Single-Cell Methods for Profiling Tumor & Microenvironment Responses to Therapeutic Challenges

by

Sanjay Mathews Prakadan

B.A., The University of Chicago (2014)

Submitted to the Department of Chemistry
in Partial Fulfillment of the Requirements for the Degree of

Doctor of Philosophy

at the

MASSACHUSETTS INSTITUTE OF TECHNOLOGY

February 2020

Signature redacted

Signature of Author...
Department of Chemistry
January 7, 2020

Signature redacted

Certified by...
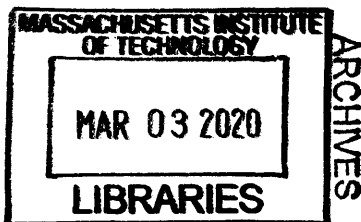Alex K. Shalek
Pfizer-Laubach Career Development Associate Professor of Chemistry
Thesis Supervisor

Signature redacted

Accepted by...
Robert W. Field
Haslam and Dewey Professor of Chemistry
Chair, Departmental Committee on Graduate Students

Single-Cell Methods for Profiling Tumor & Microenvironment Responses to Therapeutic Challenges

by

Sanjay Mathews Prakadan

Submitted to the Department of Chemistry
on January 17, 2020 in Partial Fulfillment of the
Requirements for the Degree of Doctor of Philosophy
in Chemistry

Signatures

Signature redacted

....

Alex K. Shalek
Pfizer-Laubach Career Development Associate Professor of Chemistry
Thesis Supervisor

Signature redacted

....

Adam P. Willard
Associate Professor of Chemistry
Thesis Committee Chair

Signature redacted __

....

Scott R. Manalis
Andrew and Erna Viterbi Professor of Biological Engineering
Thesis Committee Member

Signature redacted

....

Moungi G. Bawendi
Lester Wolfe Professor of Chemistry
Thesis Committee Member

Single-Cell Methods for Profiling Tumor & Microenvironment Responses to Therapeutic Challenges

by

Sanjay Mathews Prakadan

Submitted to the Department of Chemistry
on January 17, 2020 in Partial Fulfillment of the
Requirements for the Degree of Doctor of Philosophy
in Chemistry

Abstract

Heterogeneity among cells affects function and dysfunction across many complex biological systems. This heterogeneity is particularly important in cancer biology, where variation in the cells composing tumors and their surroundings can affect a patients' response to treatment and subsequent survival. While current methods, such as bulk RNA-Sequencing, are incredibly powerful, they typically measure average phenomena, mischaracterizing the distribution of behaviors within a system. Single-cell technologies – single-cell RNA Sequencing in particular – have been foundational in elucidating cellular heterogeneity from first principles, but there are limitations to their application for studying cancer and its response to treatment. Here, we detail efforts to address current needs in profiling treatment responses of tumors and their microenvironments at single-cell resolution.

Specifically, we characterize the underlying cellular diversity of tumor microenvironments, investigate the effect of drug treatment in specific cellular compartments, identify proxies of response in accessible cellular reservoirs, and investigate orthogonal cellular readouts of response. We first apply single-cell RNA Sequencing to study heterogeneity in metastatic melanoma, detailing heterogeneity and potential sources of resistance in cancer cells of profiled patients. Next, we study the effect of drug treatment in leptomeningeal carcinomatosis (LMD), extending previous strategies to utilize pre- and post-treatment patient sampling. We demonstrate the effect of immunotherapy in this microenvironment, and use longitudinal data from specific patients describe the evolution of cancer cell response to treatment. We next expand liquid biopsy profiling to other compartments, specifically circulating tumor cells (CTCs) in blood. We describe the development of a microfluidic device that captures murine CTCs with minimal sampling. We perform single-CTC RNA-Sequencing to study their response to treatment and relationship to their primary tumors. Finally, we develop a device that simultaneously measures the mass, growth rate and transcriptome of single cells, and use it to investigate the transcriptional activity of cancer cells that continue to grow after therapeutic challenge. Together, this body of work represents contributions towards extending single-cell profiling to understand how cells in naturally occurring and model cancer microenvironments respond to drug treatment.

Thesis Supervisor: Alex K. Shalek

Title: Pfizer-Laubach Career Development Associate Professor of Chemistry

Acknowledgements

I would be nowhere without the guidance of my mentors. I started my career in the laboratory of Pf. Steven J. Sibener at the University of Chicago, with future professors Ryan D. Brown and Gaby Avila-Bront. It was there that I built my first experimental platform, and where I first learned how to dissect a problem and develop solutions.

Alex, my graduate advisor, trusted me with his career as much as I trusted him with mine. He created a space at MIT where I could explore a completely new scientific field, develop my own research interests and goals, be protected from the severest consequences of my failures, and still have the opportunity to learn from them. I will be forever indebted to him as a professional and personal mentor.

I am grateful for the privilege of being in the Shalek Lab, alongside the finest scientists and colleagues, all of whom set a standard of excellence that I have tried to live up to. In particular, the post-doctoral fellows – Drs. Ben Mead, Britt Goods, and Pf. Jose Ordovas-Montañes – who lent their experience during my moments of doubt, and were instrumental in my development as a scientist.

I have worked alongside brilliant clinicians, engineers, physicists, chemists, and biologists in every one of my projects, and without them, we would not have accomplished anything written in these pages. I am particularly grateful for the company of Pfs. Ben Izar, Itay Tirosh, Drs. Bashar Hamza, Robert Kimmerling, Samuel Markson, and Pfs. Scott Manalis, Scott Carter, and Priscilla Brastianos.

I have been accompanied in the past 6 years by a team of friends and family, on whose compassion and support I have regularly depended. My fellow PChemists – Drs. Daniel Banks, Brian Michael, Katie Shulenberger, Nicole Moody, and Pf. Sam Kazer – have been the longest, most consistent friends that I have ever had, and I am grateful that our traditions will continue even after these PhD years.

I have found a remarkable community through Latin dance in Boston; the individual and collective connections we have formed are among the most important in my life. Sherry Zeng, Angela Balcom, Gabriella Bellemo, Dr. Zamyla Chan, Tina Cavicchio, Fardin Aryan, and Dr. Yasmin Chau – you have brought me joy beyond words, sometimes without so much as a word spoken.

Finally, I thank my parents Mat & Cissy Prakadan, who sacrificed their entire lives so that I would have the opportunities to pursue what mattered to me. They taught me that there is nothing more rewarding than working tirelessly and without compromise on behalf of those who depend on you.

Lay Summary

Cells are the building blocks that make up all biological systems and their microscopic functions can influence the macroscopic behavior we observe in tissues and organisms. In particular, the dysfunction of a tumor – the spontaneous and uncontrollable multiplication of particular cells in a tissue – is driven not only by cancer cells, but also by immune cells and other non-diseased cells in their surroundings that can aid or hinder cancer progression. Specifically, the response of a tumor after treatment with drugs that eradicate cancer cells is influenced by how the individual cancer and surrounding cells react to the treatment.

While cells tend to behave according to their "type" (immune, cancer, tissue, etc.) they sometimes exhibit subtle variation in function, even from the same population. This variation can be incredibly important because the different "subpopulations" of cells can have different roles and contribute to macroscopic behavior in different ways. In cancer and cancer treatment, for example, a drug may only target certain subpopulations in a tumor, and new groups can arise which adapt to treatment in unpredictable ways. Resolving, identifying, and cataloging this variation and its effects is important to fully understand a biological system, and in the case of cancer, how to properly treat and manage it.

One strategy for identifying the variation among individual cells is to collect them one at a time and profile them independently. This strategy, known as single-cell biology, has challenged our previous understanding of how cells work together, perform tasks, and respond to stimuli. In particular, single-cell RNA Sequencing, which measures the RNA transcripts from individual cells in order to infer their behavior, allows us to investigate the expression level of all the genes in a cell rather than just ones we target. This technique has had a profound impact on a number of fields, from immunology to neuroscience to cancer. However, applying single-cell RNA Sequencing to study cancer treatment and its effects is complicated by unique problems. For example, we often want to sample the same tumor before and after treatment, which is hard to do with samples from patients. Additionally, many samples contain too few cells to apply existing single-cell methods (such as liquid biopsies from the clinic – often the only ones available). In cases where we want to study rare cell populations, we must detect, capture, and enrich them in a single-cell manner. Finally, we sometimes want to use RNA Sequencing together with other techniques, and the development of these multimodal tools is still underway.

In this thesis, we describe work that addresses some of the aforementioned needs in single-cell applications to cancer and cancer therapeutics. We begin with one of the first applications of single-cell RNA Sequencing to study patient tumors, particularly surgical resections of metastatic skin cancer, or melanoma. In our investigations, we discovered that the strongest variation of cancer cells is primarily driven by which patient they came from, while the variation in non-cancer cells is driven more by the type of cell. Furthermore, even within a patient, we find that cancer cells exhibit a range of functions. We focused on two pathways, which were previously thought to be mutually exclusive in any given tumor. We discovered that although individual cells tend to activate only one of these pathways, a given tumor could possess both subpopulations of cells in their tumor, which could be responsible for resistance to drugs that target only one pathway at a time.

We next used pre- and post-treatment single-cell measurements to study the effect of cancer treatment, specifically in patients whose cancer has metastasized to their cerebrospinal fluid (CSF) and surrounding brain membranes, a complication known as leptomeningeal disease (LMD). Here, we used a new nanowell-based technology that was designed to perform single-

cell RNA Sequencing from small numbers of cells, such as the clinical liquid biopsies used in this study. Our efforts represent the first single-cell exploration of LMD, and we extensively catalog the cell types and states that are present in the CSF microenvironment of this disease. Furthermore, our study was paired with a clinical trial of a novel immunotherapy drug, which sought to reactivate a specific immune cell known as a CD8+ T cell to attack cancer cells in LMD. Our single-cell investigation was able to detail the effect of this drug on the abundance and behavior of its target cells, demonstrating its effectiveness in the CSF. Finally, we were able to use multiple samples from a subset of these patients to study the evolution of their disease in response to the drug, and hypothesize potential mechanisms that drove these responses.

In the last two chapters, we describe two new devices that are able to capture and measure single cells, and enable novel, powerful investigations of their biology using single-cell RNA Sequencing. First, we extend our previous work using CSF liquid biopsies as proxies for treatment response to blood. We develop a device to capture rare circulating tumor cells (CTCs) in the blood of mice with cancer. This device draws blood from the circulation of a mouse, enriches enriches and collects CTCs, and then returns the blood back to the same mouse, thereby only removing a very small volume of blood and leaving the mouse in a similar overall health. We use this device to study the response of CTCs to drug treatment, and compare the biology and response of CTCs to that of the primary tumor in the same mouse. Finally, we consider new strategies that measure other features of single cells along with RNA to develop a more complete picture of cellular function. We apply a device that can measure the mass and rate of mass accumulation of an individual cell, and equip it with the ability to simultaneously capture that cell for single-cell RNA Sequencing. This kind of measurement provides a window into a cell's function from both a gene expression and physiological perspective at the same time. We use this tool to predict which cancer cells from a patient-derived cell line of glioblastoma will continue to grow after being treated with a drug, and we determine which genes and pathways are active in cells that continue to grow in the presence of drug relative to cells that stopped growing.

Together, this body of work aims to address some of the complications of using single-cell techniques to study the response of cancer to drug treatment, and provides new methods to continue learning how variation in cancer (and other) cells can change the response of a tumor to treatment.

8

Table of Contents

Chapter 1: Introduction

Heterogeneity among individual cells affects function and dysfunction in complex biological systems across a variety of different contexts, including immune[1], differentiation[2], embryonic[3], and neoplastic[4,5]. Resolving and investigating this variation is vital to identifying the cell types and states that compose a cellular ensemble[5-8], exploring the molecular drivers that underlie macroscopically observable phenomena[1,9] and understanding how cellular components work together in response to changes in their environment or external stimuli[3,10].

Cellular heterogeneity is of particular concern in cancer biology, where variation in the malignant and non-malignant cells composing tumors and their surroundings can affect cancer patients' response to treatment and subsequent survival[11]. For example, genetic and phenotypic variation of cancer cells in the tumor microenvironment can complicate the sensitivity of an entire tumor to chemo-[10], radio-[12], or immunotherapy[13,14]. Furthermore, the resident non-malignant cells can also influence tumor response to therapy through direct (cell-cell communication) and indirect (cytokine signaling) interactions that modulate cancer cell phenotype[15,16]. While recent studies have demonstrated that the underlying variation of cells in the microenvironment can modulate resistance to many forms of therapy[5,10,14,17-19], the mechanisms driving these phenomenon have remained difficult to characterize in both human and model systems.

Among the most powerful strategies for profiling cellular phenotype is to capture and sequence the RNA transcripts of a cell, providing a window into its instantaneous gene expression[20]. By detecting the transcript-level expression of all possible genes in a sample, RNA-Sequencing enables novel discovery of active genes and pathways in a manner that targeted profiling methods (such as quantitative polymerase chain reaction (qPCR), mass spectrometry, and

microscopy) do not. Since its development and widespread adoption, RNA-Sequencing has been an effective tool to detail active pathways in a biological system, and has even been applied across numerous clinical settings, such as testing for HIV infection[21,22] and drug efficacy in cancer[23-25]. While transcriptomics is incredibly powerful, these measurements are typically performed in a bulk setting, detecting the average abundance of a transcript across many cells. By measuring the average phenomena, this and other bulk techniques may mischaracterize the distribution of behavioral states within a system or even mischaracterize them altogether[26]. Average misrepresentation is particularly problematic when studying the effect of perturbations, such as drug treatment, as changes in subpopulation representation, rare or transient phenotypes, or other unappreciated diversity are lost. Even while isolation methods, such as flow cytometry, can homogenize analyzed cells into well-defined, seemingly "identical" populations, heterogeneity persists and complicates the interpretation of their bulk measurement[27].

Over the past decade, technologies have been developed that can resolve cellular identity and function at the single-cell level. These tools have been foundational in elucidating genetic, epigenetic, transcriptomic, and proteomic heterogeneity in biological systems[27-29] (see Chapter 2). Single-cell RNA sequencing[7,27], in particular, has benefited from considerable experimental and computational advancements in recent years[32], and has transformed our understanding of cellular biology in health and disease[30,31]. In contrast to techniques that require previous knowledge of the comprising cells and their biomarkers in a biological system, single-cell RNA Sequencing enables true discovery of these components by uncovering patterns in gene expression that distinguish groups of cells from first principles. We can subsequently use this information to identify novel axes of variation among different cell populations and even profile diversity across seemingly "identical" cells. Furthermore, comparing these profiles across a perturbation (such as drug treatment in a tumor) can learn induced changes in composition,

activated pathways, and cellular interactions, better characterizing its effect on a system *de novo*. However, there is considerable need to adjust these techniques to study therapeutic resistance in cancer, including enabling longitudinal investigation to track the short- and long-term effect of a drug in a system, handling of smaller input samples to utilize clinical samples such as biopsies, and increased utilization of the data from any captured single cell when the transcriptome does not provide enough information.

In this thesis, we detail our efforts to address current needs in profiling response dynamics of tumors and their microenvironments at single-cell resolution. We present an overview of the state of the field of single-cell RNA sequencing, describe our work applying these tools to uncover single-cell heterogeneity and its consequences for studying cancer therapeutics, and demonstrate opportunities for, and advantages of, performing multiple measurements on a single cell.

1.1: Challenges and improvements in tool development for single-cell profiling.

Current techniques in single-cell genomics, and particularly in single-cell RNA Sequencing, are challenged by stochasticity in recovery of molecular material and inherent biological noise. As such, the measurement of a given analyte (such as transcript, protein, variant, etc.) in any particular cell can be unreliable due to significant dropout[33,34]. Improved statistical power in zero-inflated, high dropout data can be achieved by increasing confidence in the rate of detection of variables (cells across genes) and the number of measurements in a single observation (genes across cells) (**Figure 1-1**). The genome-wide nature of RNA-Sequencing has maximized the number of gene expression measurements per cell (although additional measurements are possible as covered in Chapters 1.3, 2, and 5). Therefore, much of the current work in single-cell RNA Sequencing has focused on optimizing the interpretation of biological signal, with substantial effort tasked towards improving the number of cells profiled in

an individual experiment. The application of microfluidic devices to aid in cell capture and lysis has expanded the scale of single-cell experiments by several orders of magnitude in the past decade, especially those which utilize early bead-based barcoding strategies[7,35,36] (2013 – 10^2 cells; 2019 – 10^{6+} cells). The development and widespread adoption of these tools has launched large-scale single-cell investigations of many previously uncharacterized systems such as HIV[37], allergy[9,38], and Zika virus[39]. Moreover, their preliminary success and promise has encouraged the establishment of resources towards profiling entire organisms at single-cell resolution[40-42].
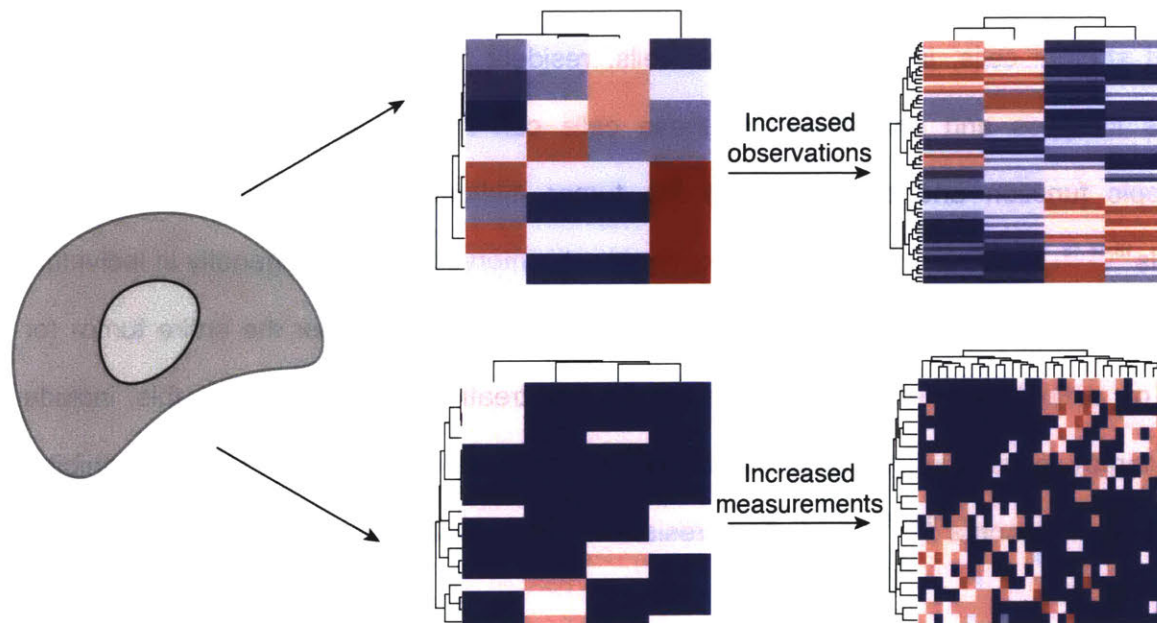

Figure 1-1: Improving biological signal in single-cell measurements. Paucity in both the number of cells profiled and their analytes as well as decreased observation due to technical artifacts (see technical noise, **Figure 2-1 a**) can lead to substantial and fundamentally limiting noise in single-cell measurements. This noise can be overcome by 1) increasing the number of cells profiled (observations) to bolster confidence in true signal across many cells and 2) increasing the number of measurements (of one or multiple types of analyte) to increase the interpretation of true signal from any one cell.

As these techniques have matured and been applied across numerous biological systems, new challenges have surfaced. Some of the general considerations for single-cell experiments include the tradeoff between complexity and cell number and the optimization of rate and

detection of multiple-cell capture. Others are unique to particular circumstances, such as longitudinal and/or spatial transcriptomics of tissue morphology, or the utilization of progressively smaller starting cellular material in clinical settings. Further application of single-cell transcriptomics will build on and tailor current experimental and computational frameworks to increasingly nuanced biological phenomena.

1.2: Consequences of heterogeneity in understanding the effects of cancer therapeutics.

Both solid and liquid tumors consist of a combination of malignant tumor cells and non-malignant stromal cells, infiltrating immune cells, resident immune cells, and others[4,15,16,43]. Variations in types and phenotypes of these cells can have a profound impact on the macroscopic function and dysfunction of the tumor microenvironment within and across individuals[18,44-46]. Moreover, when subject to therapeutic intervention, heterogeneity in individual cells' responses can influence systems-level behaviors, including whether the entire tumor (or subject, or patient) is sensitive or resistant to the treatment. Examples of this include immunotherapy resistance from point mutations[47-49], metabolic underpinnings of BRAF-inhibitor sensitivity in melanoma[10,50,51], and drug resistance in breast cancer resulting from clonal evolution[46,52].

