on observed cell morphology and DNA laddering. Khosla and Salomon later demonstrated apoptolidin A induced apoptosis in LYas mouse lymphoma cells via mitochondrial pathway (*405, 406*). In early 2015, Ishmael and co-workers demonstrated apoptolidin A treatment to several tumor cell lines in culture, caused a shift in cellular metabolism and status of AMP-activated protein kinase (AMPK) stress pathway leading to autophagy as characterized by increased phosphorylation of AMPK (Thr 172), acetyl-CoA carboxylase (ACC), and Unc-51 like autophagy activating kinase (ULK) (*407*). The observed effects were consistent with ATP synthase (mitochondrial respiration) inhibition as a mode of action (*405, 406*).

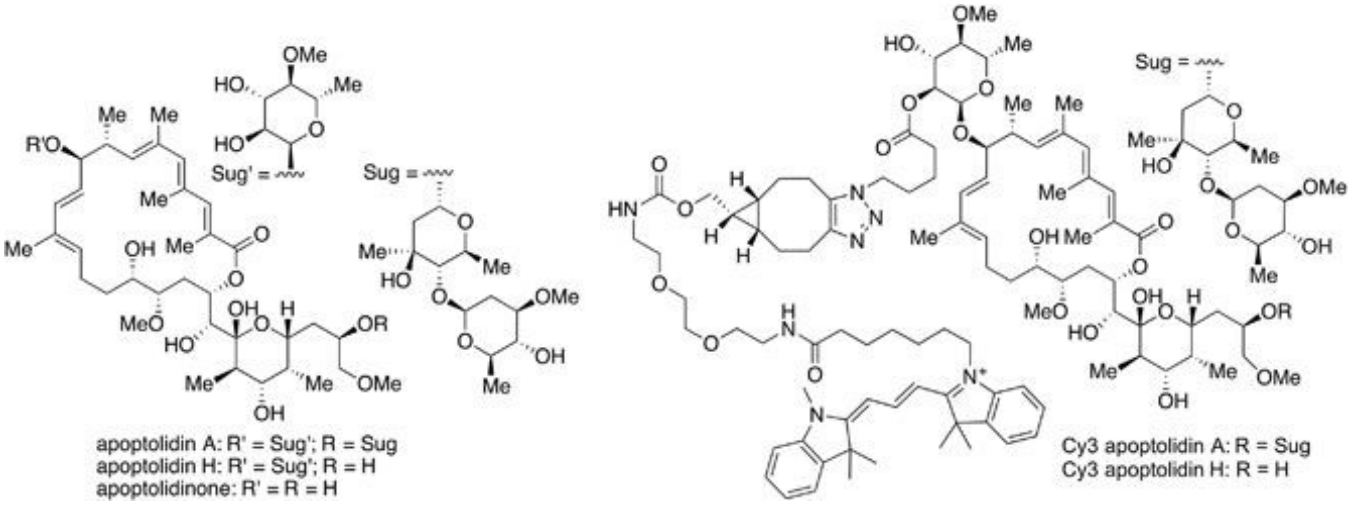


Figure C-1 Apoptolidins and fluorescent derivatives

Since the original isolation of apoptolidin A, over eight congeners have been produced by fermentation (*408-413*) mutasynthesis, precursor directed biosynthesis (*414, 415*), and chemical synthesis (Figure C-1) (*416-418*). Accompanying cell cytotoxicity studies have demonstrated the need for deoxy sugars to observe cellular activity, as the aglycone (apoptolidinone) has shown no activity (EC50 >10 μm) in a variety of tumor cell lines (for example, H292, breast, and HCT 116) (*417, 418*).

Dye labeled small molecules enable visualization of cellular uptake, localization and selective response studies using confocal microscopy. We had prepared cyanine-3 (Cy3) conjugates of apoptolidin A and H enabling cellular localization studies that demonstrated localization in the mitochondria using H292 cells (*419*). Recently, Hecht and co-workers demonstrated using dye-labeled conjugates of bleomycin and its corresponding aglycone (deglycobleomycin), the selective uptake of the former by cancer cells in cell culture (*420-423*). We reasoned similar studies could be conducted using Cy3 apoptolidins A/H in cell culture leading to a possible explanation of the observed cell selectivity as described in the original isolation paper. Furthermore, we employed fluorescent flow cytometry to monitor cellular uptake of fluorescent apoptolidin probes in a high-throughput, single-cell format.

Materials and methods

Production and chemical synthesis of apoptolidins and fluorescent derivatives

Apoptolidins A and H were produced by fermentation of the apoptolidin producer FU 40 and a mutant strain (ApoGT2 knockout) at Vanderbilt University. Cyanine-3 derivatives of apoptolidin A and H were prepared by semi synthesis as described earlier.

Uptake of apoptolidins A and H in various cell types

Human cancer cell lines and peripheral blood mononuclear cells (PBMCs) were used to characterize uptake of apoptolidin A and apoptolidin H. The following cell lines were included: SW620 (colon cancer), U87-MG (glioblastoma), LN229 (glioblastoma), and A549 (lung adenocarcinoma). Cell lines were cultured under ATCC recommended protocols. Cells were detached using Trypsin and resuspended in recommended culture media at 1×10^6 cells ml^{-1} prior to drug treatment. Human PBMCs were collected from a healthy donor, following the protocols approved by Vanderbilt University Medical Center Institutional Review Board, processed by

standard Ficoll preparation protocol, and cryopreserved in liquid nitrogen. PBMCs were thawed and resuspended in warm RPMI 1640 media containing 10% FBS at 1×10^6 cells ml^{-1} prior to drug treatment. Cells were treated with either vehicle dimethyl sulfoxide (DMSO), 1 μm of Cy3 apoplolidin A, or 1 μm of Cy3 apoptolidin H for 1 h at 37 °C. Cells were washed twice in PBS and fixed with 1.6% paraformaldehyde for 10 min at room temperature, and were permeabilized with ice-cold methanol for 30 min.

Fluorescence flow cytometry

After methanol permeabilization, cells were stained with 1:250 anti p-ACC antibody (cell signaling) for 30 min in the dark at room temperature. Cells were then stained with 1:1000 Donkey anti-Rabbit Ax647 (Life Technologies, Carlsbad, CA, USA) for 30 min in the dark at room temperature and were washed and resuspended in PBS for analysis on 5-laser BD LSRII (BD Biosciences, San Jose, CA, USA) at the Vanderbilt Flow Cytometry Shared Resource and evaluated using the Cytobank software (Mountain View, CA, USA). Untreated cells were stained with only the secondary antibody and used as negative control.

Confocal microscopy

The stained cell suspensions described above were incubated with diaminophenylindole at 1 μg ml^{-1} for 10 min at room temperature and placed on glass slides for imaging on an LSM 710 META inverted microscope (Zeiss, Oberkdochen, Germany) at the Vanderbilt Cell Imaging Shared Resource. Data were analyzed using Zen 2011 software.

Results and Discussion

In earlier work, we determined (using Cy3) that apoptolidin A and H localized in the mitochondria of human lung (H292) tumor cells using confocal microscopy.16 This observation is in complete agreement with earlier reports describing apoptolidin A as a potential ATP synthase (oxidative phosphorylation) inhibitor (*405, 406*).

In the original isolation paper and later when evaluated against the National Cancer Institute 60 (NCI-60) human cancer cell line panel, apoptolidin A was described as a selective inhibitor of cell growth (*404-406*). Cell

selectivity can be imparted by either selective cell uptake and/or targeting of cell phenotype (for example, metabolic state of a tumor cell). Hecht and co-workers have demonstrated cyanine tagged bleomycin is selectively taken up in most cancer cell lines in comparison with 'normal' cell counterparts in cell culture (*422, 423*). However, these studies determined the activity of compounds, including apoptolidin, after long-term treatment (2–6 days) of cells *in vitro*. Here, we aimed to characterize the immediate uptake of apoptolidin A and H by different human cell types, as well as their signaling responses to the treatment.

We used confocal fluorescent microscopy to characterize the uptake of Cy3 apoptolidin A and H in healthy PBMCs and human lung adenocarcinoma (A549) and human glioblastoma (U87) tumor cells after 1-hour treatment (Figure C-2). The confocal images revealed minimal uptake of Cy3 apoptolidins by healthy PBMCs, but higher uptake of Cy3 apoptolidins by A549 and U87 tumor cells.

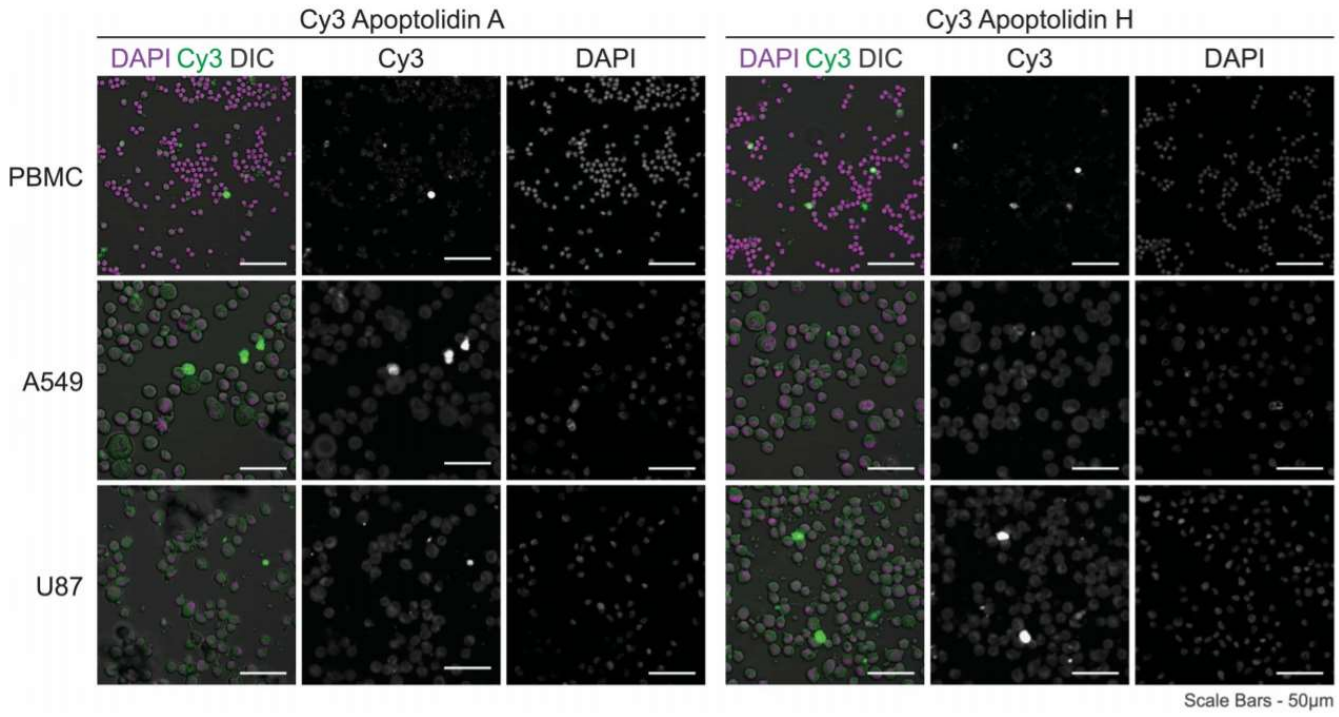


Figure C-2 Differential uptake of Cy3 apoptolidins by tumor cells relative to healthy blood cells

Confocal images of different human cell types showing minimal Cy3 apoptolidins uptake (green) by healthy PBMCs and higher uptake in A549 and U87 cells. Nuclei were stained with diaminophenylindole (pink). Scale bars=50 μm. DIC, differential interference contrast.

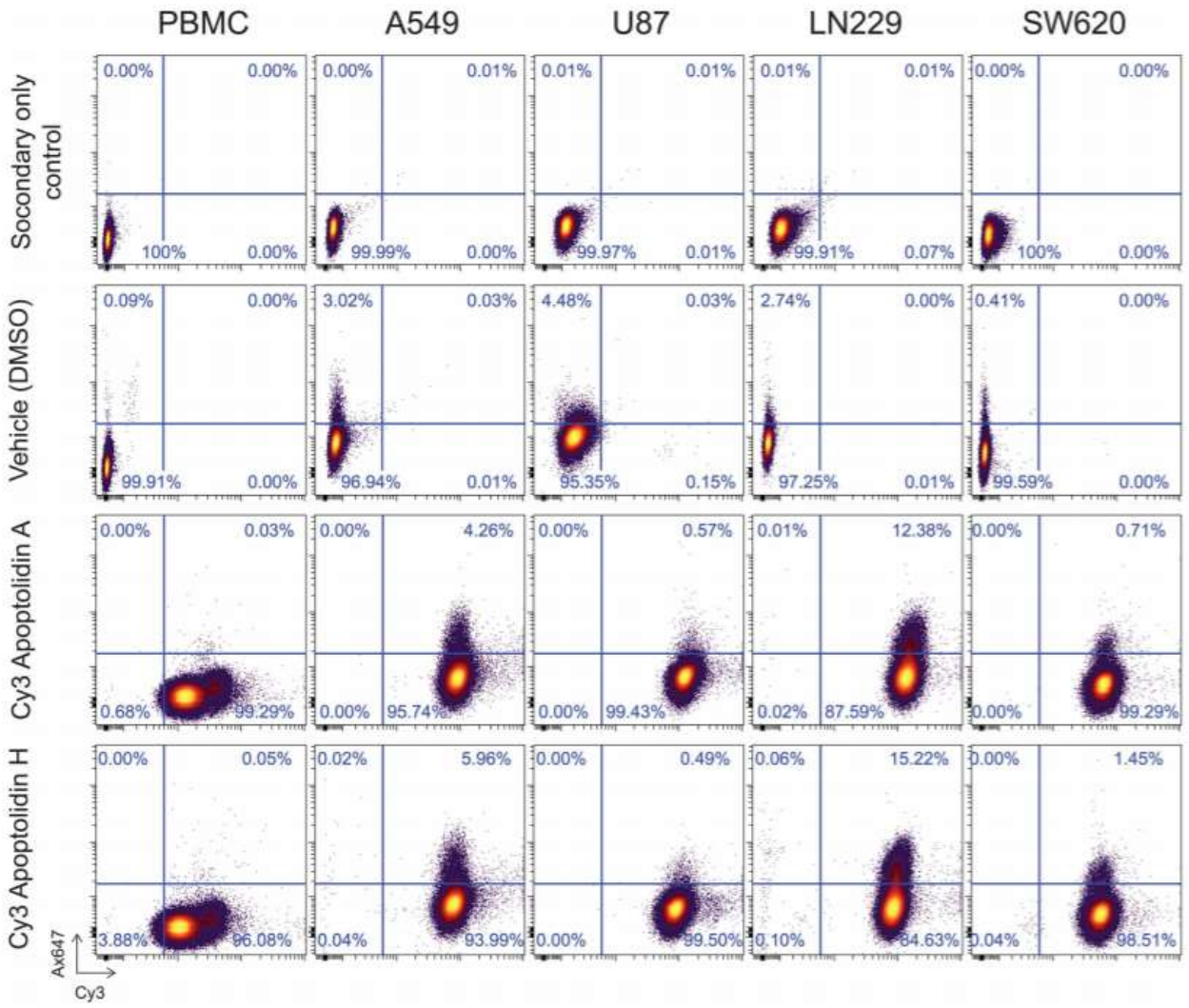


Figure C-3 Fluorescent flow cytometry used to monitor cellular uptake of Cy3 apoptolidins and phosphorylation of ACC (Ser79)

Biaxial contour plots are shown to represent cell density (heat).

Fluorescent phospho-specific flow cytometry (phospho-flow) employs fluorescently tagged antibodies to dissect activation of cell signaling pathways in single cells in response to treatment with small molecules including natural products (*402*). As fluorophores with different emission wavelengths can be monitored in different channels the uptake of fluorescent small molecules can be monitored as well as cell response. For example, the cellular uptake of fluorescent anticancer agents such as daunomycin as well as fluorescent nanoparticles has been monitored by traditional flow cytometry (*424, 425*). Phospho-flow allows quantitative measurement of activated signaling responses of individual cells to external stimuli, such as drug treatment. Often patient-derived tumor cells contain a heterogeneous mix of different cell subsets, which can harbor distinct sensitivity to

exogenous agents. Apoptolidin A and C sensitive cell lines including glioblastoma (U87), were reported to undergo autophagy by way of AMPK activation, as indicated by increased phosphorylation of AMPK (Thr 172), ACC (Ser79), and ULK1 (Ser555).4 In earlier publications colon cancer SW620 cells were reported as apoptolidin A sensitive, while A549 cells appeared apoptolidin A insensitive when evaluated in the NCI-60 cell line.2 Using phospho-flow, we monitored cell uptake of Cy3 apoptolidins and phosphorylation of ACC after short-term (1 h) treatment with vehicle (DMSO) or apoptolidins (Figure C-3). We examined the response of healthy PBMCs, glioblastoma (U87 and LN-229), lung adenocarcinoma (A549), and colorectal (SW620) tumor cell lines to Cy3 apoptolidins A and H.

After 1 h of treatment, healthy PBMCs and all four cancer cell lines showed almost complete (>98%) uptake of Cy3 apoptolidins A and H (x-axis). Cancer cells showed higher Cy3 signal compared with healthy PBMCs, suggesting greater uptake of Cy3 apoptolidin A and H, corresponding to imaging data in Figure C-2.

We also measured phosphorylation-specific ACC (p-ACC; y-axis), a marker indicative of autophagy (*406*). In all cancer cell lines, we observed a proportion of cell subset that showed high p-ACC signal at baseline (DMSO treatment). LN229 glioblastoma cells showed the highest increase in the abundance of this subset, from 2.74% (DMSO) to 12.39% and 15.28% after Cy3 apoptolidin A and H treatments, respectively. The majority (>99%) of p-ACC expressing LN229 cells after apoptolidin treatments were among the Cy3-positive cells, or cells that had Cy3 apoptolidin-A and H uptake. In contrast, healthy human PBMCs did not show an increase in p-ACC expression in response to apoptolidin treatments.

A549 and SW620 cells showed only a minimal increase in abundance of cells expressing p-ACC after treatment with Cy3 apoptolidin-A and H, suggesting that A549 and SW620 cells were relatively insensitive to apoptolidins than LN229 cells at 1 h. However, U87 cells showed no increase in p-ACC activity after short-term treatment with apoptolidins A/H.

In summary, we have shown the utility of fluorescently labeled glycomacrolides (apoptolidins) in monitoring cellular uptake and response, using confocal microscopy and fluorescent flow cytometry. All human cell types tested showed more than 98% uptake of apoptolidins after 1 h of treatment. LN229 cells responded with a markedly increase in p-ACC expressing cells, suggesting their sensitivity to apoptolidins. Even though the responses were not as striking, A549 and SW620 cells also showed minimal increase in p-ACC after 1-hour of

apoptolidin treatment, whereas healthy human PBMCs and U87 cells did not. Additional kinetics studies would be required for in-depth characterization of the effects of apoptolidins in other cell types. While the current work was conducted using cell lines, we anticipate this approach will be very useful in the study of heterogeneous cell populations using single cell proteomics as this will provide quantitative data on healthy versus tumor cell small molecule uptake and response.

Conflict of interest

The authors declare no conflict of interest.

APPENDIX D

Discovery of Human Cell Selective Effector Molecules Using Single Cell Multiplexed Activity Metabolomics

Authors: David C. Earl, P. Brent Ferrell, Jr., Nalin Leelatian, Jordan T. Froese, Benjamin J. Reisman, Jonathan M. Irish, and Brian O. Bachmann

Data presented in this chapter was published in *Nature Communications* 2018 (Earl, Ferrell, Leelatian et al. 2018, open access).

Preface

This work presents an application of fluorescence cell barcoding (FCB) for high-throughput drug screening. Acute myeloid leukemia (AML) patient samples were subjected to *ex vivo* treatments with crude biological extracts from bacteria that were separated using reverse-phase liquid chromatography. This platform was coupled with a 48-plex FCB format that is described in detail in Appendix B. This array led to a rapid, high-throughput, discovery of novel compounds that differentially targeted leukemic blasts and non-malignant blood cells.

Fluorescence flow cytometry-based high-throughput drug screening can be tailored based on the highly flexible layout of FCB. This work is an excellent example of how coupling phosphorylation-specific flow cytometry with FCB can effectively and accurately characterize the cell subset-specific mechanisms that occurred as a result of treatments with biological extracts, now identified as potential novel therapeutics. Additionally, the liquid chromatography was shown here to be an excellent approach for first-pass cytotoxic screening of crude biological extracts, bypassing the need for laborious compound isolations. This technique pairs exceptionally well with FCB and could be the ideal initial screening platform for future drug discoveries.

Abstract

Discovering bioactive metabolites within a metabolome is challenging because there is generally little foreknowledge of metabolite molecular and cell-targeting activities. Here, single-cell response profiles and primary human tissue comprise a response platform used to discover novel microbial metabolites with cell-type-selective effector properties in untargeted metabolomic inventories. Metabolites display diverse effector mechanisms, including targeting protein synthesis, cell cycle status, DNA damage repair, necrosis, apoptosis, or phosphoprotein signaling. Arrayed metabolites are tested against acute myeloid leukemia patient bone marrow and molecules that specifically targeted blast cells or nonleukemic immune cell subsets within the same tissue biopsy are revealed. Cell-targeting polyketides are identified in extracts from biosynthetically prolific bacteria, including a previously unreported leukemia blast-targeting anthracycline and a polyene macrolactam that alternates between targeting blasts or nonmalignant cells by way of light-triggered photochemical isomerization. High-resolution cell profiling with mass cytometry confirms response mechanisms and is used to validate initial observations.

Introduction

A metabolome is the sum of primary and secondary metabolites produced by an organism in its environment. Constitutive metabolites are capable of interacting intra and extracellularly with receptors and active sites within DNA (*426-428*), RNA (*429, 430*), and proteins (*431, 432*), and metabolites are therefore close partners in growth, homeostasis, and signaling in heterogeneous environments (*433-437*). Chemical communication mediated via the inventory of an organism's cellular metabolites therefore defines an important molecular axis of interaction within and between organisms (*438*). Tapping into this communication system has become a central empirical activity in chemical biology and has repeatedly illuminated molecular solutions to problems with significant clinical relevance, such as the discovery of new bioeffector antibiotics and chemotherapeutics (*439, 440*). Tools to map novel bioeffector molecules to functional roles in responding cells —i.e., to identify bioeffector mechanism class—have been adapted into single-cell assays (*10, 11*) that stratify clinical outcomes and predict treatment responses (*12, 13, 15, 80, 375, 376*). Together with cellular barcoding (*344, 346*) and single-cell chemical biology assays (*345*), the recent advances in cytomics have raised the

exciting possibility of undertaking personalized metabolomic response profiling and bioeffector mechanism class identification in primary human tissue biopsies obtained for clinical research (*11, 284, 441*).

Despite the centrality of metabolite functional analysis, the development of a generalizable omics-scale solution for uncovering the functional roles of secondary metabolites within disease-relevant cellular contexts remains a substantial challenge (*442*). It is now possible to convert biological extracts (e.g., of microbial culture, plant/tissue origin) into highly characterized chromatographic microtiter arrays by split flow liquid chromatographic mass spectrometry (*443*). The biological characterization of such untargeted metabolomic arrays results in the generation of “bioactivity chromatograms”, and correlation analysis to matched extracted ion current (EIC) mass chromatograms identifies candidate metabolites linked to measured bioassay targets. However, per-well single-assay modalities greatly limit the efficiency of this approach, and targeted biochemical assays or phenotypic assays against cell lines reveal only a fraction of significant roles of metabolites in arrays.

Signaling profiles of primary cancer cells measured using phospho-specific flow cytometry (phospho-flow) have been shown to stratify the outcome of acute myeloid leukemia (AML) (*12, 80*) and B cell non-Hodgkin’s lymphoma (*13, 15*) based on signaling network responses to environmental cues, such as cytokines. Single-cell chemical biology assays have also been developed for fluorescence cytometry (*345*) and mass cytometry (*300*) to characterize pathway and cell-type-specific responses to small molecules. Fluorescence cytometry has the advantage of cellular throughput and more robust barcoding potential (*359*), whereas mass cytometry has the power to track more than 35 key markers of AML cell phenotype and function simultaneously (*12, 290, 374*). These assays rely on cellular barcoding to multiplex a large number of variables representing stimulation conditions, compounds, dosages, or timepoints (*344, 346, 400*). Such cytomic approaches are further strengthened by recently developed computational tools to reveal and characterize changes in cell subsets (*290, 314, 321*).

Here, a combination of 1) phospho-flow, 2) single-cell chemical biology, and 3) cellular barcoding was matched with 4) metabolomic arrays to identify natural product secondary metabolites that specifically target primary human leukemia cells and spare adjacent nonmalignant immune cells. This activity metabolomics platform is termed multiplexed activity metabolomics (MAM) and comprises a system for single-cell metabolome-scale analysis of bioactivity using human cells from primary tissue biopsies in a high-throughput screening-

compatible microtiter format. This untargeted assay modality samples a cross section of biological responses in a heterogeneous mixture of cells representing an in vivo human tissue environment and has the potential to identify disease-relevant bioactive metabolites within metabolomic arrays.

Results

Cytometry-enabled MAM platform

The MAM workflow (Figure D-1) first generates a metabolomic array in microtiter plate format via reversed-phase liquid chromatographic separation of a crude biological extract produced by a “stimulus” organism. A portion of the effluent is diverted to a polarity-switching electrospray mass spectrometric analyzer (ESI-MS) and the remainder of the effluent to a microtiter plate fraction collector after passing through a UV/Vis diode array detector. Following evaporation and resuspension of collected fractions, cell preparations from a “response” organism (e.g., humans, represented by tissue cells) are added to the microtiter wells for incubation with the metabolomic fractions to induce cellular responses (Figure D-1a). Cells within wells are then stained for viability, fixed, and permeabilized, and fluorescent cell barcoding (FCB, Figure D-1b) is used to label the well contents via differential staining of cells with N-hydroxysuccinimide (NHS) ester-functionalized fluorescent dyes. Thus, “barcoded” cells in the microtiter wells are then pooled and stained with multiple fluorescent antibodies to quantitate cell status and targeted cell-type-specific responses to metabolites. Critically, flow cytometric gating based on the barcoding fluorophores facilitates the assignment of cells to their original coordinates on the microtiter plate metabolite array (i.e., “deconvolutes” treatment conditions for each cell), yielding simultaneous bioassay marker quantitation per well for each targeted antibody–fluorophore conjugate. Barcoding enables high-throughput antibody assays by using a fraction of antibody reagents compared to a microtiter format, and pooling also ensures the uniformity of antibody staining of cells across all wells, decreasing experimental variation. The result is a multiplexed series of well coordinate-linked immunoassay profiles running through the metabolomic fraction array. Metabolomic features (positive and negative m/z, UV, and retention time), which are generated via EIC analysis of all significant detected metabolites, are then manually or automatically correlated with bioactivity features, which are peaks generated from bioassay profiles across wells.

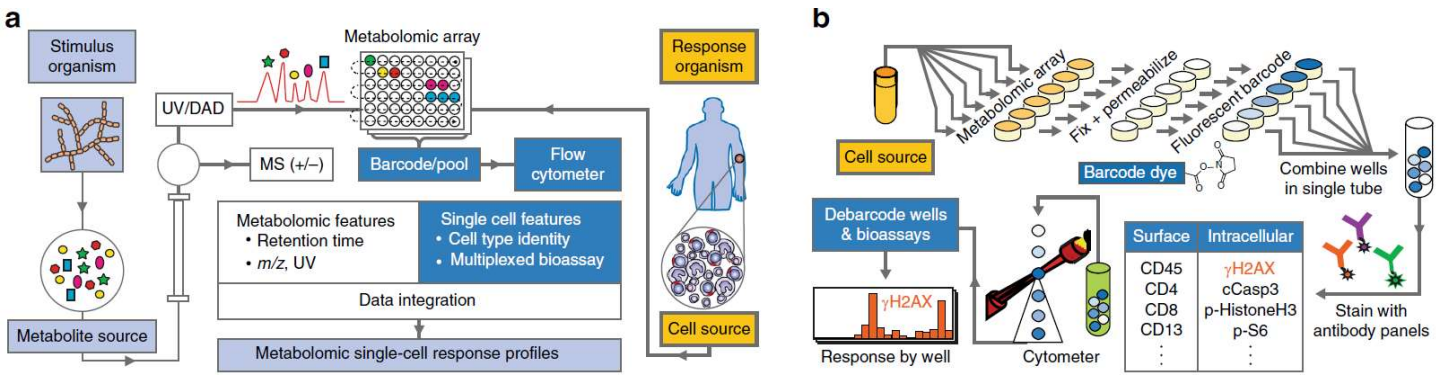


Figure D-1 Metabolome functional analysis by multiplexed activity metabolomics (MAM). a High-data-content metabolomic arrays are generated in replicate from a "stimulus" organism via split-flow polarity switching chromatography mass spectrometry. A suspension of disaggregated tissue cells from a "response" organism (human) is added to the metabolomic array. b Flow cytometric cell barcoding and multiplexed immunoassays are used to identify multiple cell type/subtype-specific biological responses to metabolites in the array. Correlation analysis of the resulting bioactivity and UV/ESI/MS(±) data generate putative functional activities for metabolites. An example of MAM with a natural product producing actinomycete and a cell preparation derived from an AML patient including, but not limited to lymphocytes, monocytes, and leukemia blasts, as well as their subtypes is illustrated here.

To maximize available fluorescence channels for multiparameter flow cytometry, FCB was adapted to barcode 48 wells with two fluorescent NHS ester-activated dye gradients of NHS-Pacific Orange and NHS-Pacific Blue. After two-dimensional barcoding, wells were pooled into a single tube and stained with fluorescently tagged antibodies. To test the robustness of the FCB assay, Kasumi-1 cells were incubated in 48 wells in a checkerboard fashion with vehicle dimethylsulfoxide (DMSO) or one of two benchmark natural products: the podophyllotoxin derivative etoposide (*444*), a potent topoisomerase inhibitor and inducer of double-strand DNA breaks, or the bisindole alkaloid staurosporine (*445*), a classical inducer of apoptosis. After treatment, cells were stained with a permeability/viability indicator, Alexafluor 700 (Ax700) (*446*), barcoded, combined into a single tube, and then stained with fluorescently labeled antibodies specific to either cleaved caspase-3 (cCasp3), a protein activated in apoptosis (*447*), or γH2AX, a histone phosphorylated during genomic damage (*448, 449*). A representative workflow and data for etoposide are shown in Figure D-2. Analysis of single-cell events revealed two populations for each readout, and biaxial plots of Pacific Orange vs. Pacific Blue yielded 48 distinct populations (Figure D-2a). Recovery of the well coordinates and determination of antibody binding in debarcoded populations was accomplished using Cytobank, a cloud-based cytometric analysis platform, to confirm compound-specific effects (Supplementary Figure D-1a, b). In the case of etoposide, gating for γH2AX and then

debarcoding illustrated the bifurcated response within the checkerboard (Figure D-2b), and biaxially gated percent changes reflect bioassay results (Figure D-2c). Comparable results were obtained for staurosporine (Supplementary Figure D-1c), and these results were used to calculate the standard deviation of each assay plate, which conformed to levels in standard practice in high-throughput screening analysis (Z factor>0.77) (Figure D-2b and Supplementary Figure D-1c). As an additional evaluation of barcoding, cCasp3 and γH2AX expression induced by staurosporine and etoposide, respectively, was demonstrated to be dose dependent within assay conditions by separate cytometric barcoding of concentration response curves and quantitating antibody binding (Supplementary Figure D-1d).

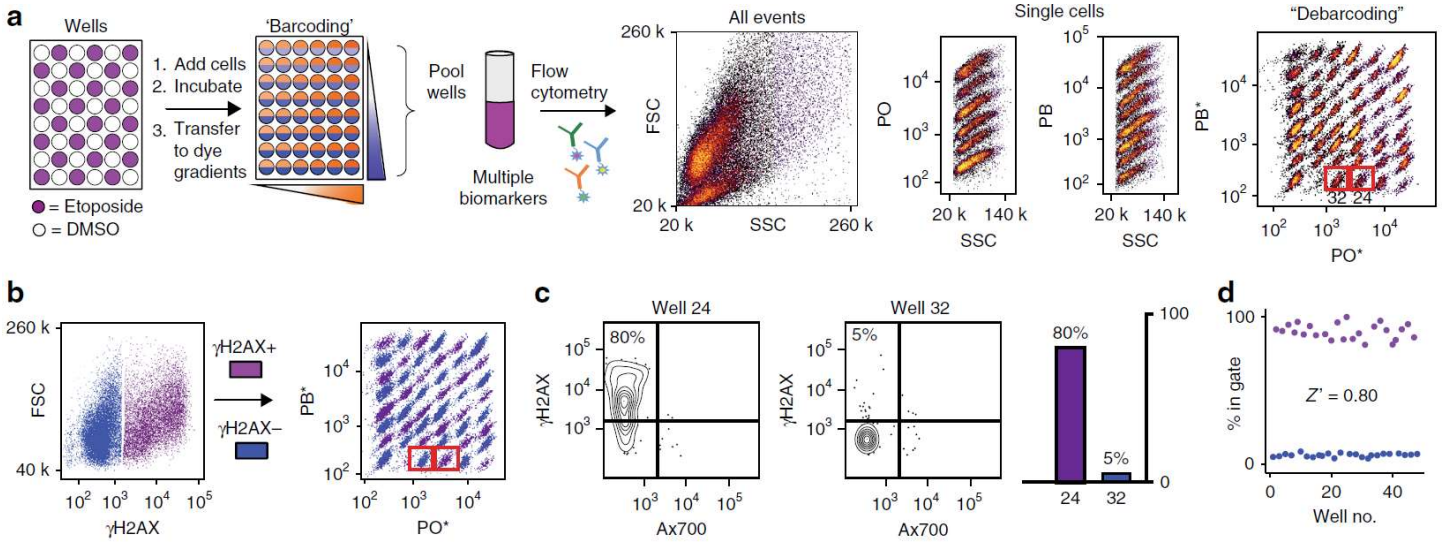


Figure D-2 Validation of a 48-well single-cell chemical biology assay for DNA damage response mechanism class. Design of a checkerboard validation experiment using Kasumi cells and a DNA-active natural product. a Overview of 48-well fluorescent cell barcoding and debarcoding validation. Compounds and vehicle are added in a checkerboard pattern to 48 wells, and cells are added and incubated prior to being barcoded using dye gradients of N-hydroxysuccinimide functional Pacific Orange (PO) and Pacific Blue (PB). Cells are stained with Ax700, fixed, permeabilized, and pooled prior to immunoassay with antibodies tagged with nonoverlapping fluorescent dyes. Cells are analyzed by flow cytometry and gated selecting (i) intact, single cells, (ii) PO to reveal columns, and (iii) PB to reveal rows and generate populations for each well. PO* and PB* denote dye intensity after correcting for cell morphology using a protocol from ref. 67. A data set using Kasumi cells for an etoposide checkerboard assayed with γH2AX binding, which is a marker for DNA damage-associated activity is shown here. b Graphical representation of well coordinates generated from pooled cells by gating on forward scatter and γH2AX and then debarcoding as above to reveal the checkerboard pattern. c Biaxial plots of individual wells representing a condition, as in Figure D-1a. d Percent γH2AX positive for each well was calculated using a cutoff of 103.2 to determine statistical effect size (Z′). A similar demonstration with etoposide vs. cCasp3 is shown in Supplementary Figure D-1.