Advances in single-cell genomics, as described above, present a unique opportunity to understand the mechanistic bases for the drug responses observed in complex cancer systems. However, application of these techniques to profile cancer treatment also presents new challenges. In particular, experiments that examine the same tumor microenvironment before and after treatment are necessary to properly interpret the effect of a drug. Additionally, strategies to sample perturbed systems longitudinally can enable study of the long-term effect of treatment. Such experiments will require the development or creative implementation of existing

14

single-cell tools. Moreover, as these investigations grow increasingly translational in nature, methods that accomplish these tasks with low-input, clinical samples hold particular value.

1.3: Opportunities to improve single-cell measurements

Applications of genomic tools to profile single cells have achieved important advances in understanding cellular phenotype, many of which have been noted above. However, while gene expression provides one perspective of cellular behavior, there are inherent limitations to the biological interpretability of transcriptomics. The abundance and activity of gene products (proteins) may represent a more tangible proxy for cellular behavior than gene transcript as they perform the direct chemical and physical biological functions within a cell, and previous work has demonstrated the complications of interpreting protein function from transcript-level readouts[53]. Furthermore, physical properties of individual cells, such as their growth, mobility, and organization, provide additional information about instantaneous cellular function, and can be particularly valuable in understanding a cell's response trajectory[54-57]. As single-cell transcriptomics remains arguably the most widely applicable genome-wide phenotyping option, opportunities abound to collect additional information from single cells simultaneously with transcriptome capture.

In addition to increasing cell numbers and utilization of low-input samples, many novel techniques are currently under development seeking to increase the number of measurements performed on any given single cell. These include protein expression[53], genetic variation[58], epigenetic variation[59,60], biophysical properties[61], temporal properties[62], and spatial organization[63]. These techniques have already uncovered previously unknown information regarding the stochasticity of transcription and translation, structure-function relationships in tissues, and changing clonal architecture of tumor cells over time. As these and other multiplexed modalities continue to improve, single-cell technologies have incredible potential to

resolve widespread biological problems complicated by cellular heterogeneity, particularly in the context of tumor resistance to therapy.

1.4: Contributions of this work

In this thesis, we describe our efforts towards the advancement and utilization of single-cell technologies to profile the response of tumor cells and their microenvironments to therapeutic challenge. This body of work addresses the need for: 1) investigation of single-cell heterogeneity in cancer and its contribution to therapeutic resistance, 2) longitudinal investigation of tumor and microenvironment response to therapy with single-cell genomics, 3) utilization of small cellular inputs, namely liquid biopsies, to profile tumor and microenvironment response at a single-cell level, and 4) incorporation of multiple measurements of single tumor cells to more comprehensively profile their response to challenge.

In Chapter 2, we explore recent developments in single-cell transcriptome profiling, with particular emphasis on the advances that microfluidic tools and their application have made towards achieving more robust and cost-effective single-cell RNA Sequencing experiments. We highlight tools and efforts over the past decade that have significantly improved the number of single cells measured in an experiment and extended the utility of a given captured single cell to include multiple measurements.

In Chapter 3, we detail our early efforts to utilize plate-based single-cell transcriptome profiling to study the tumor microenvironments of patients with melanoma. This study represents among the first explorations of a human tumor microenvironment using a genome-wide single-cell approach, enabling novel, unsupervised discovery of previously uncharacterized heterogeneity and its consequences for patient care. This work has helped promoted an explosion of single-cell investigations in other tumor systems[64-66], and these data have been further explored in

16

studies of tumor heterogeneity[67], algorithm design[68], immunotherapy[69] and immune tolerance[19]. The particular results of our work demonstrate, among other findings, that the presence of patient-specific malignant cell heterogeneity may predict the response of patients in this study to specific chemotherapy agents. We further explore cancer cell subpopulation heterogeneity and its consequences on therapeutic resistance in the following chapters of this thesis.

In Chapter 4, we apply the strategies described in the Chapter 3 to explore the response of a tumor microenvironment to immunotherapy. In particular, we study the effect of Pembrolizumab, a blocking antibody of the immune checkpoint molecule *PD-1* on CD8+ T cells, on the tumor microenvironment of leptomeningeal carcinomatosis (LMD), the metastasis of malignant cells to the cerebrospinal fluid (CSF) and surrounding meninges in the central nervous system (CNS). In this study, we utilize a nanowell-based technology, Seq-Well, to conduct single-cell RNA Sequencing using sparse liquid biopsy material (clinical CSF draw from a clinical trial studying the efficacy of Pembrolizumab to treat LMD (NCT02886585)). Our work represents the first single-cellular description of the LMD microenvironment, elucidating the cell types and states present in the CSF during leptomeningeal invasion. Additionally, the clinical trial NCT02286585 demonstrated clinical efficacy of intravenously administered Pembrolizumab in LMD. Our data demonstrate that intravenous Pembrolizumab produces not only a systemic response, but also a shift in immune composition and phenotype of specific cells directly in the CSF microenvironment. Finally, we use longitudinal single-cell RNA Sequencing and CSF-derived cell-free DNA Sequencing (cfDNA) to develop case studies on the response in individual patients, highlighting adaptive selection and tumor phenotype as potential mechanisms of resistance, and demonstrating the utility of longitudinal single-cell measurements from liquid biopsies to identify subpopulations, their drivers, and their implications in patient resistance.

In Chapter 5, we describe work applying longitudinal liquid biopsy profiling in a controlled murine model system to study the response of rare circulating tumor cells (CTCs), in blood. CTCs

present an opportunity for diagnostic measurement and biomarker discovery in patients with inaccessible tumors. However, the relationship of CTCs to primary tumors, their response to treatment, and their heterogeneity are difficult to characterize. Model organisms such as murine models enable more targeted investigation of rare cells such as CTCs, but CTC studies in murine systems are hindered by low cell numbers and capture efficiency. Furthermore, treatment studies in CTCs are severely limited because the volume of blood required to capture CTCs cannot be drawn across the acute window of drug action while maintaining the physiological condition of the mouse. We describe a novel microfluidic device to enrich and capture CTCs from a single mouse, enabling, for the first time, sampling over acute time intervals of drug response. We perform single-cell RNA Sequencing on these cells, investigating their transcriptional heterogeneity across different mice and their response to JQ1, a selective bromodomain inhibitor. Additionally, we use paired single-cell profiling of matched primary tumor resections to investigate how CTCs differ from solid malignancies in their heterogeneity and response to therapy.

Finally, in Chapter 6, we describe work extending single-cell transcriptome profiling to include additional metadata to analyze the therapeutic response of tumor cells. As described previously, transcriptomic data from a single-cell may not totally capture its phenotype. Other phenomena, especially biophysical features such as mass and change in mass over time, can better elucidate the instantaneous condition of a cell. Mass and growth rate in particular have been recently shown to be powerful predictors in some cancers of the sensitivity or resistance of individual tumor cells following treatment. This observation presents a unique window into the eventual fate of single cells, which is functionally undetectable by profiling the end state by genomic means. To take advantage of this orthogonal readout of cell function, we develop and apply a novel single-cell platform to simultaneously profile the mass, growth rate, and transcriptome of individual cells, demonstrating correlative features between their gene

expression and biophysical properties. We pilot the utility of this platform to investigate the molecular underpinnings of chemotherapy response, applying this technique to patient-derived cell lines (PDCLs) of glioblastoma treated with the transcription inhibitor MDM2. We demonstrate that cells continuing to grow following treatment upregulate resistance mechanisms to MDM2-inhibition, and elucidate the active pathways in these cells that might enable escape, highlighting the potential extension of the results from this study to other cancer systems.

Together, this body of work represents our contributions towards extending single-cell profiling to understand how individual malignant and non-malignant cells in naturally-occurring and model cancer microenvironments interact with and change their behavior following therapeutic intervention, and how single-cell variability can modulate the response of an entire tumor microenvironment.

1.5: References

1   Shalek, A. K. *et al.* Single-cell RNA-seq reveals dynamic paracrine control of cellular variation. *Nature* **510**, 363-369, doi:10.1038/nature13437 (2014).

2   Trapnell, C. *et al.* The dynamics and regulators of cell fate decisions are revealed by pseudotemporal ordering of single cells. *Nat Biotechnol* **32**, 381-386, doi:10.1038/nbt.2859 (2014).

3   Satija, R., Farrell, J. A., Gennert, D., Schier, A. F. & Regev, A. Spatial reconstruction of single-cell gene expression data. *Nat Biotechnol* **33**, 495-502, doi:10.1038/nbt.3192 (2015).

4   Dalerba, P. *et al.* Single-cell dissection of transcriptional heterogeneity in human colon tumors. *Nat Biotechnol* **29**, 1120-1127, doi:10.1038/nbt.2038 (2011).

5   Patel, A. P. *et al.* Single-cell RNA-seq highlights intratumoral heterogeneity in primary glioblastoma. *Science* **344**, 1396-1401, doi:10.1126/science.1254257 (2014).

6   Jones, K. B. *et al.* Quantitative Clonal Analysis and Single-Cell Transcriptomics Reveal Division Kinetics, Hierarchy, and Fate of Oral Epithelial Progenitor Cells. *Cell Stem Cell* **24**, 183-192 e188, doi:10.1016/j.stem.2018.10.015 (2019).

7   Macosko, E. Z. *et al.* Highly Parallel Genome-wide Expression Profiling of Individual Cells Using Nanoliter Droplets. *Cell* **161**, 1202-1214, doi:10.1016/j.cell.2015.05.002 (2015).

8   Zheng, C. *et al.* Landscape of Infiltrating T Cells in Liver Cancer Revealed by Single-Cell Sequencing. *Cell* **169**, 1342-1356 e1316, doi:10.1016/j.cell.2017.05.035 (2017).

9   Ordovas-Montanes, J. *et al.* Allergic inflammatory memory in human respiratory epithelial progenitor cells. *Nature* **560**, 649-654, doi:10.1038/s41586-018-0449-8 (2018).

10   Tirosh, I. *et al.* Dissecting the multicellular ecosystem of metastatic melanoma by single-cell RNA-seq. *Science* **352**, 189-196, doi:10.1126/science.aad0501 (2016).

11   Dagogo-Jack, I. & Shaw, A. T. Tumour heterogeneity and resistance to cancer therapies. *Nat Rev Clin Oncol* **15**, 81-94, doi:10.1038/nrclinonc.2017.166 (2018).

12   Barker, H. E., Paget, J. T., Khan, A. A. & Harrington, K. J. The tumour microenvironment after radiotherapy: mechanisms of resistance and recurrence. *Nat Rev Cancer* **15**, 409-425, doi:10.1038/nrc3958 (2015).

13   Van Allen, E. M. *et al.* Genomic correlates of response to CTLA-4 blockade in metastatic melanoma. *Science* **350**, 207-211, doi:10.1126/science.aad0095 (2015).

14   Riaz, N. *et al.* Tumor and Microenvironment Evolution during Immunotherapy with Nivolumab. *Cell* **171**, 934-949 e916, doi:10.1016/j.cell.2017.09.028 (2017).

15   Joyce, J. A. & Fearon, D. T. T cell exclusion, immune privilege, and the tumor microenvironment. *Science* **348**, 74-80, doi:10.1126/science.aaa6204 (2015).

16   Gajewski, T. F., Schreiber, H. & Fu, Y. X. Innate and adaptive immune cells in the tumor microenvironment. *Nat Immunol* **14**, 1014-1022, doi:10.1038/ni.2703 (2013).

17   Suva, M. L. *et al.* Reconstructing and reprogramming the tumor-propagating potential of glioblastoma stem-like cells. *Cell* **157**, 580-594, doi:10.1016/j.cell.2014.02.030 (2014).

18   Hugo, W. *et al.* Genomic and Transcriptomic Features of Response to Anti-PD-1 Therapy in Metastatic Melanoma. *Cell* **168**, 542, doi:10.1016/j.cell.2017.01.010 (2017).

19   Nirschl, C. J. *et al.* IFNgamma-Dependent Tissue-Immune Homeostasis Is Co-opted in the Tumor Microenvironment. *Cell* **170**, 127-141 e115, doi:10.1016/j.cell.2017.06.016 (2017).

20   Wang, Z., Gerstein, M. & Snyder, M. RNA-Seq: a revolutionary tool for transcriptomics. *Nat Rev Genet* **10**, 57-63, doi:10.1038/nrg2484 (2009).

21   Vargo, J. *et al.* Clinical specificity and sensitivity of a blood screening assay for detection of HIV-1 and HCV RNA. *Transfusion* **42**, 876-885, doi:10.1046/j.1537-2995.2002.00130.x (2002).

22   Stramer, S. L. *et al.* Detection of HIV-1 and HCV infections among antibody-negative blood donors by nucleic acid-amplification testing. *N Engl J Med* **351**, 760-768, doi:10.1056/NEJMoa040085 (2004).

23   Byron, S. A., Van Keuren-Jensen, K. R., Engelthaler, D. M., Carpten, J. D. & Craig, D. W. Translating RNA sequencing into clinical diagnostics: opportunities and challenges. *Nat Rev Genet* **17**, 257-271, doi:10.1038/nrg.2016.10 (2016).

24   Li, P. *et al.* Using circular RNA as a novel type of biomarker in the screening of gastric cancer. *Clin Chim Acta* **444**, 132-136, doi:10.1016/j.cca.2015.02.018 (2015).

25   Cancer Genome Atlas Research, N. Comprehensive genomic characterization defines human glioblastoma genes and core pathways. *Nature* **455**, 1061-1068, doi:10.1038/nature07385 (2008).

26   Elowitz, M. B. & Leibler, S. A synthetic oscillatory network of transcriptional regulators. *Nature* **403**, 335-338, doi:10.1038/35002125 (2000).

27   Shalek, A. K. *et al.* Single-cell transcriptomics reveals bimodality in expression and splicing in immune cells. *Nature* **498**, 236-240, doi:10.1038/nature12172 (2013).

28   Navin, N. *et al.* Tumour evolution inferred by single-cell sequencing. *Nature* **472**, 90-94, doi:10.1038/nature09807 (2011).

29   Fan, H. C., Wang, J., Potanina, A. & Quake, S. R. Whole-genome molecular haplotyping of single cells. *Nat Biotechnol* **29**, 51-57, doi:10.1038/nbt.1739 (2011).

30   Picelli, S. *et al.* Full-length RNA-seq from single cells using Smart-seq2. *Nat Protoc* **9**, 171-181, doi:10.1038/nprot.2014.006 (2014).

31   Trombetta, J. J. *et al.* Preparation of Single-Cell RNA-Seq Libraries for Next Generation Sequencing. *Curr Protoc Mol Biol* **107**, 4 22 21-17, doi:10.1002/0471142727.mb0422s107 (2014).

32   Anikeeva, P. *et al.* Voices in methods development. *Nat Methods* **16**, 945-951, doi:10.1038/s41592-019-0585-6 (2019).

33   Vallejos, C. A., Richardson, S. & Marioni, J. C. Beyond comparisons of means: understanding changes in gene expression at the single-cell level. *Genome Biol* **17**, 70, doi:10.1186/s13059-016-0930-3 (2016).

34   Wagner, A., Regev, A. & Yosef, N. Revealing the vectors of cellular identity with single-cell genomics. *Nat Biotechnol* **34**, 1145-1160, doi:10.1038/nbt.3711 (2016).

35   Klein, A. M. *et al.* Droplet barcoding for single-cell transcriptomics applied to embryonic stem cells. *Cell* **161**, 1187-1201, doi:10.1016/j.cell.2015.04.044 (2015).

36   Gierahn, T. M. *et al.* Seq-Well: portable, low-cost RNA sequencing of single cells at high throughput. *Nat Methods* **14**, 395-398, doi:10.1038/nmeth.4179 (2017).

37   Golumbeanu, M., Mohammadi, P. & Beerenwinkel, N. BMix: probabilistic modeling of occurring substitutions in PAR-CLIP data. *Bioinformatics* **32**, 976-983, doi:10.1093/bioinformatics/btv520 (2016).

38   Croote, D., Darmanis, S., Nadeau, K. C. & Quake, S. R. High-affinity allergen-specific human antibodies cloned from single IgE B cell transcriptomes. *Science* **362**, 1306-1309, doi:10.1126/science.aau2599 (2018).

39   Zanini, F., Pu, S. Y., Bekerman, E., Einav, S. & Quake, S. R. Single-cell transcriptional dynamics of flavivirus infection. *Elife* **7**, doi:10.7554/eLife.32942 (2018).

40   Wagner, D. E. *et al.* Single-cell mapping of gene expression landscapes and lineage in the zebrafish embryo. *Science* **360**, 981-987, doi:10.1126/science.aar4362 (2018).

41   Tabula Muris, C. *et al.* Single-cell transcriptomics of 20 mouse organs creates a Tabula Muris. *Nature* **562**, 367-372, doi:10.1038/s41586-018-0590-4 (2018).

42   Regev, A. *et al.* The Human Cell Atlas. *Elife* **6**, doi:10.7554/eLife.27041 (2017).

43   Joyce, J. A. Therapeutic targeting of the tumor microenvironment. *Cancer Cell* **7**, 513-520, doi:10.1016/j.ccr.2005.05.024 (2005).

44   Hugo, W. *et al.* Non-genomic and Immune Evolution of Melanoma Acquiring MAPKi Resistance. *Cell* **162**, 1271-1285, doi:10.1016/j.cell.2015.07.061 (2015).

45   Sherman, M. H. *et al.* Vitamin D receptor-mediated stromal reprogramming suppresses pancreatitis and enhances pancreatic cancer therapy. *Cell* **159**, 80-93, doi:10.1016/j.cell.2014.08.007 (2014).

46   Kim, C. *et al.* Chemoresistance Evolution in Triple-Negative Breast Cancer Delineated by Single-Cell Sequencing. *Cell* **173**, 879-893 e813, doi:10.1016/j.cell.2018.03.041 (2018).

47   Zaretsky, J. M. *et al.* Mutations Associated with Acquired Resistance to PD-1 Blockade in Melanoma. *N Engl J Med* **375**, 819-829, doi:10.1056/NEJMoa1604958 (2016).

48   Shin, D. S. *et al.* Primary Resistance to PD-1 Blockade Mediated by JAK1/2 Mutations. *Cancer Discov* **7**, 188-201, doi:10.1158/2159-8290.CD-16-1223 (2017).

49   Sade-Feldman, M. *et al.* Resistance to checkpoint blockade therapy through inactivation of antigen presentation. *Nat Commun* **8**, 1136, doi:10.1038/s41467-017-01062-w (2017).

50   Van Allen, E. M. *et al.* The genetic landscape of clinical resistance to RAF inhibition in metastatic melanoma. *Cancer Discov* **4**, 94-109, doi:10.1158/2159-8290.CD-13-0617 (2014).

51   Muller, J. *et al.* Low MITF/AXL ratio predicts early resistance to multiple targeted drugs in melanoma. *Nat Commun* **5**, 5712, doi:10.1038/ncomms6712 (2014).

52   Zhang, Y. *et al.* Somatic Trp53 mutations differentially drive breast cancer and evolution of metastases. *Nat Commun* **9**, 3953, doi:10.1038/s41467-018-06146-9 (2018).

53   Genshaft, A. S. *et al.* Multiplexed, targeted profiling of single-cell proteomes and transcriptomes in a single reaction. *Genome Biol* **17**, 188, doi:10.1186/s13059-016-1045-6 (2016).

54   Son, S. *et al.* Direct observation of mammalian cell growth and size regulation. *Nat Methods* **9**, 910-912, doi:10.1038/nmeth.2133 (2012).

55   Stevens, M. M. *et al.* Drug sensitivity of single cancer cells is predicted by changes in mass accumulation rate. *Nat Biotechnol* **34**, 1161-1167, doi:10.1038/nbt.3697 (2016).

56   Cermak, N. *et al.* High-throughput measurement of single-cell growth rates using serial microfluidic mass sensor arrays. *Nat Biotechnol* **34**, 1052-1059, doi:10.1038/nbt.3666 (2016).

57   Godin, M. *et al.* Using buoyant mass to measure the growth of single cells. *Nat Methods* **7**, 387-390, doi:10.1038/nmeth.1452 (2010).

58   Macaulay, I. C. *et al.* G&T-seq: parallel sequencing of single-cell genomes and transcriptomes. *Nat Methods* **12**, 519-522, doi:10.1038/nmeth.3370 (2015).

59   Cheow, L. F. *et al.* Single-cell multimodal profiling reveals cellular epigenetic heterogeneity. *Nat Methods* **13**, 833-836, doi:10.1038/nmeth.3961 (2016).

60   Bian, S. *et al.* Single-cell multiomics sequencing and analyses of human colorectal cancer. *Science* **362**, 1060-1063, doi:10.1126/science.aao3791 (2018).

61   Kimmerling, R. J. *et al.* Linking single-cell measurements of mass, growth rate, and gene expression. *Genome Biol* **19**, 207, doi:10.1186/s13059-018-1576-0 (2018).

62   Sharma, A. *et al.* Longitudinal single-cell RNA sequencing of patient-derived primary cells reveals drug-induced infidelity in stem cell hierarchy. *Nat Commun* **9**, 4931, doi:10.1038/s41467-018-07261-3 (2018).

63   Rodriques, S. G. *et al.* Slide-seq: A scalable technology for measuring genome-wide expression at high spatial resolution. *Science* **363**, 1463-1467, doi:10.1126/science.aaw1219 (2019).

64   Venteicher, A. S. *et al.* Decoupling genetics, lineages, and microenvironment in IDH-mutant gliomas by single-cell RNA-seq. *Science* **355**, doi:10.1126/science.aai8478 (2017).

65   Tirosh, I. *et al.* Single-cell RNA-seq supports a developmental hierarchy in human oligodendroglioma. *Nature* **539**, 309-313, doi:10.1038/nature20123 (2016).

66   Puram, S. V. *et al.* Single-Cell Transcriptomic Analysis of Primary and Metastatic Tumor Ecosystems in Head and Neck Cancer. *Cell* **171**, 1611-1624 e1624, doi:10.1016/j.cell.2017.10.044 (2017).

67   Racle, J., de Jonge, K., Baumgaertner, P., Speiser, D. E. & Gfeller, D. Simultaneous enumeration of cancer and immune cell types from bulk tumor gene expression data. *Elife* **6**, doi:10.7554/eLife.26476 (2017).

68   DePasquale, E. A. K. *et al.* DoubletDecon: Deconvoluting Doublets from Single-Cell RNA-Sequencing Data. *Cell Rep* **29**, 1718-1727 e1718, doi:10.1016/j.celrep.2019.09.082 (2019).

69   Jerby-Arnon, L. *et al.* A Cancer Cell Program Promotes T Cell Exclusion and Resistance to Checkpoint Blockade. *Cell* **175**, 984-997 e924, doi:10.1016/j.cell.2018.09.006 (2018).

Chapter 2: On the development & utilization of microfluidics for single-cell profiling

Recent advances in cellular profiling have demonstrated substantial heterogeneity in the behavior of cells once deemed 'identical', challenging fundamental notions of cell 'type' and 'state'. Not surprisingly, these findings have elicited substantial interest in deeply characterizing the diversity, interrelationships and plasticity among cellular phenotypes. Such phenotypic heterogeneity is particularly important to assess the biological response of malignant and non-malignant cells in the microenvironment of cancer systems to therapeutic intervention. To explore these questions, experimental platforms are needed that can extensively and controllably profile many individual cells and/or many components within an individual cell. Here, microfluidic structures have an important role because they can facilitate increased capture of single cells and opportunities to profile multiple components within a single cell. Here, I review the current state-of-the-art methodologies with respect to microfluidics for improving mammalian single-cell transcriptomics.

Forward

In this chapter, we provide a brief overview of some of the techniques used to profile single-cell phenotypes, and the underlying challenges associated with this task. We begin with a discussion of the specific problems associated with noise in single-cell data. We then survey some of the advances in technologies tailored for single-cell genomics, focusing on improvements in cell capture enabled by microfluidic technologies and techniques which extend single-cell profiling beyond one analyte to simultaneous measurement of multiple modalities, genomic and otherwise, from a particular single cell. The further development and application of these strategies to understand the phenotypic response of malignant and non-malignant cells in the tumor microenvironment provides the framework for the contributions detailed in subsequent chapters.

2.1: Introduction

The phenotypic identity of a cell is informed by many factors, including the abundance, distribution and dynamics of its internal components and the spatiotemporal pattern of signals it receives from its environment. Scientists have long attempted to classify cells into distinct types based upon defining characteristics, extending from, at first, macroscopic observables (such as anatomical location, gross morphology, origin, or distinct behaviors) to eventually more nuanced molecular ones (like what proteins or mRNAs they express). However, recent advances in the processing and profiling of cellular components have uncovered previously unappreciated heterogeneities within both seemingly 'uniform' cell populations and complex tissues[1-8]. In many instances, these findings have altered existing cellular classification schemes (introducing new categories, redefining their breadth, uncovering more informative features, or suggesting

previously unappreciated interrelationships); in others, they have challenged some of our atomistic operating assumptions and long-held rubrics[9,10].

Accurate cellular classification itself is complicated by the considerable difficulties associated with characterizing the properties of single cells. Indeed, the resolving power of any individual measurement is limited by technical problems associated with handling and profiling the minute inputs obtained from just one cell, as well as the stochasticity inherent in biological processes[26] (**Figure 2-1**). Small processing losses (technical noise) that are inconsequential at the population level can be disastrous when attempting to accurately score single cells (**Figure 2-1 a**). Similarly, differences in the timing of individual cellular events, driven by the biological, physical, and temporal properties that control their generation (intrinsic noise[27]), can average cleanly at the ensemble level, but render any single measurement an unreliable marker of a specific cell's identity (**Figure 2-1 b**). Moreover, given the broad range of factors that can potentially impact cellular phenotype (and hence a cell's classification), several variables can be required for accurate description.
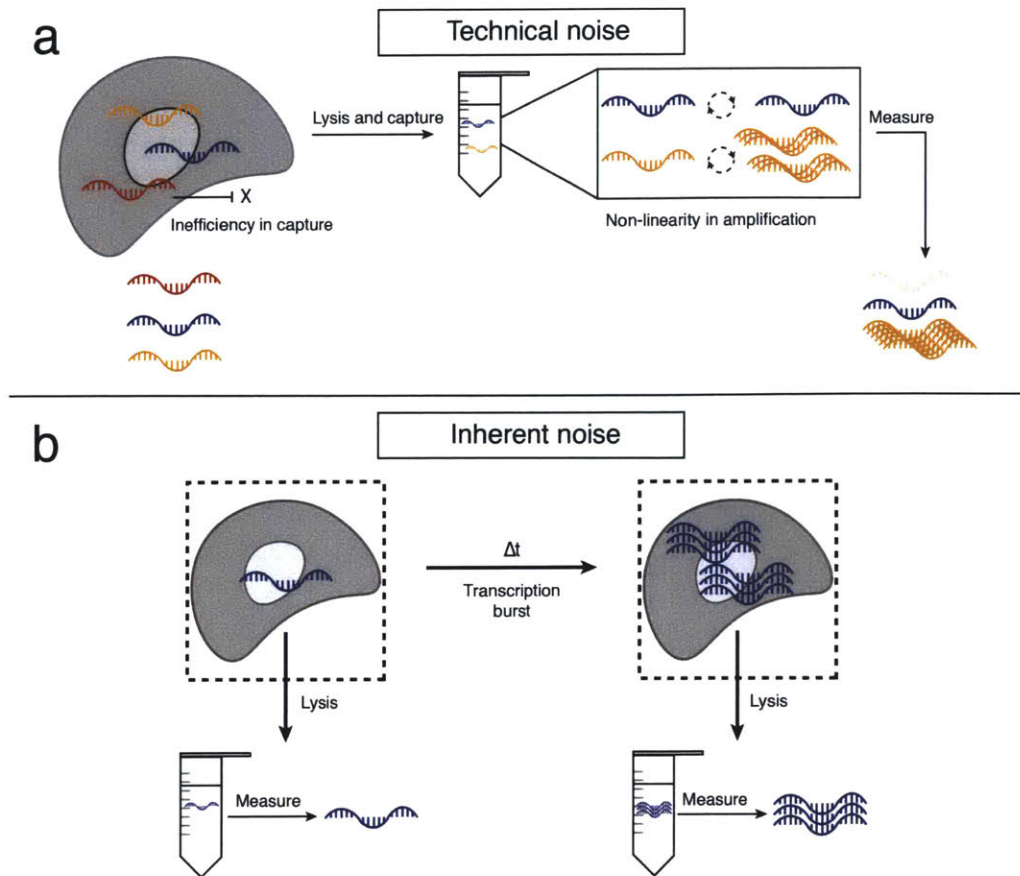
Figure 2-1. Technical and biological noise in single-cell measurements. a) Technical errors in cellular processing ('technical noise'), such as failure to reverse transcribe an mRNA transcript or over-amplification during the ensuing PCR, can dramatically impact the utility of any single measurement in a single-cell experiment. **b)** Similarly, the physical, spatial, and temporal processes governing biological phenomena ('intrinsic noise'), such as the burstiness of mRNA transcription[26], can limit the information content in any single instantaneous end-point measurement.

One strategy for overcoming the noise that is inherent in single-cell measurements is to increase the number of cells profiled. Although any given cellular measurement is subject to systematic (technical noise) and random (intrinsic noise) artifacts, improved throughput, coupled with a fundamental understanding of the limitations of the specific assay in use, can empower studies of the distribution of a variable across a population. Microfluidic devices, tailored to approximately the size of individual cells, can help to achieve this, enhancing experimental

scale by miniaturizing, parallelizing, and integrating. This substantially reduces labor and reagent costs, simplifies workflows, and improves consistency.

A second approach is to increase the number of variables that are measured from a single cell so that a more coherent picture can be achieved. Whereas the expression of any single gene may be an unreliable indicator, the average value of a collection of genes whose expression co-vary across cells is more buffered from noise and thus may more effectively reveal a cell's type, state, or properties[3,6,28,29]. Furthermore, the measurement of additional, orthogonal analytes or metadata of a given cell, including cross-'omic', longitudinal, or biophysical variables can improve the interpretation of biological signal from a given single cell. Over the past few years, several new technologies have been developed that exploit this principle, driven, in part, by the reduced cost and improved accessibility of next-generation sequencing (NGS), a currently preferred method for investigating several variables at once. Microfluidic devices can also considerably improve the preparation of single-cell analytes for NGS-based readouts.

In this chapter, I will describe the common microfluidic approaches for profiling single-cell transcriptomes, as well as advances in techniques that improve the number of single cells profiled and profiling potential of each single cell.

2.2: Increasing the number of single cells measured

Arguably the most common molecular variable used for examining cellular phenotype at the genomic scale is messenger RNA (mRNA). The presence of polyadenylated tails on mRNAs allows universal priming via a single oligo(dT)-based scheme while avoiding excess ribosomal RNA contamination. This enables transcriptome-wide profiling of cellular state, and the identification of distinct cell types, states, and circuits through patterns in gene expression covariation[1,2,4,28,30-35].

In most instances, single-cell transcriptional profiling is performed by first isolating and lysing individual cells of interest, reverse transcribing their mRNA into cDNA and amplifying that material. Afterwards, amplified cDNA, whether obtained by specific target amplification (STA) or whole-transcriptome amplification (WTA), can be profiled using established methods, such as qPCR, dPCR, or sequencing (mRNA-seq). One of the first demonstrations of single-cell transcriptomics using microfluidics was performed by Warren and colleagues[36]: following FACS-based sorting of single hematopoietic stem cells and FLK2$^+$ and FLK2$^-$ common myeloid progenitor cells into strip-tubes for reverse-transcription PCR (RT-PCR), they used a custom valve-based IFC to array each single-cell cDNA sample and perform on-chip quantification of *Pu.1* and *Gapdh* levels by dPCR.

Whereas qPCR and dPCR work has continued to find niche applications, sequencing now dominates. Importantly, foundational work[30,37-40] has helped demonstrate that tens to hundreds of thousands of sequencing reads are sufficient for gene expression analyses in end-tagging and full-length protocols, respectively. Thus, efforts in the field, which were once dominated by protocols performed by FACS-sorting single cells into multiwell plates[1,2,4,41] for tens to hundreds of cells, have now, in part, begun to embrace valve-based microfluidics for profiling hundreds of cells, and droplet- and nanowell-based approaches for studying thousands of cells. For all methods, further work will be needed to realize methods for studying non-polyadenylated RNAs, such as small RNAs[33,35,42].

As mentioned above, valve-based methods initially focused on processing and profiling single-cell transcriptional information. For example, White and coworkers[43] developed a valve-based IFC for single-cell reverse-transcription qPCR (RT-qPCR) that enabled up to two transcripts to be measured on-chip for up to 50 single cells from each of 6 independent sample loading lanes. Here, the number of genes that could be simultaneously assayed was limited due to the decision to detect on-chip, although off-chip detection using other microfluidic chips[14] or

sequencing[4] is possible. The C_1 Auto Prep System's IFC, meanwhile, has transcriptome-compatible protocols that facilitate an order of magnitude improvement in throughput over plate-based methods with an order of magnitude reduction in cost. Leveraging the scale afforded by this platform, researchers have been able to uncover rare immune cell states[3], survey neuronal diversity[34], assess cellular hierarchy within lung epithelia[44], and more.

However, further scaling of this and related systems, which will be necessary to study increasingly complex cellular ensembles, has been hindered by the reliance on separate microfluidic channels for delivering processing reagents to each single cell in parallel. Moreover, for single-cell RNA-Seq, multiwell-plate based methods are still superior with respect to transcript capture[45]. For instances where emphasis is placed on efficiency, defined as the percent of input mRNAs recovered, plate-based methods are preferred. Still, in many instances, more cells, even if captured with lower fidelity, are preferred since, collectively, they minimize the impact of the technical and biological noise associated with any single cell or measurement on global analyses. Among the microfluidic strategies, valve-based systems still provide the greatest molecular efficiency and have utility in applications for which the highest-quality transcriptomes are needed at moderate scale[46]; newer methods that enable on-chip pooling should help to decrease the throughput gap between valve-based and droplet- or well-based approaches.

To address the fundamental shortcomings of scale presented by valve-based microfluidic systems and plate-based methods, several platforms have emerged that leverage the power of early cellular barcoding to achieve high throughput single-cell transcriptomics. By tagging the mRNAs from each cell with a unique barcode during reverse transcription, these methods simplify library preparation by enabling ensemble processing with single-cell resolution via barcode-based computational deconvolution. To perform early barcoding, in one version, microfluidic devices have been used to capture single cells in droplets with uniquely barcoded

30

mRNA capture beads[30,37]. Illustratively, Macosko et al presented a method called Drop-Seq[30] that employs barcoded acrylic mRNA capture beads to achieve bead-bound cDNA replicas of the individual cells with which they are co-encapsulated; this allows user-control over the number of cells sequenced and archiving for subsequent re-querying. Using a mixture of human and mouse cell lines (species mixing experiments), the authors demonstrated the feasibility of Drop-Seq, obtaining >95% cell-of-origin specificity, >7,000 recovered genes per cell, and >12% RNA capture. By collecting 44,808 single-cell profiles from the mouse retina over 4 days of experiments, they uncovered, with high reproducibility, ~39 subtypes and their molecular markers, and subsequently validated select observations *in situ*. However, to avoid cell or bead doublets, Drop-Seq requires that both cells and beads be loaded at low densities into the co-flow device used to confine cells, lyse them, and capture cellular mRNAs by hybridization, prior to breaking the emulsion. As a result, only a small fraction of the cells encapsulated are effectively used.
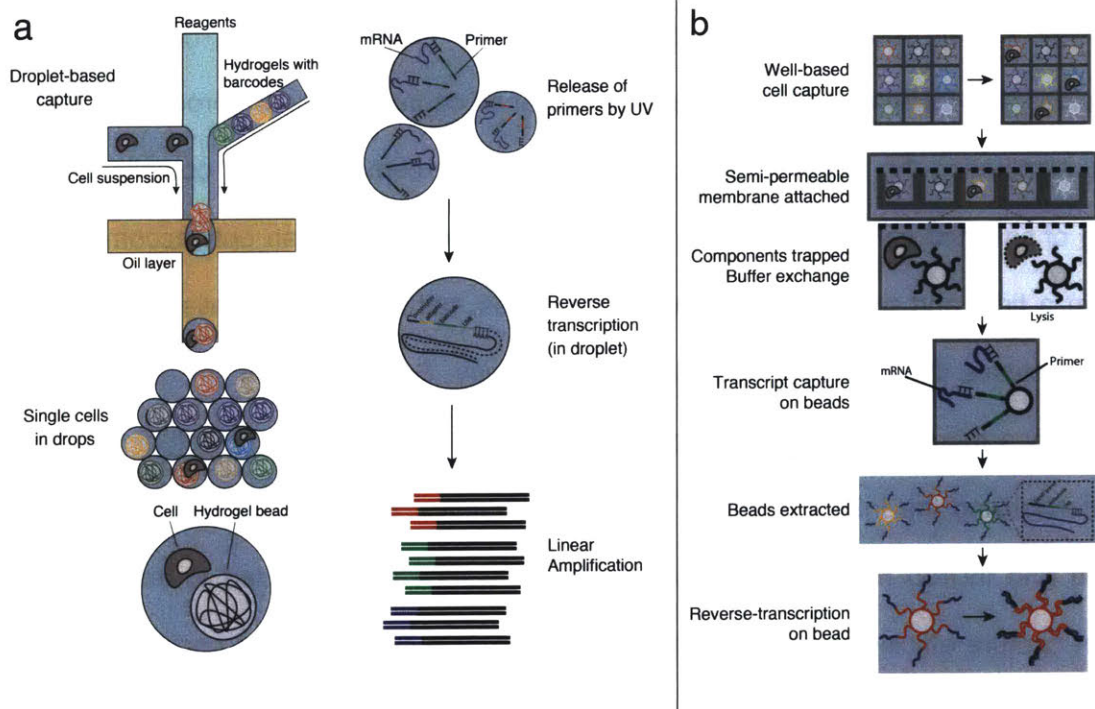

Figure 2-2. Selected examples of microfluidic devices used to measure single-cell transcriptomes and proteomes. a) InDrop from Klein et al.[37], a droplet-based single-cell

31

transcriptomics method, works by co-confining single cells with hydrogel beads containing uniquely barcoded primers (top view). Upon UV-mediated release of those primers, reverse transcription and barcoding is performed in-droplet, and the resulting cDNA is collected after the droplets are broken for subsequent processing. **b)** Seq-Well, a massively-parallel single-cell RNA-sequencing method from Gierahn et al[47], combines early bead-based barcoding with nanowells to generate thousands of single-cell libraries (top: top view; below: side view). Single cells are gravity-loaded onto an array that has been preloaded with uniquely barcoded oligo(dT) capture beads, sealed with a membrane that permits buffer exchange but not mRNA escape, and lysed; mRNAs are then captured by oligo(dT) primers bound to the surface of the barcoded beads, and the beads are removed for off chip reverse transcription, amplification, library preparation and sequencing.

A second technique developed by Klein et al called InDrop[37] overcomes this inefficiency through the use of hydrogel capture particles than can be more efficiently loaded, filling about 90% of the drops (**Figure 2-2 a**). As a result, most encapsulated cells are efficiently used. In InDrop, barcodes are released after cell lysis, and reverse transcription occurs in-droplet prior to the breaking of the emulsion. Species mixing validations showed 96% cell-of-origin specificity, 5,000-15,000 genes, and >7% RNA capture. Klein and coworkers then used this method to probe population structure and cellular heterogeneity in 11,149 mouse embryonic stem cells during their differentiation in response to leukemia inhibitory factor (LIF) withdrawal[37]. Nevertheless, the InDrop workflow necessitates that all captured cells be processed in a single reaction, eliminating the ability to tune sequencing library complexity and complicating the re-sampling of selected cells. Additionally, both Drop-Seq and inDrop can only accommodate one sample at a time and process and lyse cells in series, rather than in parallel, leading to time-dependent biases that can obfuscate underlying biology. Some of these limitations have been overcome by a more user-controllable device introduced by 1-Cell (http://1cell-bio.com/), which is commercializing the InDrop method, and 10x Genomics, which has brought a hybrid of the Drop-Seq and inDrop methods to market[48]. Future work will be needed to improve cell and transcript capture efficiency, and the uniformity of the final library so that each cell gets similar coverage.

Arrays of nanowells can also be used to mate single cells and barcoded beads. Wells rely on gravity-assisted cell and bead loading, reducing the need for peripherals. By matching bead and well size, more efficient bead loading can be achieved, which minimizes the frequency of cells without beads and improves sampling efficiency. In a method they dubbed CytoSeq, Fan et al[29] demonstrated a system for co-confining cells and beads in unsealed nanowell arrays, enabling targeted transcriptional profiling from approximately 5,000 human PBMCs. However, the use of an open well design considerably limited capture efficiency and increased cross-contamination between individual libraries[47].