The integrated analysis of a high-performance liquid chromatography (HPLC)/MS-generated chromatographic array in conjunction with FCB and cellular response data (MAM) was validated using a chemically defined mixture of bioactive compounds. A mixture of six structurally and mechanistically diverse cytotoxic small molecules was chromatographically arrayed and assayed against a human myeloid leukemia-derived cell line (KG1) using the MAM platform. EIC chromatograms for the six compounds can be readily compared to bioactivity chromatograms and demonstrated specific and mechanistically expected responses to multiplexed immunoassays (Supplementary Figure D-2a). For instance, the EIC peak for the known apoptosis-inducing secondary metabolite staurosporine (m/z = 467.5), was the highest correlating peak in the well 25 bioactivity bin for cCasp3. Similarly, the largest response for γH2AX occurred in well 20, matching the retention time of the potent topoisomerase inhibitor etoposide (m/z = 606.5). Of note, in this experiment of modest complexity, a single cytometric flow run generates an aggregate of 240 individual raw immunoassays, which may be further combined into additional function assays that can be compared to arrayed compound elution profiles. Importantly, MAM successfully identified and differentiated compounds in a mixture based on their elution profile and differential response to a multiplexed antibody panel.

MAM finds bioactive metabolites within complex extracts

The identification of bioactive molecules within complex cellular (e.g., microbial) metabolomes using MAM requires that barcoding and bioassay cytometric measurements be stable to potential interferences present under typical secondary metabolite-producing conditions, such as soluble extractable cellular metabolites, cell wall components, and spent growth medium species. The robustness of MAM was therefore tested by fractionating and analyzing a concentrated methanolic microbial extract generated from a Streptomyces strain grown in complex media and spiked with etoposide and staurosporine prior to chromatography. Prior to spiking, the extract possessed no measurable bioactivity. After fractionation, wells were evaporated, and KG1 cells were added to the plate and incubated for 16 h. Subsequent to fixation and permeabilization, cells were barcoded, pooled, and assayed using antibodies against γH2AX and cCasp3. Bioactivity chromatograms for these markers were generated from the debarcoded data set and formatted for correlation analysis. Cells were effectively assigned as distinct populations to the 48 wells according to the dye-gradient selection, demonstrating no

cytometric interference with FCB from extract components. Moreover, as shown in Supplementary Figure D-2b, lots of median fluorescence intensity for cCasp3, and γH2AX expression per debarcoded population generated bioactivity chromatograms for correlation analysis. Importantly, although the EIC abundance of staurosporine and etoposide was below the threshold of the average intensity of the TIC (Supplementary Figure D-2c), both were the highest Pearson-correlating components in the bioactive fractions. Despite the presence of high-abundance products of cellular metabolism and media components, no additional potent cCasp3 or γH2AXmodulating activities were observed in the test metabolome.

Finding metabolites with anticancer activity in human tissue

In addition to quantitating intracellular events and cell status immunomarkers, single-cell characterization via cytometry facilitates the differentiation of cell types within heterogeneous mixtures based on cell size, shape, complexity (via differential light scattering), and the detection of cell-type-selective surface markers (*399, 450*). This enables characterization of the ways in which the components of metabolomic arrays affect molecular phenotypic changes in mixtures of cells, including primary cell preparations that more closely approximate a native cancerous microenvironment than pure immortalized cell lines. Acute myeloid leukemia patient bone marrow samples were selected as an advantageous system for MAM due to their beneficial cytometric properties and clinical significance. AML remains a deadly adult cancer, and treatments have not greatly improved the 5-year overall survival rate, which is 21.3% overall and remains under 5% for patients who are 65 and older (*451*). Bone marrow biopsies that are routinely obtained from patients being treated for AML contain a complex mixture of multiple cancer and normal cell types. These tissues are fully “suspended”, require minimal processing (e.g., disaggregation) for cytometric analysis, and contain a mixture of cell types representative of in vivo therapeutic contexts. Cytometric characterization of AML via immunophenotyping is widespread in the diagnosis and management of this disease, providing a strong basis for biomarker selection and analysis.

To test the ability of MAM to assess the effects of a bioactive metabolomic array against a heterogeneous cell mixture, microbial metabolomic arrays were incubated with cell preparations derived from AML biopsy samples from two separate patients. The patient samples used in this experiment represent two common underlying genetic mutational profiles occurring in AML. Patient 001 was a 23-year-old female with a gene

translocation (MLL-MLLT3) correlated to intermediate prognosis but without other tested common molecular mutations. Patient 015 was a 68-year-old male with the FMS-like tyrosine kinase 3 internal tandem duplication (FLT3-ITD) strongly associated with poor prognosis (*374*), but with otherwise normal cytogenetics. Subsequent to aspiration from bone marrow, red blood cells and platelets were removed from the patient samples via density gradient separation, resulting in bone marrow mononuclear cells (BMMCs) containing a mixture of heterogeneous AML blasts and nonmalignant myeloid and lymphoid cells and their progenitors (*452*). These heterogeneous mixtures served as the response organism system for multiplexed cellular and biochemical analysis. For the microbial metabolomic array source organism, we selected an actinomycete strain designated Streptomyces specus that we had isolated from Blue Springs cave in Sparta, Tennessee. S. specus was of particular relevance, as it had been observed via depreplication analysis of HPLC/MS data to produce a family of anthracycline natural products related to the clinically employed AML drug daunorubicin, including baumycins, unusual natural acetal functionalized congeners (*436*), and related anthracycline functional metabolites with apparent masses not previously reported.

In the case of the S. specus metabolomic array interacting with AML biopsy samples, and gating for the three major cell types, the strongest bioactivity was observed in the viable (Ax700−) and γH2AX+ and cCasp3+ subsets in both patients (Figure D-3, Supplementary Figure D-3). The sample derived from Patient 015 contained the three readily discernable subpopulations of leukemia blasts, myeloid cells, and lymphocytes, which could be separately debarcoded to yield defined cell-type response profiles in each well (Supplementary Figure D-3a). Patient 001's sample was comprised of predominantly leukemia blasts and lymphocytes. Individual biaxial plots (Supplementary Figs. 4–5) were used to generate well bioactivity profiles based on set positivity thresholds. A subset of these bioactivity chromatograms is shown in Figure D-3d and indicates the presence of bioactive molecules in the S. specus extract, which can be preliminarily identified via comparison of EIC to bioactivity profiles. Bioactivity profiles represent averages of thousands of single-cell measurements and are highly reproducible between biological replicates (Supplementary Figure D-3b). Two observations result from this data set. First, apparent cell-type selectivity was demonstrated for several features in the metabolomics array. For instance, a 2.5- and 47-fold increase in selective blast-targeting bioactivity vs. leukocytes was observed eluting in wells 17 and 24, respectively. Second, patient-specific activities were evident in bioactivity profiles.

Examination of HPLC/MS and bioactivity profiles of well 17, containing the most abundant m/z = 442, revealed a 30-fold increase in apoptosis and fivefold increase in DNA damage in patient 001 vs. patient 015. A similar trend was observed in later eluting wells containing anthracycline chromophores. Notably, patient 015 possessed the FLT3-ITD phenotype, which is an internal tandem deletion in kinaseencoding gene FLT3 demonstrated to confer resistance to anthracyclines (*374*). Therefore, these observations are consistent with substantially enhanced resistance to anthracyclines in patient 015.

To validate bioactive features putatively identified using MAM, most abundant correlating mass features were isolated. The most potent bioactivity peak was observed in well 21 and correlated to the most abundant eluting anthracycline (m/z = 674), termed specumycin A1. Specumycin A1 was isolated in scale-up fermentations and its structure was determined by multidimensional nuclear magnetic resonance experiments (Supplementary Table D-2). The planar structure of specumycin A1 is identical to the structures of baumycin A1/2, which contain an unusual acetal appending the 3′-O-methyl on the daunosamine sugar (*415*). The next most abundant feature, and the primary feature in well 20, was specumycin B1 (Supplementary Table D-3), a previously unreported 11-deoxy congener. Specumycin B1 was observed to be as active to A1 under assay conditions but threefold less abundant, suggesting a potentially more potent congener. Comprehensive isolation of low-abundance bioactive species is beyond the scope of this study. However, cell-type and patient-specific responses identified by MAM, such as bioactive metabolites demonstrating enhanced activity against a FLT3-ITD AML sample and selective activity for leukemia cells (e.g., well 24, m/z = 1054, Figure D-3), demonstrate the potential of this platform for performing preliminary analysis and prioritization of activity differences within a natural product family for common AML subclasses.

The cell-targeting potential of secondary metabolites within metabolomic arrays in the anthracycline-resistant phenotype sample (015), was further explored employing the soil actinobacterium Nocardiopsis sp. FU40 as a source organism. This strain produces a family of bioactive compounds called apoptolidins (A–H) (*453*), which are cytotoxic glycosylated macrolides, and a pair of cytotoxic glycosylated polyene macrolactams, ciromicins A and B (*454*). Thus, a metabolome array was generated from Nocardiopsis sp. FU40, AML Patient 015-derived anthracycline-resistant cell preparations were incubated with the array and subjected the samples to MAM analysis. In Figure D-4 a selection of bioactivity profiles generated from this single data set is shown,

indicating how the arrayed metabolome is obtained from Nocardiopsis sp. FU40 can be mined for bioeffectors that have selective activity against different cell types present in an AML patient. For instance, apoptolidins selectively induced caspase-dependent apoptosis in lymphocytes, whereas ciromicins induced apoptosis most prominently in leukemic cells. Similarly, apoptolidin A induced γH2AX, apoptosis, and decreased p-Histone H3 signaling selectively in lymphocytes, and ciromicins induced more DNA damage in monocytes and blast cells. Taken together, these data demonstrate the identification of differential cell targeting of secondary metabolites against primary cell mixtures in the background of an extracted microbial metabolome. In contrast to the specumycins, ciromicins demonstrate potent selectivity for blasts in comparison to lymphocytes in the anthracycline-resistant phenotype.

Light-modulated secondary metabolite cell-type targeting

Expansion of the bioactivity chromatogram in the region of ciromicin elution revealed that the isobaric metabolites ciromicin A and ciromicin B were resolved into separate wells with strikingly distinct biological phenotypes. Specifically, ciromicin A displayed maximal apoptosis markers in leukemia blast cells, whereas its photoisomerization product ciromicin B stimulated monocyte apoptosis. We recently reported the discovery of ciromicins in an apoptolidin polyketide synthase knockout strain of Nocardopsis sp. FU40, and demonstrated that ciromicin B is the product of an unexpected visible light-initiated 12-π electron photoisomerization of ciromicin A58. The identification of ciromicins here in the wild-type monoculture of Nocardopsis sp. FU40 was surprising because it is produced in low levels in the wild-type strain. Thus, the sensitivity of the MAM platform using primary cells was capable of effectively identifying the bioactivity of this low abundant secondary metabolite family, and the modest resolution of 48-well binning of fractions was sufficient to resolve bioactivities of closely eluting species.

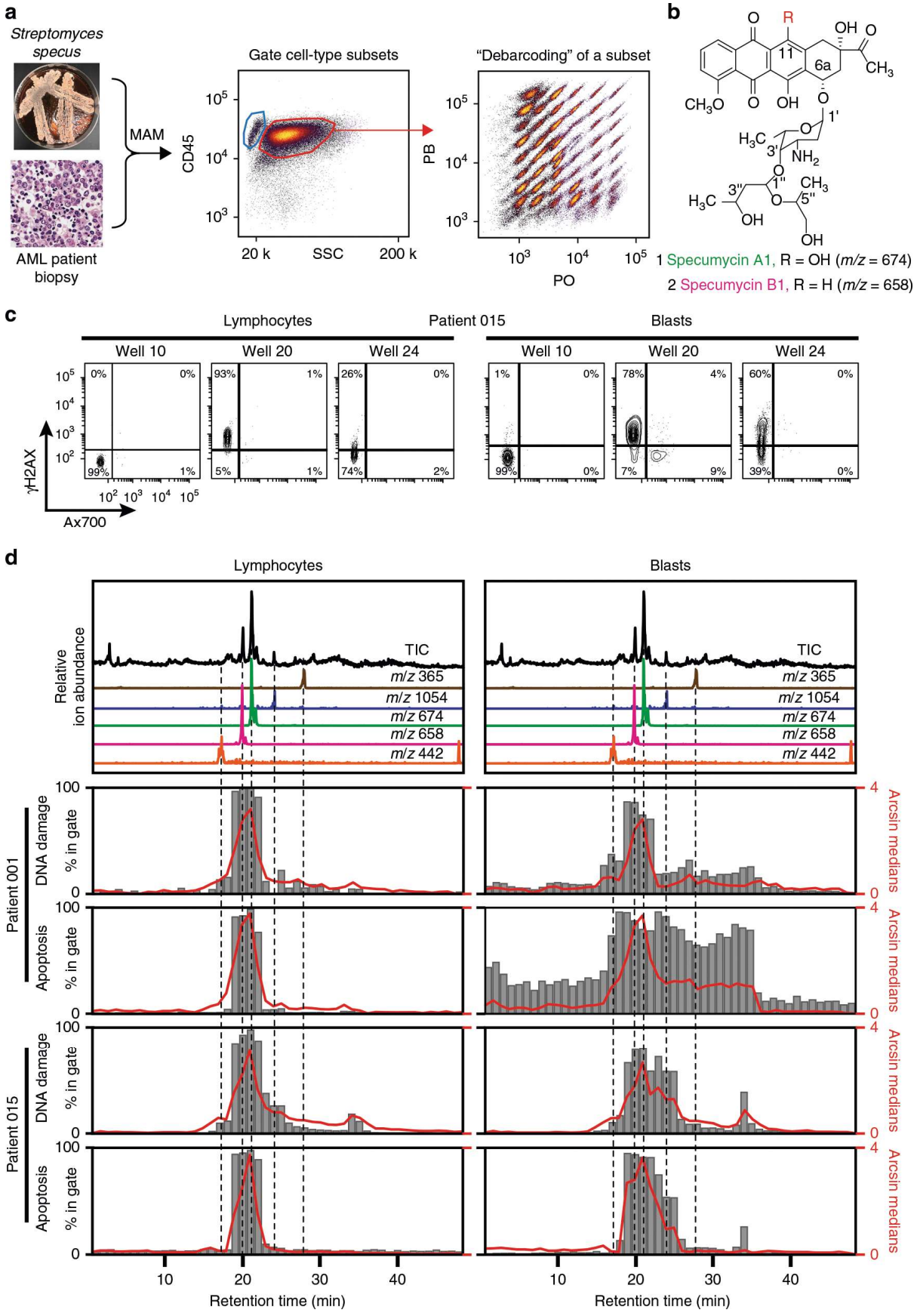


Figure D-3 A structurally novel acetal-functional anthracycline selectively targets leukemic blast cells and not nonmalignant lymphocytes within a human bone marrow biopsy. a Chromatographic arraying of Streptomyces specus

was performed using split flow HPLC/UV/MS with polarity-switching mass scanning resulting in an array of highly characterized fractions from a crude extract of the baumycin producer S. specus. Primary cell preparations were prepared from AML patient biopsy. The metabolite array was incubated with heterogeneous cell samples, and the cells were then viability stained, fixed, barcoded, and stained with antibodies for biomarker and surface marker expression. Standard gating was performed using biaxial plots of CD45 expression vs. SSC was used to determine cell-type subsets (blue gate: lymphocytes, red gate: blasts) that were each individually analyzed for γH2AX and cCasp3 expression. b Structures of specumycin A1 and B1. c Biaxial plots of selected wells gated for lymphocytes and leukemia blasts. d Total ion current and selected extracted ion currents of metabolites within the metabolome of S. specus correlating to bioactive wells from assays against two patient samples. Bar graphs of percent of cells in the upper left quadrant of marker/viability gate (marker positive and viable cells). The solid red line is the arcsinh transformed median of the marker. For patient 015, the average cells collected per well were 3996 blasts (minimum 240) and 764 lymphocytes (minimum 66). For patient 001, the average collected per well was 1655 blasts (minimum 189) and 53 lymphocytes (minimum 17).

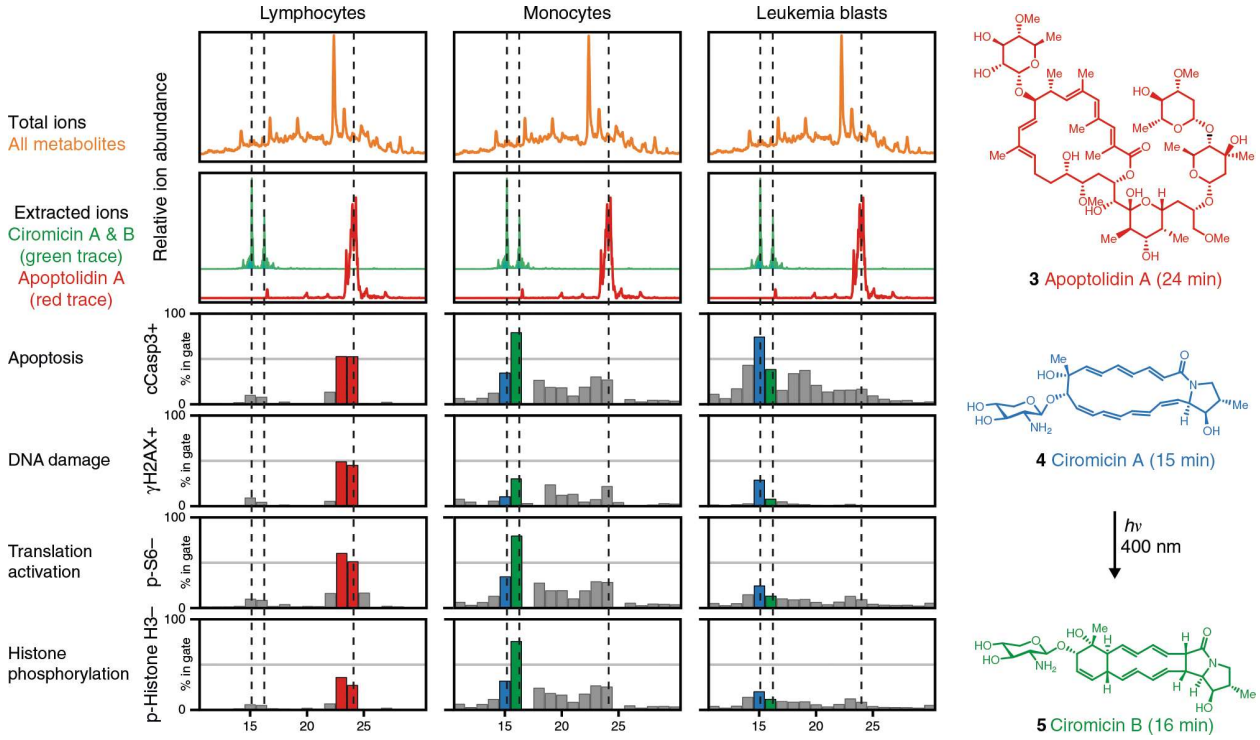


Figure D-4 An optochemical cell selectivity switching natural product in Nocardiopsis revealed by primary cell MAM. The top metabolome row (orange) shows total ion current and extracted ions for ciromicins A and B (m/z = 515, 15, and 16 min) and apoptolidin (m/z = 1129.5, 23.4 min) with their elution times shown in dotted lines. The next four rows show selected bioactivity chromatograms from a single flow experiment, which were generated by adding an aspirated preparation of bone marrow mononuclear cells from an AML patient, barcoding, immunostaining, and debarcoding. The immunostaining panel contained CD45 (leukocyte-common antigen), cCasp3 (cleaved caspase), γH2AX (DNA damage), p-Histone H3 (cell cycle marker upregulated during mitosis), and p-S6 (marker for active translation). The average collected per well was 5400 blasts (minimum 209), 750 monocytes (minimum 19), and 291 lymphocytes (minimum 20). Histograms for each marker for highlighted wells are shown in Supplementary Figure D-6

The discovery of putative cell-type-specific cellular responses to ciromicin isomers using MAM may be considered as a primary screening “hit”, describing a multidimensional response of a metabolomics fraction with associated correlation coefficientranked metabolite features. To validate the unusual photochemically triggered modulation of primary cell selectivity, pure ciromicins A and B were isolated from scaled-up cultures and assayed against the same biopsy sample using an enhanced panel of 29 cell surface markers that classifies all myeloid cell populations when paired with unsupervised machine-learning tools. The viSNE algorithm, which allows robust identification of both nonmalignant and leukemia cell subsets (*452*), was applied to data sets collected by mass cytometry after a 48-h treatment of patient-derived BMMC with ciromicin A, ciromicin B, or DMSO. In Figure D-5a viSNE maps showing the overall changes in the cellular landscape of the primary BMMC after this treatment are shown. Proximity in viSNE space corresponds to similarity in cell-type identity, while differences in immunophenotype drive separation of cells (dots) on a viSNE map. To quantify the overall shifts in cellular subsets, gates were drawn on the viSNE map corresponding to prominent populations based on abundance (Figure D-5a). The relative abundance of phenotypically distinct cell subsets present in different treatment conditions and the enriched features of these populations were characterized (Figure D-5a). Per cell marker expression and median values allowed assignment of cellular identity to populations (Supplementary Figure D-7). Changes in the relative abundance of each population demonstrated that photoisomerization polarizes overall cellular immunophenotype within leukemia cells. Based on subset gating within viSNE, ciromicin B reduced the relative abundance of leukemia stem cells and hematopoietic stem cells and smaller blast subsets, while ciromicin A reduced the largest blast subset (subset 9, Figure D-5a). Both isomers had comparatively little impact on lymphoid cells, which were comprised largely of CD4+ and CD8+ T cells (subsets 2 and 4, Figure D-5a and Supplementary Table D-1). Next, marker enrichment modeling (MEM) was used to characterize feature enrichment in comparison to a population of CD34+CD38−Lin− cells within the leukemia sample. MEM identified changes in populations between ciromicin A and ciromicin B, including greater enrichment for CD13, a marker of myeloid differentiation, and CD43, or sialophorin, which is commonly expressed on more mature myeloid cells such as granulocytes and monocytes, after treatment with ciromicin A within a major leukemia population (subset 9, Figure D-5b). Overall, the pattern of cell-type selectivity expanded the depth of the initial screen, validating MAM’s ability to identify selectively bioactive compounds.

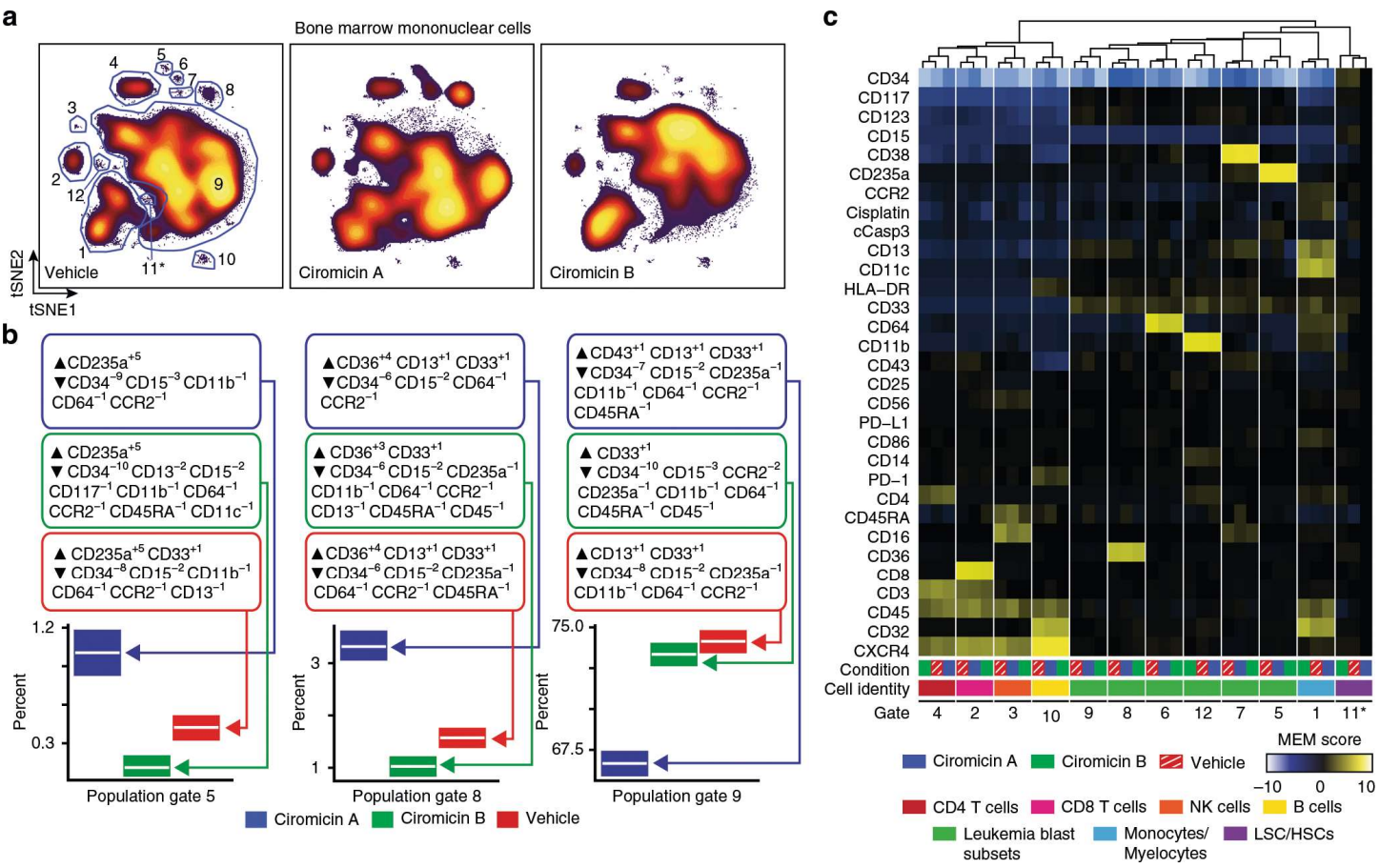


Figure D-5 Photochemical isomers ciromicins A and B selectively target different cell subsets within the heterogeneous mixture of patient biopsy cells. Mass cytometry uses DOTA-chelated metals detected by inductively coupled plasma mass spectrometry (ICP-MS) to eliminate spectral overlap, expanding the feature range to 29 antibody-quantified features per cell. a viSNE maps of 20,000 individual cells from each condition are organized according to differences in their surface marker profiles for each treatment condition. b MEM labels for three blasts subsets and plots of population prevalence with observed prevalence in white and 95% binomial confidence interval represented by a box. c Marker enrichment modeling (MEM) was used to characterize major populations within the samples and highlight differences in marker expression relative to a gated population of phenotypic hematopoietic stem cells (gate 11). Heat maps of hierarchical clustered MEM labels reveal subsets specific differences and cellular identification. Heat maps of median marker expression are shown in Supplementary Figure D-7. Dose–response data against PMBCs from a healthy donor and patient sample 015 are shown in Supplementary Figure D-8.

Discussion

Cancer is challenging to study and to therapeutically manipulate, due in part to the complexity of cell signaling processes affected by pharmacological interactions, and system heterogeneity as seen in the polyclonal nature of cancer cells, the complexity of the supporting stroma, and the infiltrating immune cells (*11, 80*). MAM as implemented here provides a generalizable system to link metabolomic feature data from one

organism or system to functional targets or their causally related networks within another heterogeneous cellular environment. A key feature of the cytometric strategy underpinning MAM is its ability to analyze heterogeneous mixtures of cells, which more closely approximate a native cellular milieu than immortalized cell lines, using multiplexed markers of cell status and type. Specifically, MAM was employed here to study the interkingdom interactions of metabolomes of two secondary metabolite-producing soil bacteria with primary cell preparations from two phenotypically distinct patient-derived AML cell samples. The combination of metabolomics, single-cell biology, and cheminformatics used here identified biologically active secondary metabolites produced at low levels that mediate apoptosis, DNA damage, and cell signaling in a cohort of cells present in AML patient's bone marrow samples. In this primary cytological screen, differential activities were identified for secondary metabolites present within complex and concentrated microbial extracts.

The cave-derived bacterium S. specus is a producer of multiple compounds that share the anthracycline core of daunorubicin, which is used in combination therapy with nucleoside analog cytarabine as the standard of care in the treatment of AML, but differ in decorating glycosides (*455*). Specumycins described here are daunorubicin variants similar to baumycins, appended with an unusual acid-labile acetal moiety on the 3′-hydroxyl of daunosamine (*456*), and are reported to demonstrate broad cytotoxicity comparable to that of daunorubicin (*405*). Despite its clinical significance, daunorubicin is actually a low-abundance biosynthetic intermediate en route to baumycin in most daunorubicin producers, and is typically isolated by acid-catalyzed degradation of baumycin glycosides (*456*). However, though being the major product of most daunorubicin biosynthetic pathways, the potential role of the baumycin acetal moiety in cytotoxicity and cell targeting has remained untested prior to this study. Applying MAM to the metabolomic array of S. specus revealed activities of a spectrum of specumycin polyketides and related metabolites against divergent primary cell phenotypes. Along with the discovery of the previously unreported and more potent compound specumycin B1, these data suggest previously unnoticed potential for the 3′-acetal functional in AML anthracycline therapy. Validating the observed bioactivity trends, the two most abundant features demonstrated potent activity against leukemia blasts and leukocyte cells and were isolated and structurally elucidated. Numerous less abundant species displayed remarkable differential cell-type targeting between patients, suggesting an untapped potential for discovery of

more selective pharmacological agents within the anthracycline family in biosynthetically competent actinomycete strains.

Nocardiopsis sp. FU40 was selected as a subject for the MAM platform as its metabolomic array is complex, both in terms of the sheer number of apoptolidin analogs it produces (denoted A–H), and in its capacity to simultaneously produce polyene macrolactams and aromatic polyketides that were previously reported to possess moderate-to-potent cytotoxicity against cell lines. Applying MAM to test Nocardiopsis arrays against AML primary cell preparations successfully deconvoluted apoptolidins from ciromicins and revealed distinct cell-targeting phenotypes. Apoptolidin A and its isobaric analogs present in the extract (isoapoptolidins A and G) correlated to the most potent lymphocyte-targeting activity across all markers. The induction of cCasp3 in lymphocytes is consistent to prior studies of this compound performed in cell lines, which also present evidence in support of mitochondrial FoF1-ATPase targeting within this family (*457*). Other apoptolidin congeners are generally chromatographically dispersed from apoptolidin A, but did not display this degree of activity. Notably, the apoptolidins only nominally affected marker expression in leukemia blasts and monocytes, demonstrating how MAM readily identifies first-pass cell-targeting activity in primary tissue samples. The distinctive blast/myeloid cell-type targeting observed for ciromicins A and B was notable, and careful examination of their elution region revealed a remarkable switch of cell specificity between blast and nonblast myeloid lineages for the two compounds.

As a follow-up to using MAM as a primary assay for lead discovery, mass cytometry was used as a secondary validation and deep cell-profiling assay. A 29-marker mass cytometry panel was used to classify the cellular effects of purified ciromicins A and B on subsets in primary cell preparations. Mass cytometry revealed changes in differentiated immunophenotypic subsets and demonstrated that visible light-induced photoisomerization of ciromicin A to B induces wholesale shifts in cell-type targeting and indicating the importance of aglycone structure and geometry to the mechanism of action of this family of macrolactams. For instance, bicyclic Michael acceptor containing ciromicin A exerted its greatest influence on the largest subset of AML cells, whereas tetracyclic potentially less electrophillic ciromicin B targeted stem-like myeloid progenitors, a subset that may be beneficial to address in therapy (*458, 459*). Mass cytometry also revealed that ciromicins target leukemia blasts in a patient with an anthracycline-resistant leukemia phenotype and, unlike anthracyclines

in the previous study, have little negative effect on lymphoid cells. Finally, mass cytometric findings, performed in concert with MAM using patient samples, validated and provided a deeper profiling of bioactive compounds discovered. Overall, the multiplexed single-cell approaches used here represent a paradigm shift in comparison to typical discovery efforts using monoclonal immortalized cell lines or other research models that do not accurately reflect the cell diversity and composition of primary human tumors and leukemic tissues. That ciromicins A and B represent photoswitching natural products with distinct cell subtype-targeting phenotypes provides potential tools for investigating the pharmacology of this family and the effects of targeting subtypes. Notably, molecular photopharmacological switches currently find broad application toward understanding cellular function by leveraging the spatiotemporal control afforded by such compounds (*460*).

In summary, a general method is demonstrated for searching preliminary structure–activity relationships in secondary metabolite families in producing organisms, without the need for compound isolation, and provides insight into how bioactive lead compounds affect diseased and normal cell types in major patient phenotypes using clinical samples. Given that there are a limited number of distinct clinical subsets, automated cytometric analysis of untargeted metabolomic inventories against sets of relevant patient phenotypes provides a process for “personalized” natural product discovery. This is a proof-of-principle study of a viable drug discovery platform. In a full-scale implementation, cells derived from multiple patients, including cells derived from healthy individuals, would be necessary to realize the full scope of lead-compound preclinical assessment. While applied here for the case of identifying bioactive secondary metabolites within metabolomes, the MAM platform enables the discovery of cellular responses to molecular inventories, regardless of sources. Given the importance of all chemical communications in mediating life processes within and between organisms, a generalizable method for identifying functional roles for metabolites has significant potential in applications spanning a broad range of applications in cellular chemical biology.

Methods

Preparation of microbial crude extracts

Streptomyces strains were maintained on ISP2 agar (yeast extract 4 g/L, malt extract 10 g/L, glucose 4 g/L, and agar 20 g/L, pH 7.2). Loops of mycelia were used to inoculate 5-mL seed cultures in ISP2 medium

(yeast extract 4 g/L, malt extract 10 g/L, and glucose 4 g/L, pH 7.2) for Streptomyces strains, incubating them for 3 days at 30 °C. Seed cultures were then transferred to 250-mL Erlenmeyer flasks containing 25 mL of BA medium (soybean powder 15 g/L, glucose 10 g/L, soluble starch 10 g/L, NaCl 3 g/L, MgSO4 1 g/L, K2HPO4 1 g/L, and trace elemental solution 1 mL/L, pH 7.2) and grown for 7 days at 30 °C with shaking. Aqueous fermentation broth was extracted by shaking with Diaion HP20 synthetic absorbent resin (Alfa Aesar) (125 mL of HP20 bead/H2O slurry per 500 mL of aqueous broth) for 2 h. Fermentation broth was then centrifuged (3700 × g, 30 min) and the supernatant was decanted. Metabolites were eluted from absorbent resin and cells with methanol (250 mL of methanol/125 mL of HP20 bead/H2O slurry) by shaking for 1.5 h, followed by centrifugation (3700 × g, 30 min) and decanting of the methanol extract. Further extraction was performed with acetone (250 mL of acetone/125 mL of HP20 bead/H2O slurry) by shaking for 1.5 h, followed by centrifugation (3700 × g, 30 min) and decanting of the acetone extract. Nocardiopsis strains were cultured and extracted as previously described (*453*). Purified ciromicins A and B were isolated from co-cultures as previously described (*454*). Kasumi-1 and KG-1 cell lines were obtained from ATCC and identified using mass cytometry analysis of 35 myeloid proteins as reported previously (*374*).

Specumycin A1 and B1 isolation

Crude acetone extract was concentrated and fractionated with Sephadex LH-20 resin (GE Healthcare Bio-Sciences) with methanol as the eluent. Fractions were analyzed by analytical HPLC/MS, and fractions containing the compound(s) of interest were pooled and further purified by preparative HPLC (Waters, XBridge C18 Prep, 5 µM) (10 mL/min, 0–1 min: 100% solution A, 5 min: 85% solution A; 15% solution B, 65 min: 15% solution A; 85% solution B, and 70 min: 100% solution B) (Solution A = 95:5, H2O:MeCN, 10mM NH4OAc; Solution B: 5:95 H2O:MeCN, 10mM NH4OAc). In order to obtain analytical purity, fractions containing the compound of interest (34–35 min) were pooled and purified by flash column chromatography (98:2 CH2Cl2:MeOH to 95:5 CH2Cl2:MeOH). The structure of specumycins A1 and B1 was elucidated using a combination of mass spectrometry and two-dimensional nuclear magnetic resonsance spectroscopy data. Mass spectrometry data produced with electrospray ionization and collected in both positive and negative modes provided the molecular weight of specumycins A1 and B1. Correlated nuclear magnetic spectroscopy (COSY)

allowed for the assignment of the spin systems present in the aglycone, amino sugar, and acetal moieties of specumycins A1 and B1. Multiplicity-edited heteronuclear single-quantum coherence spectroscopy allowed for assigned 1H shifts to be correlated to their corresponding 13C shifts, as well as for the assignment of shifts as corresponding to methylenes or methines. Full structure elucidation was completed with heteronuclear multiple-bond correlation spectroscopy, which allowed for the assignment of remaining shifts based upon their proximity to assigned shifts.

Generation of metabolomic arrays

Mass spectrometry was performed by using a TSQ Triple Quantum mass spectrometer equipped with an electrospray ionization source and Surveyor PDA Plus detector. For positive ion mode, the following settings were used: capillary temperature was 270 °C; spray voltage 4.2 kV; spray current 30 mA; capillary voltage 35 V; tube lens 119 V; and skimmer offset 15 V. For negative ion mode, capillary temperature 270 °C; spray voltage 30 kV; spray current 20 mA; capillary voltage 35 V; tube lens 119 V; and skimmer offset 15 V. Fraction plates were prepared by injecting 20 µL of purified compounds in methanol or concentrated extract via a Thermo PAL auto injector onto a phenomenex luna 5 µm C18(2) reverse-phase HPLC column. The sample was fractionated using a gradient of 100% Buffer A (95% H2O, 5% acetonitrile) to 100% Buffer B (5% acetonitrile, 95% H2O) over 48 min at a flow rate of 1 mL/min and a fixed splitter with a 3:1 ratio with three parts going to the photodiode array detector and fraction collector and one part going to the MS. Fractions were collected in 1-min intervals in a 96 deep well plate. A volume of 150 µL of eluent from each well was transferred to four replica plates and dried in vacuo using a Genevac HT-2 system at 30 °C.

Fluorescent cell barcoding of cell-seeded metabolomic arrays

Eight serial 1:2.14 dilutions of Pacific Blue were prepared, covering a concentration range from 0.038 to 7.67 µg/mL. Six serial 1:2.5 dilutions of Pacific Orange were prepared, covering a concentration range from 0.22 to 21 µg/mL. Each dilution of Pacific Blue was added to all wells in a single row of a 96-well plate (10 µL/well), so that the dye concentration in each row decreased from the top to the bottom of the plate. Similarly, each dilution of Pacific Orange was added to all wells in a column of the same 96-well plate (10 µL/well), so that the

concentration in each column decreased from columns 1 to 6 and from columns 7 to 12. This procedure yielded two sets of 48 barcoded wells per plate. Approximately 200,000 cells (180 µL suspended in phosphate-buffered saline (PBS)) were added to each well and incubated in the dark at room temperature for 30 min. Staining was then quenched by addition of 75 µL of 1% BSA (Sigma) in PBS.

Antibody staining

Cells were stained with antibodies in 100-µL staining medium for 30 min in the dark, unless otherwise noted. Individual antibodies were added in accordance with the manufacturer's instructions. Staining was quenched with 1% BSA in PBS, and stained cells were washed with PBS prior to analysis.

Validation checkerboard

Kasumi-1 cells were incubated with either 20 µM etoposide or 1 µM staurosporine and/or DMSO in a checkerboard pattern overnight. After treatment, cells were stained with Alexa 700, fixed, permeabilized, barcoded, pooled, and then stained with anti-γH2AX-PerCP-Cy5.5 (clone N1-431, BD) or anti-cleaved caspase-3-PE (clone C92-605, BD). Subsequent to staining, samples were run on a five-laser BD Fortessa flow cytometer. Upon gating single-cell events, wells were debarcoded, and the percent of positive cells for each respective marker was determined for each of the 48 populations. Z scores were calculated according to the formula $Z = 1 - 3(\sigma p + \sigma n)/|\mu p - \mu n|$.

Dose–response curves

Serial dilutions of etoposide and staurosporine were prepared from DMSO stocks (10 mM) covering a range from 100 µM to 100 nM. A volume of 1 µL of each dilution point was added to a well. Each compound titration was handled individually on a separate plate. KG1 cells [150,000 in 199 µL of culture medium (RPMI1640 + 20% FBS + 1% penicillin/streptomycin)] were added to each well and mixed by pipetting. After incubation for 16 h, cells were stained with Alexa 700, fixed with 1.6% paraformaldehyde, and permeabilized in methanol for 20 min at –20 °C. Wells were then barcoded as described above, combined, and then stained with antibodies specific for anti-cleaved caspase-3-PE (clone C92-605, BD), or anti-γH2AX-PerCP-Cy5.5 (clone N1-431, BD).

MAM protocol with six pure compounds

DMSO stocks (10mM) of etoposide, staurosporine, CL994, PF04708671, PCI34051, mevastatin, and NU7441 were added (1 µL each) to 43 µL of methanol and fractionated as described above. Before addition of cells, compounds in the wells were suspended by addition of 1 µL of DMSO and mixing by Vortex. KG1 cells [150,000 in 199 µL of culture medium (RPMI1640 + 20% fetal bovine serum (FBS) + 1% penicillin/streptomycin)] were added to each well and mixed by pipetting. After incubation for 16 h, cells were stained with Alexa 700, fixed with 1.6% paraformaldehyde, and permeabilized in methanol for 20min at −20 °C. Wells were then barcoded as described above, combined, and then stained with the following antibodies: anti-cleaved caspase-3-PE (clone C92-605, BD), anti-p-Histone H3-PE-Cy7 (clone HTA28, BioLegend), anti-γH2AX-PerCP-Cy5.5 (clone N1-431, BD), and anti-p-S6-Ax647 (clone D57.22E, CST).

MAM using crude extract with internal standards

DMSO stocks (10 mM) of etoposide and staurosporine were added (1 µL each) to 48 µL of a crude extract (200 mg/mL in 50% methanol/water) and fractionated as described above. The extract was generated from a Streptomyces cave strain grown in BA medium, and extracted with 50% methanol prior to evaporation in vacuo. Before treatment, compounds in the wells were suspended by addition of 1 µL of DMSO and mixing by vortexing. KG1 cells [150,000 in 199 µL of culture medium (RPMI1640 + 20% fetal bovine serum (FBS) + 1% penicillin/streptomycin)] were added to each well and mixed. After incubation for 16 h, cells were stained with Ax700, fixed with 1.6% paraformaldehyde, and permeabilized in methanol for 20 min at −20 °C. Wells were then barcoded as described above, combined, and then stained with anti-cleaved caspase-3-PE (clone C92-605, BD) and anti-γH2AX-PerCP-Cy5.5 (clone N1-431, BD).

AML patient samples

All specimens were obtained in accordance with the Declaration of Helsinki following protocols approved by the Vanderbilt University Medical Center Institutional Review Board. Details of patients and sample acquisition were previously published49. Briefly, consent was obtained via an approved written consent form, and eligibility criteria included ≥18 years of age with suspected AML undergoing clinical evaluation at Vanderbilt. Samples

analyzed here were collected from bone marrow prior to any treatment. Once obtained, samples underwent immediate (within <30 min) density gradient separation of mononuclear cells using a BD Vacutainer CPT Cell Preparation Tube with Sodium Heparin (BD Biosciences, Franklin Lakes, NJ). The separated mononuclear cells were then pelleted with low-speed centrifugation (200 × g) and aliquoted into multiple cryotubes in an 88% FBS + 12% DMSO solution. Samples were stored at −80 °C for 24–72 h prior to long-term storage in liquid nitrogen. Patient 1,305,001 (001) was found to have the MLL-MLLT3 t(9;11)(p22;q23) translocation in all cells by karyotype analysis and was without other tested common molecular mutations (which included FLT3-ITD, NPM1, CEPBA, and c-KIT)66. Patient 1305015 (015) had a normal karyotype, but was found to have both a FLT3-ITD and an NPM1 mutation.

MAM of bacterial extracts against primary cell preparations

Primary cell preparations were thawed, and 200,000 cells were added to each well of a fraction plate containing a metabolite array that was generated from a crude extract from S. specus or Nocardiopsis sp. FU40. After a 16-h incubation, cells were stained for viability, fixed, permeabilized, barcoded, and stained with the following antibodies: anti-Human CD45-Ax488 (clone HI30, BioLegend), anti-cleaved caspase-3-PE (clone C92-605, BD), anti-p-Histone H3-PE-Cy7 (clone HTA28, BioLegend), anti-γH2AX-PerCP-Cy5.5 (clone N1-431, BD), and anti-p-S6-Ax647 (clone D57.22E, CST). SSC and CD45 expression were used to define lymphocyte, monocyte, and blast populations. Each population was then debarcoded, and readouts were determined for 48 wells per cell type.

Deep profiling of ciromicins A and B against primary cell preparations

Ciromicins were purified by separation on a Water 600 HPLC system with a reverse-phase column using a linear gradient of water/acetonitrile containing 0.1% formic acid. Fractions with UV absorbance indicative of ciromicins were then combined and applied on a size exclusion Sephadex LH-20 column for a gravity elution in methanol. Ciromicin A and B were then separated by a secondary HPLC purification. Approximately 6 million cells (2 million per condition) from a thawed primary AML sample were incubated in culture medium [80% RPMI 1640 (Mediatech, Inc., Manassas, VA) + 20% FBS (Gibco standard FBS, Life Technologies, Grand Island, NY)

with 10 μM ciromicin A, ciromicin B, or DMSO] for 48 h. Mass cytometry experiments were performed as previously described17. Briefly, after incubation with ciromicin A, B, or vehicle, samples were pelleted by centrifugation at 200 × g, resuspended and washed with PBS (HyClone, HyClone Laboratories, Logan, UT), pelleted, and resuspended in PBS. They were then stained with Cell-ID Cisplatin (Fluidigm, South San Francisco, CA) as per the manufacturer's recommended protocol. The cells were washed and resuspended in staining medium [CSM: PBS + 1% BSA (Fisher Scientific, Fair Lawn, NJ)]. Cells were then stained with a mass cytometry antibody panel of 29 extracellular antibodies designed to characterize AML blasts and most non-AML peripheral blood mononuclear cells consisting of anti-human CD235a-141 (clone HIR2, Fluidigm), anti-human CD117-143 (clone 104D2, Fluidigm), anti-human CD11b-144 (clone ICRF44, Fluidigm), anti-human CD4-145 (clone RPAT4, Fluidigm), anti-human CD64-146 (clone 10.1, Fluidigm), anti-human CD36-147 (clone 5-271, BioLegend), anti-human CD34-148 (clone 581, Fluidigm), anti-human CCR2-149 (clone K036C2, BioLegend), anti-human CD43-150 (clone 84-3C1, Fluidigm), antihuman CD123-151 (clone 6H6, Fluidigm), anti-human CD13-152 (clone WM15, Fluidigm), anti-human CD45RA-153 (clone HI100, Fluidigm), anti-human CD45-154 (clone HI30, Fluidigm), anti-human CD86-156 (clone IT2.2, Fluidigm), antihuman CD33-158 (clone WM53, Fluidigm), anti-human CD11c-159 (clone BU15, Fluidigm), anti-human CD14-160 (clone M5E2, Fluidigm), anti-human CD32-161 (clone FUN-2, BioLegend), anti-human CDHLA-DR-163 (clone L243, BioLegend), anti-human CD15-164 (clone W6D3, Fluidigm), anti-human CD16-165 (clone 3G8, Fluidigm), anti-human CD38-167 (clone HIT2, Fluidigm), anti-human CD8-168 (clone SK1, Fluidigm), anti-human CD25-169 (clone 2A3, Fluidigm), antihuman CD3-170 (clone UCHT1, Fluidigm), anti-human CD184-173 (clone 12G5, Fluidigm), anti-human PD1-174 (clone EH12.2H7, Fluidigm), anti-human PD-L1-175 (clone 29E.2A3, Fluidigm), and anti-human CD56-176 (clone CMSSB, Fluidigm) 50. A master mix of these antibodies was added to each sample to give a final staining volume of 50 μL and incubated at room temperature for 30 min. Cells were then washed twice, first with CSM and then with PBS and then permeabilized in –20 °C 100% methanol for 20 min. Following permeabilization, cells were washed, stained with 250 nM iridium intercalator (Fluidigm, South San Francisco, CA) for 30 min at 4 °C, washed twice in PBS, and then resuspended in 500 μL of ddH2O for CyTOF analysis. Samples were analyzed using a CyTOF 1.0 cytometer (Fluidigm, South San Francisco, CA) (*452*).

Data availability

Kasumi validation data (Figure D-2) and AML patient sample responses to S. specus (Figure D-3) and Nocardiopsis (Figure D-4) were generated by fluorescence cytometry as described by Irish et al (*15, 80*). and Krutzik et al (*344, 345, 400*). and are available as FCS files in FlowRepository (https://flowrepository.org/experiments/1476, https://flowrepository.org/experiments/1477, https://flowrepository.org/experiments/1478, and https://flowrepository.org/experiments/1479). AML patient sample responses to ciromicin A and ciromicin B (Figure D-5) were generated by mass cytometry as described by Leelatian et al (*81*). and Ferrell et al (*374*). and are available as FCS files in FlowRepository (https://flowrepository.org/experiments/1480). MEM enrichment scores (Figure D-5b) were generated as in Diggins et al (*321*). using stem-like population 11 as a reference.

Supplemental information

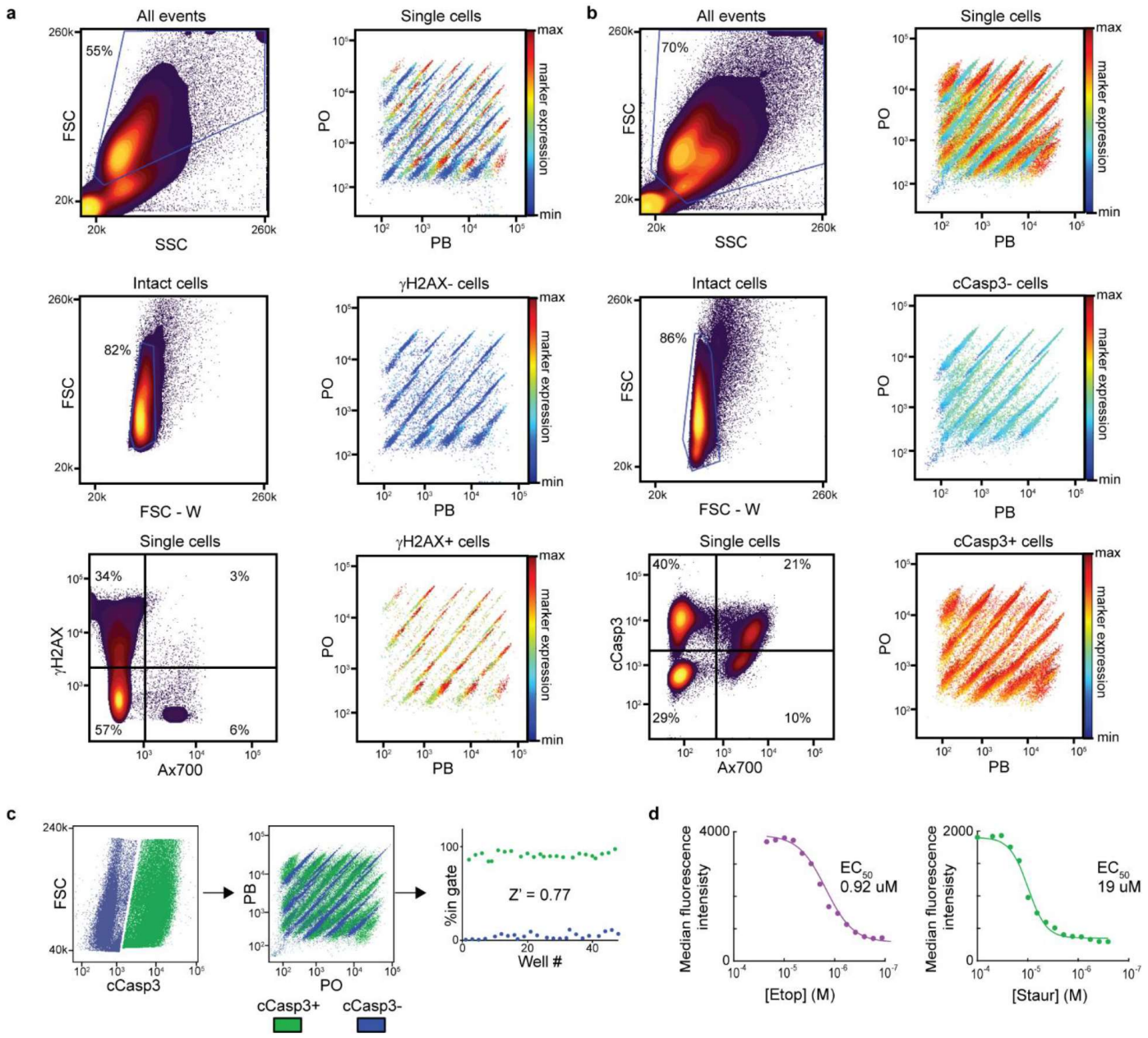


Figure SD-1. Gating strategy for assay validation and demonstration of dose dependent behavior in FCB array assays. a) All collected events from etoposide checkerboard experiment were gated for intact cells (FSC vs SSC), then single cells (FSC vs FSC-W) and finally for maker expression. Biaxial plots of single cells (PO vs PB) colored by γH2AX expression visually reflect the checkboard pattern. Plots of only γH2AX- cells or γH2AX+ cells show populations whose FCB coordinates match assay wells with vehicle or compound respectively. b) Same as (a) for staurosporine checkerboard experiment with cCasp3 used as the marker readout. c) Determination of z score from staurosporine checkerboard. d) A titration series of etoposide and staurosporine was prepared on a microtiter plate, incubated with KG1 cells, barcoded and stained. EC50 values were calculated using the median fluorescent intensity of cell populations from each well. The average collected cells per well used to calculate z score was 1016 events (minimum 79). The average collected cells per well used to calculate EC50 values was 3122 events (minimum 794).

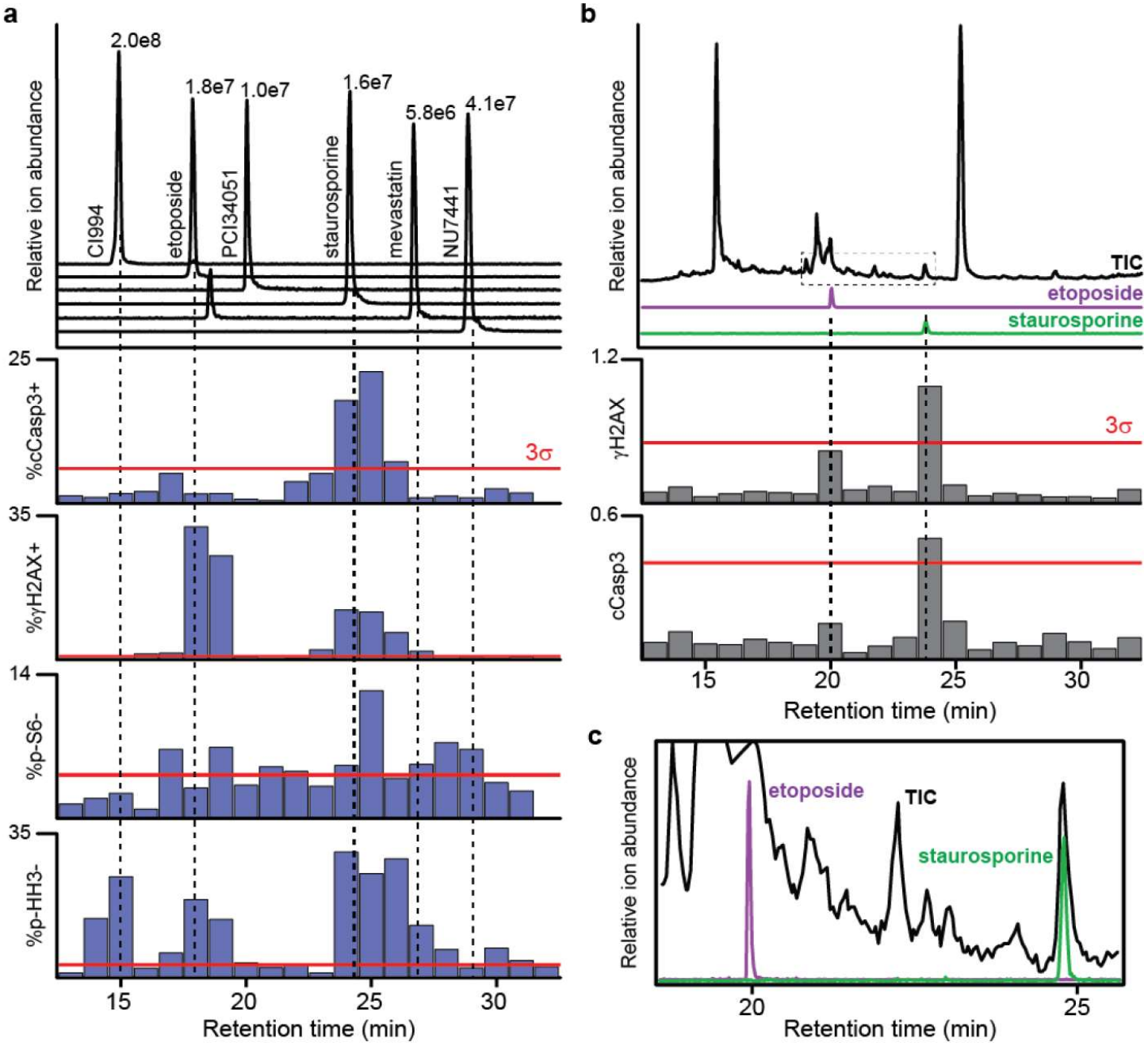


Figure SD-2. Integration and validation of chromatographic arrays and FCB. Chromatographic arraying is performed using split flow HPLC/UV/MS with polarity switching mass scanning resulting in an array of highly characterized fractions. a) A mixture six bioactive small molecules was arrayed onto a microtiter plate via split flow HPLC/MS fractionation and solvent was evaporated. Subsequently KG1 cells were added to the wells of the plate for incubation with the various toxicants. Cells were stained with Alexa-700 dye to indicate cell viability, fixed, permeabilized, barcoded, pooled and then immuno-stained with antibody-dye conjugates for DNA damage and apoptosis using anti-γH2AX and anti-cCasp3, respectively, and additional conjugates directed against phosphorylated Histone H3 (p-HH3), and phosphorylated ribosomal protein S6K (p-S6). The sample was analyzed via flow cytometry, and reconstructed bioactivity chromatograms were generated by gating on viable and marker positive (γH2AX and cCasp3) or marker negative (p-Histone H3 and p-S6) cells. Selective ion traces are aligned with bioactivity chromatograms. b) crude extract spiked with etoposide and staurosporine prior to fractionation. Bioactivity chromatograms were constructed using the arcsinh transformed median of all cells per well. c. expansion of TIC, and EIC for etoposide and staurosporine. Red line denotes threshold for signal greater than 3 standard deviations of the readout from 4 blank control wells. The average collected cells per well in a was 656 events (minimum 89) and in b was 1277 (minimum 102).

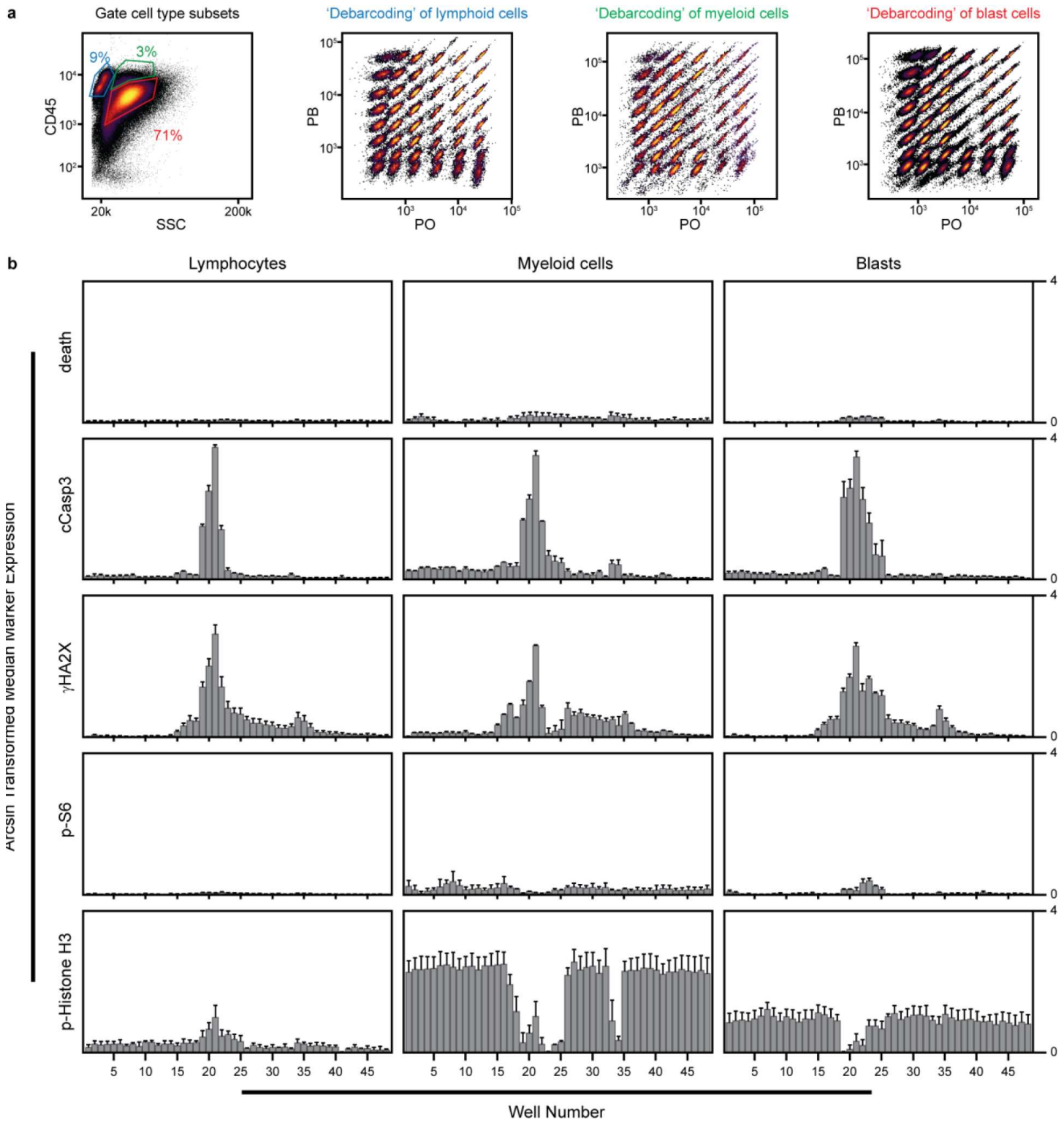


Figure SD-3. Comparison of replicates of MAM with primary patient samples. a) Samples from patient 015 were incubated with two replicate plates of the fractionated S. specus extract. b) Bar graphs show the average of the arcsin transformed medians for each marker for the two experiments. Error bars are standard deviations. For each of the 48 lymphocyte populations from each well, an average of 1360 events were recorded in each replicate experiment. For myeloid populations the average collected cells per well was 595 and for blast populations was 5755.

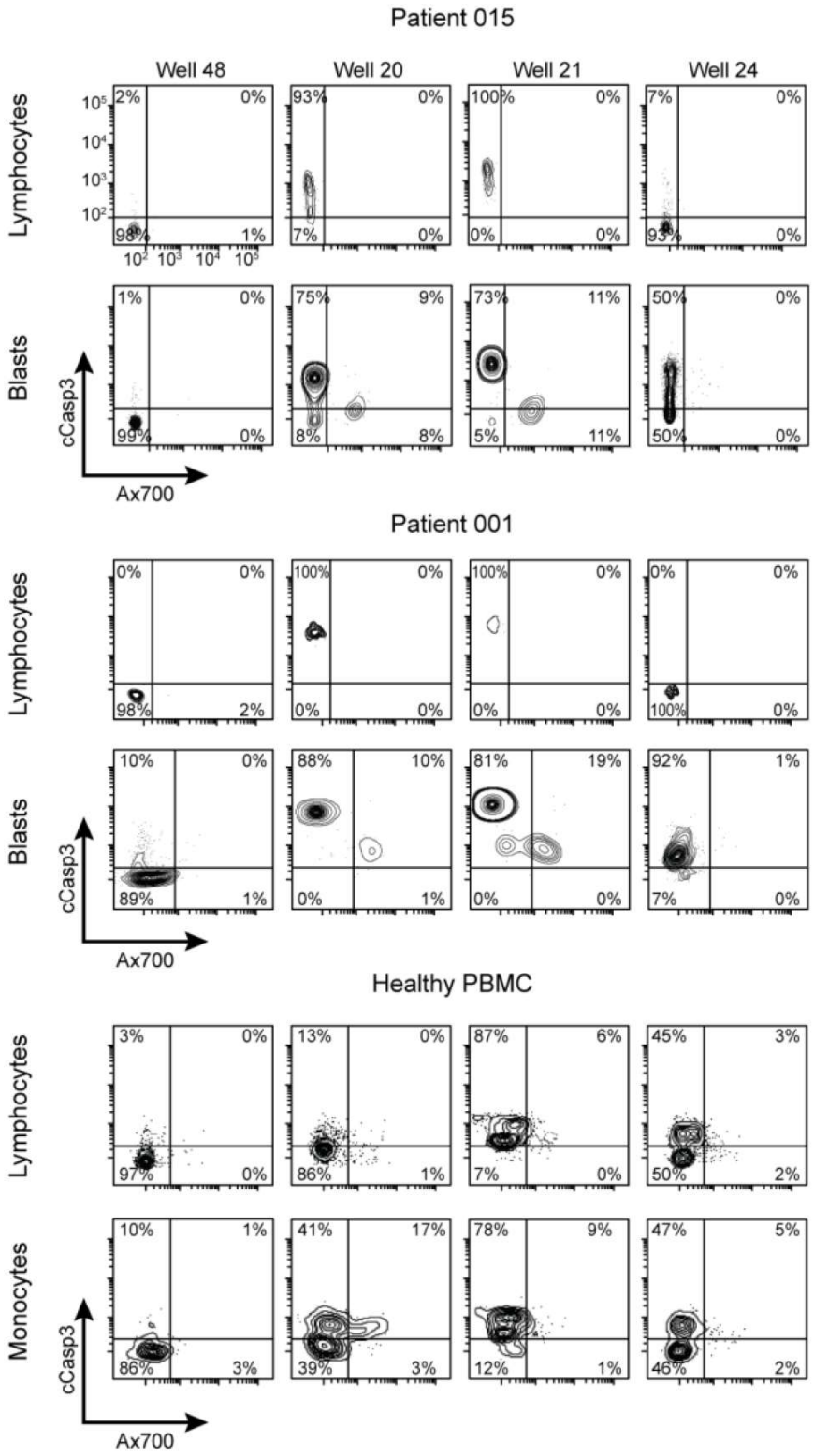


Figure SD-4. Biaxial plots of cCasp3 vs Ax700 from fraction wells containing specumycins. The predominant analogs of the specumycins elute from 20 to 21 minutes with well 21 containing the highest amount of specumycin A. Also shown are the biaxial plots from well 48 which contained only elution buffer and well 24 containing the m/z 1054.