DeKosky et al[49] utilized a sealed well-based device to isolate and capture mRNA from single B cells with pools of magnetic poly(dT) beads. Following bead removal, the heavy and light chain sequences on each bead were linked using emulsion PCR and sequenced to determine the B cell receptor (BCR) pairs found across individual cells. To enable unique determination of the full transcriptomes of single cells, Bose et al[50] utilized a single uniquely-barcoded bead per well, as introduced in CytoSeq, Drop-Seq and InDrop[29,30,37,51]. Unlike in CytoSeq, here, the authors integrated their nanowell array into a microfluidic circuit to enable oil-based well sealing, allowing them to profile hundreds of cells from a cancer cell line with reduced cross-contamination. A more recent improvement by Yuan and Sims further increased cell and transcript capture efficiency[52]. Still, in both implementations, operation required integrated temperature and pressure controllers, and the use of oil-based sealing limited reagent exchange. To address this requirement, most recently, Gierahn et al[47] (**Figure 2-2 b**) demonstrated improved transcript capture and portability using a simple-to-implement, semi-permeable membrane-based nanowell sealing approach (Seq-Well) that envisions global application of single-cell RNA-Seq to precious samples. In each instance, further improvements can be made in transcript capture. Similarly, in instances where multiple measurements have

been made on the same cell (e.g. immunophenotyping and transcriptional profiling in Seq-Well[47]), experimental strategies are needed to link these mutually informative metrics.

2.3: Increasing and extending the utility of a single cell measurement

In addition to enabling studies of host–pathogen and host–host interactions, microfluidic devices are now beginning to empower multi-'omic' studies that examine the interrelationships among distinct classes of cellular components. Illustratively, recent efforts have begun to characterize how RNA expression patterns correlate with and are driven by the levels and activity of various protein species[10,53-55,56]. Whereas early studies featured fluorescent- or isotope-labelling strategies for proteomic readout, including FACS and mass cytometry[6], a more recent approach leverages PEA-based protein detection which can be coupled with STA using the valve-based design of the C_1 IFC[56]. Similar methodologies, focused on simultaneously profiling mutually informative classes of variables, such as DNA and RNA or RNA and the epigenome[57], appear to be just around the corner (**Box 1**), and may soon guide our understanding of cellular function. Importantly, many other informative properties of cells have been measured with microfluidic devices — for example, morphology[58], proliferation dynamics[59], motility[60], invasiveness[61], interactions[62], calcium dynamics[63], and familial relationships[64] — all of which may help determine cellular identity in the future.

The precision of microfluidic devices holds promise for simultaneously profiling multiple analytes across thousands of single cells. New strategies for improved multi-omics are likely on the horizon, as are approaches for collecting additional, relevant cellular metadata, such as mass[65-68]. Additionally, high throughput methods, such as droplets and nanowells, provide a powerful means of probing and selecting single cells from large ensembles. For example, using droplets, it is possible to directly identify antibody-producing B-cells without the need to first immortalize them through the production of hybridomas[69]. This enables a much faster and deeper search of

34

the antibodies produced by an immunized animal. Similarly, these methods make it possible to screen enzymatic activity and search large libraries of enzymes, produced either in a cell-free manner or expressed in cells, to perform directed evolution for iteratively exploring a much larger library size[70]. Relatedly, microfluidic devices that couple perturbations and '-omic' readouts in many single cells have begun to fundamentally transform our studies of cellular phenotype and function by allowing us to uncover how different cellular components impact cell cellular behaviours[31,71,72,73]. Future work, both experimental and computational, will be required to integrate these profiling and screening approaches to fully leverage the comprehensive views of cellular phenotype they afford.

Before diving in, it is important to first weigh the costs and benefits of porting any single-cell assay into a microfluidic device (**Box 1**). As these advances continue to mature, there are likely to be substantial improvements in the efficiency of processing both cells and cellular analytes. As this occurs, an important question will be how to best balance the number of cells profiled and the depth to which each is sampled, given a fixed number of reads per sequencing run. Further progress will result in overcoming the technical obstacles of preparing different classes of single-cell 'omics.' This will lead to a deeper understanding of how a cell's identity is influenced by its components and their interactions, and the cellular microenvironment, as well as questions of how to balance the number of reads afforded to each different analyte. In parallel, improved computational algorithms, enhanced statistical frameworks, and advanced visualizations will be needed to truly maximize the insights that single-cell 'omic' profiling promises to afford, and guide the development of future devices[6,28,74,75]. Overall, the coming years should to be an extremely exciting time for single-cell biology and the interdisciplinary science that enables.

Note 2-1: **Considerations for porting single-cell 'omic' assays into microdevices.**

Before attempting to develop a new microfluidic device or port an established assay to an existing one, it is important to consider the relative utility of miniaturization and integration in the context of the problem at hand. Indeed, each cellular analyte poses unique challenges with respect to detection and readout; similarly, each assay has nuances that inform if and how it should be adapted.

A first means by which microfluidic devices can improve assay implementations concerns scale. In some instances, such as constructing cellular atlases with single-cell RNA-Seq[1,76], a researcher may need to profile thousands of cells to build as comprehensive a catalogue as sequencer bandwidth allows. In such cases, considerations of scale may surpass others, such as capture efficiency, especially given the limited resolving power of any single cell in light of technical and intrinsic noise sources[4,39,77]. Alternatively, for whole-genome or whole-exome sequencing, a sequencing run may only support full coverage of a few cells[78].

Outside of scale, microdevices can benefit biological assays in other ways, such as by increasing control and reducing background. For example, in the context of whole-genome sequencing, valve-based microdevices can be used to isolate individual chromosomes from a single cell, enabling direct analysis of haplotypes, which would be difficult using a multiwell-plate-based approach[79]. Similarly, for assays that have a constant background per unit volume due to non-specific hybridization, such as PEA[56,80], miniaturization can reduce the noise floor by increasing the relative frequency of positive detection events to spurious ones. Thus, even when the need for scale is not the driving factor, microfluidic devices may have utility.

Ultimately, decisions on when and how to deploy microfluidic devices must carefully weigh the pertinent costs and benefits. At present, valve-, droplet-, and well-based devices each offer unique advantages, making them preferentially suited for different biological tasks. Valve-based devices enable exquisite control over cells and their components, and are often most

appropriate when careful manipulation of analytes is imperative, such as PEA and direct deterministic phasing[56,79,80]. Droplet-based methods do not engender the same degree of precision but facilitate dramatic scaling, making them most pertinent when attempting to profile thousands of cells[30,37,48,76]. Well-based platforms, meanwhile, offer simplicity at intermediate scale with defined spatial locations that ease the coupling of multiple discrete readouts from the same cell[47].

Collectively, these and related considerations will influence how existing assays are ported and the choice of device for future efforts. While it is hard to fully envision what the future holds, we predict that future applications will include new single-cell measurements (such as chromatin conformation via Hi-C (probably using valves))[81,82], biophysical parameters such as cell mass[65,66,68,83] (with requires tracking but not isolation), lineage determination[64] (using hydrodynamic traps), integrated multi-'omic' profiling (probably most easily implemented using valve-based[56] or well-based[47] devices), and studies of how microenvironmental considerations (soluble factors and other cells)[84,85] impact cellular behaviours.

2.4: References

1   Tirosh, I. *et al.* Dissecting the multicellular ecosystem of metastatic melanoma by single-cell RNA-seq. *Science* **352**, 189-196, doi:10.1126/science.aad0501 (2016).

2   Patel, A. P. *et al.* Single-cell RNA-seq highlights intratumoral heterogeneity in primary glioblastoma. *Science* **344**, 1396-1401, doi:10.1126/science.1254257 (2014).

3   Shalek, A. K. *et al.* Single-cell RNA-seq reveals dynamic paracrine control of cellular variation. *Nature* **510**, 363-369, doi:10.1038/nature13437 (2014).

4   Shalek, A. K. *et al.* Single-cell transcriptomics reveals bimodality in expression and splicing in immune cells. *Nature* **498**, 236-240, doi:10.1038/nature12172 (2013).

5   Lohr, J. G. *et al.* Whole-exome sequencing of circulating tumor cells provides a window into metastatic prostate cancer. *Nat Biotechnol* **32**, 479-484, doi:10.1038/nbt.2892 (2014).

6   Bendall, S. C. *et al.* Single-cell mass cytometry of differential immune and drug responses across a human hematopoietic continuum. *Science* **332**, 687-696, doi:10.1126/science.1198704 (2011).

7   Cohen, A. A. *et al.* Dynamic proteomics of individual cancer cells in response to a drug. *Science* **322**, 1511-1516, doi:10.1126/science.1160165 (2008).

8   Feinerman, O. *et al.* Single-cell quantification of IL-2 response by effector and regulatory T cells reveals critical plasticity in immune response. *Mol Syst Biol* **6**, 437, doi:10.1038/msb.2010.90 (2010).

9   Wurtzel, O. *et al.* A Generic and Cell-Type-Specific Wound Response Precedes Regeneration in Planarians. *Dev Cell* **35**, 632-645, doi:10.1016/j.devcel.2015.11.004 (2015).

10   Bjorklund, A. K. *et al.* The heterogeneity of human CD127(+) innate lymphoid cells revealed by single-cell RNA sequencing. *Nat Immunol* **17**, 451-460, doi:10.1038/ni.3368 (2016).

11   Joyce, J. A. Therapeutic targeting of the tumor microenvironment. *Cancer Cell* **7**, 513-520, doi:10.1016/j.ccr.2005.05.024 (2005).

12   Joyce, J. A. & Fearon, D. T. T cell exclusion, immune privilege, and the tumor microenvironment. *Science* **348**, 74-80, doi:10.1126/science.aaa6204 (2015).

13   Gajewski, T. F., Schreiber, H. & Fu, Y. X. Innate and adaptive immune cells in the tumor microenvironment. *Nat Immunol* **14**, 1014-1022, doi:10.1038/ni.2703 (2013).

14   Dalerba, P. *et al.* Single-cell dissection of transcriptional heterogeneity in human colon tumors. *Nat Biotechnol* **29**, 1120-1127, doi:10.1038/nbt.2038 (2011).

15   Hugo, W. *et al.* Non-genomic and Immune Evolution of Melanoma Acquiring MAPKi Resistance. *Cell* **162**, 1271-1285, doi:10.1016/j.cell.2015.07.061 (2015).

16   Hugo, W. *et al.* Genomic and Transcriptomic Features of Response to Anti-PD-1 Therapy in Metastatic Melanoma. *Cell* **168**, 542, doi:10.1016/j.cell.2017.01.010 (2017).

17   Sherman, M. H. *et al.* Vitamin D receptor-mediated stromal reprogramming suppresses pancreatitis and enhances pancreatic cancer therapy. *Cell* **159**, 80-93, doi:10.1016/j.cell.2014.08.007 (2014).

18   Kim, C. *et al.* Chemoresistance Evolution in Triple-Negative Breast Cancer Delineated by Single-Cell Sequencing. *Cell* **173**, 879-893 e813, doi:10.1016/j.cell.2018.03.041 (2018).

19   Zaretsky, J. M. *et al.* Mutations Associated with Acquired Resistance to PD-1 Blockade in Melanoma. *N Engl J Med* **375**, 819-829, doi:10.1056/NEJMoa1604958 (2016).

20   Shin, D. S. *et al.* Primary Resistance to PD-1 Blockade Mediated by JAK1/2 Mutations. *Cancer Discov* **7**, 188-201, doi:10.1158/2159-8290.CD-16-1223 (2017).

21   Sade-Feldman, M. *et al.* Resistance to checkpoint blockade therapy through inactivation of antigen presentation. *Nat Commun* **8**, 1136, doi:10.1038/s41467-017-01062-w (2017).

22   Tirosh, I. *et al.* Dissecting the multicellular ecosystem of metastatic melanoma by single-cell RNA-seq. *Science* **352**, 189-196, doi:10.1126/science.aad0501 (2016).

23   Van Allen, E. M. *et al.* The genetic landscape of clinical resistance to RAF inhibition in metastatic melanoma. *Cancer Discov* **4**, 94-109, doi:10.1158/2159-8290.CD-13-0617 (2014).

24   Muller, J. *et al.* Low MITF/AXL ratio predicts early resistance to multiple targeted drugs in melanoma. *Nat Commun* **5**, 5712, doi:10.1038/ncomms6712 (2014).

25   Zhang, Y. *et al.* Somatic Trp53 mutations differentially drive breast cancer and evolution of metastases. *Nat Commun* **9**, 3953, doi:10.1038/s41467-018-06146-9 (2018).

26   Raj, A., van den Bogaard, P., Rifkin, S. A., van Oudenaarden, A. & Tyagi, S. Imaging individual mRNA molecules using multiple singly labeled probes. *Nat Methods* **5**, 877-879, doi:10.1038/nmeth.1253 (2008).

27   Elowitz, M. B., Levine, A. J., Siggia, E. D. & Swain, P. S. Stochastic gene expression in a single cell. *Science* **297**, 1183-1186, doi:10.1126/science.1070919 (2002).

28   Satija, R., Farrell, J. A., Gennert, D., Schier, A. F. & Regev, A. Spatial reconstruction of single-cell gene expression data. *Nat Biotechnol* **33**, 495-502, doi:10.1038/nbt.3192 (2015).

29   Fan, H. C., Fu, G. K. & Fodor, S. P. Expression profiling. Combinatorial labeling of single cells for gene expression cytometry. *Science* **347**, 1258367, doi:10.1126/science.1258367 (2015).

30   Macosko, E. Z. *et al.* Highly Parallel Genome-wide Expression Profiling of Individual Cells Using Nanoliter Droplets. *Cell* **161**, 1202-1214, doi:10.1016/j.cell.2015.05.002 (2015).

31   Dixit, A. *et al.* Perturb-Seq: Dissecting Molecular Circuits with Scalable Single-Cell RNA Profiling of Pooled Genetic Screens. *Cell* **167**, 1853-1866 e1817, doi:10.1016/j.cell.2016.11.038 (2016).

32   Trapnell, C. *et al.* The dynamics and regulators of cell fate decisions are revealed by pseudotemporal ordering of single cells. *Nat Biotechnol* **32**, 381-386, doi:10.1038/nbt.2859 (2014).

33   Yu, M. *et al.* RNA sequencing of pancreatic circulating tumour cells implicates WNT signalling in metastasis. *Nature* **487**, 510-513, doi:10.1038/nature11217 (2012).

34   Zeisel, A. *et al.* Brain structure. Cell types in the mouse cortex and hippocampus revealed by single-cell RNA-seq. *Science* **347**, 1138-1142, doi:10.1126/science.aaa1934 (2015).

35   Grun, D. *et al.* Single-cell messenger RNA sequencing reveals rare intestinal cell types. *Nature* **525**, 251-255, doi:10.1038/nature14966 (2015).

36   Warren, L., Bryder, D., Weissman, I. L. & Quake, S. R. Transcription factor profiling in individual hematopoietic progenitors by digital RT-PCR. *Proc Natl Acad Sci U S A* **103**, 17807-17812, doi:10.1073/pnas.0608512103 (2006).

37   Klein, A. M. *et al.* Droplet barcoding for single-cell transcriptomics applied to embryonic stem cells. *Cell* **161**, 1187-1201, doi:10.1016/j.cell.2015.04.044 (2015).

38   Pollen, A. A. *et al.* Low-coverage single-cell mRNA sequencing reveals cellular heterogeneity and activated signaling pathways in developing cerebral cortex. *Nat Biotechnol* **32**, 1053-1058, doi:10.1038/nbt.2967 (2014).

39   Heimberg, G., Bhatnagar, R., El-Samad, H. & Thomson, M. Low Dimensionality in Gene Expression Data Enables the Accurate Extraction of Transcriptional Programs from Shallow Sequencing. *Cell Syst* **2**, 239-250, doi:10.1016/j.cels.2016.04.001 (2016).

40   Bacher, R. & Kendziorski, C. Design and computational analysis of single-cell RNA-sequencing experiments. *Genome Biol* **17**, 63, doi:10.1186/s13059-016-0927-y (2016).

41   Kowalczyk, M. S. *et al.* Single-cell RNA-seq reveals changes in cell cycle and differentiation programs upon aging of hematopoietic stem cells. *Genome Res* **25**, 1860-1872, doi:10.1101/gr.192237.115 (2015).

42   Faridani, O. R. *et al.* Single-cell sequencing of the small-RNA transcriptome. *Nat Biotechnol* **34**, 1264-1266, doi:10.1038/nbt.3701 (2016).

43   White, A. K. *et al.* High-throughput microfluidic single-cell RT-qPCR. *Proc Natl Acad Sci U S A* **108**, 13999-14004, doi:10.1073/pnas.1019446108 (2011).

44   Treutlein, B. *et al.* Reconstructing lineage hierarchies of the distal lung epithelium using single-cell RNA-seq. *Nature* **509**, 371-375, doi:10.1038/nature13173 (2014).

45   Ziegenhain, C. *et al.* Comparative analysis of single-cell RNA-sequencing methods. *bioRxiv*, doi:10.1101/035758 (2016).

46   Stubbington, M. J. *et al.* T cell fate and clonality inference from single-cell transcriptomes. *Nat Methods* **13**, 329-332, doi:10.1038/nmeth.3800 (2016).

47   Gierahn, T. M. *et al.* Seq-Well: portable, low-cost RNA sequencing of single cells at high throughput. *Nat Methods* **14**, 395-398, doi:10.1038/nmeth.4179 (2017).

48   Zheng, G. X. *et al.* Massively parallel digital transcriptional profiling of single cells. *Nat Commun* **8**, 14049, doi:10.1038/ncomms14049 (2017).

49   DeKosky, B. J. *et al.* High-throughput sequencing of the paired human immunoglobulin heavy and light chain repertoire. *Nat Biotechnol* **31**, 166-169, doi:10.1038/nbt.2492 (2013).

50   Bose, S. *et al.* Scalable microfluidics for single-cell RNA printing and sequencing. *Genome Biol* **16**, 120, doi:10.1186/s13059-015-0684-3 (2015).

51   Farlik, M. *et al.* Single-cell DNA methylome sequencing and bioinformatic inference of epigenomic cell-state dynamics. *Cell Rep* **10**, 1386-1397, doi:10.1016/j.celrep.2015.02.001 (2015).

52   Yuan, J. & Sims, P. A. An Automated Microwell Platform for Large-Scale Single Cell RNA-Seq. *Sci Rep* **6**, 33883, doi:10.1038/srep33883 (2016).

53   Taniguchi, Y. *et al.* Quantifying E. coli proteome and transcriptome with single-molecule sensitivity in single cells. *Science* **329**, 533-538, doi:10.1126/science.1188308 (2010).

54   Satija, R. & Shalek, A. K. Heterogeneity in immune responses: from populations to single cells. *Trends Immunol* **35**, 219-229, doi:10.1016/j.it.2014.03.004 (2014).

55   Jovanovic, M. *et al.* Immunogenetics. Dynamic profiling of the protein life cycle in response to pathogens. *Science* **347**, 1259038, doi:10.1126/science.1259038 (2015).

56   Genshaft, A. S. *et al.* Multiplexed, targeted profiling of single-cell proteomes and transcriptomes in a single reaction. *Genome Biol* **17**, 188, doi:10.1186/s13059-016-1045-6 (2016).

57   Cheow, L. F. *et al.* Single-cell multimodal profiling reveals cellular epigenetic heterogeneity. *Nat Methods* **13**, 833-836, doi:10.1038/nmeth.3961 (2016).

58   Frisk, T. W., Khorshidi, M. A., Guldevall, K., Vanherberghen, B. & Onfelt, B. A silicon-glass microwell platform for high-resolution imaging and high-content screening with single cell resolution. *Biomed Microdevices* **13**, 683-693, doi:10.1007/s10544-011-9538-2 (2011).

59   Zaretsky, I. *et al.* Monitoring the dynamics of primary T cell activation and differentiation using long term live cell imaging in microwell arrays. *Lab Chip* **12**, 5007-5015, doi:10.1039/c2lc40808b (2012).

60   Khorshidi, M. A. *et al.* Analysis of transient migration behavior of natural killer cells imaged in situ and in vitro. *Integr Biol (Camb)* **3**, 770-778, doi:10.1039/c1ib00007a (2011).

61   Yao, X. *et al.* Functional analysis of single cells identifies a rare subset of circulating tumor cells with malignant traits. *Integr Biol (Camb)* **6**, 388-398, doi:10.1039/c3ib40264a (2014).

62   Yamanaka, Y. J. *et al.* Single-cell analysis of the dynamics and functional outcomes of interactions between human natural killer cells and target cells. *Integr Biol (Camb)* **4**, 1175-1184, doi:10.1039/c2ib20167d (2012).

63   Wheeler, A. R. *et al.* Microfluidic device for single-cell analysis. *Anal Chem* **75**, 3581-3586 (2003).

64   Kimmerling, R. J. *et al.* A microfluidic platform enabling single-cell RNA-seq of multigenerational lineages. *Nat Commun* **7**, 10220, doi:10.1038/ncomms10220 (2016).