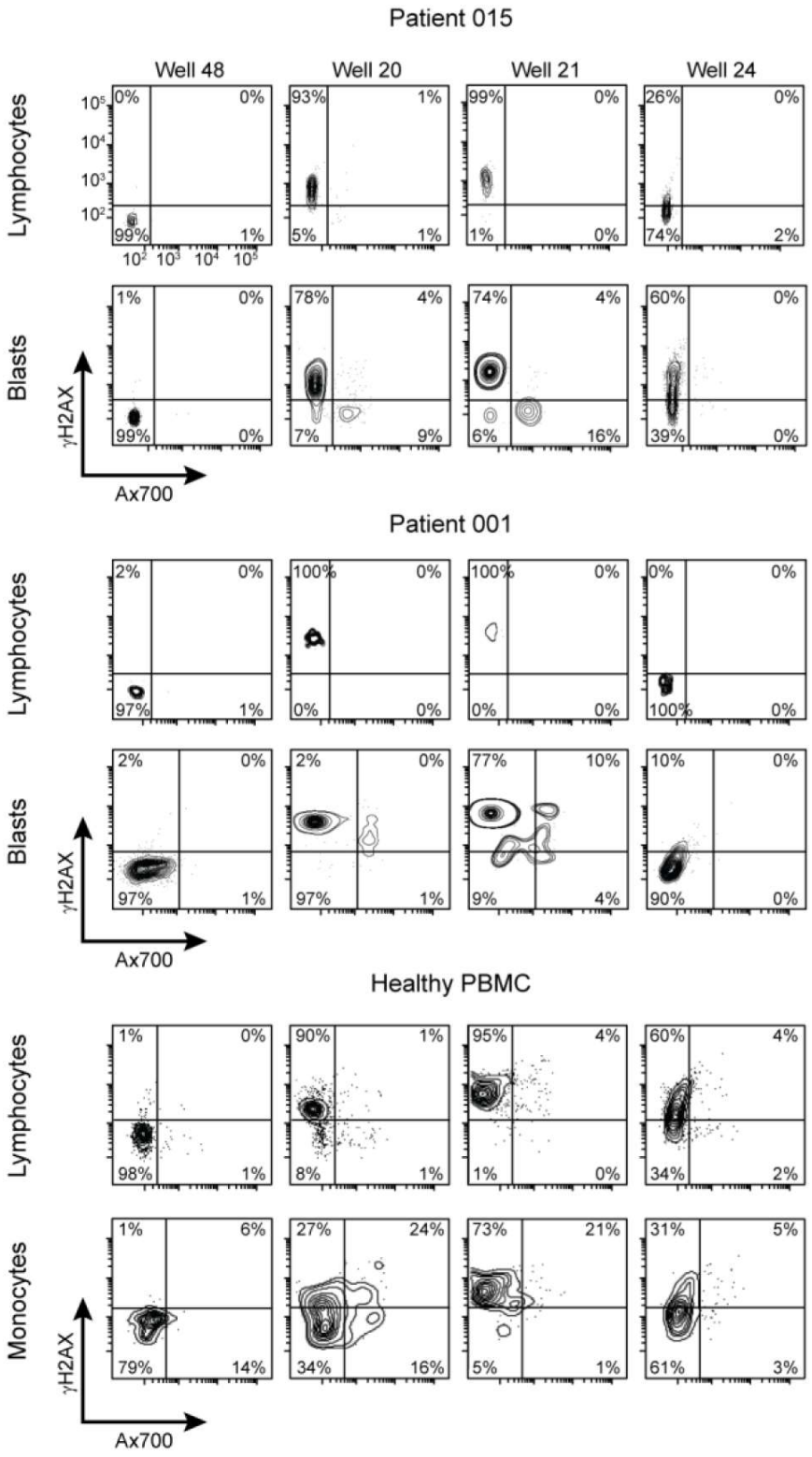


Figure SD-5. Biaxial plots of γH2AX vs Ax700 from fraction wells containing specumycins. Same as Supplementary Figure D-4. except corresponding plots for γH2AX are shown instead.

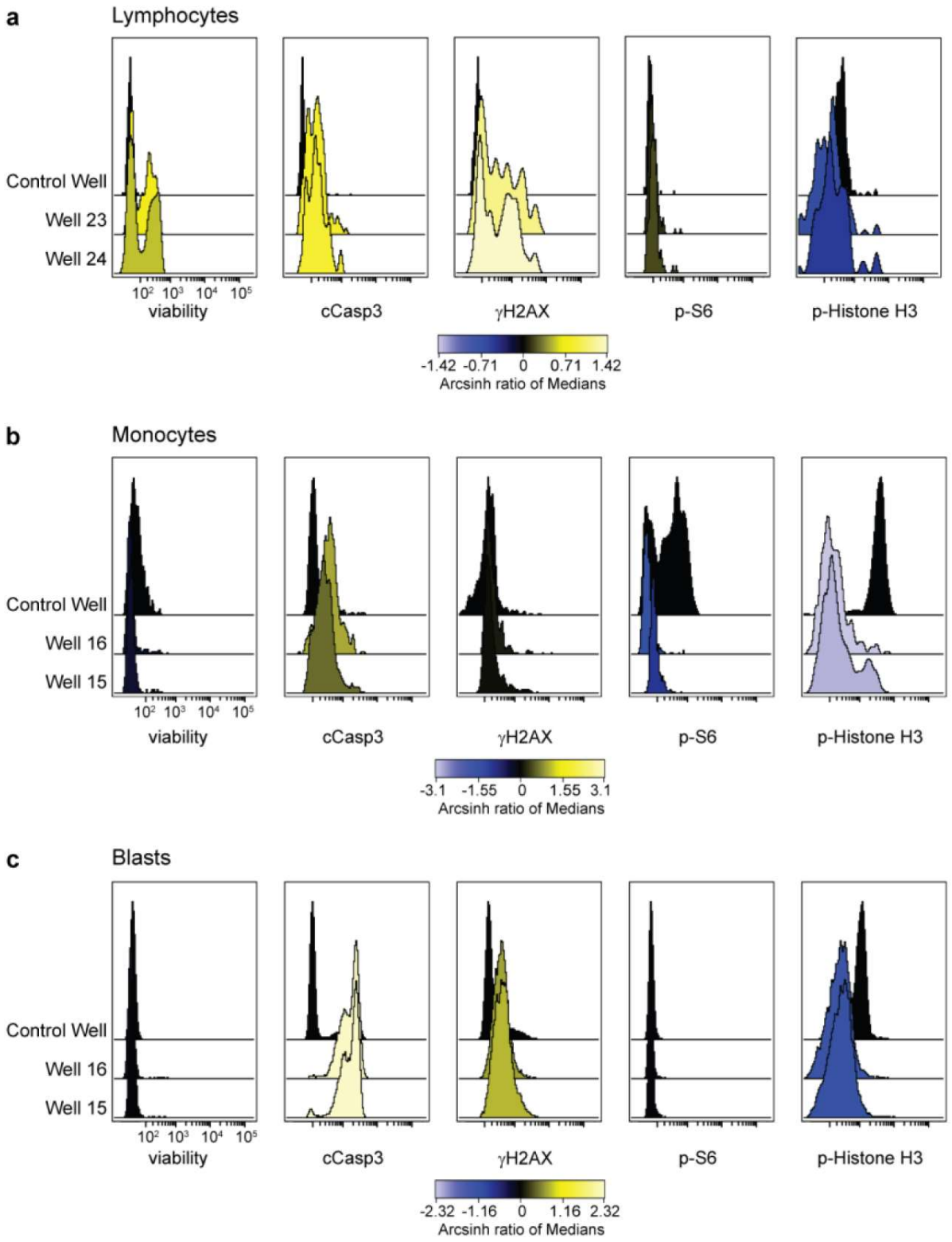


Figure SD-6. Histogram plots of each marker in control wells and highlighted wells in Figure D-4. a) Marker distribution in wells 23 and 24 which had the maximal response in lymphocytes and contained the apoptolidins. b) and c) Marker distribution in wells 15 and 16 which contained the ciromicins which induced the largest response in monocytes and blasts.

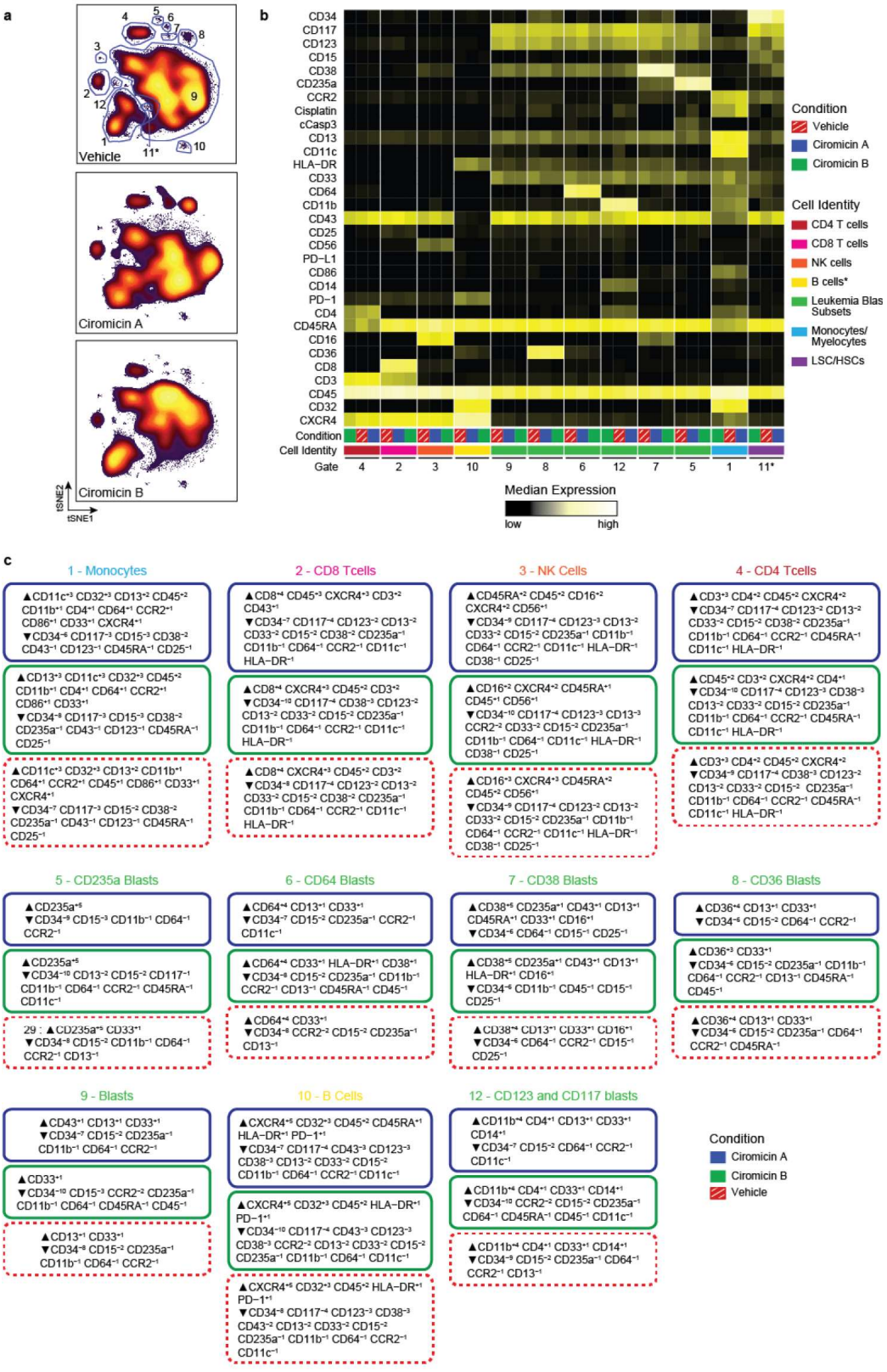


Figure SD-7. Median marker expression of gated viSNE populations. a) 12 major populations were identified after viSNE analysis and gated. b) Median marker expression for each gated population after treatment with vehicle, ciromicin A or B. c) MEM text labels for gated populations with population 11 (LSC/HSCs) used as reference.

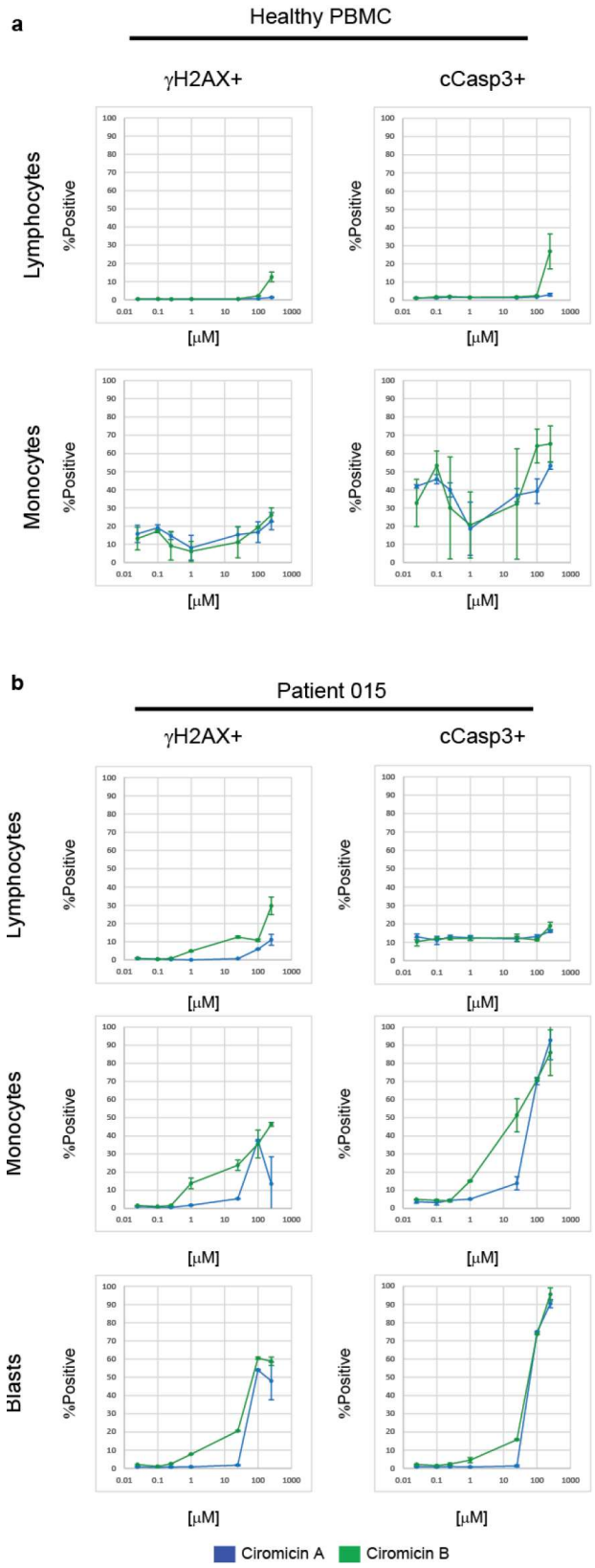


Figure SD-8. 24 hour titration of ciromicins assayed against a PBMCs from a healthy donor and b a AML patient sample. Points represent the average of two replicate experiments.

Table SD-1. Changes in Population (Percent of Cells) in Response to Ciromicins

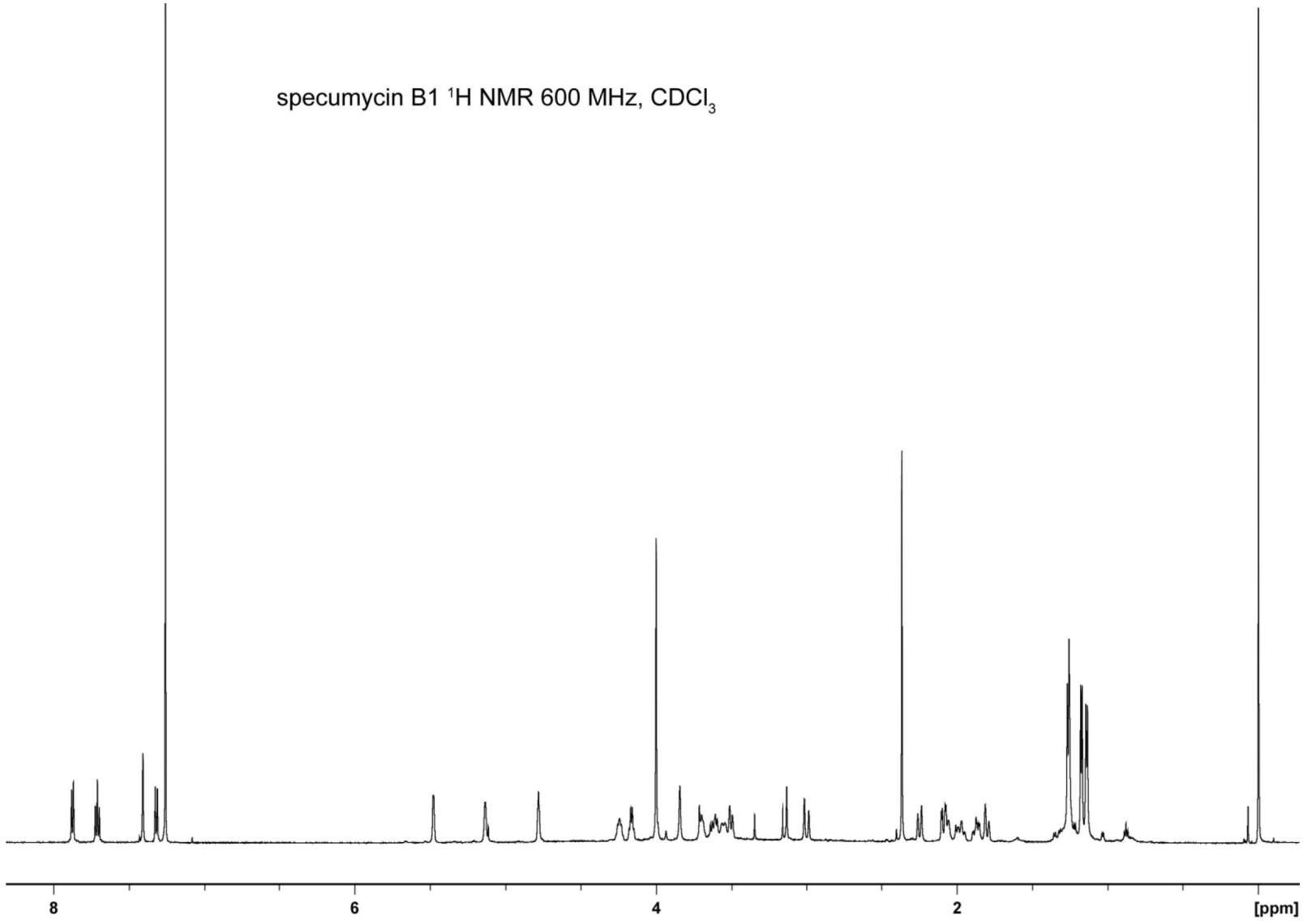
Gate	Vehicle	Ciromicin A	Ciromicin B
1	13.6	16.0	17.1
2	2.8	2.9	2.4
3	0.3	0.4	0.3
4	4.3	3.6	4.1
5	0.4	1.0	0.1
6	0.5	0.8	0.2
7	0.3	1.2	0.2
8	3.6	3.3	1.0
9	73.5	66.0	72.9
10	0.9	1.0	0.9
11	0.4	1.3	0.1
12	0.5	1.0	0.4

Table SD-2. NMR Shift Assignments for specumycin A1

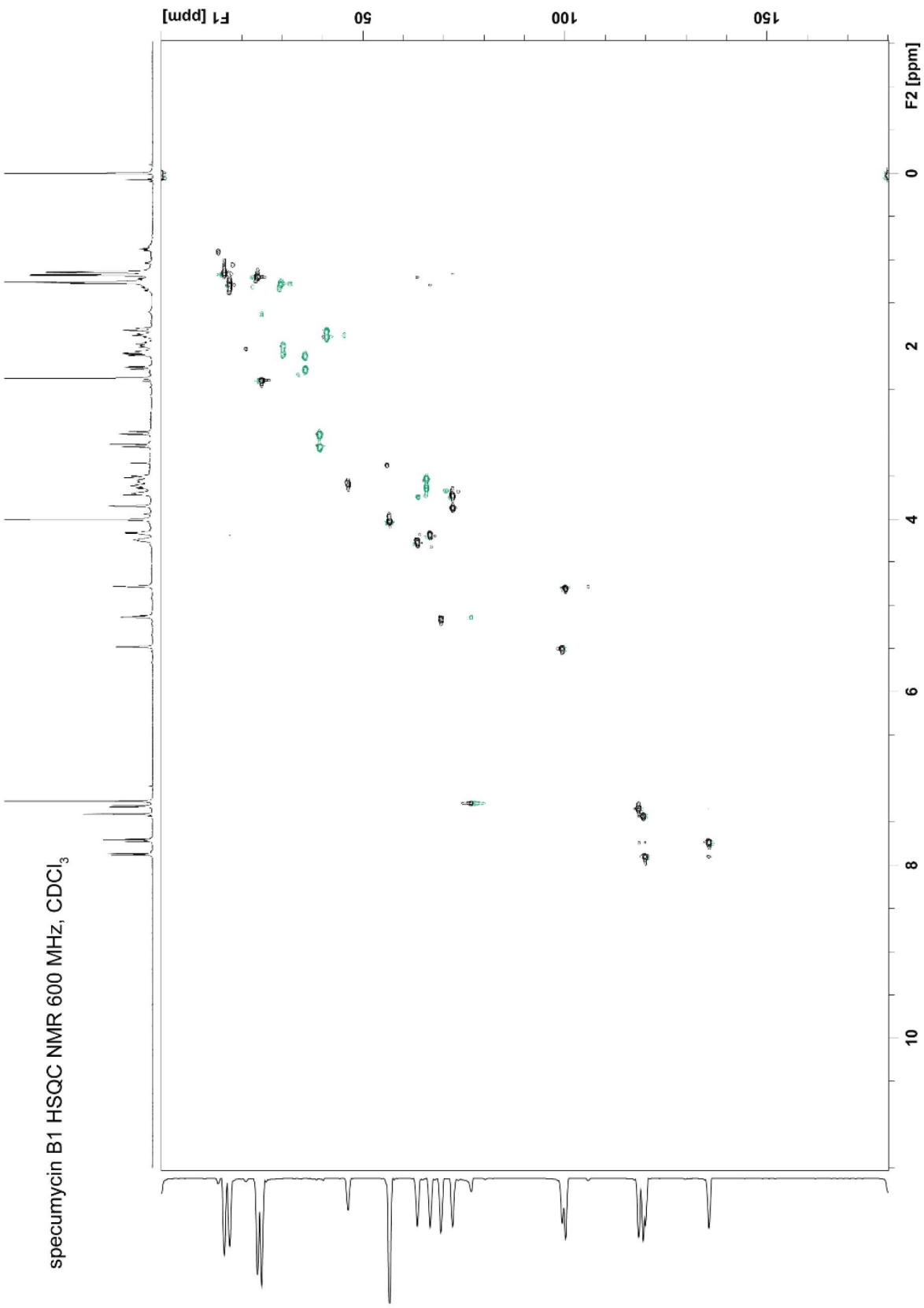
Position	^{1}H NMR δ_H *(J in Hz)*	^{13}C NMR δ_C	^{1}H - ^{1}H COSY NMR δ_H	^{1}H - ^{13}C HMBC NMR δ_C
1	7.89 d (5.9)	120.1	7.71	187.2, 136.1, 121.6, 118.9
2	7.71 t (7.8)	136.0	7.89, 7.32	161.8, 136.1
3	7.32 d (7.8)	118.9	7.71	187.2, 161.8, 120.1
7	5.17	69.8	2.27, 2.08	135.3, 77.6
8	2.27 d (14.6), 2.08 dd (14.8, 4.2)	35.4	5.17, 3.16	77.6, 69.8, 33.8
10	3.16 d(18.6), 2.89 d (18.6)	33.8	2.27	212.6, 156.3, 135.1, 77.6, 35.4
14	2.40 s	25.4		212.6, 77.6
1'	5.5	100.5	3.82, 1.93	67.8, 46.8
2'	1.93-1.88 m	32.7	5.50, 3.35	46.8
3'	3.35	46.8	3.82, 1.93	74.2
4'	3.82	74.2	5.50, 4.15, 3.35, 1.93	100.5, 46.8, 32.7, 17.6
5'	4.15 q (6.4)	67.8	3.82, 1.29	100.5, 74.2, 46.8, 17.6
6'	1.29 d (6.7)	17.6	4.15	74.2, 67.8
1"	4.84	101.2	1.86	
2"	1.86-1.81 m	42.4	4.84, 4.23	101.2, 64.4
3"	4.23 m	64.4	1.86, 1.20	
4"	1.20 d (6.2)	24.2	4.23	64.4, 42.4
5"	3.76	73.0	3.54, 3.50, 1.13	73.0, 16.6
6"	3.54 dd (11.8, 2.1) 3.50 dd (11.8, 7.1)	66.5	3.76	
7"	1.13 d (6.3)	16.6	3.76	73.0, 66.5
4		161.8	7.71, 7.32	
5				
6				
9		77.6	5.17, 3.16, 2.89, 2.40	
11		156.3	3.16, 2.89	
12		187.2	7.89, 7.32	
13		212.6	3.16, 2.89, 2.40	
4a		121.6	7.89, 7.32	
5a				
6a		135.3	5.17	
10a		135.1	3.18, 2.89	
11a				
12a		136.1	7.89, 7.71	
4-OMe	4.01	57.2	7.32	161.8

Table SD-3. NMR Shift Assignments for specumycin B1

Position	^{1}H NMR δ_H (*J in Hz*)	^{13}C NMR δ_C	^{1}H - ^{1}H COSY NMR δ_H	^{1}H - ^{13}C HMBC NMR δ_C
1	7.88 d (7.6)	120.5	7.71, 7.32	182.8, 120.9, 118.8
2	7.71 t 8.0)	136.1	7.88, 7.32	161.3
3	7.32 d (8.4)	118.8	7.88, 7.71	188.7, 161.3, 120.5
7	5.14	69.8	2.25, 2.09	130.0, 77.8
8	2.25 d (14.6), 2.09 dd (14.8, 4.1)	36.1	5.14, 3.00	130.0, 77.8, 69.8, 39.7
10	3.15 d (17.4), 3.00 d (17.4)	39.7	7.41, 2.25	212.6, 143.1, 130.0, 120.0, 77.8, 36.1
14	2.37 s	25.4		212.6, 77.8
1'	5.48	99.8	3.85, 2.06, 1.99	69.8, 67.1, 46.8
2'	2.06, 1.99	30.2	5.48, 3.56	
3'	3.56	46.8	3.85, 2.06, 1.99	
4'	3.85	72.6	4.17, 3.56	100.7, 46.8, 30.2
5'	4.17 q (5.9)	67.1	1.27	99.8, 72.6, 17.5
6'	1.27 d (6.6)	17.5	4.17	72.6, 67.1, 30.2
1''	4.78	100.7	1.87, 1.80	
2''	1.87, 1.80	41.4	4.78, 4.25	100.7, 64.0
3''	4.25	64	1.87, 1.80, 1.18	
4''	1.18 d (6.1)	24.4	4.25	64.0, 41.4
5''	3.7	72.6	3.62, 3.51, 1.14	100.7, 66.1
6''	3.62 dd (18.7, 8.0), 3.51 d (10.8)	66.1	3.7	72.6
7''	1.14 d (6.0)	16.2	3.7	72.6, 66.1
4		161.3	7.71, 7.32, 4.00	
5		188.7	7.32	
6				
9		77.8	5.14, 3.15, 3.00, 2.37, 2.25, 2.09	
11	7.41 s	120	3.15, 3.00	188.7, 182.8, 130.0, 39.7
12		182.8	7.88	
13		212.6	3.15, 3.00, 2.37	
4a				
5a				
6a		130	5.14, 3.15, 3.00, 2.25, 2.09	
10a		143.1	3.15, 3.00	
12a		120.9	7.88	
4-OMe	4.00 s	57.1	7.32	161.3, 118.8


specumycin B1 ^{1}H NMR 600 MHz, $CDCl_3$
8
6
4
2
[ppm]

specumycin B1 COSY NMR 600 MHz, $CDCl_3$


specumycin B1 HSQC NMR 600 MHz, $CDCl_3$
F1 [ppm]
50
100
150
F2 [ppm]
0
2
4
6
8
10

specumycin B1 HMBC NMR 600 MHz, $CDCl_3$
F1 [ppm]
50
100
150
200
F2 [ppm]
0
2
4
6
8
10

APPENDIX E

Characterizing Cell Subsets Using Marker Enrichment Modeling

Authors: Kirsten E. Diggins, Allison R. Greenplate, Nalin Leelatian, Cara E. Wogsland, and Jonathan M. Irish

Data presented in this chapter was published in *Nature Methods* 2017 (Diggins, Greenplate Leelatian et al. 2017)

Preface

The most differentially or most highly expressed cellular features are often the presumed significant characteristics of a given cell population. However, these features can show certain variability even within a seemingly uniform cell type. The work presented here introduces a new quantitative measure of cellular features, Marker Enrichment Modeling (MEM), which emphasizes not only the magnitude of feature expression but also feature variability. MEM determines, quantitatively, how specific a feature is to a given cell subset, compared to a reference point. This was proven to be powerful for cell subset identification compared to the qualitative description of canonical lineage markers, even in the context of a well-described human tissue type. Specifically, MEM effectively distinguished glioblastoma cells from tumor-infiltrating immune cells even without the information of canonical identity proteins.

This work is incorporated as a component of the automated risk-stratification workflow that led to the discovery of novel stratifying glioblastoma cell subsets, described in Chapter 6. Importantly, MEM was key to revealing the enriched features that specifically and critically defined the clinically significant glioblastoma cells. This led to the development of a clinically applicable manual analysis workflow and suggested the essential features of clinically unfavorable cells that should be tested as targets for innovative therapy in glioblastoma.

Abstract

Learning cell identity from high-content single-cell data presently relies on human experts. We present Marker Enrichment Modeling (MEM), an algorithm that objectively describes cells by quantifying contextual feature enrichment and reporting a human and machine-readable text label. MEM outperforms traditional metrics in describing immune and cancer cell subsets from fluorescence and mass cytometry. MEM provides a

quantitative language to communicate characteristics of new and established cytotypes observed in complex tissues.

Main Text

Quantitative cytometry workflows have developed diverse approaches to grouping cells into populations and visualizing results in graphs that arrange populations based on phenotype (*314, 394*). Important features of populations are typically assumed to be those most highly or differentially expressed. This approach works well when feature variability is low and cells match established types, but computational analysis of single cell data routinely reveals novel cells with non-canonical phenotypes (*5, 294, 441*). This is especially common in diseases where abnormal expression profiles and signaling responses distinguish clinically significant cell subsets (*12, 15, 80, 360, 375*). Existing statistical approaches can be used to characterize a population's degree of difference from a reference, but may be limited to a normal distribution or may not account for intra- and inter-population variability in a single metric.

The MEM equation (equation (1)) produces a signed value for each population feature by quantifying positive and negative, population-specific, contextual feature enrichment relative to a reference cell population (Supplementary Note 1).

$$\text{MEM score} = \left| MAG_{POP} - MAG_{REF} \right| + \left(\frac{IQR_{REF}}{IQR_{POP}} \right) - 1, \quad (MAG_{POP}\text{-}MAG_{REF}) < 0 \rightarrow \text{MEM} = \text{-MEM} \qquad \text{(Eq. 1)}$$

In Eq. 1, POP denotes the population of interest, REF denotes the reference population to which POP will be compared, MAG is feature magnitude (here, median protein expression detected by mass or fluorescence flow cytometry), and IQR indicates the interquartile range. A reference population (REF) is chosen based on a biological comparison of interest (Supplementary Note 1, Supplementary Table 1, Supplementary Figure E1). MEM was designed to quantify enrichment, whereas other metrics used in cytometry, such as Kolmogorov-Smirnov (K-S) (*461*), area under the ROC curve (AUC) (*462*), and Earth Mover's Distance (EMD) (*463*), capture other differences between frequency distributions (Supplementary Note 1). In datasets including healthy human blood, bone marrow, and tonsil, murine tissues, and human tumors, MEM identified key proteins used by experts to distinguish rare and novel cell subsets.

Four cytometry studies, Dataset A (*81*), Dataset B (*83*), Dataset C (*294*), and Dataset D, collected as described by Leelatian and Doxie, et al. (*284*), were used to evaluate the ability of MEM to identify biological features of expert and machine identified cell subsets. For datasets A, B, and C, populations had been previously identified by experts and by computational tools including viSNE (*290*) and SPADE (*282*), which are used in mass cytometry for dimensionality reduction and cell clustering (*314*), respectively.

Dataset A was mass cytometry data quantifying expression of 25 proteins on healthy human peripheral blood mononuclear cells (PBMC) (*81*). This dataset was chosen for two reasons: 1) the 7 cell subsets present are well-established, phenotypically distinct populations that served as a gold standard of biological 'truth' and 2) the cells in each of the 7 subsets were characterized for 25 proteins that displayed varying homogeneous and heterogeneous expression patterns. Populations were expert gated following viSNE analysis and each population was compared to the other cells in the sample (Figure E1, Supplementary Table 2). MEM returned labels that matched prior expert analysis (*81*) and correctly assigned high positive enrichment values to canonical protein features of each subset (Figure E1b), including CD4 on $CD4^+$ T cells (▲$CD4^{+6}$ $CD3^{+5}$ ▼$CD8a^{-4}$ $CD16^{-3}$), IgM on IgM^+ B cells (▲MHC II^{+8} IgM^{+6} $CD19^{+5}$ ▼$CD4^{-6}$ $CD3^{-5}$), CD11c and MHC II on monocytes (▲$CD11c^{+8}$ $CD33^{+7}$ $CD14^{+6}$ $CD61^{+6}$ MHC II^{+4} $CD44^{+3}$ ▼$CD3^{-5}$ $CD4^{-4}$), and CD16 on NK cells (▲$CD16^{+9}$ $CD56^{+2}$ $CD11c^{+2}$ ▼$CD4^{-7}$ $CD3^{-4}$ $CD44^{-3}$). Proteins that were not significantly enriched on any of the 7 subsets of mature human blood mononuclear cells were correctly assigned near-zero MEM scores (e.g. CD34 and CD117 proteins expressed on hematopoietic stem cells, Figure E1b). Similarly, proteins with little variability across cell subsets were assigned low, near-zero MEM scores, even for highly expressed proteins (e.g. CD45 on all subsets, CD45RA on non-T cells, Figure E1b). Incorporating information about feature variability allowed MEM to capture negative enrichment that was not reflected in magnitude difference (MAGDIFF, Supplementary Note 2). Highly enriched proteins were more important to accurate population identification than proteins characterized by high median expression alone (Figure E1c; Supplementary Figure E2; Supplementary Figure E3).