65   Cermak, N. *et al.* High-throughput measurement of single-cell growth rates using serial microfluidic mass sensor arrays. *Nat Biotechnol* **34**, 1052-1059, doi:10.1038/nbt.3666 (2016).

66   Stevens, M. M. *et al.* Drug sensitivity of single cancer cells is predicted by changes in mass accumulation rate. *Nat Biotechnol* **34**, 1161-1167, doi:10.1038/nbt.3697 (2016).

67   Godin, M. *et al.* Using buoyant mass to measure the growth of single cells. *Nat Methods* **7**, 387-390, doi:10.1038/nmeth.1452 (2010).

68   Burg, T. P. *et al.* Weighing of biomolecules, single cells and single nanoparticles in fluid. *Nature* **446**, 1066-1069, doi:10.1038/nature05741 (2007).

69   Love, J. C., Ronan, J. L., Grotenbreg, G. M., van der Veen, A. G. & Ploegh, H. L. A microengraving method for rapid selection of single cells producing antigen-specific antibodies. *Nat Biotechnol* **24**, 703-707, doi:10.1038/nbt1210 (2006).

70 Agresti, J. J. *et al.* Ultrahigh-throughput screening in drop-based microfluidics for directed evolution. *Proceedings of the National Academy of Sciences* **107**, 4004-4009, doi:10.1073/pnas.0910781107 (2010).

71 Adamson, B. *et al.* A Multiplexed Single-Cell CRISPR Screening Platform Enables Systematic Dissection of the Unfolded Protein Response. *Cell* **167**, 1867-1882 e1821, doi:10.1016/j.cell.2016.11.048 (2016).

72 Bodenmiller, B. *et al.* Multiplexed mass cytometry profiling of cellular states perturbed by small-molecule regulators. *Nat Biotechnol* **30**, 858-867, doi:10.1038/nbt.2317 (2012).

73 Jaitin, D. A. *et al.* Dissecting Immune Circuits by Linking CRISPR-Pooled Screens with Single-Cell RNA-Seq. *Cell* **167**, 1883-1896 e1815, doi:10.1016/j.cell.2016.11.039 (2016).

74 Bendall, S. C. *et al.* Single-cell trajectory detection uncovers progression and regulatory coordination in human B cell development. *Cell* **157**, 714-725, doi:10.1016/j.cell.2014.04.005 (2014).

75 Wagner, A., Regev, A. & Yosef, N. Revealing the vectors of cellular identity with single-cell genomics. *Nat Biotechnol* **34**, 1145-1160, doi:10.1038/nbt.3711 (2016).

76 Baron, M. *et al.* A Single-Cell Transcriptomic Map of the Human and Mouse Pancreas Reveals Inter- and Intra-cell Population Structure. *Cell Syst* **3**, 346-360 e344, doi:10.1016/j.cels.2016.08.011 (2016).

77 Kim, J. K., Kolodziejczyk, A. A., Ilicic, T., Teichmann, S. A. & Marioni, J. C. Characterizing noise structure in single-cell RNA-seq distinguishes genuine from technical stochastic allelic expression. *Nat Commun* **6**, 8687, doi:10.1038/ncomms9687 (2015).

78 Fu, Y. *et al.* Uniform and accurate single-cell sequencing based on emulsion whole-genome amplification. *Proc Natl Acad Sci U S A* **112**, 11923-11928, doi:10.1073/pnas.1513988112 (2015).

79 Fan, H. C., Wang, J., Potanina, A. & Quake, S. R. Whole-genome molecular haplotyping of single cells. *Nat Biotechnol* **29**, 51-57, doi:10.1038/nbt.1739 (2011).

80 Darmanis, S. *et al.* Simultaneous Multiplexed Measurement of RNA and Proteins in Single Cells. *Cell Rep* **14**, 380-389, doi:10.1016/j.celrep.2015.12.021 (2016).

81 Nagano, T. *et al.* Single-cell Hi-C for genome-wide detection of chromatin interactions that occur simultaneously in a single cell. *Nat Protoc* **10**, 1986-2003, doi:10.1038/nprot.2015.127 (2015).

82 Nagano, T. *et al.* Single-cell Hi-C reveals cell-to-cell variability in chromosome structure. *Nature* **502**, 59-64, doi:10.1038/nature12593 (2013).

83 Burg, T. P. & Manalis, S. R. Suspended microchannel resonators for biomolecular detection. *Applied Physics Letters* **83**, 2698-2700, doi:doi:http://dx.doi.org/10.1063/1.1611625 (2003).

84 Zhao, W. *et al.* Cell-surface sensors for real-time probing of cellular environments. *Nat Nanotechnol* **6**, 524-531, doi:10.1038/nnano.2011.101 (2011).

85 Todhunter, M. E. *et al.* Programmed synthesis of three-dimensional tissues. *Nat Methods* **12**, 975-981, doi:10.1038/nmeth.3553 (2015).

Chapter 3: Dissecting the multicellular ecosystem of metastatic melanoma by single-cell RNA-seq

Tirosh I*, Izar B*[†‡], Prakadan SM, Wadsworth MH II, Treacy D, Trombetta JJ, Rotem A, Rodman C, Lian C, Murphy G, Fallahi-Sichani M, Dutton-Regester K, Lin J, Cohen O, Shah P, Lu D, Genshaft AS, Hughes TK, Ziegler CGK, Kazer SW, Gaillard A, Kolb KE, Villani AC, Johannessen CM, Andreev AY, Van Allen EM, Bertagnolli M, Sorger PK, Sullivan RJ, Flaherty KT, Frederick DT, Jané-Valbuena J, Yoon CH[†], Rozenblatt-Rosen O[†], Shalek AK[†], Regev A[†‡] and Garraway LA[†‡]. "Dissecting the multicellular ecosystem of metastatic melanoma by single-cell RNA-seq," *Science* **352**, 189-196 (2016).

[*,†] Denote equal contribution
[‡] Denote corresponding authorship

To explore the distinct genotypic and phenotypic states of melanoma tumors we applied single-cell RNA-seq to 4,645 single cells isolated from 19 patients, profiling malignant, immune, stromal and endothelial cells. Malignant cells within the same tumor displayed transcriptional heterogeneity associated with the cell cycle, spatial context, and a drug resistance program. In particular, all tumors harbored malignant cells from two distinct transcriptional cell states, such that "MITF-high" tumors also contained "AXL-high" tumor cells. Single-cell analyses suggested distinct tumor micro- environmental patterns, including cell-to-cell interactions. Analysis of tumor-infiltrating T cells revealed exhaustion programs, their connection to T cell activation and to clonal expansion, and their variability across patients. Overall, we begin to unravel the cellular ecosystem of tumors and how single cell genomics offers insights with implications for both targeted and immune therapies.

Key Contributions

BI led the design and implementation of the single-cell experiments. IT led the analysis of single-cell RNA-Sequencing data. SMP, with MWH and others, led the single-cell RNA-Sequencing and pre-processing efforts.

Forward

Tumor biology is influenced by variation among the behaviors of its component cells. This impacts overall tumor function and dysfunction to an extent that has not yet been fully appreciated, largely due to difficulties associated with acquiring primary patient samples and heterogeneity in cell type composition, both within and across samples. Here, we apply single-cell RNA Sequencing to study freshly resected tissue samples of metastatic melanoma across 19 patients. Our results represent among the first applications of single-cell genomics to examine the tumor microenvironment of patient-derived tissue samples, and detail specific consequences of tumor cell heterogeneity on resistance of patients to therapeutic intervention. Together, these investigations reveal important information about how cellular variation drives tumor function, and introduce new problems regarding the application of single-cell genomics to cancer systems, which we seek to address in future chapters.

3.1: Introduction

Tumors are complex ecosystems defined by spatiotemporal interactions between heterogeneous cell types, including malignant, immune and stromal cells[1]. Each tumor's cellular composition, as well as the interplay between these components, may exert critical roles in cancer development[2]. However, the specific components, their salient biological functions, and the means by which they collectively define tumor behavior remain incompletely characterized.

44

Tumor cellular diversity poses both challenges and opportunities for cancer therapy. This is exemplified by the varied clinical efficacy achieved in malignant melanoma with targeted therapies and immunotherapies. Immune checkpoint inhibitors can produce clinical responses in some patients with metastatic melanomas[3-7]; however, the genomic and molecular determinants of response to these agents remain poorly understood. Although tumor neoantigens and the PD-L1 antibody clearly contribute to this response[8-10], it is likely that other factors from subsets of malignant cells, the microenvironment, and tumor-infiltrating lymphocytes (TILs) also play essential roles[11].

Melanomas that harbor the $BRAF^{V600E}$ mutation are commonly treated with RAF/MEK-inhibition prior to or following immune checkpoint inhibition. Although this regimen improves survival, virtually all tumors eventually develop resistance to these drugs[12,13]. Unfortunately, no targeted therapy currently exists for patients whose tumors lack BRAF mutations—including NRAS mutant tumors, those with inactivating NF1 mutations, or rarer events (e.g., RAF fusions). Collectively, these factors highlight the need for a deeper understanding of melanoma composition and its impact on clinical course.

The next wave of therapeutic advances in cancer will likely be accelerated by technologies that assess the malignant, micro-environmental, and immunologic states most likely to inform treatment response and resistance. Ideally we would be able to assess salient cellular heterogeneity by quantifying variation in oncogenic signaling pathways, drug-resistant tumor cell subsets, and the spectrum of immune, stromal and other cell states that may inform immunotherapy response. Toward this end, single-cell genomic approaches enable detailed evaluation of genetic and transcriptional features present in 100s-1000s of individual cells per tumor[14-16]. In principle, this approach may identify all major cellular components simultaneously, determine their individual genomic and molecular states[16], and ascertain which of these features may predict or explain clinical responses to anticancer agents. In order to explore this question

we used single-cell RNA-seq to examine intra- and inter-tumoral heterogeneities in both malignant and non-malignant cell types and states, their drivers and interrelationships in the complex tumor cellular ecosystem.

Results

3.2: Profiles of individual cells from patient-derived melanoma tumors

We measured single-cell RNA-seq profiles from 4,645 malignant, immune and stromal cells isolated from 19 freshly procured melanoma tumors that span a range of clinical and therapeutic backgrounds (**Table 3-1**). These included ten metastases to lymphoid tissues (nine to lymph nodes and one to the spleen), eight to distant sites (five to sub-cutaneous/intramuscular tissue and three to the gastrointestinal tract) and one primary acral melanoma. Genotypic information was available for 17 of 19 tumors, of which four had activating mutations in *BRAF* and five in *NRAS* oncogenes; eight patients were *BRAF/NRAS* wild-type (**Table 3-1**).

To isolate viable single cells suitable for high-quality single-cell RNA-seq, we developed and implemented a rapid translational workflow (**Figure 3-1**)[16]. Tumor tissues were processed immediately following surgical procurement, and single-cell suspensions were generated within ~45 minutes with an experimental protocol optimized to reduce artifactual transcriptional changes introduced by disaggregation, temperature, or time. Once in suspension, individual viable immune (CD45+) and non-immune (CD45-) cells (including malignant and stromal cells) were recovered by flow cytometry (FACS). Next, cDNA was prepared from the individual cells, followed by library construction and massively parallel sequencing. The average number of mapped reads per cell was ~150,000, with a median library complexity of 4,659 genes for malignant cells and 3,438 genes for immune cells, comparable to previous studies of only malignant cells from fresh glioblastoma tumors[16].
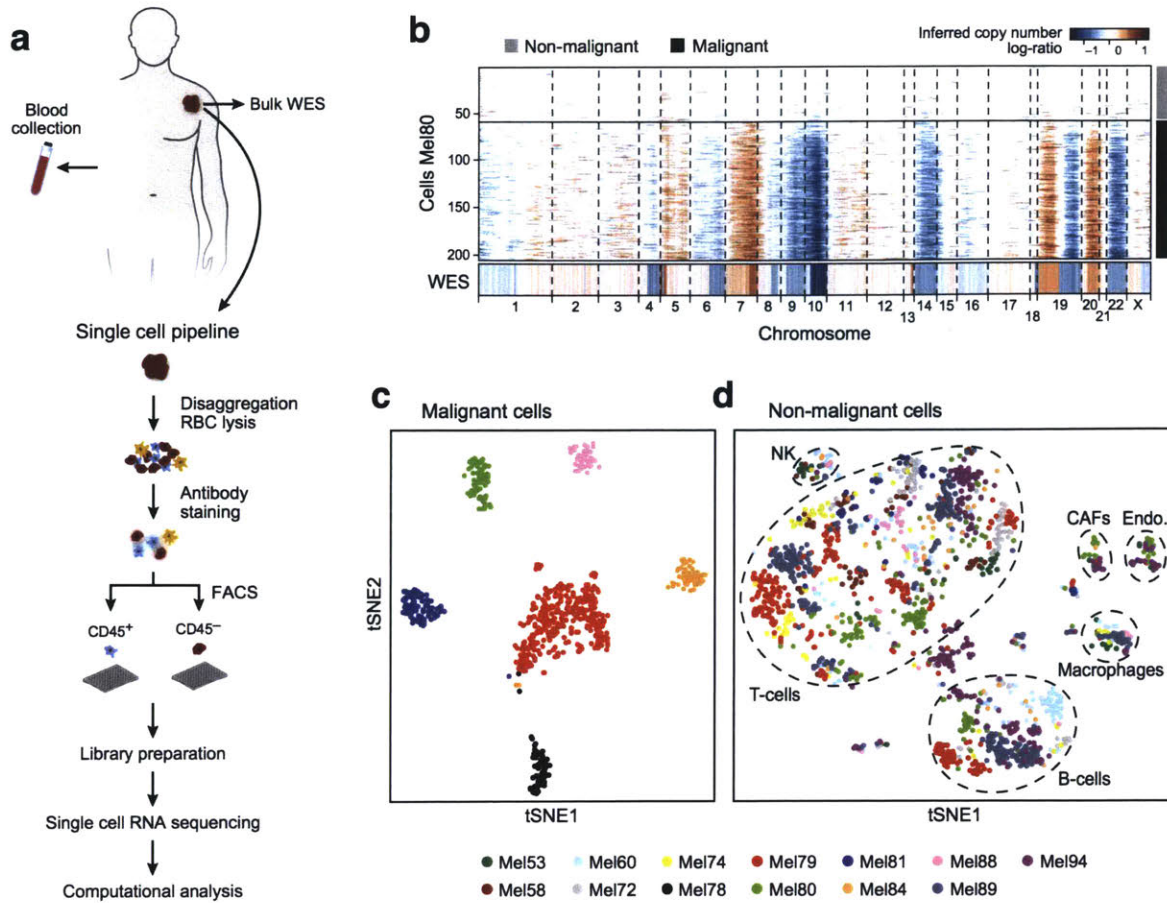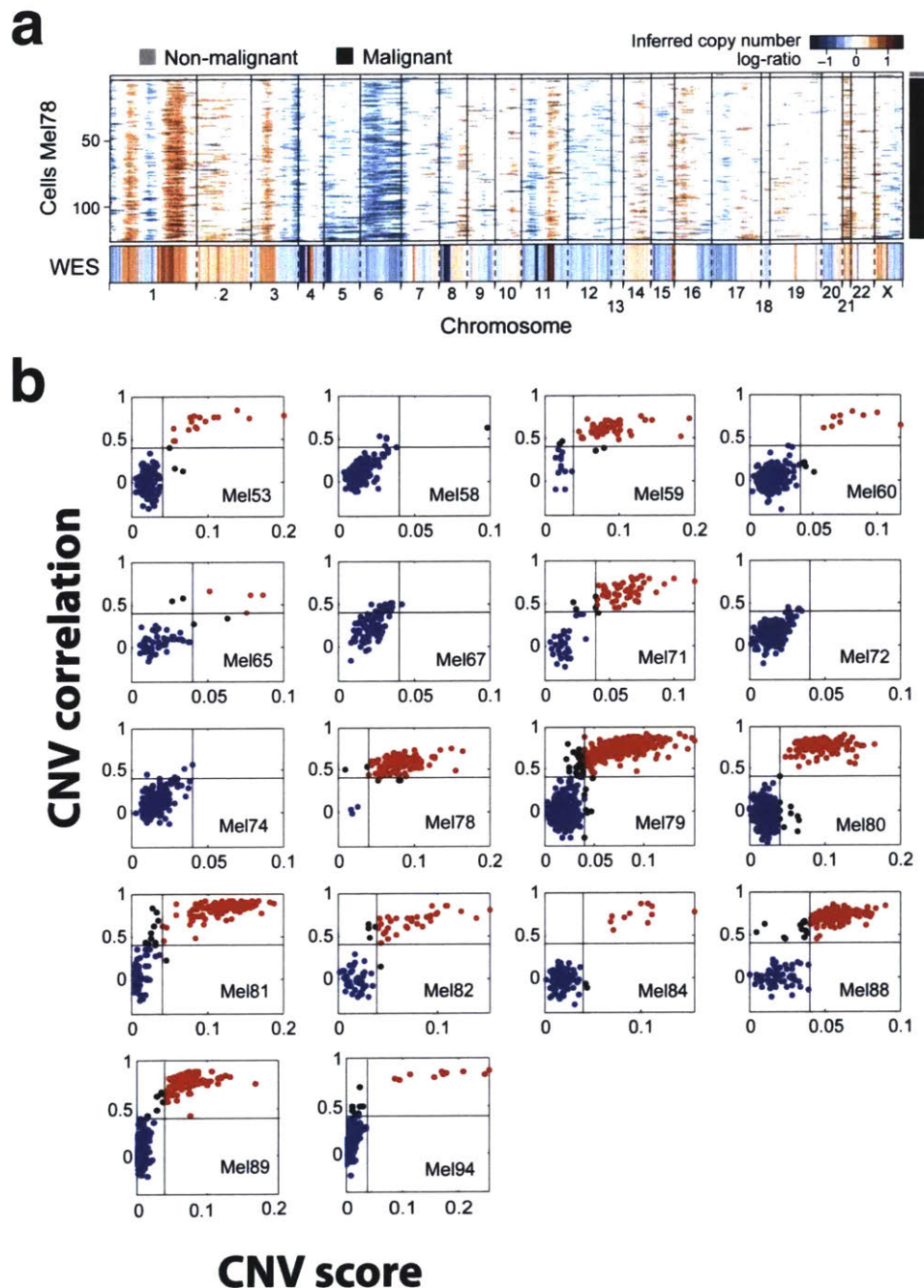
46

Figure 3-1: Dissection of melanoma with single-cell RNA-seq. (a) Overview of workflow. **(b)** Chromosomal landscape of inferred large-scale copy number variations (CNVs) distinguishes malignant from non-malignant cells. The Mel80 tumor is shown with individual cells (y-axis) and chromosomal regions (x-axis). Amplifications (red) or deletions (blue) were inferred by averaging expression over 100-gene stretches on the respective chromosomes. Inferred CNVs are concordant with calls from whole-exome sequencing (WES, bottom). **(c,d)** Single cell expression profiles distinguish malignant and non-malignant cell types. Shown are t-SNE plots of malignant (c, shown are the six tumors each with >50 malignant cells) and non-malignant (d) cells (as called from inferred CNVs as in b) from 11 tumors with >100 cells per tumor (color code). Clusters of non-malignant cells (called by DBScan) are marked by dashed ellipses and were annotated as T cells, B cells, macrophages, CAFs and endothelial cells, from preferentially expressed genes (**Figure 3-3**).

3.3: Single-cell transcriptome profiles distinguish cell states in malignant & non-malignant cells

We used a multi-step approach to distinguish the different cell types within melanoma tumors on the basis of both genetic and transcriptional states (**Figure 3-1 a**). First, we inferred large-scale

copy number variations (CNVs) from expression profiles by averaging expression over 100-gene stretches on their respective chromosomes[16] (**Figure 3-1 b**). For each tumor, this approach revealed a common pattern of aneuploidy, which we validated in two tumors by bulk whole-exome sequencing (WES, **Figure 3-1 b**, **Figure 3-2**). Cells in which aneuploidy was inferred were classified as malignant cells (**Figure 3-1 b**, **Figure 3-2**).