To test the hypothesis that features with high MEM scores would be important for computational cluster formation, the 25 proteins measured in Dataset A (Figure 1b) were sorted in six ways: 1) high to low MEM score, 2) high to low median value, 3) high to low MAGDIFF, 4) high to low z-score, 5) high to low K-S statistic, and 6) randomly (Supplementary Table 3). Z-score and K-S statistic values are shown in Supplementary Table 4. The

proteins were then sequentially, cumulatively excluded from use in k-means clustering and f-measure was calculated to measure clustering accuracy (Figure E1c and Supplementary Figure E2). The order in which markers were excluded is shown in Supplementary Table 3. Random exclusion was performed 15 times and the average result is shown (Figure E1c). Clustering accuracy was most impacted by excluding proteins based on MEM score. F-measure dropped to 0.75 after removing the proteins with the top 6 MEM scores, whereas a comparable F-measure decrease was only observed after removing the 14 highest markers based on MAGDIFF, the 13 highest markers based on z-score, and the 12 highest markers based on K-S statistic values (Supplementary Figure E2). Removing markers based on median was not significantly different from removing markers randomly until the 15 markers with the highest median signal intensity were excluded (Supplementary Figure E2). The same analysis was performed with viSNE in place of k-means clustering to visualize loss of population resolution (Supplementary Figure E3c). In this case, loss of accuracy was reflected in the viSNE map as a loss of separation between "islands" of cells. These results indicated that MEM enrichment scores captured markers that were important to cell identity better than traditional comparisons based solely on median protein expression.

Dataset B was mass cytometry data quantifying expression of 31 proteins on healthy human bone marrow (*83*). Computational and expert analysis had previously identified 23 populations of cells that were analyzed here by MEM (Supplementary Note 3). For example, the cell subset labeled as HSCs was highly enriched for CD34 ($CD34^{+6}$) and negatively enriched for CD45 ($CD45^{-5}$). Dataset B also illustrated the general rule that MEM scores will approach median values as feature variability within populations decreases (Supplementary Figure E4). MEM captured feature enrichment and heterogeneity better than median in diverse populations, as in Figure E1c.

Dataset C was mass cytometry data quantifying expression of 38 proteins on murine cells from eight tissues4 (Supplementary Note 4). In this dataset, "cluster 28" was a novel population identified as CD11bint NK cells. The MEM label for cluster 28 within ILCs was ▲$CD11b^{+5}$ $CD62L^{+3}$ ▼$CD4^{-7}$ $CD103^{-4}$ $Terr119^{-3}$ (Supplementary Note 4 and Supplementary Figure E5). This MEM label captured the key feature of this novel innate lymphoid cell subset (CD11bint) and highlighted additional features that can be used to match this subset to cells identified by others (i.e., to cytotype the population). These results indicate that MEM labels complement

unbiased population discovery and effectively characterize cyto incognito (*296*) by providing unbiased descriptions that correctly capture key features of novel cell types.

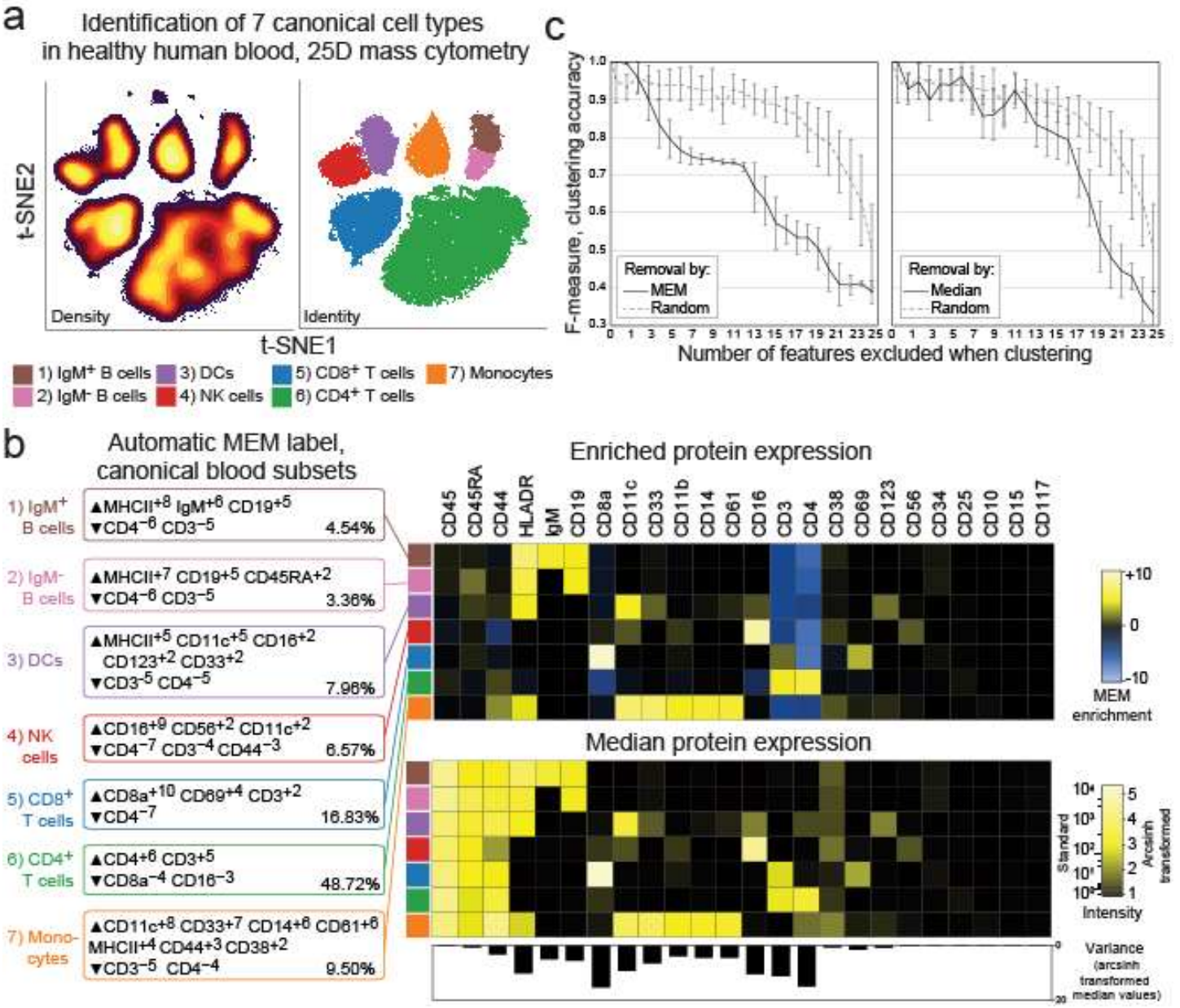


Figure E-1 Marker enrichment modeling (MEM) automatically labels human blood cell populations in Dataset A. a) Cells from normal human blood grouped into 7 canonical populations using viSNE analysis and expert review of 25D mass cytometry data. b) MEM labels computationally generated for each canonical cell subset. Heatmaps show protein enrichment values used to generate MEM labels and the median protein expression values for each protein on each cell subset. Variability in protein expression across the 7 canonical cell populations is shown below to highlight proteins that were expressed homogeneously (low variability, e.g. CD45) and those that were expressed heterogeneously (high variability, e.g. CD8a, CD4). c) Graphs show decreasing f-measure (clustering accuracy) as markers were excluded from k-means cluster analysis based on high to low absolute MEM or median values, compared to random exclusion.

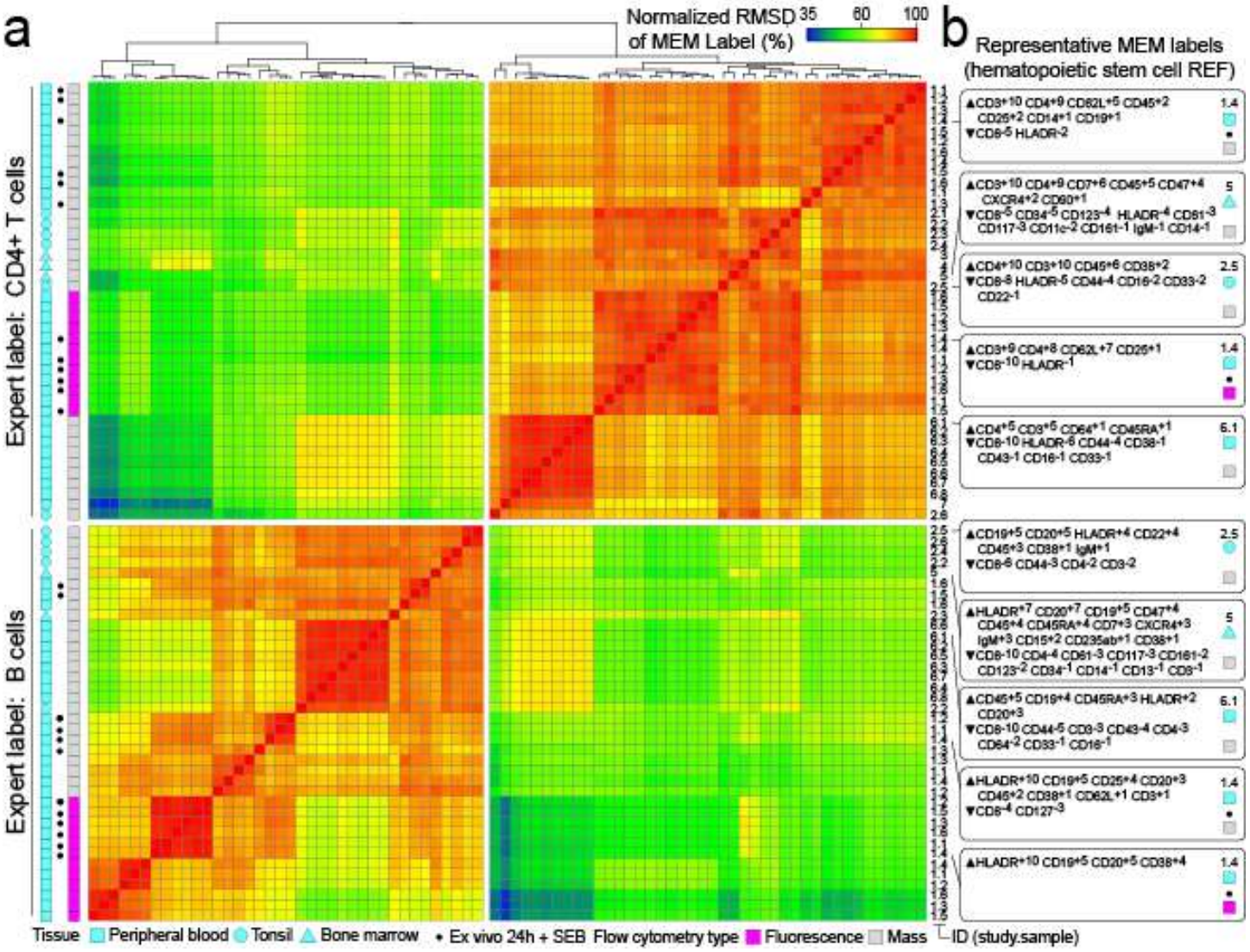


Figure E-2 Hierarchical clustering based solely on MEM label groups T cells and B cells measured in diverse studies using different cytometry platforms. A) MEM label values were compared for each of 80 populations ($CD4^+$ T cells and B cells) from 3 human tissues representing 6 mass cytometry studies and 1 fluorescence flow cytometry study. Populations are shown clustered according to MEM label percent similarity. Tissue type, source study (numbered 1-7 and referenced in online methods), and individual sample IDs are indicated to the right. *indicates samples stimulated by bacterial superantigen Staphlococcus enterotoxin B (SEB). B) Representative MEM labels for $CD4^+$ T cells (top) and B cells (bottom) from SEB-stimulated normal human blood (1.4, top, mass cytometry), normal human bone marrow (5, mass cytometry), normal human tonsil (2.5, mass cytometry), SEB-stimulated normal human blood (1.4, bottom, fluorescence flow cytometry), and normal human blood (6.1, mass cytometry).

An important aspect of MEM is generation of machine-readable quantitative labels that can be used to register population identities across samples and studies. A MEM label for a newly discovered population can be compared quantitatively against a reference set of established MEM labels or a MEM label reported in a

paper. To illustrate this idea, the pairwise, normalized root-mean-squared distance (RMSD) of MEM scores was calculated as a measure of similarity between 80 populations of cells from 7 different studies including healthy $CD4^+$ T cell and B cell (Figure E2). Cells had highly similar MEM scores within each major cell type, regardless of platform (mass or fluorescence flow cytometry), study, or tissue source. For example, T cells run on mass cytometry from different blood donors were 97% ± 1.3 similar to each other, 85% ± 1.9 similar to T cells from blood run on fluorescence flow cytometry, and 87% ± 2.1 similar to T cells from tonsil run on mass cytometry (Figure E2, Supplementary Table 5). However, these cells were 66.9% ± 13 similar to any B cell population. This indicates that MEM scores provide a way to communicate cell identity and to quantify similarities of cell types from the text label alone.

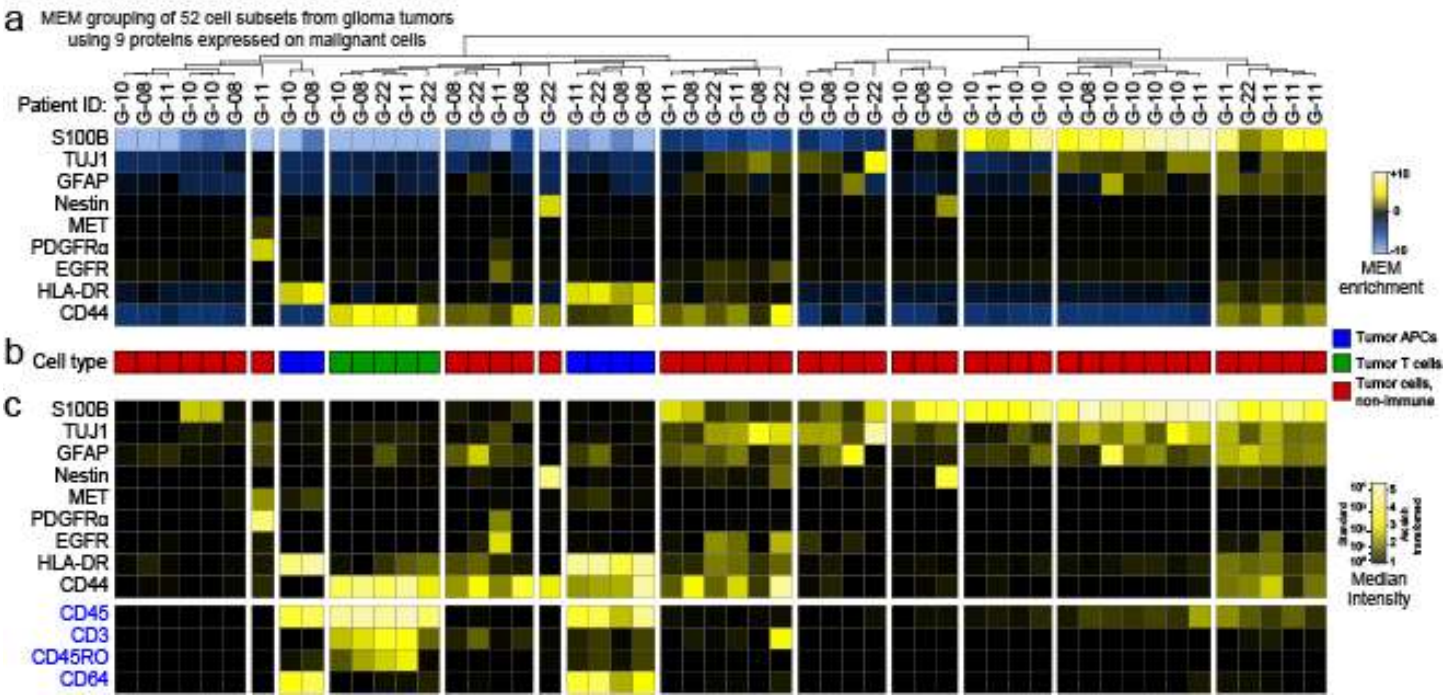


Figure E-3 MEM correctly grouped immune and cancer cell populations from glioma tumors using nine proteins expressed on cancer cells in Dataset D. (A) A heatmap of MEM enrichment scores is shown for 52 populations of cells identified in tumors from 4 glioblastoma patients (G-08, G-10, G-11, G22) in an unsupervised manner using viSNE. (B) Each population was annotated for a cell type based on review of the MEM label and classified as tumor infiltrating APCs (blue), tumor infiltrating T cells (green), or non-immune tumor cells (red). (C) A heatmap of median intensity values is shown for the 13 measured proteins from each of the 52 tumor cell populations.

Dataset D included 52 populations of tumor infiltrating APCs, tumor infiltrating T cells, and non-immune malignant tumor cells identified in human glioma tumors (*284*). To obtain these populations, each tumor was

analyzed by viSNE and cell subsets were expert gated solely on t-SNE cluster density (Supplementary Figure E6). To determine whether MEM could distinguish immune cell subsets from other tumor cell types with limited information, MEM scores were calculated using only 9 markers that were expected to be expressed on cancer cells (S100B, TUJ1, GFAP, Nestin, MET, PDGFRα, EGFR, HLA-DR, and CD44, Figure E3a). The 52 populations were grouped into 13 major cell types based on MEM enrichment of 9 analyzed proteins, and these groups were interpreted as tumor infiltrating APCs (Figure E3b, blue), tumor infiltrating T cells (Figure E3b, green), or non-immune tumor cells (Figure E3b, red). To confirm cell identity, four protein features that had been excluded from MEM analysis were assessed (Figure E3c, CD45, CD3, CD45RO, and CD64). CD45 and CD3 were used to confirm T cell identity and CD45 and CD64 were used to confirm APC identity. MEM correctly identified both immune cell subsets from all tumor types without using key immune lineage markers and without using healthy populations (e.g. APCs from blood or tonsil) to guide the clustering. Thus, MEM labels distinguished populations of cells based on non-traditional features and in a disease context.

MEM labels provided a quantitative language to objectively communicate characteristics of new and established cell types observed in complex tissue microenvironments. Algorithmic comparison of MEM labels correctly identified 80 cell populations from 7 studies of 3 human tissues measured using different instrumentation and distinguished tumor-infiltrating immune cell subsets and malignant cell populations from human glioma tumors. Following additional validation in other cell types, tissues, and instrumentation platforms, it may be possible for machines and humans to use MEM labels to learn and clearly communicate cell identity (cytotype). Given widespread adoption and reporting, MEM labels could be used to communicate cytotypes in a manner analogous to cluster of differentiation (CD) naming of antigen targets of antibodies (*396*). MEM can compare populations against a common reference (Supplementary Note 5) and guide feature selection for computational and experimental analysis. MEM can also be used to monitor changes in tissues over time during treatment. Deviation from a stable MEM score for peripheral blood cell subsets would be expected in the case of emerging malignant cells (*360*), and lack of change towards a healthy set of MEM scores for blood or bone marrow cell subsets might indicate a lack of response to chemotherapy for a leukemia patient. MEM is expected to assist in machine learning applications by providing quantitative text descriptions of cytotype that can be algorithmically parsed and used to classify newly identified cell subpopulations.

Methods

Code availability

Software for generating MEM scores is available as Supplementary Software.

CyTOF data preprocessing and analysis

Data analysis was performed using the online analysis platform Cytobank (*301*) and the statistical programming environment R. Raw median intensity (MI) values were transformed to a hyperbolic arcsine scale. A cofactor of 15 was used for the PBMC data set (Figure E1), and a cofactor of 5 was used for the normal human bone marrow data set and for the murine myeloid data set. Single, intact cells were gated based on event length and nucleic acid intercalator (iridium). Major PBMC subsets were gated based on CD45 expression (leukocytes) and on canonical lineage marker expression to identify major blood cell subsets.

FCS files were exported from Cytobank as FCS or tab-delimited text files that were parsed for expression intensity information using the R package flowCore (*464*). MEM was calculated using the arcsinh-transformed MI values, as described above. Heatmaps were generated using the heatmap.2 function in the gplots R package (*465*).

Fluorescence Phospho-Flow AML data analysis

Data were downloaded from Cytobank as FCS files and processed in R as described above. MFI values were transformed to a log-normal scale. For each AML patient, a median value and an IQR value was calculated for each marker in the unstimulated condition and for the stimulated conditions. The unstimulated median values were subtracted from the stimulated median values, and likewise for the IQR values. MEM was then calculated by comparing each patient's subtracted median and IQR values to those of the other patients. This enabled a comparison of fold-change signaling values rather than raw values.

Marker enrichment modeling

Marker enrichment modeling (MEM) analysis begins after populations have been identified and is designed to provide a simple way to compare findings from experts working with different platforms or performing

analysis using different computational tools for population discovery (*282, 295, 466-468*) and graphical visualization (*12, 80, 83, 288, 469*). These tools have differing strengths that depend greatly on the structure of the data sets and controls, the biological goals of the study, and the quality of the existing knowledge in the field (*314, 359, 394*).

MEM equation. The MEM equation is implemented as an R package (Supplementary Software). Currently, MEM uses medians as the magnitude value; however, depending on the data type, mean may be a more appropriate magnitude statistic, and mean could be substituted for median in the equation. Similarly, other statistics, such as variance, might be substituted for IQR. The MEM equation was developed with the intention of capturing and quantifying population-specific feature enrichment in a simple equation that avoids overfitting or unnecessary computation. The primary goal of this equation is to scale magnitude differences depending on distribution spread. While other distribution features such as skew or shape could be informative, incorporating only two pieces of information—magnitude and spread—into the equation captured enough information to be useful in quantifying both positive and negative population-specific feature enrichment.

MEM output and score scaling. The MEM R script outputs a heatmap of MEM values with a text label summary of feature enrichment as the population (row) names. The + or − value provided along with the marker name is converted to a −10 to +10 scale and rounded to the nearest integer. As implemented here, the maximum of the scale was set using the highest absolute value MEM score observed across all markers and populations. All values in the matrix are divided by this maximum value and multiplied by 10 to achieve the −10 to +10 scaling. After scaling, the original sign value is reapplied to each MEM score. Scaling the output this way is intended to generate MEM values and labels that are intuitive to human readers and to facilitate comparison of feature enrichment across experiments, samples, batches, timepoints, and data types.

IQR threshold. Because MEM uses a ratio of IQR values, near-zero values in the denominator, IQRPOP, will greatly increase MEM scores. For each measurement type, it is important to identify a minimum significant IQR value so that small IQR values below the platform's ability to distinguish signal from noise do not inappropriately increase MEM scores. To automatically determine a minimum threshold for IQRPOP, the algorithm here calculated the average of the IQR values that were associated with the lowest quartile of population and reference medians. For the mass and fluorescence cytometry data sets used, the automatically

calculated IQR threshold was on average 0.5 ± X, and so the IQR threshold for all studies here was set to 0.5. The default IQR threshold in the algorithm is also set to 0.5. To have the IQR threshold recalculated, investigators should specify the 'auto' option for the IQR.thresh argument in the MEM function. It is recommended that investigators applying MEM to data sets from different instruments or who are testing MEM for the first time determine whether a change in the IQR threshold is needed.

Reference population selection. MEM scores are contextual; a population's MEM score depends on the reference population(s) to which it is compared. Selection of a reference population should be made deliberately depending on the biological question being addressed. When populations in a MEM analysis arise from different experimental sources, it may be necessary in some cases to normalize measurements before MEM analysis to avoid artifacts from experimental variation.

PBMC processing and mass cytometry

PBMC were isolated and cryopreserved as described by Greenplate et al. (*360*). PBMC were stained with metal-conjugated antibodies and prepared for the mass cytometry as previously described (*360*). The following antibodies were used in the staining panel at a 1:2,000 dilution: CD19-142, CCR5-144, CD4-145, CD64-146, CD20-147, CCR4-149, CD43-150, CD14-151, TCRγδ-152, CD45RA-153, CD45-154, CXCR3-156, CD33-158, CCR7-159, CD28-169, CD29-162, CD45RO-164, CD16-165, CD44-166, CD27-167, CD8-168, CD25-169, CD3-170, CD57-172, PD-L1-175, and CD56-176 (Fluidigm Sciences). In addition, the following purified antibodies from Biolegend were labeled using MaxPar DN3 kits (Fludigm Sciences), stored at 4 °C in antibody stabilization buffer (Candor Bioscience GmbH) and used in the same panel at a 1:100 dilution: ICOS-141, TIM-143, CD38-148, CD32-161, HLA-DR-163, CXCR5-171, and PD-1-174.

Cell subpopulation MEM score similarity calculations

Comparison of CD4+T cells to B cells in Figure E2. In order to assess the robustness of MEM across tissue sample types, donors, experimental runs, and flow cytometry platforms (fluorescence and mass cytometry), MEM scores were calculated for cell subsets from seven different experiments that included three healthy human bone marrow samples (*83, 290, 374*), nine healthy human PBMC samples (*81, 470*), and six

healthy human tonsil samples (*82*). MEM scores were calculated for each population using as the reference population a combination of hematopoietic stem cells gated as $CD34^{+}$ $CD38^{lo/-}$ from two studies of healthy human bone marrow (*83, 374*). Population similarity was calculated using root mean squared distance (r.m.s. deviation) calculated on all population MEM scores in a pairwise fashion. MEM scores were calculated using all markers in common between each data set and the HSC reference (Supplementary Table 5).

r.m.s. deviation was calculated here as the square root of the average in squared distance between all MEM values in common for each pair of populations (Supplementary Table 5) and then converted into percent maximum possible r.m.s. deviation. Given the −10 to +10 MEM scale, an r.m.s. deviation of 20 was the maximum possible difference and corresponded to 0% similarity, whereas an r.m.s. deviation of 0 between MEM labels indicated 100% similarity. This approach emphasized differences in marker expression when comparing populations. Calculated statistics for $CD4^{+}$ T cell comparisons included average MEM value +/− s.d. and P value calculated using an unpaired, two-tailed Student's t-test.

Human glioma and normal immune cell MEM analysis. Glioblastoma data (G-08, G-10, G-11, and G-22) were collected following a published protocol16. Cells were stained with isotope-tagged antibodies to detect surface and intracellular targets following established protocols (*81, 284*). MEM analysis of glioblastoma patient samples was performed with nine markers (S100B, TUJ1, GFAP, Nestin, MET, PDGFRα, EGFR, HLA-DR, and CD44), using arcsinh transformation of original median intensity values with a cofactor of 5. Each cell subset was the POP, and the remaining cell subsets were the REF in the analysis.

Z-score and K–S statistic calculations

Z-score was calculated between POP and REF as (MEANpop-MEANref)/STDEVref for each marker.

The K–S statistic (*461, 471*) was calculated comparing the distribution for each marker on POP and REF using the function ks.test() in R.

F-measure analysis

PBMC populations were defined by expert human gating on canonical markers. For F-measure analysis (Figure E1c and Supplementary Figure E2), the 25 measured markers from the CyTOF analysis of healthy PBMC

were sorted based on absolute MEM scores, median values, median difference, Z-score, and K–S statistic (shown in Supplementary Figure E2), or randomly across all PBMC populations and the 25 measured proteins. The 5 × 25 matrix was converted into an ordered vector (length 25 × 5) and then sorted by absolute value. The first occurrence of each marker in the list was kept, and subsequent occurrences of that marker in the list (i.e., that marker's scores on other populations) were discarded. The order of markers excluded by MEM, median, median difference, Z-score, and K–S statistic are shown in Supplementary Table 3. Markers were then sequentially, cumulatively excluded from k-means clustering of cells from high to low absolute for each statistic or score. F-measure was calculated as:

$$\text{Sensitivity} = \text{true positives}/(\text{true positives} + \text{false negatives})$$
$$\text{Specificity} = \text{true negatives}/(\text{true negatives} + \text{false positives})$$
$$\text{F-measure} = 2 \times (\text{sensitivity} \times \text{specificity})/(\text{sensitivity} + \text{specificity})$$

An F-measure was calculated for each round of clustering, where truth was the cell cluster ID resulting from clustering on all 25 markers. The moving average of F-measure with an interval of 3 was calculated in Microsoft Excel. The F-measures for random marker exclusion are the average at each point of 15 different rounds of random marker exclusion from clustering.

Data Availability Statement

The normal human PBMC dataset (Figure E1) were generated by CyTOF analysis as described by Leelatian, et al. (*81*) and is available as an FCS file in Flow Repository (https://flowrepository.org/experiments/1043).

The normal human bone marrow data set from Bendall and Simonds, et al15 (Dataset B, Supplementary Note 3) was downloaded from Cytobank (*466*) as FCS files that included the cell population IDs defined by Bendall and Simonds, et al. (*83*) (https://reports.cytobank.org/1/v1). MEM enrichment scores from Dataset B were compared to the authors' analysis and prior studies of proteins marking stem cells, progenitor cells, and mature cells (*467, 468*).

The murine myeloid CyTOF dataset from Becher, et al (*294*) (Dataset C, Supplementary Note 4) was downloaded from Cytobank as FCS files that contained gated cell events and cluster IDs as designated by automated analysis conducted by Becher et al (*294*). MEM enrichment scores from Dataset C were compared to the authors' analysis and prior studies of neutrophils (*288, 295*).

Datasets for Figure E2 were generated in 7 separate fluorescence and mass cytometry studies by 1) Nicholas et al. (*470*), 2) Polikowsky et al. (*82*), 3) Ferrell et al. (*374*), 4) Amir et al. (*290*), 5) Bendall and Simonds et al. (*83*), 6) Greenplate et al., previously unpublished data, and 7) Leelatian et al (*81*).

The phospho-flow AML data set generated by Irish et al.6 (Supplementary Note 5-Fig.2) was downloaded from Cytobank as FCS files.

The human GBM mass cytometry dataset (Figure E3) was generated and analyzed as described by Leelatian and Doxie et al. (*284*) and are available on Flow Repository as text files (https://flowrepository.org/experiments/1044/).

Acknowledgements

This study was supported by R25 CA136440-04 (K.E.D.), F31 CA199993 (A.R.G.), R00 CA143231-03 (J.M.I.), the Vanderbilt-Ingram Cancer Center (VICC, P30 CA68485), VICC Ambassadors, a VICC Hematology Helping Hands award (J.M.I. and K.E.D.), and the Vanderbilt International Scholars Program (N.L.). Thanks to Mikael Roussel for helpful discussions of myeloid cell identity markers, to Deon Doxie for helpful discussions of MEM analysis of tumor and immune cell subsets, and to Lola Chambless and Rebecca Ihrie for use of glioma tumor data generated by N.L.

Author Contributions

All authors designed experiments, discussed data visualization and contributed intellectually to the manuscript, and approved the final manuscript. J.M.I. and K.E.D. performed computational analyses, developed analytical tools and protocols, conceived and designed the study, and wrote the manuscript. A.R.G. contributed to Figure E2 and Figure E3 and assisted with manuscript revisions. N.L. contributed to Figure E3 and manuscript revisions. C.E.W. contributed to R code implementation and manuscript revisions.

Competing Financial Interests Statement

J.M.I. is co-founder and board member and Cytobank Inc. and received research support from Incyte Corp.

Supplemental Data

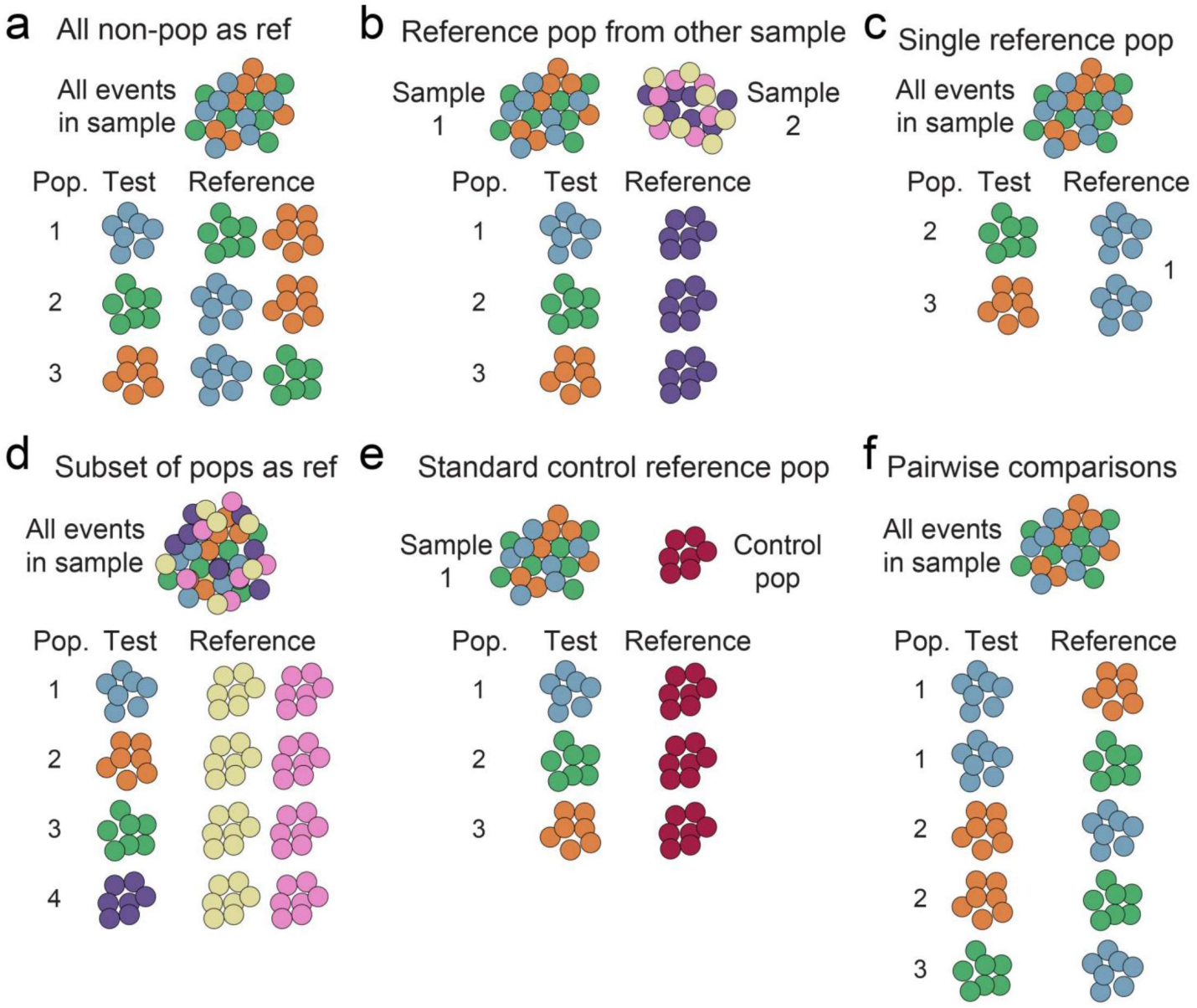


Figure SE-1 Examples of MEM reference population selection to capture different contexts

Alternative reference populations (REF) can be used to capture how features of the test population (POP) are enriched in different contexts. Reference comparisons include a) all non-population cells in the sample or experiment (default), b) a population from another sample in the same study, c) a population from the same sample, d) multiple subsets of non-population cells from the same sample, e) a standard control population, and f) pairwise comparison between all populations in a sample.

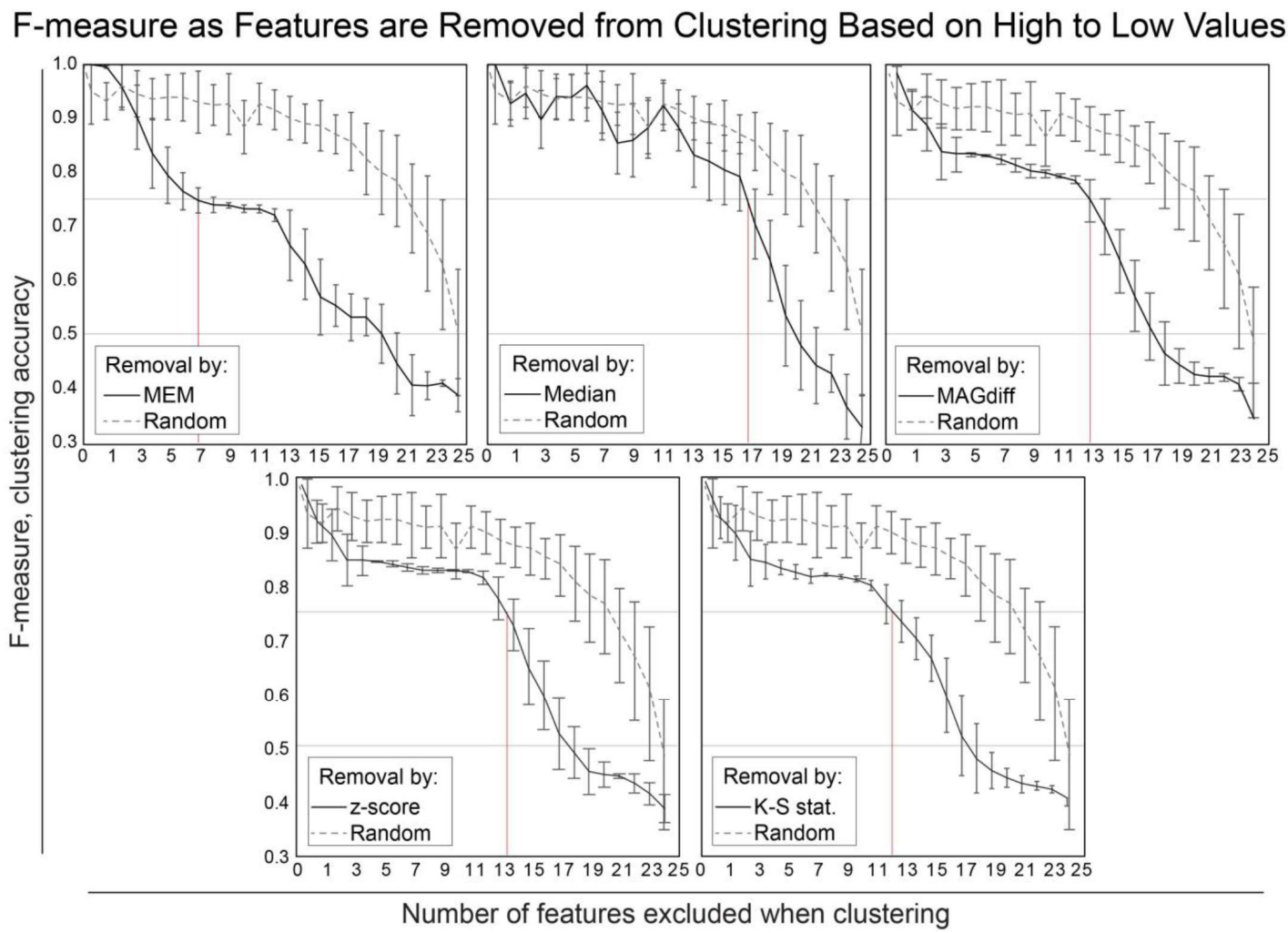


Figure SE-2 MEM highly scores markers that are important to clustering accuracy

Markers were sequentially and cumulatively excluded from k-means cluster analysis of Dataset A, from high to low, sorted based on 5 different statistics or scores (marker order shown in Supplementary Table 3): MEM, median, median difference (MAGDIFF), z-score, and Kolmogorov-Smirnov (K-S) statistic. Clustering accuracy was quantified as the f-measure where true cluster identity was assumed to be the clusters formed by clustering on all 25 markers in the dataset. The moving average of the f-measure is shown. Error bars represent the standard error. The vertical red line indicates the number of excluded features at which the f-measure reached 0.75.

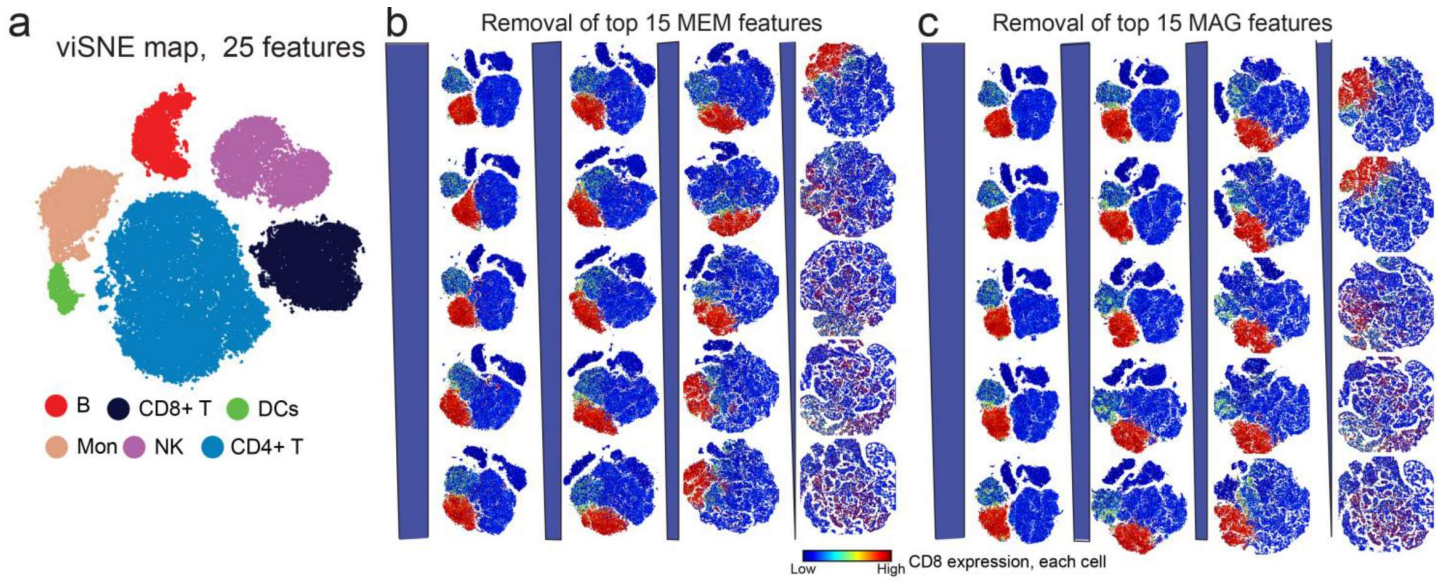


Figure SE-3 MEM highly scores markers that are important to viSNE mapping

a) viSNE map for healthy human blood, built using 25 surface protein markers. Populations were identified by expert analysis and color coded.

b). Top to bottom, left to right: viSNE maps generated as markers were iteratively, cumulatively excluded based on their MEM scores (high to low absolute value). Heat intensity for each cell indicates CD8 expression.

c) Top to bottom, left to right: viSNE maps generated as markers were iteratively, cumulatively excluded based on their median scores (high to low absolute value). Heat intensity for each cell indicates CD8 expression.

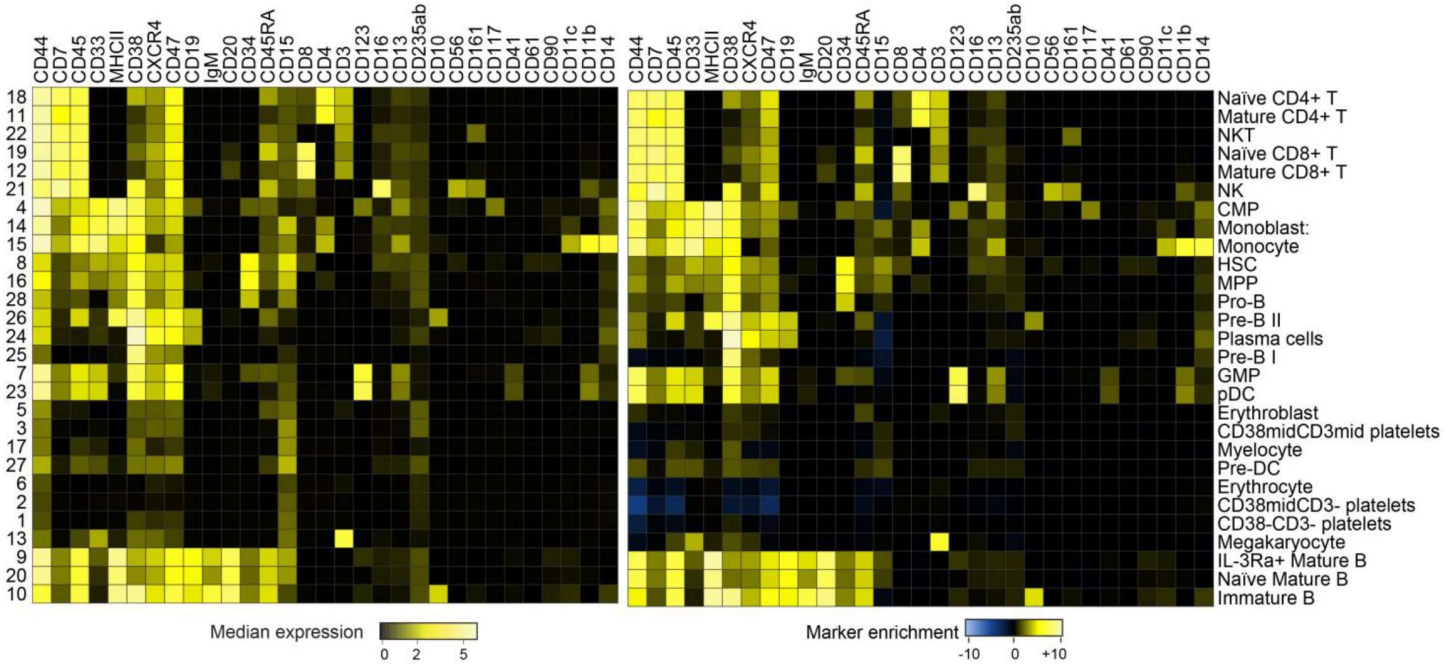


Figure SE-4 MEM scores largely reflect median expression values for relatively homogenous populations

Heatmaps show median intensity of protein expression (left) and protein enrichment by MEM (right) for measured proteins in 28 populations characterized as relatively homogeneous for established cell types by expert analysis (rows). Each population was compared to the other 27 subsets for the MEM analysis. MEM scores approach median expression values in homogeneous populations because the contribution of variance approaches zero.

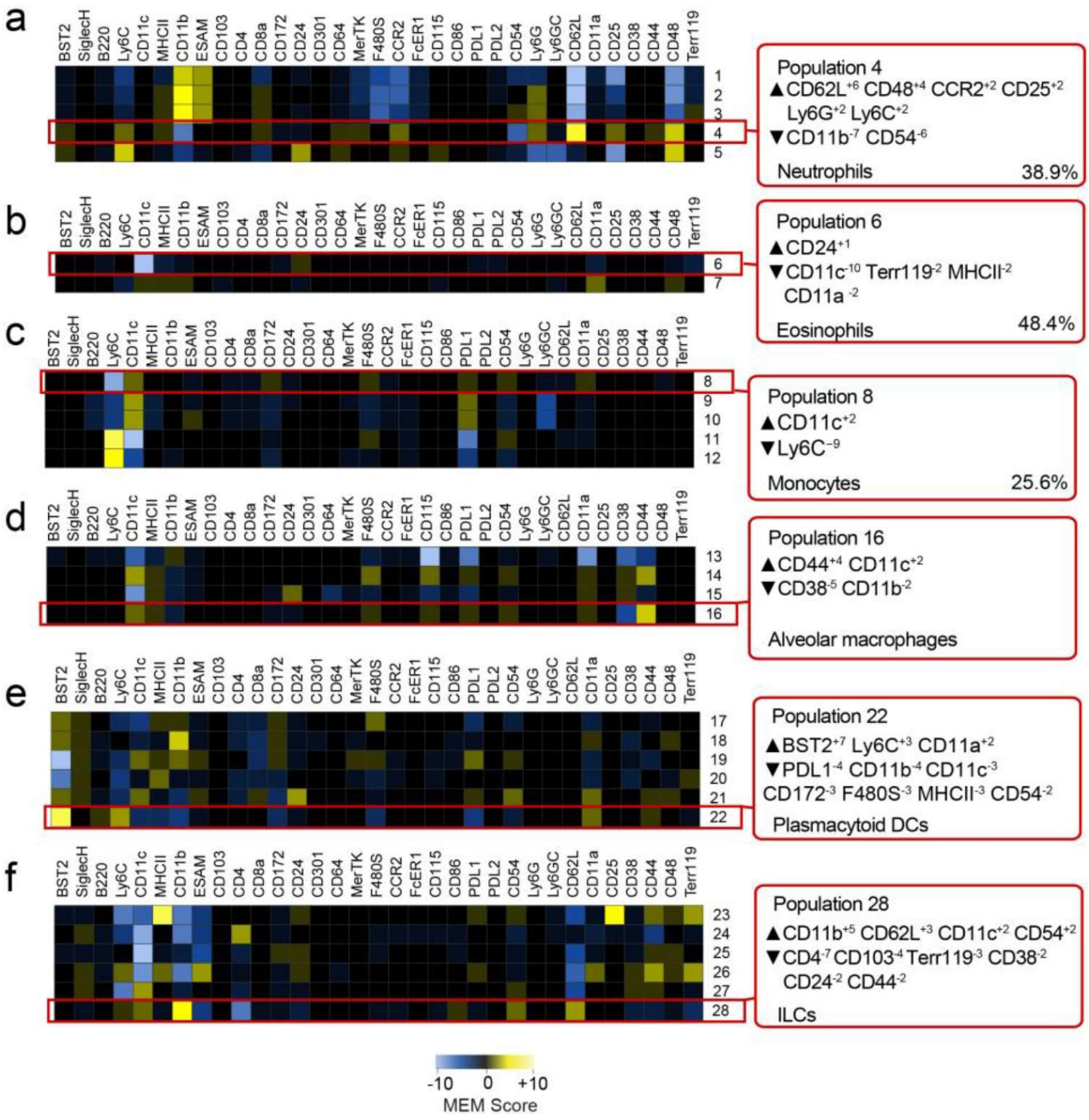


Figure SE-5 Focused MEM analysis quantifies feature enrichment within phenotypically similar groups of cells

a-f) Focused MEM analysis on murine myeloid cell subsets. A MEM label for one population within each group is shown as an example. Groups were defined as the 6 major murine subgroups identified by t-SNE and DensVM by Becher et al. (*294*).

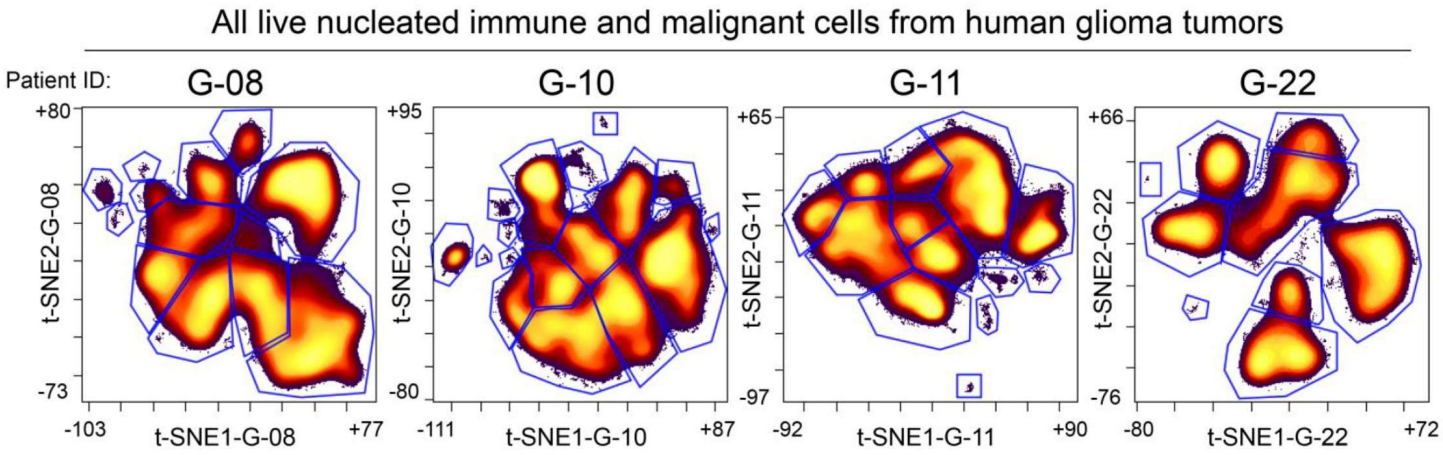


Figure SE-6 Unsupervised clustering and gating of 52 populations of malignant and immune cells in glioma

Live nucleated immune and malignant cells were gated from glioma tumors as described in Leelatian and Doxie et al., Cytometry B 20164. Patient-specific t-SNE axes were created in separate viSNE analyses of each tumor (e.g. t-SNE1-G-08 for glioma tumor G-08). Shown here is density of cells on t-SNE1 vs. t-SNE2 from each tumor-specific viSNE analysis. Expert analysis of density was then used to identify 52 cell clusters from the 4 glioma tumors. These 52 populations were subsequently grouped by MEM in Figure E5a using 9 proteins expressed on malignant cells.

BIBLIOGRAPHY

1. R. Stupp *et al.*, Radiotherapy plus concomitant and adjuvant temozolomide for glioblastoma. *N Engl J Med* **352**, 987-996 (2005).

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

3. C. W. Brennan *et al.*, The somatic genomic landscape of glioblastoma. *Cell* **155**, 462-477 (2013).

4. H. S. Phillips *et al.*, Molecular subclasses of high-grade glioma predict prognosis, delineate a pattern of disease progression, and resemble stages in neurogenesis. *Cancer Cell* **9**, 157-173 (2006).

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

6. M. A. Dawson, T. Kouzarides, Cancer epigenetics: from mechanism to therapy. *Cell* **150**, 12-27 (2012).

7. D. Hanahan, R. A. Weinberg, Hallmarks of Cancer: The Next Generation. *Cell* **144**, 646-674 (2011).

8. D. Hanahan, R. A. Weinberg, The hallmarks of cancer. *Cell* **100**, 57-70 (2000).

9. G. S. Martin, Cell signaling and cancer. *Cancer Cell* **4**, 167-174 (2003).

10. J. M. Irish, N. Kotecha, G. P. Nolan, Mapping normal and cancer cell signalling networks: towards single-cell proteomics. *Nature reviews. Cancer* **6**, 146-155 (2006).

11. J. M. Irish, D. B. Doxie, High-dimensional single-cell cancer biology. *Curr Top Microbiol Immunol* **377**, 1-21 (2014).

12. J. H. Levine *et al.*, Data-Driven Phenotypic Dissection of AML Reveals Progenitor-like Cells that Correlate with Prognosis. *Cell* **162**, 184-197 (2015).

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

14. N. Kotecha *et al.*, Single-cell profiling identifies aberrant STAT5 activation in myeloid malignancies with specific clinical and biologic correlates. *Cancer Cell* **14**, 335-343 (2008).

15. J. M. Irish *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* **107**, 12747-12754 (2010).

16. G. H. Heppner, Tumor heterogeneity. *Cancer Res* **44**, 2259-2265 (1984).

17. A. Marusyk, V. Almendro, K. Polyak, Intra-tumour heterogeneity: a looking glass for cancer? *Nat Rev Cancer* **12**, 323-334 (2012).

18. V. Almendro, A. Marusyk, K. Polyak, Cellular heterogeneity and molecular evolution in cancer. *Annu Rev Pathol* **8**, 277-302 (2013).

19. D. A. Arber *et al.*, The 2016 revision to the World Health Organization classification of myeloid neoplasms and acute leukemia. *Blood* **127**, 2391-2405 (2016).

20. S. H. Swerdlow *et al.*, The 2016 revision of the World Health Organization classification of lymphoid neoplasms. *Blood* **127**, 2375-2390 (2016).

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

22. H. Noushmehr *et al.*, Identification of a CpG island methylator phenotype that defines a distinct subgroup of glioma. *Cancer Cell* **17**, 510-522 (2010).

23. F. M. G. Cavalli *et al.*, Intertumoral Heterogeneity within Medulloblastoma Subgroups. *Cancer Cell* **31**, 737-754 e736 (2017).

24. P. A. Northcott *et al.*, The whole-genome landscape of medulloblastoma subtypes. *Nature* **547**, 311-317 (2017).

25. D. Sturm *et al.*, Hotspot mutations in H3F3A and IDH1 define distinct epigenetic and biological subgroups of glioblastoma. *Cancer Cell* **22**, 425-437 (2012).

26. H. Yan *et al.*, IDH1 and IDH2 mutations in gliomas. *N Engl J Med* **360**, 765-773 (2009).

27. C. M. Perou, J. S. Parker, A. Prat, M. J. Ellis, P. S. Bernard, Clinical implementation of the intrinsic subtypes of breast cancer. *Lancet Oncol* **11**, 718-719; author reply 720-711 (2010).

28. T. Sorlie *et al.*, Gene expression patterns of breast carcinomas distinguish tumor subclasses with clinical implications. *Proc Natl Acad Sci U S A* **98**, 10869-10874 (2001).

29. Q. Wang *et al.*, Tumor Evolution of Glioma-Intrinsic Gene Expression Subtypes Associates with Immunological Changes in the Microenvironment. *Cancer Cell* **32**, 42-56 e46 (2017).

30. M. Meyer *et al.*, Single cell-derived clonal analysis of human glioblastoma links functional and genomic heterogeneity. *Proc Natl Acad Sci U S A* **112**, 851-856 (2015).

31. L. R. Yates, P. J. Campbell, Evolution of the cancer genome. *Nat Rev Genet* **13**, 795-806 (2012).

32. L. Ding *et al.*, Clonal evolution in relapsed acute myeloid leukaemia revealed by whole-genome sequencing. *Nature* **481**, 506-510 (2012).

33. M. J. Lee *et al.*, Sequential application of anticancer drugs enhances cell death by rewiring apoptotic signaling networks. *Cell* **149**, 780-794 (2012).

34. S. Kobayashi *et al.*, EGFR mutation and resistance of non-small-cell lung cancer to gefitinib. *N Engl J Med* **352**, 786-792 (2005).

35. E. R. Fearon, B. Vogelstein, A genetic model for colorectal tumorigenesis. *Cell* **61**, 759-767 (1990).

36. I. J. Fidler, Tumor heterogeneity and the biology of cancer invasion and metastasis. *Cancer Res* **38**, 2651-2660 (1978).

37. I. Dagogo-Jack, A. T. Shaw, Tumour heterogeneity and resistance to cancer therapies. *Nat Rev Clin Oncol* **15**, 81-94 (2018).

38. N. H. Stoecklein, C. A. Klein, Genetic disparity between primary tumours, disseminated tumour cells, and manifest metastasis. *Int J Cancer* **126**, 589-598 (2010).

39. T. E. Becker *et al.*, The genomic heritage of lymph node metastases: implications for clinical management of patients with breast cancer. *Ann Surg Oncol* **15**, 1056-1063 (2008).

40. T. Kuukasjarvi *et al.*, Genetic heterogeneity and clonal evolution underlying development of asynchronous metastasis in human breast cancer. *Cancer Res* **57**, 1597-1604 (1997).

41. H. Bissig *et al.*, Evaluation of the clonal relationship between primary and metastatic renal cell carcinoma by comparative genomic hybridization. *Am J Pathol* **155**, 267-274 (1999).

42. A. Sottoriva *et al.*, Intratumor heterogeneity in human glioblastoma reflects cancer evolutionary dynamics. *Proc Natl Acad Sci U S A* **110**, 4009-4014 (2013).

43. A. S. Morrissy *et al.*, Spatial heterogeneity in medulloblastoma. *Nat Genet* **49**, 780-788 (2017).

44. M. Snuderl *et al.*, Mosaic amplification of multiple receptor tyrosine kinase genes in glioblastoma. *Cancer Cell* **20**, 810-817 (2011).

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

46. M. J. Son, K. Woolard, D. H. Nam, J. Lee, H. A. Fine, SSEA-1 is an enrichment marker for tumor-initiating cells in human glioblastoma. *Cell Stem Cell* **4**, 440-452 (2009).

47. N. J. Szerlip *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* **109**, 3041-3046 (2012).

48. W. Wei *et al.*, Single-Cell Phosphoproteomics Resolves Adaptive Signaling Dynamics and Informs Targeted Combination Therapy in Glioblastoma. *Cancer Cell* **29**, 563-573 (2016).

49. S. Bao *et al.*, Glioma stem cells promote radioresistance by preferential activation of the DNA damage response. *Nature* **444**, 756-760 (2006).

50. M. H. Spitzer, G. P. Nolan, Mass Cytometry: Single Cells, Many Features. *Cell* **165**, 780-791 (2016).

51. D. Hanahan, R. A. Weinberg, Hallmarks of cancer: the next generation. *Cell* **144**, 646-674 (2011).

52. S. L. Carter, A. C. Eklund, I. S. Kohane, L. N. Harris, Z. Szallasi, A signature of chromosomal instability inferred from gene expression profiles predicts clinical outcome in multiple human cancers. *Nat Genet* **38**, 1043-1048 (2006).

53. B. Vogelstein, K. W. Kinzler, Cancer genes and the pathways they control. *Nat Med* **10**, 789-799 (2004).

54. N. McGranahan, C. Swanton, Clonal Heterogeneity and Tumor Evolution: Past, Present, and the Future. *Cell* **168**, 613-628 (2017).

55. S. Nik-Zainal *et al.*, Landscape of somatic mutations in 560 breast cancer whole-genome sequences. *Nature* **534**, 47-54 (2016).

56. N. Cancer Genome Atlas Research, Comprehensive molecular characterization of urothelial bladder carcinoma. *Nature* **507**, 315-322 (2014).

57. G. Ciriello *et al.*, Emerging landscape of oncogenic signatures across human cancers. *Nat Genet* **45**, 1127-1133 (2013).

58. I. Bozic *et al.*, Accumulation of driver and passenger mutations during tumor progression. *Proc Natl Acad Sci U S A* **107**, 18545-18550 (2010).

59. A. M. Arias, P. Hayward, Filtering transcriptional noise during development: concepts and mechanisms. *Nat Rev Genet* **7**, 34-44 (2006).

60. T. Mazor, A. Pankov, J. S. Song, J. F. Costello, Intratumoral Heterogeneity of the Epigenome. *Cancer Cell* **29**, 440-451 (2016).

61. R. A. Irizarry *et al.*, The human colon cancer methylome shows similar hypo- and hypermethylation at conserved tissue-specific CpG island shores. *Nat Genet* **41**, 178-186 (2009).

62. A. E. Jaffe, R. A. Irizarry, Accounting for cellular heterogeneity is critical in epigenome-wide association studies. *Genome Biol* **15**, R31 (2014).

63. S. B. Baylin, DNA methylation and gene silencing in cancer. *Nat Clin Pract Oncol* **2 Suppl 1**, S4-11 (2005).

64. M. Ehrlich, DNA methylation in cancer: too much, but also too little. *Oncogene* **21**, 5400-5413 (2002).

65. S. Turcan *et al.*, IDH1 mutation is sufficient to establish the glioma hypermethylator phenotype. *Nature* **483**, 479-483 (2012).

66. F. Gaudet *et al.*, Induction of tumors in mice by genomic hypomethylation. *Science* **300**, 489-492 (2003).

67. Y. Qu *et al.*, Cancer-specific changes in DNA methylation reveal aberrant silencing and activation of enhancers in leukemia. *Blood* **129**, e13-e25 (2017).

68. S. X. Pfister, A. Ashworth, Marked for death: targeting epigenetic changes in cancer. *Nat Rev Drug Discov* **16**, 241-263 (2017).

69. R. M. Campbell, P. J. Tummino, Cancer epigenetics drug discovery and development: the challenge of hitting the mark. *J Clin Invest* **124**, 64-69 (2014).

70. S. S. Dey, L. Kester, B. Spanjaard, M. Bienko, A. van Oudenaarden, Integrated genome and transcriptome sequencing of the same cell. *Nat Biotechnol* **33**, 285-289 (2015).

71. D. A. Jaitin *et al.*, Massively parallel single-cell RNA-seq for marker-free decomposition of tissues into cell types. *Science* **343**, 776-779 (2014).

72. I. Tirosh *et al.*, Dissecting the multicellular ecosystem of metastatic melanoma by single-cell RNA-seq. *Science* **352**, 189-196 (2016).

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

74. B. Artegiani *et al.*, A Single-Cell RNA Sequencing Study Reveals Cellular and Molecular Dynamics of the Hippocampal Neurogenic Niche. *Cell Rep* **21**, 3271-3284 (2017).

75. B. Schwanhausser *et al.*, Global quantification of mammalian gene expression control. *Nature* **473**, 337-342 (2011).

76. T. Maier, M. Guell, L. Serrano, Correlation of mRNA and protein in complex biological samples. *FEBS Lett* **583**, 3966-3973 (2009).

77. M. Grzmil, B. A. Hemmings, Translation regulation as a therapeutic target in cancer. *Cancer Res* **72**, 3891-3900 (2012).

78. J. Schlessinger, Cell signaling by receptor tyrosine kinases. *Cell* **103**, 211-225 (2000).

79. A. Califano, M. J. Alvarez, The recurrent architecture of tumour initiation, progression and drug sensitivity. *Nat Rev Cancer* **17**, 116-130 (2017).

80. J. M. Irish *et al.*, Single cell profiling of potentiated phospho-protein networks in cancer cells. *Cell* **118**, 217-228 (2004).

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

82. H. G. Polikowsky, C. E. Wogsland, K. E. Diggins, K. Huse, J. M. Irish, Cutting Edge: Redox Signaling Hypersensitivity Distinguishes Human Germinal Center B Cells. *J Immunol* **195**, 1364-1367 (2015).

83. S. C. Bendall *et al.*, Single-cell mass cytometry of differential immune and drug responses across a human hematopoietic continuum. *Science* **332**, 687-696 (2011).

84. P. O. Krutzik, J. M. Irish, G. P. Nolan, O. D. Perez, Analysis of protein phosphorylation and cellular signaling events by flow cytometry: techniques and clinical applications. *Clin Immunol* **110**, 206-221 (2004).

85. O. D. Perez, G. P. Nolan, Phospho-proteomic immune analysis by flow cytometry: from mechanism to translational medicine at the single-cell level. *Immunol Rev* **210**, 208-228 (2006).

86. A. Ivey *et al.*, Assessment of Minimal Residual Disease in Standard-Risk AML. *N Engl J Med* **374**, 422-433 (2016).

87. E. Piovan *et al.*, Direct reversal of glucocorticoid resistance by AKT inhibition in acute lymphoblastic leukemia. *Cancer Cell* **24**, 766-776 (2013).

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

89. A. P. Kyritsis, M. L. Bondy, J. S. Rao, C. Sioka, Inherited predisposition to glioma. *Neuro Oncol* **12**, 104-113 (2010).

90. F. P. Li *et al.*, A cancer family syndrome in twenty-four kindreds. *Cancer Res* **48**, 5358-5362 (1988).

91. D. H. Gutmann *et al.*, Molecular analysis of astrocytomas presenting after age 10 in individuals with NF1. *Neurology* **61**, 1397-1400 (2003).

92. D. G. Evans, M. Sainio, M. E. Baser, Neurofibromatosis type 2. *J Med Genet* **37**, 897-904 (2000).

93. S. R. Hamilton *et al.*, The molecular basis of Turcot's syndrome. *N Engl J Med* **332**, 839-847 (1995).

94. S. M. Chang *et al.*, Patterns of care for adults with newly diagnosed malignant glioma. *JAMA* **293**, 557-564 (2005).

95. J. N. Scott, P. M. Brasher, R. J. Sevick, N. B. Rewcastle, P. A. Forsyth, How often are nonenhancing supratentorial gliomas malignant? A population study. *Neurology* **59**, 947-949 (2002).

96. O. Ganslandt *et al.*, Magnetic source imaging supports clinical decision making in glioma patients. *Clin Neurol Neurosurg* **107**, 20-26 (2004).

97. D. N. Louis *et al.*, International Society Of Neuropathology--Haarlem consensus guidelines for nervous system tumor classification and grading. *Brain Pathol* **24**, 429-435 (2014).

98. H. Gittleman *et al.*, An independently validated nomogram for individualized estimation of survival among patients with newly diagnosed glioblastoma: NRG Oncology RTOG 0525 and 0825. *Neuro Oncol* **19**, 669-677 (2017).

99. D. N. Louis *et al.*, *WHO Classification of Tumours of the Central Nervous System*. H. O. David N. Louis, Otmar D. Weistler, Webster K. Cavenee, Ed., (International Agency for Research on Cancer (IARC), France, ed. 4th, 2016).

100. J. W. Kernohan, R. F. Mabon, et al., A simplified classification of the gliomas. *Proc Staff Meet Mayo Clin* **24**, 71-75 (1949).

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

102. M. J. van den Bent, Interobserver variation of the histopathological diagnosis in clinical trials on glioma: a clinician's perspective. *Acta Neuropathol* **120**, 297-304 (2010).

103. C. E. Napier *et al.*, ATRX represses alternative lengthening of telomeres. *Oncotarget* **6**, 16543-16558 (2015).

104. B. Wiestler *et al.*, ATRX loss refines the classification of anaplastic gliomas and identifies a subgroup of IDH mutant astrocytic tumors with better prognosis. *Acta neuropathologica* **126**, 443-451 (2013).

105. C. Koschmann *et al.*, ATRX loss promotes tumor growth and impairs nonhomologous end joining DNA repair in glioma. *Sci Transl Med* **8**, 328ra328 (2016).

106. J. E. Eckel-Passow *et al.*, Glioma Groups Based on 1p/19q, IDH, and TERT Promoter Mutations in Tumors. *N Engl J Med* **372**, 2499-2508 (2015).

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

108. I. K. Mellinghoff *et al.*, Molecular determinants of the response of glioblastomas to EGFR kinase inhibitors. *The New England journal of medicine* **353**, 2012-2024 (2005).