Figure 3-2: Classification of cells to malignant and non-malignant based on inferred CNV patterns. (a) Same as shown in Figure 1b for another melanoma tumor (Mel78). (b) Each plot compares two CNV parameters for all cells in a given tumor: (1) CNV score (X-axis) reflects the overall CNV signal, defined as the mean square of the CNV estimates across all genomic locations; (2) CNV correlation (Y-axis) is the Pearson correlation coefficient between each cell's CNV pattern and the average CNV pattern of the top 5% of cells from the same tumor with respect to CNV signal (i.e., the most confidently-assigned malignant cells). These two values were used to classify cells as malignant (red; CNV score > 0.04; correlation score > 0.4; grey

49

lines mark thresholds on plot), non-malignant (blue; CNV score < 0.04; correlation score < 0.4), or unresolved intermediates (black, all remaining cells). In four tumors (Mel58, 67, 72 and 74), we sequenced primarily the immune infiltrates (CD45+ cells) and there were only zero or one malignant cells by this definition; in those cases, CNV correlation is not indicative of malignant cells (since the top 5% cells by CNV signal are primarily non-malignant) and therefore all cells except for one in Mel58 were defined as non-malignant. Note that while these thresholds are somewhat arbitrary, this classification was highly consistent with the clustering patterns of these cells (as shown in Figure 1c) into clusters of malignant and non-malignant cells.

Second, we grouped the cells on the basis of their expression profiles (**Figure 3-1 c,d**, **Figure 3-3**). Here, we used non-linear dimensionality reduction (t-Distributed Stochastic Neighbor Embedding (*t-SNE*)), followed by density clustering[17]. Generally, cells designated as malignant by CNV analysis formed a separate cluster for each tumor (**Figure 3-1 c**), suggesting a high degree of inter-tumor heterogeneity. In contrast, the non-malignant cells clustered by cell type (**Figure 3-1 d**, **Figure 3-3**), independent of their tumor of origin and metastatic site (**Figure 3-4**). Clusters of non-malignant cells were annotated as T cells, B cells, macrophages, endothelial cells, cancer-associated fibroblasts (CAFs) and NK cells on the basis of preferentially or uniquely expressed marker genes (**Figure 3-1**, **Figure 3-3**).
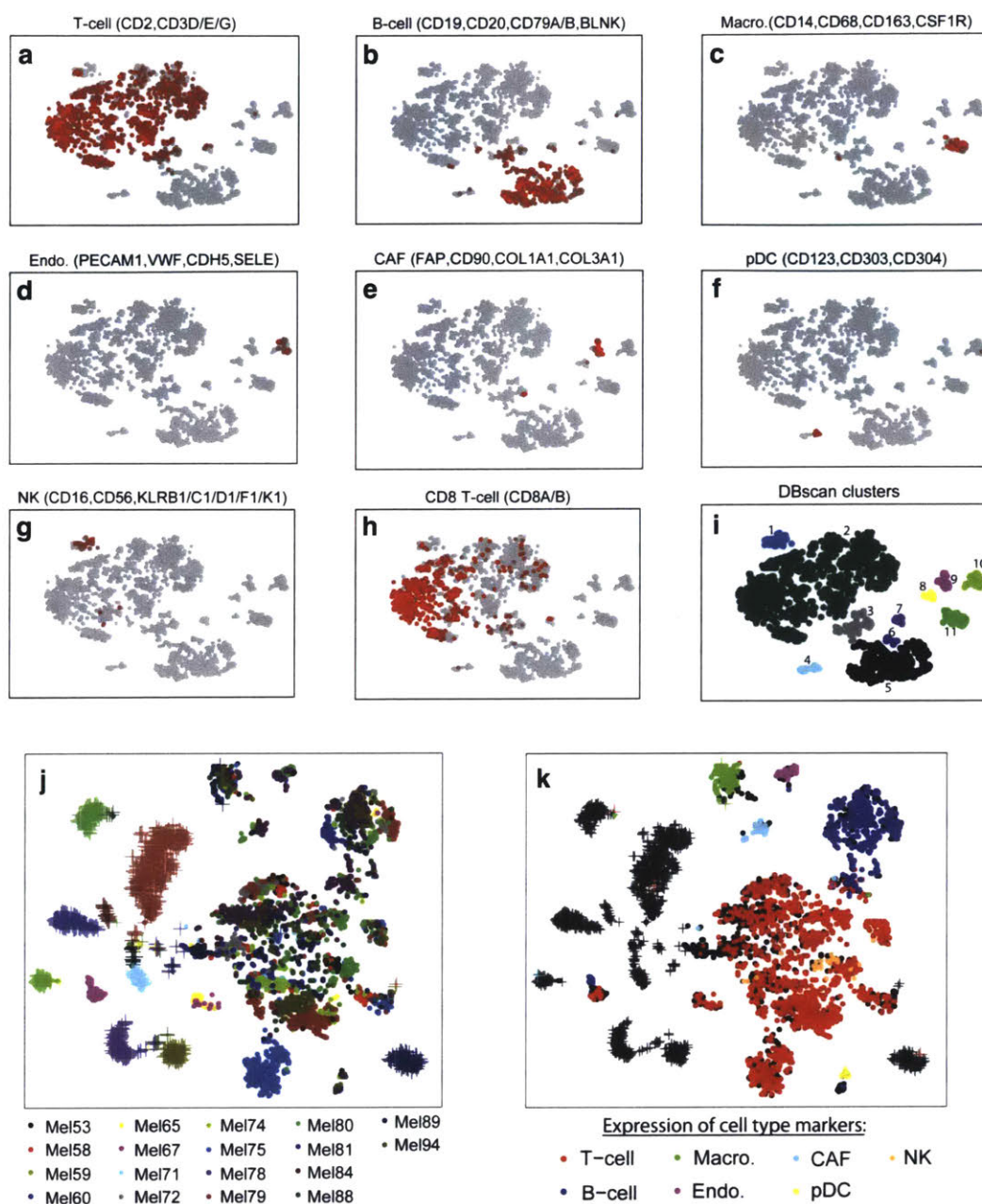
Figure 3-3: Identification of non-malignant cell types by tSNE clusters that preferentially express cell type markers. (a–h) Each plot shows the average expression of a set of known marker genes for a particular cell type (as indicated at the top) overlaid on the tSNE plot of non-malignant cells, as shown in **Figure 1c**. Gray indicates cells with no or minimal expression of the marker genes (**e**, average log2(TPM+1), below 4), dark red indicates intermediate expression (4<E<6), and light red indicates cells with high expression (E>6). (**i**) DBscan clusters derived from tSNE coordinates, with parameters eps=6 and min-points=10. Eleven clusters are indicated by numbers and colors. (**j-k**) Combined plot of all cells profiled in this work. Colors indicate tumor-of-origin in (**j**) and expression of cell type-specific marker genes (E>5) in (**k**).
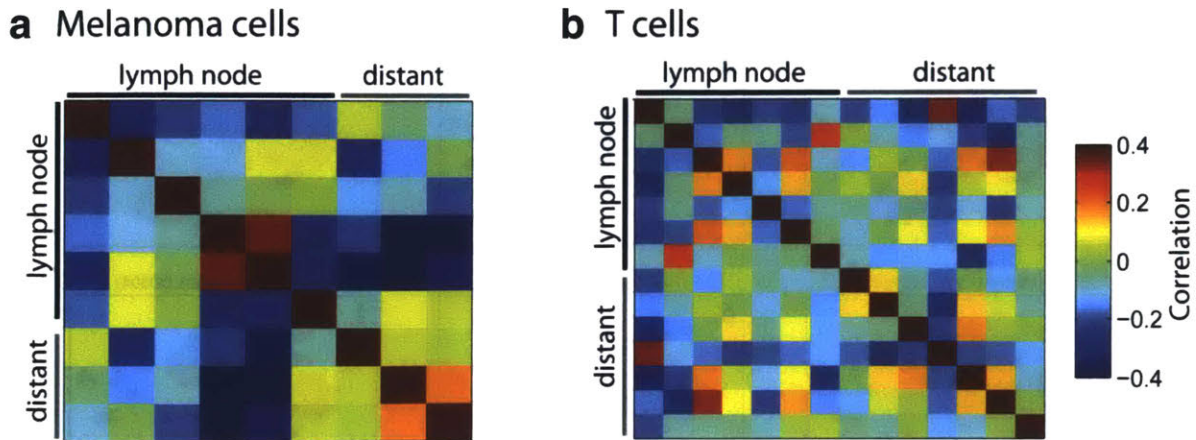
a Melanoma cells **b** T cells

Figure 3-4. Limited influence of tumor site on RNA-seq patterns. (**a–b**) Heat maps show correlations of global expression profiles between tumors, which were ordered by metastatic site. Expression levels were first averaged over melanoma (**a**) or T cells (**b**) in each tumor and then centered across the different tumors before calculating Pearson correlation coefficients. Differential expression analysis conducted between the two groups of tumors found zero differentially expressed genes with FDR of 0.05 based on a shuffling test for both T cells and melanoma cells.

3.5: Heterogeneity in the abundance of a dormant, drug-resistant melanoma subpopulation

Pre-treatment melanoma tumors may harbor malignant cell subsets less likely to respond to targeted therapy. The transcriptional programs associated with principal components PC4 and PC5 of these data were highly correlated with expression of *MITF* (microphthalmia-associated transcription factor), which encodes the master melanocyte transcriptional regulator and a melanoma lineage-survival oncogene[18]. Scoring genes by their correlation to *MITF* across single cells, we identified a "MITF-high" program consisting of several MITF targets, including *TYR*, *PMEL* and *MLANA*. A second transcriptional program, negatively correlated with the MITF program and with PC4 and PC5 (Pearson correlation, $P<10^{-24}$), included *AXL* and *NGFR* (p75NTR), a marker of resistance to various targeted therapies[19,20] and a putative melanoma cancer stem cell marker[21], respectively. Thus, these transcriptional programs resemble reported[22-25] "MITF-high" and "MITF-low/AXL-high" ("AXL-high") transcriptional profiles that can distinguish melanoma tumors, cell lines and mice models. Notably, "AXL-high" program has been linked to intrinsic resistance to RAF/MEK inhibition[22-24].
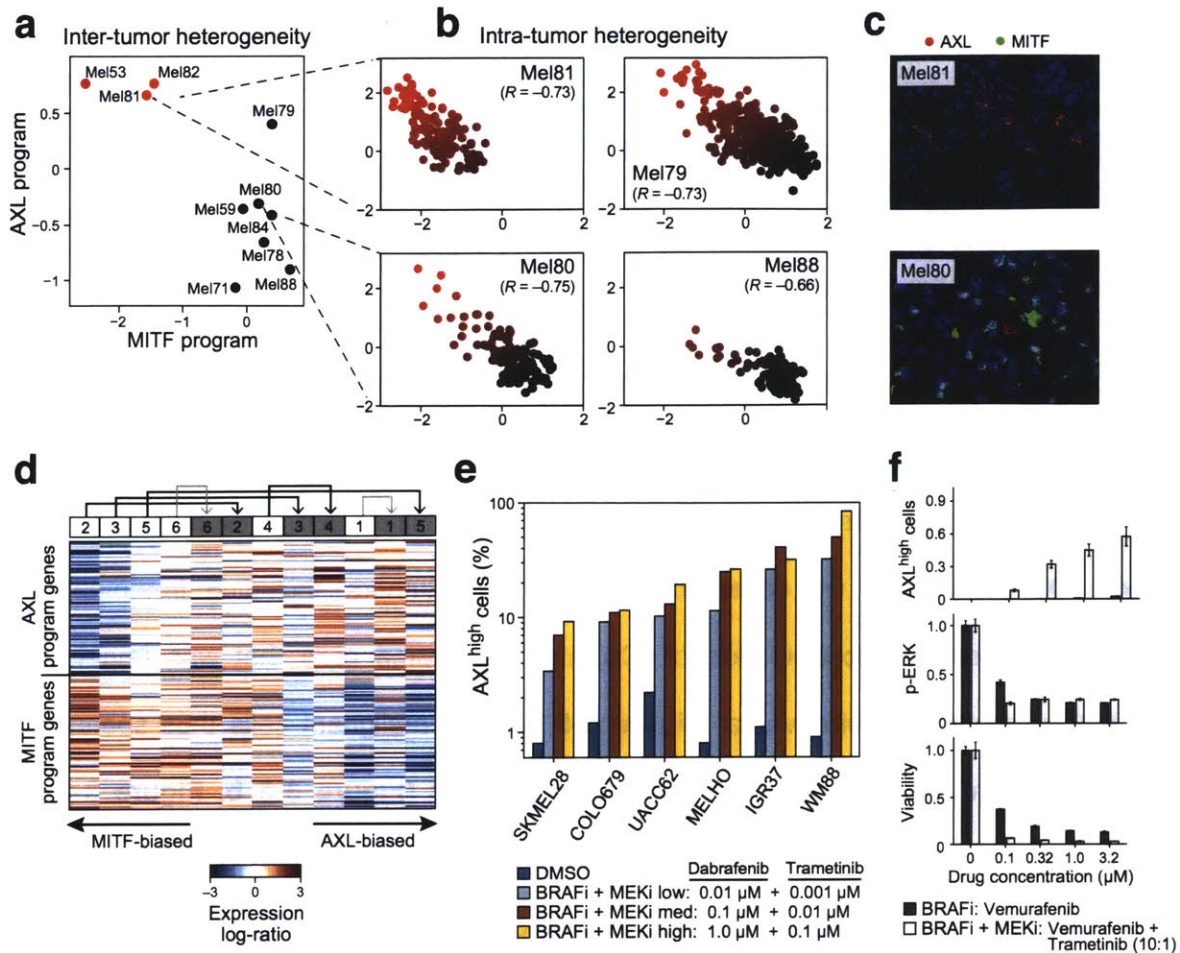
Figure 3-5: MITF- and AXL-associated expression programs vary between and within tumors, and following treatment. (**a**) Average expression signatures for the AXL program (y-axis) or the MITF program (x-axis) stratify tumors into 'MITF-high' (black) or 'AXL-high' (red). (**b**) Single-cell profiles show a negative correlation between the AXL program (y-axis) and MITF program (x-axis) across individual malignant cells within the same tumor; cells are colored by the relative expression of the MITF (black) and AXL (red) programs. Cells in both states are found in all examined tumors, including three tumors (Mel79, Mel80 and Mel81) without prior systemic treatment, indicating that dormant resistant (AXL-high) cells may be present in treatment naïve patients. (**c**) Mel81 and Mel80 immunofluorescence staining of MITF (green nuclei) and AXL (red), validating the mutual exclusivity among individual cells within the same tumor (**Figure 3-7**). (**d**) Relative expression (centered) of the AXL-program (top) and MITF-program (bottom) genes in six matched pre-treatment (white boxes) and post-relapse (gray boxes) samples from patients who progressed through RAF/MEK inhibition therapy; numbers at the top indicate patient index. Samples are sorted by the average relative expression of the AXL *vs.* MITF gene-sets. In all cases, the relapsed samples had increased ratio of AXL/MITF expression compared to their pre-treatment counterpart. This consistent shift of all six patients is statistically significant (P<0.05, binomial test), as are the individual increases in AXL/MITF for four of the six sample pairs (*P*<0.05, t-test; black and gray arrows denote increases that are individually significant or non-significant, respectively). (**e**) Flow-cytometric quantification of the relative fraction of cells with AXL-high (log-scale, y-axis) expression, when cells were treated with increasing doses of RAF/MEK-inhibition (dabrafenib and trametinib in a 10:1 ratio at

indicated doses). In all examined cell lines (x-axis), there was a dose-dependent increase in the AXL-high expressing cell fraction. **(f)** Quantitative, multiplexed single-cell immunofluorescence for AXL expression (y-axis top), MAP-kinase pathway inhibition (pERK levels, y-axis) and viability (y-axis bottom) in the example cell line WM88 treated with increasing concentrations (y-axis) of either RAF inhibitor alone (black bars) or a combination of RAF/MEK-inhibitors (yellow bars). We observe increasing relative AXL-high expressing cell fraction (top panel), consistent with flow-cytometry, as well as a dose-dependent decrease of p-ERK (middle) and viability (bottom), overall consistent with phenotypic selection (killing of MITF-high cells) as part of the shift towards the AXL-high fraction.

While each melanoma could be classified as "MITF-high" or "AXL-high" at the bulk tumor level (**Figure 3-5 a**), at the single cell level every tumor contained malignant cells corresponding to both transcriptional states. Using single-cell RNA-seq to examine each cell's expression of the MITF and AXL gene sets, we observed that MITF-high tumors, including treatment-naïve melanomas, harbored a subpopulation of AXL-high melanoma cells that was undetectable through bulk analysis, and vice versa (**Figure 3-5 b**). The malignant cells thus spanned the continuum between AXL-high and MITF-high states in both (**Figure 3-5 b**, **Figure 3-6**). We further validated the mutually exclusive expression of the MITF-high and AXL-high programs in cells from the same bulk tumors by immunofluorescence (**Figure 3-5 c**, **Figure 3-7**)
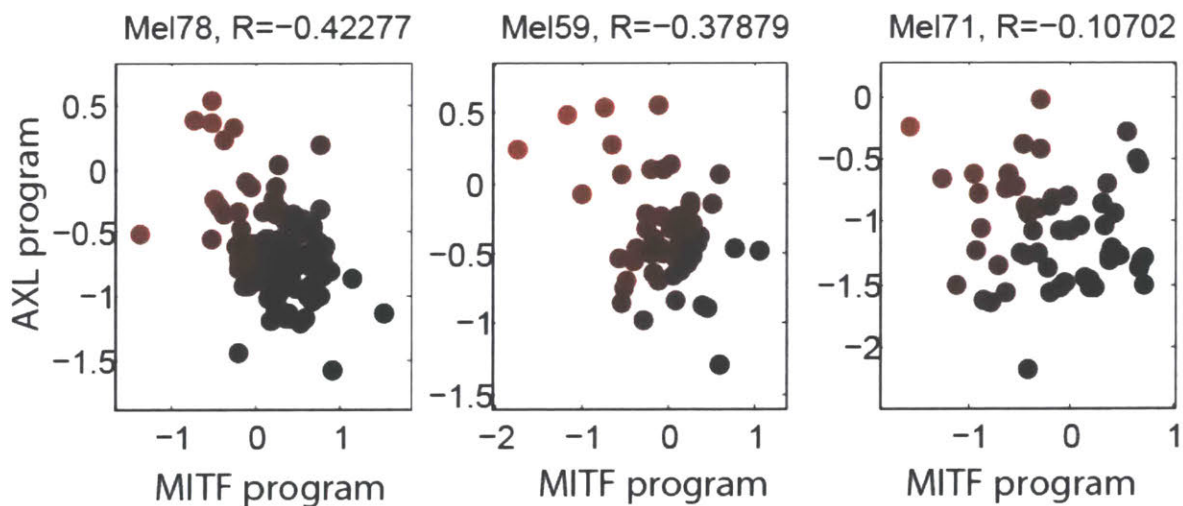

Figure 3-6: Intra-tumor heterogeneity in AXL and MITF programs. AXL-program (Y-axis) and MITF-program (X-axis) scores for malignant cells in each of the three tumors with a sufficient number of malignant cells (n>50) that were not included in **Figure 3-5**. Cells are

colored from black to red by the relative AXL and MITF scores. The Pearson correlation coefficient is denoted on top.
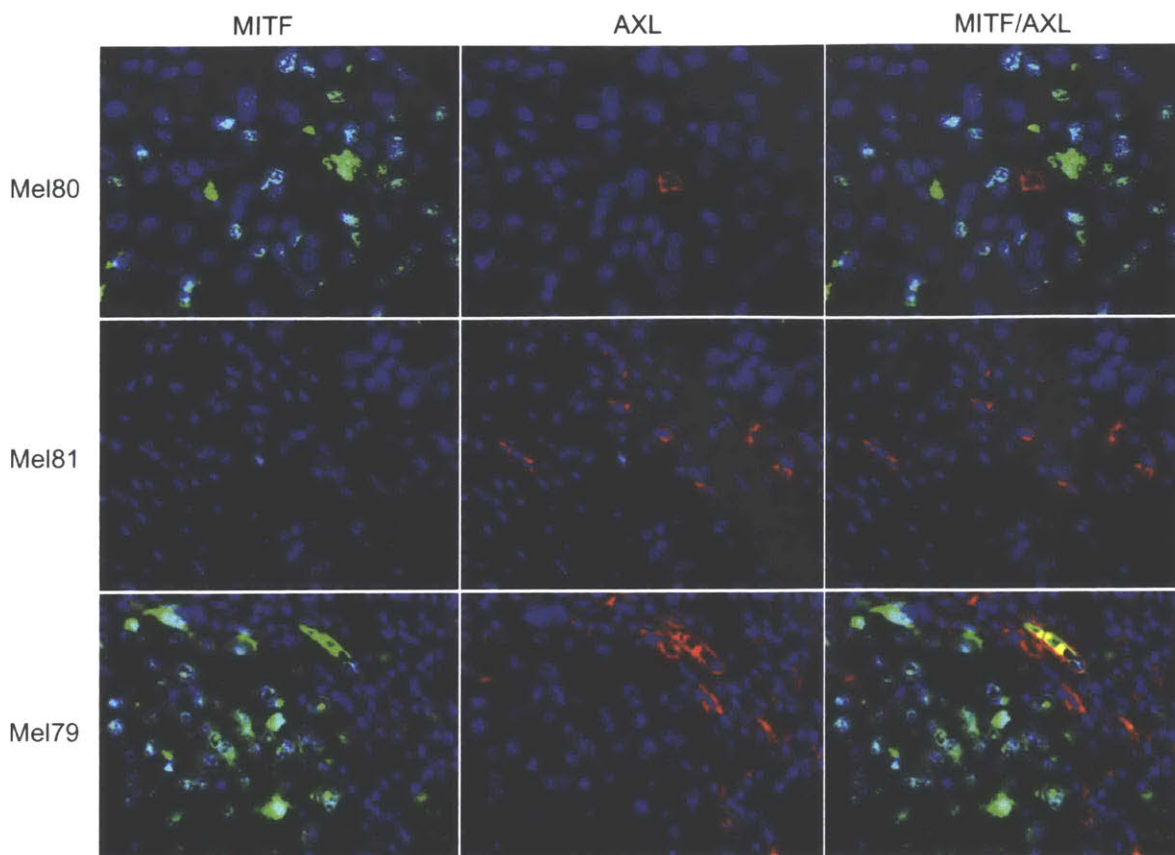

Figure 3-7: Immunofluorescence staining of MITF & AXL. AXL/MITF immunofluorescence staining of tissue slides of Mel80, Mel81 and Mel79 (40x magnification) revealed presence of AXL-expressing and MITF-expressing cells in each sample. Consistent with single-cell RNA-seq inferred frequencies of each population, Mel80 contained rare AXL-expressing cells (red, cell membrane staining) and mostly malignant MITF-positive cells (green, nuclear staining), while malignant cells of Mel81 almost exclusively consisted of AXL-expressing cells. Mel79 had a mixed population with rare cells positive for both markers, all in agreement with the inferred single-cell transcriptome data.

Since malignant cells with AXL-high and MITF-high transcriptional states co-exist in melanoma, we hypothesized that treatment with RAF/MEK inhibitors would increase the prevalence of AXL-high cells following the development of drug resistance. To test this, we analyzed RNA-seq data from a cohort[12] of six paired BRAF^{V600E} melanoma biopsies taken before treatment and after resistance to single-agent RAF inhibition (vemurafenib; n=1) or combined RAF/MEK inhibition

(dabrafenib and trametinib; n=5), respectively. We ranked the 12 transcriptomes on the basis of the relative expression of all genes in the AXL-high program compared to those in the MITF-high program. In each pair, we observed a shift towards the AXL-high program in the drug-resistant sample (**Figure 3-5 d**; P<0.05 for same effect in six out of six paired samples, binomial test; P<0.05 for four of six individual paired-sample comparisons shown by black arrows). RNA-seq data from an independent cohort[26] also showed that a subset of drug-resistant samples exhibited increased expression of the AXL program (**Figure 3-8**). Other genes previously implicated in resistance to RAF/MEK inhibition were also increased in a subset of the drug-resistant samples. PDGFRB[27] was upregulated in a similar subset as the AXL program, while MET[26] was upregulated in a mutually exclusive subset (**Figure 3-8**), suggesting that AXL and MET may reflect distinct mechanisms for drug resistance.
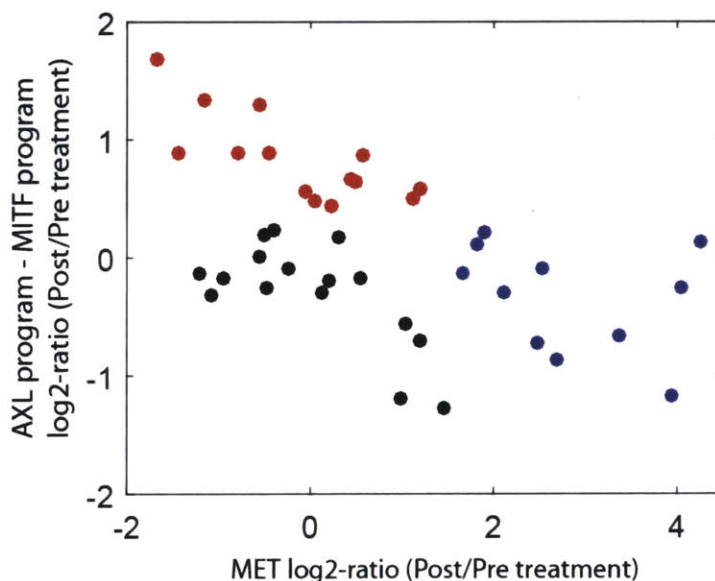

Figure 3-8. AXL upregulation in a second cohort of post-treatment melanoma samples and mutual exclusivity with MET upregulation. Each point reflects a comparison between a matched pair of pre-treatment and post-relapse samples from Hugo et al.[26], where the X-axis shows expression changes in MET, and the Y-axis shows expression changes in the AXL program minus those of the MITF program. Note that some patients are represented more than once based on multiple post-relapse samples. Fourteen out of 41 samples (34%) shown in red had significant upregulation of the AXL vs. MITF program, as determined by a modified t-test as described in Methods; these correspond to at least one sample from half (9/18) of the patients

included in the analysis. Eleven out of 41 samples (27%) shown in blue had at least 3-fold upregulation of MET; these correspond to at least one sample from a third (6/18) of the patients included in the analysis. Notably, the AXL and MET upregulated samples are mutually exclusive, consistent with the possibility that these are alternative resistance mechanism.

To further assess the connection between the AXL program and resistance to RAF/MEK inhibition, we studied single-cell AXL expression in 18 melanoma cell lines from the CCLE[28]. Flow-cytometry demonstrated a wide distribution of the proportion of AXL-positive cells, from <1% to 99% per cell line, correlated with bulk mRNA levels, that were inversely associated with sensitivity to small-molecule RAF inhibition.

We treated 10 cell lines with increasing doses of a RAF/MEK inhibitor combination (dabrafenib and trametinib) and found a rapid increase in the proportion of AXL-positive cells in six cell lines initially composed of a small (<3%) pre-treatment AXL-positive population (**Figure 3-5 e**; **Figure 3-9**). In contrast, cell lines with an intrinsically high proportion of AXL-expression, showed modest or no changes (**Figure 3-9**). Similar results were obtained by multiplexed quantitative single-cell immunofluorescence (IF), which also demonstrated that the increased fraction of AXL-positive cells following RAF/MEK inhibition are associated with rapid decreases in ERK phosphorylation (reflecting MAP-kinase signaling inhibition) (**Figure 3-5 f**). In summary, both melanoma tumors and cell lines demonstrate drug-resistant tumor cell subpopulations that become enriched during treatment with MAP-kinase targeted treatment.
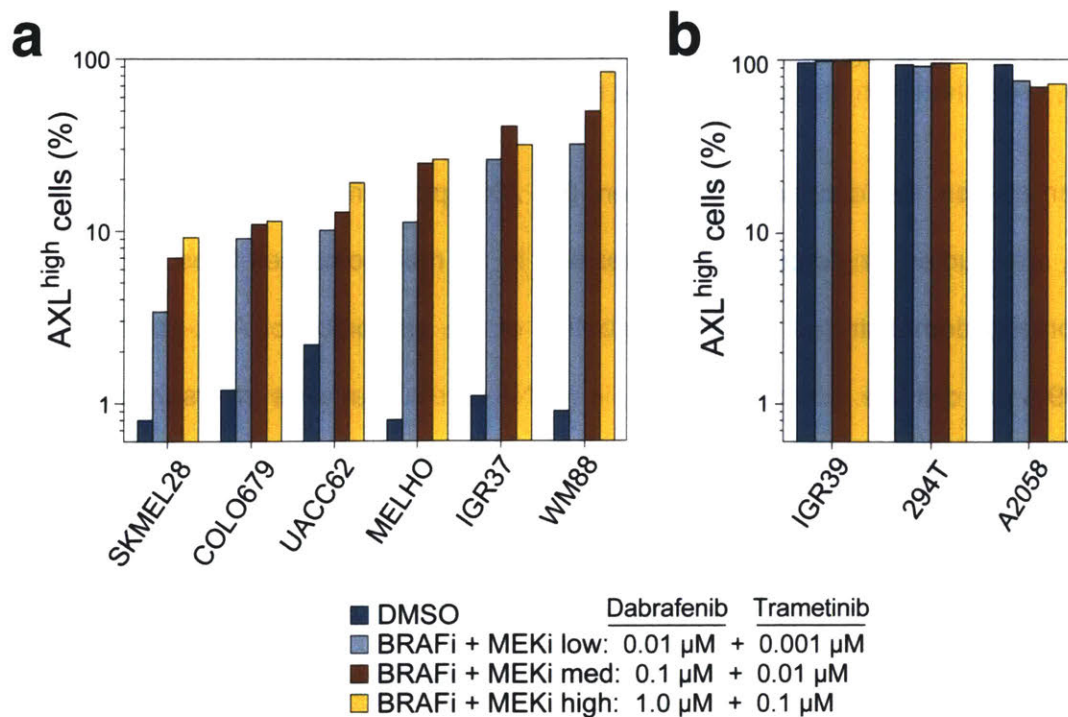
Figure 3-9: Flow-cytometry of melanoma cell lines before and after treatment with RAF/MEK-inhibition. (**a**) Sensitive cell lines show an increased proportion of AXL-positive cells while resistant cell lines (**b**) show modest or no changes following treatment with RAF/MEK-inhibitors.

3.6: Discussion

Our analysis has uncovered intra- and inter-individual, spatial, functional and genomic heterogeneity in melanoma cells and associated tumor components that shape the microenvironment. We identified a cell state in a subpopulation of all melanomas studied that is linked to resistance to targeted therapies and validated the presence of a dormant drug-resistant population in a number of melanoma cell lines using different approaches.

While future work is necessary to clarify the interplay between these cell types and functional states in space and time, the ability to carry out numerous, highly-multiplexed single cell observations within a tumor provides power to identify meaningful cell subpopulations and gene expression programs that can inform on both the analysis of bulk transcriptional data and

precision treatment strategies. Conceivably, single cell genomic profiling may soon enable a deeper understanding of the complex interplay among cells within the tumor ecosystem and its evolution in response to treatment, thereby providing a versatile new tool for future translational applications.

3.7: Methods

Tissue handling and tumor disaggregation. Resected tumors were transported in DMEM (ThermoFisher Scientific) on ice immediately after surgical procurement. Tumors were rinsed with PBS (Life Technologies). A small fragment was stored in RNA-protect (Qiagen) for bulk RNA and DNA isolation. Using scalpels, the remainder of the tumor was minced into tiny cubes <1 mm^3 and transferred into a 50 ml conical tube (BD Falcon) containing 10 ml pre-warmed M199-media (ThermoFisher Scientific), 2 mg/ml collagenase P (Roche) and 10U/µl DNase I (Roche). Tumor pieces were digested in this digestion media for 10 minutes at 37°C, then vortexed for 10 seconds and pipetted up and down for 1 minute using pipettes of descending sizes (25 ml, 10 ml and 5 ml). If needed, this was repeated twice more until a single-cell suspension was obtained. This suspension was then filtered using a 70µm nylon mesh (ThermoFisher Scientific) and residual cell clumps were discarded. The suspension was supplemented with 30 ml PBS (Life Technologies) with 2% fetal calf serum (FCS) (Gemini Bioproducts) and immediately placed on ice. After centrifuging at 580g at 4°C for 6 minutes, the supernatant was discarded and the cell pellet was re-suspended in PBS with FCS and placed on ice prior to staining for FACS.

FACS. Single-cell suspensions were stained with CD45-FITC (VWR) and Calcein-AM (Life Technologies) per manufacturer recommendations. For sorting of *ex vivo* co-cultured cancer-associated fibroblasts, we used a CD90-PE antibody (BioLegend). First, doublets were excluded based on forward and sideward scatter, then we gated on viable cells (Calcein^{high}) and sorted single cells (CD45+ or CD45- or CD45-CD90+) into 96-well plates chilled to 4°C, pre-prepared with 10µl TCL buffer (Qiagen) supplemented with 1% beta-mercaptoethanol (lysis buffer). Single-cell lysates were sealed, vortexed, spun down at 3700 rpm at 4°C for 2 minutes, immediately placed on dry ice and transferred for storage at -80°C.

60

RNA and DNA isolation from bulk specimens. RNA and DNA were isolated using the Qiagen minikit following the manufacturers recommendations.

Whole transcriptome amplification. Whole Transcriptome amplification (WTA) was performed with a modified SMART-Seq2 protocol, as described previously[29,30], with Maxima Reverse Transcriptase (Life Technologies) used in place of Superscript II.

Library preparation and RNA-seq. WTA products were cleaned with Agencourt XP DNA beads and 70% ethanol (Beckman Coulter) and Illumina sequencing libraries were prepared using Nextera XT (Illumina), as previously described[30]. The 96 samples of a multiwall plate were pooled together, and cleaned with two 0.8x DNA SPRIs (Beckman Coulter). Library quality was assessed with a high sensitivity DNA chip (Agilent) and quantified with a high sensitivity dsDNA Quant Kit (Life Technologies). Samples were sequenced on an Illumina NextSeq 500 instrument using 30bp paired-end reads.

Whole-exome sequencing and analysis. Exome sequences were captured using Illumina technology and Exome sequence data processing and analysis were performed using the Picard and Firehose pipelines at the Broad Institute. The Picard pipeline (http://picard.sourceforge.net) was used to produce a BAM file with aligned reads. This includes alignment to the hg19 human reference sequence using the Burrows-Wheeler transform algorithm[31] and estimation of base quality score and recalibration with the Genome Analysis Toolkit (GATK) (http://www.broadinstitute.org/gatk/)[32]. All sample pairs passed the Firehose pipeline including a QC pipeline to test for any tumor/normal and inter-individual contamination as previously described[33,34]. The MuTect algorithm was used to identify somatic mutations[34].

MuTect identifies candidate somatic mutations by Bayesian statistical analysis of bases and their qualities in the tumor and normal BAMs at a given genomic locus. To reduce false positive calls we additionally analyzed reads covering sites of an identified somatic mutation and realigned them with NovoAlign (www.novocraft.com) and performed additional iteration of MuTect inference on newly aligned BAM files. Furthermore, we filtered somatic mutation calls using a panel of over 8,000 TCGA Normal samples. Small somatic insertions and deletions were detected using the Strelka algorithm[35] and similarly subjected to filtering out potential false positive using the panel of TCGA Normal samples. Somatic mutations including single-nucleotide variants, insertions, and deletions were annotated using Oncotator[36]. Copy-ratios for each captured exon were calculated by comparing the mean exon coverage with expected coverage based on a panel of normal samples. The resulting copy-ratio profiles were then segmented using the circular binary segmentation (CBS) algorithm[37].

Processing of RNA-seq data. Following sequencing on the NextSeq, BAM files were converted to merged, demultiplexed FASTQs. Paired-end reads were then mapped to the UCSC hg19 human transcriptome using Bowtie[38] with parameters "-q --phred33-quals -n 1 -e 99999999 -l 25 -I 1 -X 2000 -a -m 15 -S -p 6", which allows alignment of sequences with single base changes such as due to point mutations. Expression levels of genes were quantified as $E_{i,j}=log_2(TPM_{i,j}/10+1)$, where $TPM_{i,j}$ refers to transcript-per-million (TPM) for gene i in sample j, as calculated by RSEM[39] v1.2.3 in paired-end mode. TPM values were divided by 10 since we estimate the complexity of our single cell libraries to be on the order of 100,000 transcripts and would like to avoid counting each transcript ~10 times, as would be the case with TPM, which may inflate the difference between the expression level of a gene in cells in which the gene is detected and those in which it is not detected. When evaluating the average expression of a population of cells by pooling data across cells (e.g., all cells from a given tumor or cell type) the

62

division by 10 was not required and the average expression was defined $E_p(I)=log_2(TPM(I)+1)$, where I is a set of cells.

For each cell, we quantified the number of genes for which at least one read was mapped, and the average expression level of a curated list of housekeeping genes. We then excluded all cells with either fewer than 1,700 detected genes or an average housekeeping expression (E, as defined above) below 3. For the remaining cells, we calculated the pooled expression of each gene as (Ep), and excluded genes with an aggregate expression below 4, which defined a different set of genes in different analyses depending on the subset of cells included. For the remaining cells and genes, we defined relative expression by centering the expression levels, $Er_{i,j}=E_{i,j}-average[E_{i,1...n}]$.

Data availability

Raw and processed single-cell RNA-seq data is available through the Gene Expression Omnibus (GSE72056).

CNV estimation. Initial CNVs (CNV_0) were estimated by sorting the analyzed genes by their chromosomal location and applying a moving average to the relative expression values, with a sliding window of 100 genes within each chromosome, as previously described[16]. To avoid considerable impact of any particular gene on the moving average we limited the relative expression values to [-3,3] by replacing all values above 3 by 3, and replacing values below -3 by -3. This was performed only in the context of CNV estimation. This initial analysis is based on the average expression of genes in each cell compared to all other cells and therefore does not have a proper reference which is required to define the baseline. However, we identified five subsets of cells that each had more limited high or low values of CNV_0 and which were

consistent across the genome despite the fact that these cells originate from multiple tumors. We thus considered these as putative non-malignant cells and used their CNV estimates to define the baseline. The normal cells included five cell types (see below, not including NK cells), which differed in gene expression patterns and accordingly also slightly in CNV estimates (*e.g.*, the MHC region in chromosome 6 had consistently higher values in T cells than in stromal or cancer cells). We therefore defined multiple baselines, as the average of each cell type, and based on these the maximal (*BaseMax*) and minimal (*BaseMin*) baseline at each window of 100 genes. The final CNV estimate of cell *i* at position *j* was then defined as:

$$CNV_f(i,j) = \begin{cases} CNV_0(i,j) - BaseMax(j), & if\ CNV_0(i,j) > BaseMax(j) + 0.2 \\ CNV_0(i,j) - BaseMin(j), & i\ f\ CNV_0(i,j) < BaseMin(j) - 0.2 \\ 0, & if\ BaseMin(j) - 0.2 < CNV_0(i,j) < BaseMin(j) + 0.2 \end{cases}$$

To quantitatively evaluate how likely each cell is to be a malignant or non-malignant cell we summarized the CNV pattern of each cell by two values: (1) overall CNV signal, defined as the sum of squares of the CNV_f estimates across all windows; (2) the correlation of each cells' CNV_f vector with the average CNV_f vector of the top 10% of cells from the same tumor with respect to CNV signal (*i.e.*, the most confidently-assigned malignant cells). These two values were used to classify cells as malignant, non-malignant, and intermediates that were excluded from further analysis, as shown in **Figure 3-2**.

T-SNE analysis and cell type classification. A Matlab implementation of the tSNE method was downloaded from http://lvdmaaten.github.io/tsne/ and applied with dim=15 to the relative expression data of malignant and to that of non-malignant cells. Since the complexity of tSNE visualization increases with the number of tumors we restricted the analysis presented in **Figure 3-1** to the 13 tumors with at least 100 cells, and for the malignant cell analysis we further restricted the analysis to 6 tumors with >50 malignant cells. To define cell types from the non-

64

malignant tSNE analysis we used a density clustering method, DBscan[17]. This process revealed six clusters for which the top preferentially expressed genes (p<0.001, permutation test) included multiple known markers of particular cell types. In this way, we identified T cell, B-cell, macrophage, endothelial, CAF (cancer-associated fibroblast) and NK cell clusters, as marked in **Figure 3-1** (dashed ellipses). To ensure the specificity of our assignment of individual cells to each cell type cluster, while avoiding potential doublet cells (which might be composed of two cells from distinct cell types), cells with low-quality data, and cells that spuriously cluster with a certain cell type, we next scored each non-malignant cell (by CNV estimates, as described above) by the average expression of the identified cell type marker genes. Cells were classified as each cell type only if they express the marker genes for that cell type much more than those for any other cell type (average relative expression, E_r, of markers for one cell type higher by at least 3 than those of other cell types, which corresponds to 8-fold expression difference).

Principal component analysis. In order to decrease the impact of inter-tumoral variability on the combined analysis of cancer cells we re-centered the data within each tumor separately, such that the average of each gene was zero among cells from each tumor. The covariance matrix used for PCA was generated using an approach outlined in Shalek et al[40] to decrease the weight of less reliable "missing" values in the data. This approach aims to address the challenge that arises due to the limited sensitivity of single-cell RNA-seq, where many genes are not detected in a particular cell despite being expressed. This is particularly pronounced for genes that are more lowly expressed, and for cells that have lower library complexity (*i.e.*, for which relatively fewer genes are detected), and results in non-random patterns in the data, whereby cells may cluster based on their complexity and genes may cluster based on their expression levels, rather than "true" co-variation. To mitigate this effect we assign weights to missing values, such that the weight of $E_{i,j}$ is proportional to the expectation that gene *i* will be detected

in cell *j* given the average expression of gene *i* and the total complexity (number of detected genes) of cell *j*.