109. O. L. Chinot *et al.*, Bevacizumab plus radiotherapy-temozolomide for newly diagnosed glioblastoma. *N Engl J Med* **370**, 709-722 (2014).

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

111. R. Stupp *et al.*, Effects of radiotherapy with concomitant and adjuvant temozolomide versus radiotherapy alone on survival in glioblastoma in a randomised phase III study: 5-year analysis of the EORTC-NCIC trial. *Lancet Oncol* **10**, 459-466 (2009).

112. H. Ohgaki *et al.*, Genetic pathways to glioblastoma: a population-based study. *Cancer research* **64**, 6892-6899 (2004).

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

114. M. M. Grabowski *et al.*, Residual tumor volume versus extent of resection: predictors of survival after surgery for glioblastoma. *Journal of neurosurgery* **121**, 1115-1123 (2014).

115. W. Stummer *et al.*, Fluorescence-guided resection of glioblastoma multiforme by using 5-aminolevulinic acid-induced porphyrins: a prospective study in 52 consecutive patients. *J Neurosurg* **93**, 1003-1013 (2000).

116. W. Stummer *et al.*, Fluorescence-guided surgery with 5-aminolevulinic acid for resection of malignant glioma: a randomised controlled multicentre phase III trial. *Lancet Oncol* **7**, 392-401 (2006).

117. P. P. Panciani *et al.*, Fluorescence and image guided resection in high grade glioma. *Clin Neurol Neurosurg* **114**, 37-41 (2012).

118. C. G. Hadjipanayis, G. Widhalm, W. Stummer, What is the Surgical Benefit of Utilizing 5-Aminolevulinic Acid for Fluorescence-Guided Surgery of Malignant Gliomas? *Neurosurgery* **77**, 663-673 (2015).

119. S. Santagata *et al.*, Intraoperative mass spectrometry mapping of an onco-metabolite to guide brain tumor surgery. *Proc Natl Acad Sci U S A* **111**, 11121-11126 (2014).

120. J. Balog *et al.*, Intraoperative tissue identification using rapid evaporative ionization mass spectrometry. *Sci Transl Med* **5**, 194ra193 (2013).

121. M. D. Walker *et al.*, Evaluation of BCNU and/or radiotherapy in the treatment of anaplastic gliomas. A cooperative clinical trial. *J Neurosurg* **49**, 333-343 (1978).

122. L. C. Hygino da Cruz, Jr., I. Rodriguez, R. C. Domingues, E. L. Gasparetto, A. G. Sorensen, Pseudoprogression and pseudoresponse: imaging challenges in the assessment of posttreatment glioma. *AJNR Am J Neuroradiol* **32**, 1978-1985 (2011).

123. A. A. Brandes *et al.*, MGMT promoter methylation status can predict the incidence and outcome of pseudoprogression after concomitant radiochemotherapy in newly diagnosed glioblastoma patients. *J Clin Oncol* **26**, 2192-2197 (2008).

124. W. Taal *et al.*, Incidence of early pseudo-progression in a cohort of malignant glioma patients treated with chemoirradiation with temozolomide. *Cancer* **113**, 405-410 (2008).

125. R. Stupp *et al.*, Maintenance Therapy With Tumor-Treating Fields Plus Temozolomide vs Temozolomide Alone for Glioblastoma: A Randomized Clinical Trial. *JAMA* **314**, 2535-2543 (2015).

126. R. Stupp *et al.*, Effect of Tumor-Treating Fields Plus Maintenance Temozolomide vs Maintenance Temozolomide Alone on Survival in Patients With Glioblastoma: A Randomized Clinical Trial. *JAMA* **318**, 2306-2316 (2017).

127. M. Giladi *et al.*, Mitotic Spindle Disruption by Alternating Electric Fields Leads to Improper Chromosome Segregation and Mitotic Catastrophe in Cancer Cells. *Sci Rep* **5**, 18046 (2015).

128. E. D. Kirson *et al.*, Alternating electric fields arrest cell proliferation in animal tumor models and human brain tumors. *Proc Natl Acad Sci U S A* **104**, 10152-10157 (2007).

129. A. F. Hottinger, P. Pacheco, R. Stupp, Tumor treating fields: a novel treatment modality and its use in brain tumors. *Neuro Oncol* **18**, 1338-1349 (2016).

130. F. Bernard-Arnoux *et al.*, The cost-effectiveness of tumor-treating fields therapy in patients with newly diagnosed glioblastoma. *Neuro Oncol* **18**, 1129-1136 (2016).

131. R. Brisman, E. M. Housepian, C. Chang, P. Duffy, E. Balis, Adjuvant nitrosourea therapy for glioblastoma. *Arch Neurol* **33**, 745-750 (1976).

132. C. L. Solero *et al.*, Controlled study with BCNU vs. CCNU as adjuvant chemotherapy following surgery plus radiotherapy for glioblastoma multiforme. *Cancer Clin Trials* **2**, 43-48 (1979).

133. S. S. Agarwala, J. M. Kirkwood, Temozolomide, a novel alkylating agent with activity in the central nervous system, may improve the treatment of advanced metastatic melanoma. *Oncologist* **5**, 144-151 (2000).

134. J. Dinnes, C. Cave, S. Huang, R. Milne, A rapid and systematic review of the effectiveness of temozolomide for the treatment of recurrent malignant glioma. *Br J Cancer* **86**, 501-505 (2002).

135. C. K. Park *et al.*, The Changes in MGMT Promoter Methylation Status in Initial and Recurrent Glioblastomas. *Transl Oncol* **5**, 393-397 (2012).

136. N. Cancer Genome Atlas Research, Comprehensive genomic characterization defines human glioblastoma genes and core pathways. *Nature* **455**, 1061-1068 (2008).

137. A. J. Wong *et al.*, Structural alterations of the epidermal growth factor receptor gene in human gliomas. *Proceedings of the National Academy of Sciences of the United States of America* **89**, 2965-2969 (1992).

138. H. S. Huang *et al.*, The enhanced tumorigenic activity of a mutant epidermal growth factor receptor common in human cancers is mediated by threshold levels of constitutive tyrosine phosphorylation and unattenuated signaling. *The Journal of biological chemistry* **272**, 2927-2935 (1997).

139. K. J. Hatanpaa, S. Burma, D. Zhao, A. A. Habib, Epidermal growth factor receptor in glioma: signal transduction, neuropathology, imaging, and radioresistance. *Neoplasia* **12**, 675-684 (2010).

140. Q. W. Fan *et al.*, EGFR phosphorylates tumor-derived EGFRvIII driving STAT3/5 and progression in glioblastoma. *Cancer Cell* **24**, 438-449 (2013).

141. A. M. Xu, P. H. Huang, Receptor tyrosine kinase coactivation networks in cancer. *Cancer Res* **70**, 3857-3860 (2010).

142. J. M. Stommel *et al.*, Coactivation of receptor tyrosine kinases affects the response of tumor cells to targeted therapies. *Science* **318**, 287-290 (2007).

143. G. Guo *et al.*, A TNF-JNK-Axl-ERK signaling axis mediates primary resistance to EGFR inhibition in glioblastoma. *Nat Neurosci* **20**, 1074-1084 (2017).

144. B. Neyns *et al.*, Stratified phase II trial of cetuximab in patients with recurrent high-grade glioma. *Ann Oncol* **20**, 1596-1603 (2009).

145. M. Preusser, M. Lim, D. A. Hafler, D. A. Reardon, J. H. Sampson, Prospects of immune checkpoint modulators in the treatment of glioblastoma. *Nat Rev Neurol* **11**, 504-514 (2015).

146. A. Lai *et al.*, Phase II study of bevacizumab plus temozolomide during and after radiation therapy for patients with newly diagnosed glioblastoma multiforme. *J Clin Oncol* **29**, 142-148 (2011).

147. A. B. Lassman *et al.*, Molecular study of malignant gliomas treated with epidermal growth factor receptor inhibitors: tissue analysis from North American Brain Tumor Consortium Trials 01-03 and 00-01. *Clin Cancer Res* **11**, 7841-7850 (2005).

148. Q. W. Fan, W. A. Weiss, Targeting the RTK-PI3K-mTOR axis in malignant glioma: overcoming resistance. *Curr Top Microbiol Immunol* **347**, 279-296 (2010).

149. Y. C. Chae *et al.*, Mitochondrial Akt Regulation of Hypoxic Tumor Reprogramming. *Cancer Cell* **30**, 257-272 (2016).

150. R. L. Elstrom *et al.*, Akt stimulates aerobic glycolysis in cancer cells. *Cancer Res* **64**, 3892-3899 (2004).

151. M. G. Vander Heiden, L. C. Cantley, C. B. Thompson, Understanding the Warburg effect: the metabolic requirements of cell proliferation. *Science* **324**, 1029-1033 (2009).

152. M. G. Vander Heiden *et al.*, Growth factors can influence cell growth and survival through effects on glucose metabolism. *Mol Cell Biol* **21**, 5899-5912 (2001).

153. O. Warburg, On the origin of cancer cells. *Science* **123**, 309-314 (1956).

154. A. Barthel *et al.*, Regulation of GLUT1 gene transcription by the serine/threonine kinase Akt1. *J Biol Chem* **274**, 20281-20286 (1999).

155. J. Deprez, D. Vertommen, D. R. Alessi, L. Hue, M. H. Rider, Phosphorylation and activation of heart 6-phosphofructo-2-kinase by protein kinase B and other protein kinases of the insulin signaling cascades. *J Biol Chem* **272**, 17269-17275 (1997).

156. H. Makinoshima *et al.*, Signaling through the Phosphatidylinositol 3-Kinase (PI3K)/Mammalian Target of Rapamycin (mTOR) Axis Is Responsible for Aerobic Glycolysis mediated by Glucose Transporter in Epidermal Growth Factor Receptor (EGFR)-mutated Lung Adenocarcinoma. *J Biol Chem* **290**, 17495-17504 (2015).

157. Y. Zhang *et al.*, A Pan-Cancer Proteogenomic Atlas of PI3K/AKT/mTOR Pathway Alterations. *Cancer Cell* **31**, 820-832 e823 (2017).

158. J. Sunayama *et al.*, Crosstalk between the PI3K/mTOR and MEK/ERK pathways involved in the maintenance of self-renewal and tumorigenicity of glioblastoma stem-like cells. *Stem Cells* **28**, 1930-1939 (2010).

159. J. Sunayama *et al.*, Dual blocking of mTor and PI3K elicits a prodifferentiation effect on glioblastoma stem-like cells. *Neuro Oncol* **12**, 1205-1219 (2010).

160. G. Bollag *et al.*, Loss of NF1 results in activation of the Ras signaling pathway and leads to aberrant growth in haematopoietic cells. *Nat Genet* **12**, 144-148 (1996).

161. S. Dahiya *et al.*, BRAF-V600E mutation in pediatric and adult glioblastoma. *Neuro Oncol* **16**, 318-319 (2014).

162. A. R. Grassian, C. M. Metallo, J. L. Coloff, G. Stephanopoulos, J. S. Brugge, Erk regulation of pyruvate dehydrogenase flux through PDK4 modulates cell proliferation. *Genes Dev* **25**, 1716-1733 (2011).

163. S. Tanaka, D. N. Louis, W. T. Curry, T. T. Batchelor, J. Dietrich, Diagnostic and therapeutic avenues for glioblastoma: no longer a dead end? *Nat Rev Clin Oncol* **10**, 14-26 (2013).

164. O. Bogler, H. J. Huang, W. K. Cavenee, Loss of wild-type p53 bestows a growth advantage on primary cortical astrocytes and facilitates their in vitro transformation. *Cancer Res* **55**, 2746-2751 (1995).

165. H. Ohgaki, P. Kleihues, Genetic pathways to primary and secondary glioblastoma. *Am J Pathol* **170**, 1445-1453 (2007).

166. G. Reifenberger, L. Liu, K. Ichimura, E. E. Schmidt, V. P. Collins, Amplification and overexpression of the MDM2 gene in a subset of human malignant gliomas without p53 mutations. *Cancer Res* **53**, 2736-2739 (1993).

167. R. Nishikawa *et al.*, Loss of P16INK4 expression is frequent in high grade gliomas. *Cancer Res* **55**, 1941-1945 (1995).

168. J. W. Henson *et al.*, The retinoblastoma gene is involved in malignant progression of astrocytomas. *Ann Neurol* **36**, 714-721 (1994).

169. N. Nonoguchi *et al.*, TERT promoter mutations in primary and secondary glioblastomas. *Acta Neuropathol* **126**, 931-937 (2013).

170. P. J. Killela *et al.*, TERT promoter mutations occur frequently in gliomas and a subset of tumors derived from cells with low rates of self-renewal. *Proc Natl Acad Sci U S A* **110**, 6021-6026 (2013).

171. Y. S. Cong, W. E. Wright, J. W. Shay, Human telomerase and its regulation. *Microbiol Mol Biol Rev* **66**, 407-425, table of contents (2002).

172. C. M. Heaphy *et al.*, Prevalence of the alternative lengthening of telomeres telomere maintenance mechanism in human cancer subtypes. *Am J Pathol* **179**, 1608-1615 (2011).

173. D. N. Nguyen *et al.*, Molecular and morphologic correlates of the alternative lengthening of telomeres phenotype in high-grade astrocytomas. *Brain Pathol* **23**, 237-243 (2013).

174. K. Bower *et al.*, Loss of wild-type ATRX expression in somatic cell hybrids segregates with activation of Alternative Lengthening of Telomeres. *PLoS One* **7**, e50062 (2012).

175. J. W. Leung *et al.*, Alpha thalassemia/mental retardation syndrome X-linked gene product ATRX is required for proper replication restart and cellular resistance to replication stress. *J Biol Chem* **288**, 6342-6350 (2013).

176. A. J. Levine, A. M. Puzio-Kuter, The control of the metabolic switch in cancers by oncogenes and tumor suppressor genes. *Science* **330**, 1340-1344 (2010).

177. M. Strickland, E. A. Stoll, Metabolic Reprogramming in Glioma. *Front Cell Dev Biol* **5**, 43 (2017).

178. D. W. Parsons *et al.*, An integrated genomic analysis of human glioblastoma multiforme. *Science* **321**, 1807-1812 (2008).

179. H. Yang, D. Ye, K. L. Guan, Y. Xiong, IDH1 and IDH2 mutations in tumorigenesis: mechanistic insights and clinical perspectives. *Clin Cancer Res* **18**, 5562-5571 (2012).

180. H. Yan, D. D. Bigner, V. Velculescu, D. W. Parsons, Mutant metabolic enzymes are at the origin of gliomas. *Cancer Res* **69**, 9157-9159 (2009).

181. J. G. Scott *et al.*, Production of 2-hydroxyglutarate by isocitrate dehydrogenase 1-mutated gliomas: an evolutionary alternative to the Warburg shift? *Neuro Oncol* **13**, 1262-1264 (2011).

182. L. Dang *et al.*, Cancer-associated IDH1 mutations produce 2-hydroxyglutarate. *Nature* **462**, 739-744 (2009).

183. A. L. Cohen, S. L. Holmen, H. Colman, IDH1 and IDH2 mutations in gliomas. *Curr Neurol Neurosci Rep* **13**, 345 (2013).

184. C. Lu *et al.*, IDH mutation impairs histone demethylation and results in a block to cell differentiation. *Nature* **483**, 474-478 (2012).

185. T. C. Pansuriya *et al.*, Somatic mosaic IDH1 and IDH2 mutations are associated with enchondroma and spindle cell hemangioma in Ollier disease and Maffucci syndrome. *Nat Genet* **43**, 1256-1261 (2011).

186. P. Koivunen *et al.*, Transformation by the (R)-enantiomer of 2-hydroxyglutarate linked to EGLN activation. *Nature* **483**, 484-488 (2012).

187. G. P. Dunn *et al.*, Emerging insights into the molecular and cellular basis of glioblastoma. *Genes Dev* **26**, 756-784 (2012).

188. L. C. Hou, A. Veeravagu, A. R. Hsu, V. C. Tse, Recurrent glioblastoma multiforme: a review of natural history and management options. *Neurosurg Focus* **20**, E5 (2006).

189. M. Lun, E. Lok, S. Gautam, E. Wu, E. T. Wong, The natural history of extracranial metastasis from glioblastoma multiforme. *J Neurooncol* **105**, 261-273 (2011).

190. P. C. Burger, E. R. Heinz, T. Shibata, P. Kleihues, Topographic anatomy and CT correlations in the untreated glioblastoma multiforme. *J Neurosurg* **68**, 698-704 (1988).

191. Y. Kim, S. Kumar, CD44-mediated adhesion to hyaluronic acid contributes to mechanosensing and invasive motility. *Mol Cancer Res* **12**, 1416-1429 (2014).

192. R. L. Klank *et al.*, Biphasic Dependence of Glioma Survival and Cell Migration on CD44 Expression Level. *Cell Rep* **18**, 23-31 (2017).

193. B. Kaur *et al.*, Hypoxia and the hypoxia-inducible-factor pathway in glioma growth and angiogenesis. *Neuro Oncol* **7**, 134-153 (2005).

194. J. S. Lee *et al.*, Oligodendrocyte progenitor cell numbers and migration are regulated by the zebrafish orthologs of the NF1 tumor suppressor gene. *Hum Mol Genet* **19**, 4643-4653 (2010).

195. M. Lenarduzzi *et al.*, MicroRNA-193b enhances tumor progression via down regulation of neurofibromin 1. *PLoS One* **8**, e53765 (2013).

196. V. Orian-Rousseau, L. Chen, J. P. Sleeman, P. Herrlich, H. Ponta, CD44 is required for two consecutive steps in HGF/c-Met signaling. *Genes Dev* **16**, 3074-3086 (2002).

197. M. Zoller, CD44: can a cancer-initiating cell profit from an abundantly expressed molecule? *Nat Rev Cancer* **11**, 254-267 (2011).

198. V. Samanna, T. Ma, T. W. Mak, M. Rogers, M. A. Chellaiah, Actin polymerization modulates CD44 surface expression, MMP-9 activation, and osteoclast function. *J Cell Physiol* **213**, 710-720 (2007).

199. A. C. Bellail, S. B. Hunter, D. J. Brat, C. Tan, E. G. Van Meir, Microregional extracellular matrix heterogeneity in brain modulates glioma cell invasion. *Int J Biochem Cell Biol* **36**, 1046-1069 (2004).

200. K. P. L. Bhat *et al.*, Mesenchymal differentiation mediated by NF-kappaB promotes radiation resistance in glioblastoma. *Cancer Cell* **24**, 331-346 (2013).

201. D. Sturm *et al.*, Paediatric and adult glioblastoma: multiform (epi)genomic culprits emerge. *Nat Rev Cancer* **14**, 92-107 (2014).

202. M. Esteller *et al.*, Inactivation of the DNA-repair gene MGMT and the clinical response of gliomas to alkylating agents. *N Engl J Med* **343**, 1350-1354 (2000).

203. M. E. Hegi *et al.*, Clinical trial substantiates the predictive value of O-6-methylguanine-DNA methyltransferase promoter methylation in glioblastoma patients treated with temozolomide. *Clin Cancer Res* **10**, 1871-1874 (2004).

204. W. Xu *et al.*, Oncometabolite 2-hydroxyglutarate is a competitive inhibitor of alpha-ketoglutarate-dependent dioxygenases. *Cancer Cell* **19**, 17-30 (2011).

205. M. Ceccarelli *et al.*, Molecular Profiling Reveals Biologically Discrete Subsets and Pathways of Progression in Diffuse Glioma. *Cell* **164**, 550-563 (2016).

206. S. P. Leon, R. D. Folkerth, P. M. Black, Microvessel density is a prognostic indicator for patients with astroglial brain tumors. *Cancer* **77**, 362-372 (1996).

207. D. M. Long, Capillary ultrastructure and the blood-brain barrier in human malignant brain tumors. *J Neurosurg* **32**, 127-144 (1970).

208. K. H. Plate, G. Breier, H. A. Weich, H. D. Mennel, W. Risau, Vascular endothelial growth factor and glioma angiogenesis: coordinate induction of VEGF receptors, distribution of VEGF protein and possible in vivo regulatory mechanisms. *Int J Cancer* **59**, 520-529 (1994).

209. K. H. Plate, G. Breier, H. A. Weich, W. Risau, Vascular endothelial growth factor is a potential tumour angiogenesis factor in human gliomas in vivo. *Nature* **359**, 845-848 (1992).

210. J. A. Forsythe *et al.*, Activation of vascular endothelial growth factor gene transcription by hypoxia-inducible factor 1. *Mol Cell Biol* **16**, 4604-4613 (1996).

211. N. Rohwer, C. Zasada, S. Kempa, T. Cramer, The growing complexity of HIF-1alpha's role in tumorigenesis: DNA repair and beyond. *Oncogene* **32**, 3569-3576 (2013).

212. P. H. Maxwell *et al.*, The tumour suppressor protein VHL targets hypoxia-inducible factors for oxygen-dependent proteolysis. *Nature* **399**, 271-275 (1999).

213. D. R. Senger *et al.*, Tumor cells secrete a vascular permeability factor that promotes accumulation of ascites fluid. *Science* **219**, 983-985 (1983).

214. D. T. Connolly *et al.*, Tumor vascular permeability factor stimulates endothelial cell growth and angiogenesis. *J Clin Invest* **84**, 1470-1478 (1989).

215. W. Wick *et al.*, Lomustine and Bevacizumab in Progressive Glioblastoma. *N Engl J Med* **377**, 1954-1963 (2017).

216. I. F. Dunn, O. Heese, P. M. Black, Growth factors in glioma angiogenesis: FGFs, PDGF, EGF, and TGFs. *J Neurooncol* **50**, 121-137 (2000).

217. O. Saksela, D. B. Rifkin, Release of basic fibroblast growth factor-heparan sulfate complexes from endothelial cells by plasminogen activator-mediated proteolytic activity. *J Cell Biol* **110**, 767-775 (1990).

218. D. Moscatelli, M. Presta, D. B. Rifkin, Purification of a factor from human placenta that stimulates capillary endothelial cell protease production, DNA synthesis, and migration. *Proc Natl Acad Sci U S A* **83**, 2091-2095 (1986).

219. P. Mignatti, D. B. Rifkin, Release of basic fibroblast growth factor, an angiogenic factor devoid of secretory signal sequence: a trivial phenomenon or a novel secretion mechanism? *J Cell Biochem* **47**, 201-207 (1991).

220. C. Calabrese *et al.*, A perivascular niche for brain tumor stem cells. *Cancer Cell* **11**, 69-82 (2007).

221. K. E. Hovinga *et al.*, Inhibition of notch signaling in glioblastoma targets cancer stem cells via an endothelial cell intermediate. *Stem Cells* **28**, 1019-1029 (2010).

222. T. S. Zhu *et al.*, Endothelial cells create a stem cell niche in glioblastoma by providing NOTCH ligands that nurture self-renewal of cancer stem-like cells. *Cancer Res* **71**, 6061-6072 (2011).

223. R. Wang *et al.*, Glioblastoma stem-like cells give rise to tumour endothelium. *Nature* **468**, 829-833 (2010).

224. L. Ricci-Vitiani *et al.*, Tumour vascularization via endothelial differentiation of glioblastoma stem-like cells. *Nature* **468**, 824-828 (2010).

225. M. T. Chiao, Y. C. Yang, W. Y. Cheng, C. C. Shen, J. L. Ko, CD133+ glioblastoma stem-like cells induce vascular mimicry in vivo. *Curr Neurovasc Res* **8**, 210-219 (2011).

226. Y. Soda *et al.*, Transdifferentiation of glioblastoma cells into vascular endothelial cells. *Proc Natl Acad Sci U S A* **108**, 4274-4280 (2011).

227. L. Cheng *et al.*, Glioblastoma stem cells generate vascular pericytes to support vessel function and tumor growth. *Cell* **153**, 139-152 (2013).

228. F. J. Rodriguez, B. A. Orr, K. L. Ligon, C. G. Eberhart, Neoplastic cells are a rare component in human glioblastoma microvasculature. *Oncotarget* **3**, 98-106 (2012).

229. J. J. Luke, K. T. Flaherty, A. Ribas, G. V. Long, Targeted agents and immunotherapies: optimizing outcomes in melanoma. *Nat Rev Clin Oncol* **14**, 463-482 (2017).

230. P. Sharma, S. Hu-Lieskovan, J. A. Wargo, A. Ribas, Primary, Adaptive, and Acquired Resistance to Cancer Immunotherapy. *Cell* **168**, 707-723 (2017).

231. A. Louveau *et al.*, Structural and functional features of central nervous system lymphatic vessels. *Nature* **523**, 337-341 (2015).

232. M. Schwartz, J. Kipnis, S. Rivest, A. Prat, How do immune cells support and shape the brain in health, disease, and aging? *J Neurosci* **33**, 17587-17596 (2013).

233. A. B. Heimberger, J. H. Sampson, Immunotherapy coming of age: what will it take to make it standard of care for glioblastoma? *Neuro Oncol* **13**, 3-13 (2011).

234. E. Vauleon, T. Avril, B. Collet, J. Mosser, V. Quillien, Overview of cellular immunotherapy for patients with glioblastoma. *Clin Dev Immunol* **2010**, (2010).

235. B. Ajami, J. L. Bennett, C. Krieger, W. Tetzlaff, F. M. Rossi, Local self-renewal can sustain CNS microglia maintenance and function throughout adult life. *Nat Neurosci* **10**, 1538-1543 (2007).

236. B. Ajami, J. L. Bennett, C. Krieger, K. M. McNagny, F. M. Rossi, Infiltrating monocytes trigger EAE progression, but do not contribute to the resident microglia pool. *Nat Neurosci* **14**, 1142-1149 (2011).

237. D. Hambardzumyan, D. H. Gutmann, H. Kettenmann, The role of microglia and macrophages in glioma maintenance and progression. *Nat Neurosci* **19**, 20-27 (2016).

238. P. Kivisakk *et al.*, Human cerebrospinal fluid central memory CD4+ T cells: evidence for trafficking through choroid plexus and meninges via P-selectin. *Proc Natl Acad Sci U S A* **100**, 8389-8394 (2003).

239. R. M. Ransohoff, P. Kivisakk, G. Kidd, Three or more routes for leukocyte migration into the central nervous system. *Nat Rev Immunol* **3**, 569-581 (2003).

240. I. Yang *et al.*, CD8+ T-cell infiltrate in newly diagnosed glioblastoma is associated with long-term survival. *J Clin Neurosci* **17**, 1381-1385 (2010).

241. J. Kmiecik *et al.*, Elevated CD3+ and CD8+ tumor-infiltrating immune cells correlate with prolonged survival in glioblastoma patients despite integrated immunosuppressive mechanisms in the tumor microenvironment and at the systemic level. *J Neuroimmunol* **264**, 71-83 (2013).

242. W. C. Rutledge *et al.*, Tumor-infiltrating lymphocytes in glioblastoma are associated with specific genomic alterations and related to transcriptional class. *Clin Cancer Res* **19**, 4951-4960 (2013).

243. T. Doucette *et al.*, Immune heterogeneity of glioblastoma subtypes: extrapolation from the cancer genome atlas. *Cancer Immunol Res* **1**, 112-122 (2013).

244. F. S. Hodi *et al.*, Combined nivolumab and ipilimumab versus ipilimumab alone in patients with advanced melanoma: 2-year overall survival outcomes in a multicentre, randomised, controlled, phase 2 trial. *Lancet Oncol* **17**, 1558-1568 (2016).

245. J. D. Wolchok *et al.*, Overall Survival with Combined Nivolumab and Ipilimumab in Advanced Melanoma. *N Engl J Med* **377**, 1345-1356 (2017).

246. H. Borghaei *et al.*, Nivolumab versus Docetaxel in Advanced Nonsquamous Non-Small-Cell Lung Cancer. *N Engl J Med* **373**, 1627-1639 (2015).

247. D. P. Carbone *et al.*, First-Line Nivolumab in Stage IV or Recurrent Non-Small-Cell Lung Cancer. *N Engl J Med* **376**, 2415-2426 (2017).

248. T. J. Lynch *et al.*, Activating mutations in the epidermal growth factor receptor underlying responsiveness of non-small-cell lung cancer to gefitinib. *N Engl J Med* **350**, 2129-2139 (2004).

249. P. D. Brown *et al.*, Phase I/II trial of erlotinib and temozolomide with radiation therapy in the treatment of newly diagnosed glioblastoma multiforme: North Central Cancer Treatment Group Study N0177. *J Clin Oncol* **26**, 5603-5609 (2008).

250. D. A. Reardon *et al.*, Phase 2 trial of erlotinib plus sirolimus in adults with recurrent glioblastoma. *J Neurooncol* **96**, 219-230 (2010).

251. J. N. Rich *et al.*, Phase II trial of gefitinib in recurrent glioblastoma. *J Clin Oncol* **22**, 133-142 (2004).

252. J. Schuster *et al.*, A phase II, multicenter trial of rindopepimut (CDX-110) in newly diagnosed glioblastoma: the ACT III study. *Neuro Oncol* **17**, 854-861 (2015).

253. M. Weller *et al.*, Rindopepimut with temozolomide for patients with newly diagnosed, EGFRvIII-expressing glioblastoma (ACT IV): a randomised, double-blind, international phase 3 trial. *Lancet Oncol* **18**, 1373-1385 (2017).

254. M. Bredel *et al.*, NFKBIA deletion in glioblastomas. *N Engl J Med* **364**, 627-637 (2011).

255. G. Dresemann *et al.*, Imatinib in combination with hydroxyurea versus hydroxyurea alone as oral therapy in patients with progressive pretreated glioblastoma resistant to standard dose temozolomide. *J Neurooncol* **96**, 393-402 (2010).

256. D. A. Reardon *et al.*, Multicentre phase II studies evaluating imatinib plus hydroxyurea in patients with progressive glioblastoma. *Br J Cancer* **101**, 1995-2004 (2009).

257. M. W. Pitz *et al.*, Phase II study of PX-866 in recurrent glioblastoma. *Neuro Oncol* **17**, 1270-1274 (2015).

258. S. M. Chang *et al.*, Phase II study of CCI-779 in patients with recurrent glioblastoma multiforme. *Invest New Drugs* **23**, 357-361 (2005).

259. D. J. Ma *et al.*, A phase II trial of everolimus, temozolomide, and radiotherapy in patients with newly diagnosed glioblastoma: NCCTG N057K. *Neuro Oncol* **17**, 1261-1269 (2015).

260. E. Galanis *et al.*, Phase II trial of temsirolimus (CCI-779) in recurrent glioblastoma multiforme: a North Central Cancer Treatment Group Study. *J Clin Oncol* **23**, 5294-5304 (2005).

261. M. H. Cohen, Y. L. Shen, P. Keegan, R. Pazdur, FDA drug approval summary: bevacizumab (Avastin) as treatment of recurrent glioblastoma multiforme. *Oncologist* **14**, 1131-1138 (2009).

262. B. M. Alexander, T. F. Cloughesy, Adult Glioblastoma. *J Clin Oncol* **35**, 2402-2409 (2017).

263. J. F. de Groot *et al.*, Phase II study of aflibercept in recurrent malignant glioma: a North American Brain Tumor Consortium study. *J Clin Oncol* **29**, 2689-2695 (2011).

264. E. Q. Lee *et al.*, A Multicenter, Phase II, Randomized, Noncomparative Clinical Trial of Radiation and Temozolomide with or without Vandetanib in Newly Diagnosed Glioblastoma Patients. *Clin Cancer Res* **21**, 3610-3618 (2015).

265. D. S. Chen, I. Mellman, Oncology meets immunology: the cancer-immunity cycle. *Immunity* **39**, 1-10 (2013).

266. D. M. O'Rourke *et al.*, A single dose of peripherally infused EGFRvIII-directed CAR T cells mediates antigen loss and induces adaptive resistance in patients with recurrent glioblastoma. *Sci Transl Med* **9**, (2017).

267. C. E. Brown *et al.*, Regression of Glioblastoma after Chimeric Antigen Receptor T-Cell Therapy. *N Engl J Med* **375**, 2561-2569 (2016).

268. P. E. Fecci *et al.*, Systemic CTLA-4 blockade ameliorates glioma-induced changes to the CD4+ T cell compartment without affecting regulatory T-cell function. *Clin Cancer Res* **13**, 2158-2167 (2007).

269. J. Vom Berg *et al.*, Intratumoral IL-12 combined with CTLA-4 blockade elicits T cell-mediated glioma rejection. *J Exp Med* **210**, 2803-2811 (2013).

270. J. Zeng *et al.*, Anti-PD-1 blockade and stereotactic radiation produce long-term survival in mice with intracranial gliomas. *Int J Radiat Oncol Biol Phys* **86**, 343-349 (2013).

271. A. S. Berghoff *et al.*, Programmed death ligand 1 expression and tumor-infiltrating lymphocytes in glioblastoma. *Neuro Oncol* **17**, 1064-1075 (2015).

272. C. Jones, S. J. Baker, Unique genetic and epigenetic mechanisms driving paediatric diffuse high-grade glioma. *Nat Rev Cancer* **14**, (2014).

273. M. S. Lawrence *et al.*, Mutational heterogeneity in cancer and the search for new cancer-associated genes. *Nature* **499**, 214-218 (2013).

274. B. J. Gill *et al.*, MRI-localized biopsies reveal subtype-specific differences in molecular and cellular composition at the margins of glioblastoma. *Proc Natl Acad Sci U S A* **111**, 12550-12555 (2014).

275. N. R. Parker *et al.*, Intratumoral heterogeneity identified at the epigenetic, genetic and transcriptional level in glioblastoma. *Sci Rep* **6**, 22477 (2016).

276. B. E. McLaughlin *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. Part A : the journal of the International Society for Analytical Cytology* **73**, 400-410 (2008).

277. S. C. De Rosa, L. A. Herzenberg, L. A. Herzenberg, M. Roederer, 11-color, 13-parameter flow cytometry: identification of human naive T cells by phenotype, function, and T-cell receptor diversity. *Nat Med* **7**, 245-248 (2001).

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

279. C. Giesen *et al.*, Highly multiplexed imaging of tumor tissues with subcellular resolution by mass cytometry. *Nat Methods* **11**, 417-422 (2014).

280. M. Angelo *et al.*, Multiplexed ion beam imaging of human breast tumors. *Nat Med* **20**, 436-442 (2014).

281. C. B. Moeder, J. M. Giltnane, S. P. Moulis, D. L. Rimm, Quantitative, fluorescence-based in-situ assessment of protein expression. *Methods Mol Biol* **520**, 163-175 (2009).

282. P. Qiu *et al.*, Extracting a cellular hierarchy from high-dimensional cytometry data with SPADE. *Nature biotechnology* **29**, 886-891 (2011).

283. S. C. Bendall, G. P. Nolan, M. Roederer, P. K. Chattopadhyay, A deep profiler's guide to cytometry. *Trends Immunol* **33**, 323-332 (2012).

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

285. N. Leelatian *et al.*, Preparing Viable Single Cells from Human Tissue and Tumors for Cytomic Analysis. *Curr Protoc Mol Biol* **118**, 25C 21 21-25C 21 23 (2017).

286. H. G. Fienberg, E. F. Simonds, W. J. Fantl, G. P. Nolan, B. Bodenmiller, A platinum-based covalent viability reagent for single-cell mass cytometry. *Cytometry. Part A : the journal of the International Society for Analytical Cytology* **81**, 467-475 (2012).

287. G. K. Behbehani, S. C. Bendall, M. R. Clutter, W. J. Fantl, G. P. Nolan, Single-cell mass cytometry adapted to measurements of the cell cycle. *Cytometry A* **81**, 552-566 (2012).

288. S. C. Bendall *et al.*, Single-cell trajectory detection uncovers progression and regulatory coordination in human B cell development. *Cell* **157**, 714-725 (2014).

289. G. H. Laurens van der Maaten, Visualizing Data using t-SNE. *J Mach Learn Res* **9**, 2579-2605 (2008).

290. A. D. Amir el *et al.*, viSNE enables visualization of high dimensional single-cell data and reveals phenotypic heterogeneity of leukemia. *Nat Biotechnol* **31**, 545-552 (2013).

291. S. T. Roweis, L. K. Saul, Nonlinear dimensionality reduction by locally linear embedding. *Science* **290**, 2323-2326 (2000).

292. J. B. Tenenbaum, V. de Silva, J. C. Langford, A global geometric framework for nonlinear dimensionality reduction. *Science* **290**, 2319-2323 (2000).

293. S. Van Gassen *et al.*, FlowSOM: Using self-organizing maps for visualization and interpretation of cytometry data. *Cytometry. Part A : the journal of the International Society for Analytical Cytology* **87**, 636-645 (2015).

294. B. Becher *et al.*, High-dimensional analysis of the murine myeloid cell system. *Nat Immunol* **15**, 1181-1189 (2014).

295. K. Shekhar, P. Brodin, M. M. Davis, A. K. Chakraborty, Automatic Classification of Cellular Expression by Nonlinear Stochastic Embedding (ACCENSE). *Proc Natl Acad Sci U S A* **111**, 202-207 (2014).

296. J. M. Irish, Beyond the age of cellular discovery. *Nat Immunol* **15**, 1095-1097 (2014).

297. M. Kleppe, R. L. Levine, Tumor heterogeneity confounds and illuminates: assessing the implications. *Nat Med* **20**, 342-344 (2014).

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

299. E. W. Newell, N. Sigal, S. C. Bendall, G. P. Nolan, M. M. Davis, Cytometry by time-of-flight shows combinatorial cytokine expression and virus-specific cell niches within a continuum of CD8+ T cell phenotypes. *Immunity* **36**, 142-152 (2012).

300. B. Bodenmiller *et al.*, Multiplexed mass cytometry profiling of cellular states perturbed by small-molecule regulators. *Nat Biotechnol* **30**, 858-867 (2012).

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

302. D. R. Bandura *et al.*, Mass cytometry: technique for real time single cell multitarget immunoassay based on inductively coupled plasma time-of-flight mass spectrometry. *Anal Chem* **81**, 6813-6822 (2009).

303. J. M. Irish, D. K. Czerwinski, G. P. Nolan, R. Levy, Altered B-cell receptor signaling kinetics distinguish human follicular lymphoma B cells from tumor-infiltrating nonmalignant B cells. *Blood* **108**, 3135-3142 (2006).

304. A. D. Boiko *et al.*, Human melanoma-initiating cells express neural crest nerve growth factor receptor CD271. *Nature* **466**, 133-137 (2010).

305. M. Al-Hajj, M. S. Wicha, A. Benito-Hernandez, S. J. Morrison, M. F. Clarke, Prospective identification of tumorigenic breast cancer cells. *Proceedings of the National Academy of Sciences of the United States of America* **100**, 3983-3988 (2003).

306. L. Zimmerlin, V. S. Donnenberg, 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* **699**, 251-273 (2011).

307. J. O. Richards, J. Treisman, N. Garlie, J. P. Hanson, M. K. Oaks, Flow cytometry assessment of residual melanoma cells in tumor-infiltrating lymphocyte cultures. *Cytometry. Part A : the journal of the International Society for Analytical Cytology* **81**, 374-381 (2012).

308. K. S. Chan *et al.*, Identification, molecular characterization, clinical prognosis, and therapeutic targeting of human bladder tumor-initiating cells. *Proc Natl Acad Sci U S A* **106**, 14016-14021 (2009).

309. V. S. Donnenberg, R. J. Landreneau, M. E. Pfeifer, A. D. Donnenberg, Flow cytometric determination of stem/progenitor content in epithelial tissues: an example from nonsmall lung cancer and normal lung. *Cytometry. Part A : the journal of the International Society for Analytical Cytology* **83**, 141-149 (2013).

310. V. C. Daniel *et al.*, A primary xenograft model of small-cell lung cancer reveals irreversible changes in gene expression imposed by culture in vitro. *Cancer research* **69**, 3364-3373 (2009).

311. B. A. Reynolds, W. Tetzlaff, S. Weiss, A multipotent EGF-responsive striatal embryonic progenitor cell produces neurons and astrocytes. *The Journal of neuroscience : the official journal of the Society for Neuroscience* **12**, 4565-4574 (1992).

312. G. K. Behbehani *et al.*, Transient partial permeabilization with saponin enables cellular barcoding prior to surface marker staining. *Cytometry A* **85**, 1011-1019 (2014).

313. R. Finck *et al.*, Normalization of mass cytometry data with bead standards. *Cytometry. Part A : the journal of the International Society for Analytical Cytology* **83**, 483-494 (2013).

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

315. K. J. Nicholas *et al.*, Multiparameter analysis of stimulated human peripheral blood mononuclear cells: A comparison of mass and fluorescence cytometry. *Cytometry. Part A : the journal of the International Society for Analytical Cytology*, (2015).

316. H. T. Maecker, J. P. McCoy, R. Nussenblatt, Standardizing immunophenotyping for the Human Immunology Project. *Nat Rev Immunol* **12**, 191-200 (2012).

317. Q. T. Ostrom *et al.*, CBTRUS Statistical Report: Primary Brain and Central Nervous System Tumors Diagnosed in the United States in 2008-2012. *Neuro-oncology* **17 Suppl 4**, iv1-iv62 (2015).

318. C. Brennan *et al.*, Glioblastoma subclasses can be defined by activity among signal transduction pathways and associated genomic alterations. *PloS one* **4**, e7752 (2009).

319. H. Johnson, F. M. White, Quantitative analysis of signaling networks across differentially embedded tumors highlights interpatient heterogeneity in human glioblastoma. *Journal of proteome research* **13**, 4581-4593 (2014).

320. P. H. Huang *et al.*, Quantitative analysis of EGFRvIII cellular signaling networks reveals a combinatorial therapeutic strategy for glioblastoma. *Proc Natl Acad Sci U S A* **104**, 12867-12872 (2007).

321. K. E. Diggins, A. R. Greenplate, N. Leelatian, C. E. Wogsland, J. M. Irish, Characterizing cell subsets using marker enrichment modeling. *Nat Methods* **14**, 275-278 (2017).

322. N. Leelatian *et al.*, Single cell analysis of human tissues and solid tumors with mass cytometry. *Cytometry B Clin Cytom*, (2016).

323. S. Van Gassen *et al.*, FlowSOM: Using self-organizing maps for visualization and interpretation of cytometry data. *Cytometry A*, (2015).

324. S. Baumhueter, N. Dybdal, C. Kyle, L. A. Lasky, Global vascular expression of murine CD34, a sialomucin-like endothelial ligand for L-selectin. *Blood* **84**, 2554-2565 (1994).

325. M. Marx, R. A. Perlmutter, J. A. Madri, Modulation of platelet-derived growth factor receptor expression in microvascular endothelial cells during in vitro angiogenesis. *The Journal of clinical investigation* **93**, 131-139 (1994).

326. L. Primo *et al.*, Increased expression of alpha6 integrin in endothelial cells unveils a proangiogenic role for basement membrane. *Cancer research* **70**, 5759-5769 (2010).

327. B. J. Baaten *et al.*, CD44 regulates survival and memory development in Th1 cells. *Immunity* **32**, 104-115 (2010).

328. H. C. DeGrendele, P. Estess, M. H. Siegelman, Requirement for CD44 in activated T cell extravasation into an inflammatory site. *Science* **278**, 672-675 (1997).

329. W. R. Shapiro *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. *Journal of neurosurgery* **71**, 1-9 (1989).

330. A. L. Rivera *et al.*, MGMT promoter methylation is predictive of response to radiotherapy and prognostic in the absence of adjuvant alkylating chemotherapy for glioblastoma. *Neuro-oncology* **12**, 116-121 (2010).

331. R. O. Mirimanoff *et al.*, Radiotherapy and temozolomide for newly diagnosed glioblastoma: recursive partitioning analysis of the EORTC 26981/22981-NCIC CE3 phase III randomized trial. *Journal of clinical oncology : official journal of the American Society of Clinical Oncology* **24**, 2563-2569 (2006).

332. M. D. Walker *et al.*, Randomized comparisons of radiotherapy and nitrosoureas for the treatment of malignant glioma after surgery. *The New England journal of medicine* **303**, 1323-1329 (1980).

333. M. L. Bennett *et al.*, New tools for studying microglia in the mouse and human CNS. *Proc Natl Acad Sci U S A* **113**, E1738-1746 (2016).

334. A. L. Ford, A. L. Goodsall, W. F. Hickey, J. D. Sedgwick, Normal adult ramified microglia separated from other central nervous system macrophages by flow cytometric sorting. Phenotypic differences defined and direct ex vivo antigen presentation to myelin basic protein-reactive CD4+ T cells compared. *J Immunol* **154**, 4309-4321 (1995).

335. G. Locatelli *et al.*, Primary oligodendrocyte death does not elicit anti-CNS immunity. *Nat Neurosci* **15**, 543-550 (2012).

336. T. Matsumoto *et al.*, CD44 expression in astrocytes and microglia is associated with ALS progression in a mouse model. *Neurosci Lett* **520**, 115-120 (2012).

337. J. F. Bromberg, C. M. Horvath, Z. Wen, R. D. Schreiber, J. E. Darnell, Jr., Transcriptionally active Stat1 is required for the antiproliferative effects of both interferon alpha and interferon gamma. *Proceedings of the National Academy of Sciences of the United States of America* **93**, 7673-7678 (1996).

338. Y. E. Chin *et al.*, Cell growth arrest and induction of cyclin-dependent kinase inhibitor p21 WAF1/CIP1 mediated by STAT1. *Science* **272**, 719-722 (1996).

339. S. Fujioka *et al.*, Stabilization of p53 is a novel mechanism for proapoptotic function of NF-kappaB. *The Journal of biological chemistry* **279**, 27549-27559 (2004).

340. A. Schneider *et al.*, NF-kappaB is activated and promotes cell death in focal cerebral ischemia. *Nature medicine* **5**, 554-559 (1999).

341. M. Thomas *et al.*, STAT1: a modulator of chemotherapy-induced apoptosis. *Cancer research* **64**, 8357-8364 (2004).

342. B. Korin *et al.*, High-dimensional, single-cell characterization of the brain's immune compartment. *Nature neuroscience* **20**, 1300-1309 (2017).

343. D. B. Johnson *et al.*, Melanoma-specific MHC-II expression represents a tumour-autonomous phenotype and predicts response to anti-PD-1/PD-L1 therapy. *Nature communications* **7**, 10582 (2016).

344. P. O. Krutzik, G. P. Nolan, Fluorescent cell barcoding in flow cytometry allows high-throughput drug screening and signaling profiling. *Nat Methods* **3**, 361-368 (2006).

345. P. O. Krutzik, J. M. Crane, M. R. Clutter, G. P. Nolan, High-content single-cell drug screening with phosphospecific flow cytometry. *Nat Chem Biol* **4**, 132-142 (2008).

346. E. R. Zunder *et al.*, Palladium-based mass tag cell barcoding with a doublet-filtering scheme and single-cell deconvolution algorithm. *Nat Protoc* **10**, 316-333 (2015).

347. D. C. Earl *et al.*, Discovery of human cell selective effector molecules using single cell multiplexed activity metabolomics. *Nat Commun* **9**, 39 (2018).

348. P. A. Scuffham *et al.*, Using N-of-1 trials to improve patient management and save costs. *J Gen Intern Med* **25**, 906-913 (2010).

349. S. K. Singh *et al.*, Identification of a cancer stem cell in human brain tumors. *Cancer Res* **63**, 5821-5828 (2003).

350. N. Uchida *et al.*, Direct isolation of human central nervous system stem cells. *Proc Natl Acad Sci U S A* **97**, 14720-14725 (2000).

351. D. Juric *et al.*, Convergent loss of PTEN leads to clinical resistance to a PI(3)Kalpha inhibitor. *Nature* **518**, 240-244 (2015).

352. T. Zhao *et al.*, Humanized Mice Reveal Differential Immunogenicity of Cells Derived from Autologous Induced Pluripotent Stem Cells. *Cell Stem Cell* **17**, 353-359 (2015).

353. W. Jiskoot *et al.*, Mouse Models for Assessing Protein Immunogenicity: Lessons and Challenges. *J Pharm Sci* **105**, 1567-1575 (2016).

354. T. Blankenstein, P. G. Coulie, E. Gilboa, E. M. Jaffee, The determinants of tumour immunogenicity. *Nat Rev Cancer* **12**, 307-313 (2012).

355. C. Pfirschke *et al.*, Immunogenic Chemotherapy Sensitizes Tumors to Checkpoint Blockade Therapy. *Immunity* **44**, 343-354 (2016).

356. E. Turola, R. Furlan, F. Bianco, M. Matteoli, C. Verderio, Microglial microvesicle secretion and intercellular signaling. *Front Physiol* **3**, 149 (2012).

357. F. Bianco *et al.*, Astrocyte-derived ATP induces vesicle shedding and IL-1 beta release from microglia. *J Immunol* **174**, 7268-7277 (2005).

358. T. Lawrence, The nuclear factor NF-kappaB pathway in inflammation. *Cold Spring Harb Perspect Biol* **1**, a001651 (2009).

359. P. K. Chattopadhyay, T. M. Gierahn, M. Roederer, J. C. Love, Single-cell technologies for monitoring immune systems. *Nat Immunol* **15**, 128-135 (2014).

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

361. E. W. Newell, Y. Cheng, Mass cytometry: blessed with the curse of dimensionality. *Nat Immunol* **17**, 890-895 (2016).

362. R. R. Hardy, K. Hayakawa, D. R. Parks, L. A. Herzenberg, Demonstration of B-cell maturation in X-linked immunodeficient mice by simultaneous three-colour immunofluorescence. *Nature* **306**, 270-272 (1983).

363. , (!!! INVALID CITATION !!! {}).

364. D. R. Parks, R. R. Hardy, L. A. Herzenberg, Three-color immunofluorescence analysis of mouse B-lymphocyte subpopulations. *Cytometry* **5**, 159-168 (1984).

365. J. W. Tung, D. R. Parks, W. A. Moore, L. A. Herzenberg, L. A. Herzenberg, Identification of B-cell subsets: an exposition of 11-color (Hi-D) FACS methods. *Methods Mol Biol* **271**, 37-58 (2004).

366. C. E. Wogsland *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* **92**, 79-87 (2017).

367. F. E. Craig, K. A. Foon, Flow cytometric immunophenotyping for hematologic neoplasms. *Blood* **111**, 3941-3967 (2008).

368. J. J. van Dongen *et al.*, EuroFlow antibody panels for standardized n-dimensional flow cytometric immunophenotyping of normal, reactive and malignant leukocytes. *Leukemia* **26**, 1908-1975 (2012).

369. B. L. Wood *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* **72 Suppl 1**, S14-22 (2007).

370. M. J. Borowitz *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* **111**, 5477-5485 (2008).

371. J. J. van Dongen, V. H. van der Velden, M. Bruggemann, A. Orfao, Minimal residual disease diagnostics in acute lymphoblastic leukemia: need for sensitive, fast, and standardized technologies. *Blood* **125**, 3996-4009 (2015).

372. H. T. Maecker, T. Frey, L. E. Nomura, J. Trotter, Selecting fluorochrome conjugates for maximum sensitivity. *Cytometry. Part A : the journal of the International Society for Analytical Cytology* **62**, 169-173 (2004).

373. L. Wang *et al.*, Human CD4+ lymphocytes for antigen quantification: characterization using conventional flow cytometry and mass cytometry. *Cytometry. Part A : the journal of the International Society for Analytical Cytology* **81**, 567-575 (2012).

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

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

376. S. Kordasti *et al.*, Deep phenotyping of Tregs identifies an immune signature for idiopathic aplastic anemia and predicts response to treatment. *Blood* **128**, 1193-1205 (2016).

377. P. Codega *et al.*, Prospective identification and purification of quiescent adult neural stem cells from their in vivo niche. *Neuron* **82**, 545-559 (2014).

378. E. Pastrana, L. C. Cheng, F. Doetsch, Simultaneous prospective purification of adult subventricular zone neural stem cells and their progeny. *Proceedings of the National Academy of Sciences of the United States of America* **106**, 6387-6392 (2009).

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

380. K. R. Schulz, E. A. Danna, P. O. Krutzik, G. P. Nolan, Single-cell phospho-protein analysis by flow cytometry. *Curr Protoc Immunol* **Chapter 8**, Unit 8 17 11-20 (2012).

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

382. J. Lee *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* **9**, 391-403 (2006).

383. P. O. Krutzik, M. R. Clutter, G. P. Nolan, Coordinate analysis of murine immune cell surface markers and intracellular phosphoproteins by flow cytometry. *J Immunol* **175**, 2357-2365 (2005).

384. R. Hulspas, M. R. O'Gorman, B. L. Wood, J. W. Gratama, D. R. Sutherland, Considerations for the control of background fluorescence in clinical flow cytometry. *Cytometry B Clin Cytom* **76**, 355-364 (2009).

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

386. P. K. Chattopadhyay, M. Roederer, Cytometry: today's technology and tomorrow's horizons. *Methods* **57**, 251-258 (2012).

387. Y. D. Mahnke, M. Roederer, Optimizing a multicolor immunophenotyping assay. *Clin Lab Med* **27**, 469-485, v (2007).

388. S. P. Perfetto, P. K. Chattopadhyay, M. Roederer, Seventeen-colour flow cytometry: unravelling the immune system. *Nat Rev Immunol* **4**, 648-655 (2004).

389. J. P. Robinson, Comparative overview of flow and image cytometry. *Curr Protoc Cytom* **Chapter 12**, Unit 12 11 (2005).

390. Z. B. Bjornson, G. P. Nolan, W. J. Fantl, Single-cell mass cytometry for analysis of immune system functional states. *Curr Opin Immunol* **25**, 484-494 (2013).

391. P. P. Lee *et al.*, Characterization of circulating T cells specific for tumor-associated antigens in melanoma patients. *Nature medicine* **5**, 677-685 (1999).

392. M. Roederer *et al.*, The genetic architecture of the human immune system: a bioresource for autoimmunity and disease pathogenesis. *Cell* **161**, 387-403 (2015).

393. O. I. Ornatsky *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* **80**, 2539-2547 (2008).

394. Y. Saeys, S. V. Gassen, B. N. Lambrecht, Computational flow cytometry: helping to make sense of high-dimensional immunology data. *Nat Rev Immunol* **16**, 449-462 (2016).

395. M. DuPage, J. A. Bluestone, Harnessing the plasticity of CD4(+) T cells to treat immune-mediated disease. *Nat Rev Immunol* **16**, 149-163 (2016).

396. Nomenclature for clusters of differentiation (CD) of antigens defined on human leukocyte populations. IUIS-WHO Nomenclature Subcommittee. *Bull World Health Organ* **62**, 809-815 (1984).

397. B. Wang, J. Zhu, E. Pierson, D. Ramazzotti, S. Batzoglou, Visualization and analysis of single-cell RNA-seq data by kernel-based similarity learning. *Nature methods* **14**, 414-416 (2017).

398. S. S. Shen-Orr *et al.*, Towards a cytokine-cell interaction knowledgebase of the adaptive immune system. *Pac Symp Biocomput*, 439-450 (2009).

399. M. J. Borowitz, K. L. Guenther, K. E. Shults, G. T. Stelzer, Immunophenotyping of acute leukemia by flow cytometric analysis. Use of CD45 and right-angle light scatter to gate on leukemic blasts in three-color analysis. *Am J Clin Pathol* **100**, 534-540 (1993).

400. P. O. Krutzik, M. R. Clutter, A. Trejo, G. P. Nolan, Fluorescent cell barcoding for multiplex flow cytometry. *Curr Protoc Cytom* **Chapter 6**, Unit 6 31 (2011).

401. G. Szaloki, K. Goda, Compensation in multicolor flow cytometry. *Cytometry A* **87**, 982-985 (2015).

402. P. O. Krutzik, A. Trejo, K. R. Schulz, G. P. Nolan, Phospho flow cytometry methods for the analysis of kinase signaling in cell lines and primary human blood samples. *Methods Mol Biol* **699**, 179-202 (2011).

403. S. Madler, C. Bich, D. Touboul, R. Zenobi, Chemical cross-linking with NHS esters: a systematic study on amino acid reactivities. *J Mass Spectrom* **44**, 694-706 (2009).

404. J. W. Kim, H. Adachi, K. Shin-ya, Y. Hayakawa, H. Seto, Apoptolidin, a new apoptosis inducer in transformed cells from Nocardiopsis sp. *J Antibiot (Tokyo)* **50**, 628-630 (1997).

405. A. R. Salomon, D. W. Voehringer, L. A. Herzenberg, C. Khosla, Apoptolidin, a selective cytotoxic agent, is an inhibitor of F0F1-ATPase. *Chem Biol* **8**, 71-80 (2001).

406. A. R. Salomon, D. W. Voehringer, L. A. Herzenberg, C. Khosla, Understanding and exploiting the mechanistic basis for selectivity of polyketide inhibitors of F(0)F(1)-ATPase. *Proc Natl Acad Sci U S A* **97**, 14766-14771 (2000).

407. J. D. Serrill *et al.*, Apoptolidins A and C activate AMPK in metabolically sensitive cell types and are mechanistically distinct from oligomycin A. *Biochem Pharmacol* **93**, 251-265 (2015).

408. B. O. Bachmann *et al.*, Light-induced isomerization of apoptolidin a leads to inversion of C2-C3 double bond geometry. *Org Lett* **12**, 2944-2947 (2010).

409. J. D. Pennington, H. J. Williams, A. R. Salomon, G. A. Sulikowski, Toward a stable apoptolidin derivative: identification of isoapoptolidin and selective deglycosylation of apoptolidin. *Org Lett* **4**, 3823-3825 (2002).

410. P. A. Wender, A. V. Gulledge, O. D. Jankowski, H. Seto, Isoapoptolidin: structure and activity of the ring-expanded isomer of apoptolidin. *Org Lett* **4**, 3819-3822 (2002).

411. P. A. Wender, M. Sukopp, K. Longcore, Apoptolidins B and C: isolation, structure determination, and biological activity. *Org Lett* **7**, 3025-3028 (2005).

412. P. A. Wender, K. E. Longcore, Isolation, structure determination, and anti-cancer activity of apoptolidin D. *Org Lett* **9**, 691-694 (2007).

413. P. A. Wender, K. E. Longcore, Apoptolidins E and F, new glycosylated macrolactones isolated from Nocardiopsis sp. *Org Lett* **11**, 5474-5477 (2009).

414. V. P. Ghidu *et al.*, Combined chemical and biosynthetic route to access a new apoptolidin congener. *Org Lett* **11**, 3032-3034 (2009).

415. Y. Du *et al.*, Biosynthesis of the Apoptolidins in Nocardiopsis sp. FU 40. *Tetrahedron* **67**, 6568-6575 (2011).

416. P. T. Daniel, U. Koert, J. Schuppan, Apoptolidin: induction of apoptosis by a natural product. *Angew Chem Int Ed Engl* **45**, 872-893 (2006).

417. H. Wehlan *et al.*, Apoptolidin A: total synthesis and partially glycosylated analogues. *Chemistry* **12**, 7378-7397 (2006).

418. V. P. Ghidu *et al.*, Synthesis and evaluation of the cytotoxicity of apoptolidinones A and D. *J Org Chem* **73**, 4949-4955 (2008).

419. S. M. DeGuire *et al.*, Fluorescent probes of the apoptolidins and their utility in cellular localization studies. *Angew Chem Int Ed Engl* **54**, 961-964 (2015).

420. C. Bhattacharya, Z. Yu, M. J. Rishel, S. M. Hecht, The carbamoylmannose moiety of bleomycin mediates selective tumor cell targeting. *Biochemistry* **53**, 3264-3266 (2014).

421. M. M. Madathil *et al.*, Modified bleomycin disaccharides exhibiting improved tumor cell targeting. *Biochemistry* **53**, 6800-6810 (2014).

422. B. R. Schroeder *et al.*, The disaccharide moiety of bleomycin facilitates uptake by cancer cells. *J Am Chem Soc* **136**, 13641-13656 (2014).

423. Z. Yu *et al.*, Selective tumor cell targeting by the disaccharide moiety of bleomycin. *J Am Chem Soc* **135**, 2883-2886 (2013).

424. W. Punfa, S. Yodkeeree, P. Pitchakarn, C. Ampasavate, P. Limtrakul, Enhancement of cellular uptake and cytotoxicity of curcumin-loaded PLGA nanoparticles by conjugation with anti-P-glycoprotein in drug resistance cancer cells. *Acta Pharmacol Sin* **33**, 823-831 (2012).

425. M. S. Dordal *et al.*, Flow cytometric assessment of the cellular pharmacokinetics of fluorescent drugs. *Cytometry* **20**, 307-314 (1995).

426. M. I. Fekry *et al.*, Noncovalent DNA binding drives DNA alkylation by leinamycin: evidence that the Z,E-5-(thiazol-4-yl)-penta-2,4-dienone moiety of the natural product serves as an atypical DNA intercalator. *J Am Chem Soc* **133**, 17641-17651 (2011).

427. P. L. Hamilton, D. P. Arya, Natural product DNA major groove binders. *Nat Prod Rep* **29**, 134-143 (2012).

428. B. Pang *et al.*, Drug-induced histone eviction from open chromatin contributes to the chemotherapeutic effects of doxorubicin. *Nat Commun* **4**, 1908 (2013).

429. S. M. Swaney, H. Aoki, M. C. Ganoza, D. L. Shinabarger, The oxazolidinone linezolid inhibits initiation of protein synthesis in bacteria. *Antimicrob Agents Chemother* **42**, 3251-3255 (1998).

430. S. Arenz *et al.*, Structures of the orthosomycin antibiotics avilamycin and evernimicin in complex with the bacterial 70S ribosome. *Proc Natl Acad Sci U S A* **113**, 7527-7532 (2016).

431. N. S. Hegde, D. A. Sanders, R. Rodriguez, S. Balasubramanian, The transcription factor FOXM1 is a cellular target of the natural product thiostrepton. *Nat Chem* **3**, 725-731 (2011).

432. M. Gersch, J. Kreuzer, S. A. Sieber, Electrophilic natural products and their biological targets. *Nat Prod Rep* **29**, 659-682 (2012).

433. P. S. Ward, C. B. Thompson, Metabolic reprogramming: a cancer hallmark even warburg did not anticipate. *Cancer Cell* **21**, 297-308 (2012).

434. T. Fotsis *et al.*, The endogenous oestrogen metabolite 2-methoxyoestradiol inhibits angiogenesis and suppresses tumour growth. *Nature* **368**, 237-239 (1994).

435. P. M. Smith *et al.*, The microbial metabolites, short-chain fatty acids, regulate colonic Treg cell homeostasis. *Science* **341**, 569-573 (2013).

436. N. Arpaia *et al.*, Metabolites produced by commensal bacteria promote peripheral regulatory T-cell generation. *Nature* **504**, 451-455 (2013).

437. H. Kayama, K. Takeda, Functions of innate immune cells and commensal bacteria in gut homeostasis. *J Biochem* **159**, 141-149 (2016).

438. J. Watrous *et al.*, Mass spectral molecular networking of living microbial colonies. *Proc Natl Acad Sci U S A* **109**, E1743-1752 (2012).

439. D. J. Newman, G. M. Cragg, Natural Products as Sources of New Drugs from 1981 to 2014. *J Nat Prod* **79**, 629-661 (2016).

440. D. J. Payne, M. N. Gwynn, D. J. Holmes, D. L. Pompliano, Drugs for bad bugs: confronting the challenges of antibacterial discovery. *Nat Rev Drug Discov* **6**, 29-40 (2007).

441. A. R. Greenplate, D. B. Johnson, P. B. Ferrell, Jr., J. M. Irish, Systems immune monitoring in cancer therapy. *Eur J Cancer* **61**, 77-84 (2016).

442. G. J. Patti, O. Yanes, G. Siuzdak, Innovation: Metabolomics: the apogee of the omics trilogy. *Nat Rev Mol Cell Biol* **13**, 263-269 (2012).

443. T. A. Johnson *et al.*, Natural product libraries to accelerate the high-throughput discovery of therapeutic leads. *J Nat Prod* **74**, 2545-2555 (2011).

444. G. L. Chen *et al.*, Nonintercalative antitumor drugs interfere with the breakage-reunion reaction of mammalian DNA topoisomerase II. *J Biol Chem* **259**, 13560-13566 (1984).

445. C. A. Belmokhtar, J. Hillion, E. Segal-Bendirdjian, Staurosporine induces apoptosis through both caspase-dependent and caspase-independent mechanisms. *Oncogene* **20**, 3354-3362 (2001).

446. S. P. Perfetto *et al.*, Amine reactive dyes: an effective tool to discriminate live and dead cells in polychromatic flow cytometry. *J Immunol Methods* **313**, 199-208 (2006).

447. K. Vermeulen, D. R. Van Bockstaele, Z. N. Berneman, Apoptosis: mechanisms and relevance in cancer. *Ann Hematol* **84**, 627-639 (2005).

448. L. J. Kuo, L. X. Yang, Gamma-H2AX - a novel biomarker for DNA double-strand breaks. *In Vivo* **22**, 305-309 (2008).

449. W. Z. Tu *et al.*, gammaH2AX foci formation in the absence of DNA damage: mitotic H2AX phosphorylation is mediated by the DNA-PKcs/CHK2 pathway. *FEBS Lett* **587**, 3437-3443 (2013).

450. G. Rothe, G. Schmitz, Consensus protocol for the flow cytometric immunophenotyping of hematopoietic malignancies. Working Group on Flow Cytometry and Image Analysis. *Leukemia* **10**, 877-895 (1996).

451. E. H. Estey, Acute myeloid leukemia: 2012 update on diagnosis, risk stratification, and management. *Am J Hematol* **87**, 89-99 (2012).

452. A. Dufner, G. Thomas, Ribosomal S6 kinase signaling and the control of translation. *Exp Cell Res* **253**, 100-109 (1999).

453. D. K. Derewacz, B. C. Covington, J. A. McLean, B. O. Bachmann, Mapping Microbial Response Metabolomes for Induced Natural Product Discovery. *ACS Chem Biol* **10**, 1998-2006 (2015).

454. T. Komiyana *et al.*, Baumycins, new antitumor antibiotics related to daunomycin. *J Antibiot (Tokyo)* **30**, 619-621 (1977).

455. C. Scotti, C. R. Hutchinson, Enhanced antibiotic production by manipulation of the Streptomyces peucetius dnrH and dnmT genes involved in doxorubicin (adriamycin) biosynthesis. *J Bacteriol* **178**, 7316-7321 (1996).

456. Y. Nozaki *et al.*, TAN-1120, a new anthracycline with potent angiostatic activity. *J Antibiot (Tokyo)* **46**, 569-579 (1993).

457. A. van Rhenen *et al.*, High stem cell frequency in acute myeloid leukemia at diagnosis predicts high minimal residual disease and poor survival. *Clin Cancer Res* **11**, 6520-6527 (2005).

458. T. Reya, S. J. Morrison, M. F. Clarke, I. L. Weissman, Stem cells, cancer, and cancer stem cells. *Nature* **414**, 105-111 (2001).

459. J. A. Frank *et al.*, Photoswitchable diacylglycerols enable optical control of protein kinase C. *Nat Chem Biol* **12**, 755-762 (2016).

460. G. Marcucci, T. Haferlach, H. Dohner, Molecular genetics of adult acute myeloid leukemia: prognostic and therapeutic implications. *J Clin Oncol* **29**, 475-486 (2011).

461. I. T. Young, Proof without prejudice: use of the Kolmogorov-Smirnov test for the analysis of histograms from flow systems and other sources. *J Histochem Cytochem* **25**, 935-941 (1977).

462. D. Kim, V. S. Donnenberg, J. W. Wilson, A. D. Donnenberg, The use of simultaneous confidence bands for comparison of single parameter fluorescent intensity data. *Cytometry A* **89**, 89-97 (2016).

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

464. F. Hahne *et al.*, flowCore: a Bioconductor package for high throughput flow cytometry. *BMC Bioinformatics* **10**, 106 (2009).

465. B. B. Gregory R Warnes, Lodewijk Bonebakker, Robert Gentleman, Wolfgang Huber, Andy Liaw, Thomas Lumley, Martin Maechler, Arni Magnusson, Steffen Moeller, Marc Schwartz, Bill Venables, gplots: Various R programming tools for plotting data. *R package version* **2**, 1 (2009).

466. K. Lo, F. Hahne, R. R. Brinkman, R. Gottardo, flowClust: a Bioconductor package for automated gating of flow cytometry data. *BMC Bioinformatics* **10**, 145 (2009).

467. T. R. Mosmann *et al.*, SWIFT-scalable clustering for automated identification of rare cell populations in large, high-dimensional flow cytometry datasets, part 2: biological evaluation. *Cytometry A* **85**, 422-433 (2014).

468. R. V. Bruggner, B. Bodenmiller, D. L. Dill, R. J. Tibshirani, G. P. Nolan, Automated identification of stratifying signatures in cellular subpopulations. *Proc Natl Acad Sci U S A* **111**, E2770-2777 (2014).

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

470. K. J. Nicholas *et al.*, Multiparameter analysis of stimulated human peripheral blood mononuclear cells: A comparison of mass and fluorescence cytometry. *Cytometry A* **89**, 271-280 (2016).

471. C. Cox, J. E. Reeder, R. D. Robinson, S. B. Suppes, L. L. Wheeless, Comparison of frequency distributions in flow cytometry. *Cytometry* **9**, 291-298 (1988).