Following PCA, we focused on the top six components as these were the only components that both explained a significant proportion of the variance and were significantly correlated with at least one gene, where significance was determined by comparison to the top 5% (of variance explained and of top gene correlations) from 100 control PCA analyses on shuffled data. PC1 had a high correlation (R=0.46) with the number of genes detected in each cell and we did not observe a more specific biological function that may be associated with it and thus we infer this to be a technically-driven component which is reflecting the systematic variation in the data due to the large differences in the quality and complexity of data for different cells. Subsequent analysis was focused on understanding the biological function of the next components PC2-6, which were associated with the cell cycle (PC2 and 6), regional heterogeneity (PC3) and MITF expression program (PC4 and 5).

MITF and AXL expression programs and cell scores. The top 100 MITF-correlated genes across the entire set of malignant cells were defined as the MITF program, and their average relative expression as the MITF-program cell score. The average expression of the top 100 genes that negatively correlate with the MITF program scores were defined as the AXL program and used to define AXL program cell score. To decrease the effect that the quality and complexity of each cell's data might have on its MITF/AXL scores we defined control gene-sets and their average relative expression as control scores, for both the MITF and AXL programs. These control cell scores were subtracted from the respective MITF/AXL cell scores. The control gene-sets were defined by first binning all analyzed genes into 25 bins of aggregate expression levels and then, for each gene in the MITF/AXL gene-set, randomly selecting 100 genes from the same expression bin as that gene. In this way, a control gene-sets have a comparable distribution of expression levels to that of the MITF/AXL gene-set and the control

66

gene set is 100-fold larger, such that its average expression is analogous to averaging over 100 randomly-selected gene-sets of the same size as the MITF/AXL gene-set. To calculate significance of the changes in AXL and MITF programs upon relapse, we defined the expression $\log_2$-ratio between matched pre- and post- samples for all AXL and MITF program genes (**Figure 3-5**). Since AXL and MITF programs are inversely related, we flipped the signs of the log-ratios for MITF program genes and used a t-test to examine if the average of the combined set of AXL program and (sign-flipped) MITF program genes is significantly higher than zero, which was the case four out of six matched sample pairs (**Figure 3-5**, black arrows).

Immunohistochemical staining. All melanoma specimens were formalin fixed, paraffin-embedded, sectioned, and stained with hematoxylin and eosin (H&E) for histopathological evaluation at the Brigham and Women's Pathology core facility, unless otherwise specified. Immunohistochemical (IHC) studies employed 5 mm sections of formalin-fixed, paraffin-embedded tissue. All were stained on the Leica Bond III automated platform using the Leica Refine detection kit. Sections were deparaffinized and HIER was performed on the unit using EDTA for 20 minutes at 90°C. All sections were stained per routine protocols of the Brigham and Women's Pathology core facility. Additional sections were incubated for 30 min with primary antibody Ki-67 (1:250, Vector, VP-RM04) and JunB rabbit mAb (C37F9, Cell Signaling Technologies) and were then completed with the Leica Refine detection kit. The Refine detection kit encompasses the secondary antibody, the DAB chromagen (DAKO) and the Hematoxilyn counterstain. Cell counting using an ocular grid micrometer over at least five high-power fields was performed.

Tissue immunofluorescence staining. Dual-labeling immunofluorescence was performed to complement immunohistochemistry as a means of two-channel identification of epitopes co-expressed in similar or overlapping sub-cellular locations. Briefly, 5-mm-thick paraffin sections

were incubated with primary antibodies, AXL rabbit mAb antibody (C89E7, Abcam) plus MITF mouse mAb (clone D5, ab3201, Abcam) and JARID1B rabbit mAb (ab56759, Abcam) plus Ki67 (ab8191, Abcam) that recognize the target epitopes at 4°C overnight and then incubated with Alexa Fluor 594-conjugated anti-mouse IgG and Alexa Fluor 488-conjugated anti-rabbit IgG (Invitrogen) at room temperature for 1 h. The sections were cover slipped with ProLong Gold anti-fade with DAPI (Invitrogen). Sections were analyzed with a BX51/BX52 microscope (Olympus America, Melville, NY, USA), and images were captured using the CytoVision 3.6 software (Applied Imaging, San Jose, CA, USA). The following primary antibodies were used for staining per manufactures recommendations: mouse anti-MITF (DAKO), rabbit ant-AXL (Cell Signaling), goat anti-TIM3 (R&D Systems), rabbit ant-PD1 (Sigma Aldrich), and goat anti-PD1 (R&D Systems).

Cell culture experiments and AXL flow-cytometry. Cell lines from the Cancer Cell Encyclopedia Lines[28] were used for flow-cytometry analysis of the proportion of AXL-positive cells. Based on IC50 values for vemurafenib, we selected seven cell lines that were predicted to be sensitive to MAP-kinase pathway inhibition, including WM88, IGR37, MELHO, UACC62, COLO679, SKMEL28 and A375 and three cell lines predicted to be resistant, including IGR39, 294T and A2058. These ten cell lines were used for drug sensitivity testing and pre-treatment and post-treatment analysis of the AXL-positive fraction. For WM88, IGR37, MELHO, UACC62, COLO679, SKMEL28 and A375, cells were plated at a density to be at 30-50% confluent after 16 hours post seeding. A total of four drug arms were plated for each cell line using two T75 (Corning) and two T175 (Corning) culture flasks. Approximately 16-24 hours after seeding, cells were treated with DMSO or dabrafenib (D) and trametinib (T) at the following drug doses of D/T: 0.01uM/0.001uM, 0.1uM/0.01uM and 1uM/0.1uM (T175 reserved for higher drug concentrations). Cells were maintained in drug for a total of 5 days, at which point, cells were harvested for flow sorting. For IGR39, 294T and A2058, cells were plated at a density to be at

20-30% confluent 16 hours post seeding. Cells were treated with the DMSO or D/T at using the same doses as above and maintained in drug for a total of 10 days, at which point, cells were harvested for flow sorting. For AXL-flow sorting, cells were first washed with warm PBS, followed by an addition of 10mM EDTA and incubated for 2 minutes at room temperature. Excess EDTA was then aspirated and cells incubated at 37°C until cells detached from flask. Cells were resuspended in cold PBS 2% FBS and kept on ice. Cells were counted and 500,000 cells were transferred to 15ml conical tubes (Falcon), spun down and resuspended in 100µl of cold PBS 2% FBS alone (negative control) or antibodies using manufacturers recommendations, including 1µg of AXL antibody (AF154, R&D Systems) or 1µg of normal goat IgG control (Isotype control, AB-108-C, R&D Systems). Cells were incubated on ice for 1 hour, then washed twice with cold PBS 2% FBS. Cells were pelleted and resuspended in 100µl PBS 2% FBS with 5µl of Goat IgG (H+L) APC-conjugated Antibody (F0108, R&D Systems) and incubated for 30 minutes at room temperature. Cells were then washed twice with cold PBS 2% FBS, pelleted and resuspended in 500µl of PBS 2% FBS and transferred to 5mL flow-cytometry tubes (Falcon). 1µl of SYTOX Blue Dead Stain (Thermo Fisher) was added to each sample and samples analyzed by flow-cytometry. Data was analyzed using FACSDiva Version 6.2 using viable cells only (as determined by SYTOX Blue staining) and gates for AXL-positivity were set using the Isotype control set to <1%.

Single-cell immunofluorescence staining and analysis. For single-cell immunofluorescence (single-cell IF) studies, we included the following cell lines from CCLE: WM88, MELHO, SKMEL28, COLO679, IGR39, A2058 and 294T. Cells were cultured and detached as described above, and seeded at a density of 10,000 cells per well into Costar 96-well black clear-bottom tissue culture plates (3603, Corning). Cells were treated using Hewlett-Packard (HP) D300 Digital Dispenser with vemurafenib (Selleck) alone or in combination with trametinib (Selleck) at indicated doses for 5 and 10 days. In the case of 10-day treatment, growth medium

was changed after 5 days followed by immediate drug re-treatment. Cells were then fixed in 4% paraformaldehyde for 20 minutes at room temperature and washed with PBS with 0.1% Tween 20 (Sigma-Aldrich) (PBS-T), permeabilized in methanol for 10 min at room temperature, rewashed with PBS-T, and blocked in Odyssey Blocking Buffer for 1 hour at room temperature. Cells were incubated overnight at 4°C with primary antibodies in Odyssey Blocking Buffer. The following primary antibodies with specified animal sources and catalogue numbers were used in specified dilution ratios: p-ERK$^{T202/Y204}$ rabbit mAb (clone D13.14.4E, 4370, Cell Signaling Technology), 1:800, AXL goat polyclonal antibody (AF154, R&D Systems), 1:800, MITF mouse mAb (clone D5, ab3201, Abcam), 1:400. Cells were then stained with rabbit, mouse and goat secondary antibodies from Molecular Probes (Invitrogen) labeled with Alexa Fluor 647 (A31573), Alexa Fluor 488 (A21202), and Alexa Fluor 568 (A11057). Cells were washed once in PBS-T, once in PBS and were then incubated in 250 ng/ml Hoechst 33342 and 1:800 Whole Cell Stain (blue; Thermo Scientific) solution for 20 min. Cells were washed twice with PBS and imaged with a 10× objective on a PerkinElmer Operetta High Content Imaging System. 9-11 sites were imaged in each well. Image segmentation, analysis and signal intensity quantitation were performed using Acapella software (Perkin Elmer). Population-average and single-cell data were analyzed using MATLAB 2014b software. Single-cell density scatter plots were generated using signal intensities for individual cells.

CAF-melanoma co-cultures from melanoma 80. Solid tumor sample was removed from the transport media (Day 1: date of procurement) and minced mechanically in DMEM culture media (Thermo Scientific), 10% FCS (Gemini Bioproducts), 1% pen/strep (Life Technologies) on 10 cm culture plates (Corning Inc.) and left overnight in standard culture condition (37C, humidified atmosphere, 5% CO2). The liquid media in which the procured tissue was originally placed was spun down (1500 rpm) to isolate the detached cells in solution and the pelleted cells were resuspended in fresh culture media and propagated in culture flasks (Corning Inc.) (fraction 1).

70

The minced tumor samples were removed from the 10 cm culture dishes on Day 2 and mechanically forced through 100 uM nylon mesh filters (Fisher Scientific) using syringe plungers and washed through with fresh culture media. The cells and tissue clumps were spun down in 50 ml conical tubes (BD Falcon), resuspended in fresh culture media, and propagated in culture flasks (fraction 2). The 10 cm culture dishes in which the samples had been minced and placed overnight were washed replaced with fresh culture media so that the attached cells could be propagated (fraction 3). Cells were propagated by changing culture media every 3-4 days and passaging cells in 1:3 to 1:6 ratio using 0.05% trypsin (Thermo Scientific) when the plates became 50-80% confluent.

Tissue microarray staining, image acquisition and analysis. We purchased two individual melanoma tissue microarrays (TMAs), including ME208 (US Biomax) and CC38-01-003 (Cybrdi). These contained a total of 308 core biopsies, including a total of 180 primary melanomas, 90 metastatic lesions, 18 melanomas with adjacent healthy skin and 20 healthy skin controls. Each TMA was double-stained with conjugated complement 3-FITC antibody (F0201, DAKO) and CD8-TRITC (ab17147, Abcam) per manufacturers recommendations. Image acquisition was performed on the RareCyte CyteFinder high-throughput imaging platform[41]. For each TMA-slide, the 3-channel (DAPI/FITC/TRITC) 10× images were captured and stored as Bio-format stacks. The image stacks were background-subtracted with rolling ball method and stitched into single image montage of each channel using ImageJ. For the quantification of CD8/C3 positive area and signal intensity, the gray-scale images were converted into binary images with the Otsu thresholding method[42,43]. Each tissue spot was segmented manually and DAPI, C3 and CD8-positive areas and intensities were calculated using ImageJ (NIH, MD). In order to control for sample quality, core biopsies with a DAPI staining less than 10% of total area were excluded from the correlation analysis. The raw

numerical data were then processed and Pearson's correlation coefficients were calculated between C3/CD8 area fraction and intensity using MATLAB 2014b software (MathWorks, MA).

Table S1. Characteristics of patients and samples included in this study

Sample ID	Age/sex	Mutation status	Pre-operative treatment	Site of resection	Post-op. treatment	Alive/ deceased
Melanoma_53	77/F	Wild-type	None	Subcutaneous back lesion	None	Alive
Melanoma_58	67/F	Wild-type	Ipilimumab	Subcutaneous leg lesion	None	Alive
Melanoma_59	80/M	Wild-type	None	Femoral lymph node	Nivolumab.	Deceased
Melanoma_60	69/M	BRAF V600K	Trametinib, ipilimumab	Spleen	None	Alive
Melanoma_65	65/M	BRAF V600E	None	Paraspinal intramuscular	Neovax	Alive
Melanoma_67	58/M	BRAF V600E	None	Axillary lymph node	None	Alive
Melanoma_71	79/M	NRAS Q61L	None	Transverse colon	None	Alive
Melanoma_72	57/F	NRAS Q61R	IL-2, nivolumab, ipilimumab + anti-KIR-Ab	External iliac lymph node	None	Alive
Melanoma_74	63/M	n/a	Nivolumab	Terminal Ileum	None	Alive
Melanoma_75	80/M	Wild-type	Ipilimumab + nivolumab, WDVAX	Subcutaneous leg lesion	Nivolumab	Alive
Melanoma_78	73/M	NRAS Q61L	WDVAX, ipilimumab + nivolumab	Small bowel	None	Deceased
Melanoma_79	74/M	Wild-type	None	Axillary lymph node	None	Alive
Melanoma_80	86/F	NRAS Q61L	None	Axillary lymph node	None	Alive
Melanoma_81	43/F	BRAF V600E	None	Axillary lymph node	None	Alive
Melanoma_82	81/M	Wild-type	None	Axillary lymph node	None	Alive
Melanoma_84	67/M	Wild-type	None	Acral primary	None	Alive
Melanoma_88	54/F	NRAS Q61L	Tremelimumab + MEDI3617	Cutanoues met	None	Alive
Melanoma_89	67/M	n/a	None	Axillary lymph node	None	Alive
Melanoma_94	54/F	Wild-type	IFN, ipilimumab + nivolumab	Iliac lymph node	None	Alive

3.8: References

1   Hanahan, D. & Weinberg, R. A. Hallmarks of cancer: the next generation. *Cell* **144**, 646-674, doi:10.1016/j.cell.2011.02.013 (2011).

2   Meacham, C. E. & Morrison, S. J. Tumour heterogeneity and cancer cell plasticity. *Nature* **501**, 328-337, doi:10.1038/nature12624 (2013).

3   Hodi, F. S. *et al.* Improved survival with ipilimumab in patients with metastatic melanoma. *N Engl J Med* **363**, 711-723, doi:10.1056/NEJMoa1003466 (2010).

4   Brahmer, J. R. *et al.* Phase I study of single-agent anti-programmed death-1 (MDX-1106) in refractory solid tumors: safety, clinical activity, pharmacodynamics, and immunologic correlates. *J Clin Oncol* **28**, 3167-3175, doi:10.1200/JCO.2009.26.7609 (2010).

5   Brahmer, J. R. *et al.* Safety and activity of anti-PD-L1 antibody in patients with advanced cancer. *N Engl J Med* **366**, 2455-2465, doi:10.1056/NEJMoa1200694 (2012).

6   Topalian, S. L. *et al.* Safety, activity, and immune correlates of anti-PD-1 antibody in cancer. *N Engl J Med* **366**, 2443-2454, doi:10.1056/NEJMoa1200690 (2012).

7   Hamid, O. *et al.* Safety and tumor responses with lambrolizumab (anti-PD-1) in melanoma. *N Engl J Med* **369**, 134-144, doi:10.1056/NEJMoa1305133 (2013).

8   Weber, J. S. *et al.* Safety, efficacy, and biomarkers of nivolumab with vaccine in ipilimumab-refractory or -naive melanoma. *J Clin Oncol* **31**, 4311-4318, doi:10.1200/JCO.2013.51.4802 (2013).

9   Mahoney, K. M. & Atkins, M. B. Prognostic and predictive markers for the new immunotherapies. *Oncology (Williston Park)* **28 Suppl 3**, 39-48 (2014).

10   Larkin, J., Hodi, F. S. & Wolchok, J. D. Combined Nivolumab and Ipilimumab or Monotherapy in Untreated Melanoma. *N Engl J Med* **373**, 1270-1271, doi:10.1056/NEJMc1509660 (2015).

11   Snyder, A. *et al.* Genetic basis for clinical response to CTLA-4 blockade in melanoma. *N Engl J Med* **371**, 2189-2199, doi:10.1056/NEJMoa1406498 (2014).

12   Van Allen, E. M. *et al.* The genetic landscape of clinical resistance to RAF inhibition in metastatic melanoma. *Cancer Discov* **4**, 94-109, doi:10.1158/2159-8290.CD-13-0617 (2014).

13   Wagle, N. *et al.* Dissecting therapeutic resistance to RAF inhibition in melanoma by tumor genomic profiling. *J Clin Oncol* **29**, 3085-3096, doi:10.1200/JCO.2010.33.2312 (2011).

14   Shalek, A. K. *et al.* Single-cell transcriptomics reveals bimodality in expression and splicing in immune cells. *Nature* **498**, 236-240, doi:10.1038/nature12172 (2013).

15   Macosko, E. Z. *et al.* Highly Parallel Genome-wide Expression Profiling of Individual Cells Using Nanoliter Droplets. *Cell* **161**, 1202-1214, doi:10.1016/j.cell.2015.05.002 (2015).

16   Patel, A. P. *et al.* Single-cell RNA-seq highlights intratumoral heterogeneity in primary glioblastoma. *Science* **344**, 1396-1401, doi:10.1126/science.1254257 (2014).

17   Ester, M., Kriegel, H.-P., Sander, J. & Xu, X. in *Proceedings of the Second International Conference on Knowledge Discovery and Data Mining*   226-231 (AAAI Press, Portland, Oregon, 1996).

18   Garraway, L. A. *et al.* Integrative genomic analyses identify MITF as a lineage survival oncogene amplified in malignant melanoma. *Nature* **436**, 117-122, doi:10.1038/nature03664 (2005).

19   Wu, X. *et al.* AXL kinase as a novel target for cancer therapy. *Oncotarget* **5**, 9546-9563, doi:10.18632/oncotarget.2542 (2014).

20   Zhang, Z. *et al.* Activation of the AXL kinase causes resistance to EGFR-targeted therapy in lung cancer. *Nat Genet* **44**, 852-860, doi:10.1038/ng.2330 (2012).

21   Boiko, A. D. *et al.* Human melanoma-initiating cells express neural crest nerve growth factor receptor CD271. *Nature* **466**, 133-137, doi:10.1038/nature09161 (2010).

22   Konieczkowski, D. J. *et al.* A melanoma cell state distinction influences sensitivity to MAPK pathway inhibitors. *Cancer Discov* **4**, 816-827, doi:10.1158/2159-8290.CD-13-0424 (2014).

23   Hoek, K. S. *et al.* In vivo switching of human melanoma cells between proliferative and invasive states. *Cancer Res* **68**, 650-656, doi:10.1158/0008-5472.CAN-07-2491 (2008).

24   Muller, J. *et al.* Low MITF/AXL ratio predicts early resistance to multiple targeted drugs in melanoma. *Nat Commun* **5**, 5712, doi:10.1038/ncomms6712 (2014).

25   Li, F. Z., Dhillon, A. S., Anderson, R. L., McArthur, G. & Ferrao, P. T. Phenotype switching in melanoma: implications for progression and therapy. *Front Oncol* **5**, 31, doi:10.3389/fonc.2015.00031 (2015).

26   Hugo, W. *et al.* Non-genomic and Immune Evolution of Melanoma Acquiring MAPKi Resistance. *Cell* **162**, 1271-1285, doi:10.1016/j.cell.2015.07.061 (2015).

27   Nazarian, R. *et al.* Melanomas acquire resistance to B-RAF(V600E) inhibition by RTK or N-RAS upregulation. *Nature* **468**, 973-977, doi:10.1038/nature09626 (2010).

28   Barretina, J. *et al.* The Cancer Cell Line Encyclopedia enables predictive modelling of anticancer drug sensitivity. *Nature* **483**, 603-607, doi:10.1038/nature11003 (2012).

29   Picelli, S. *et al.* Smart-seq2 for sensitive full-length transcriptome profiling in single cells. *Nat Methods* **10**, 1096-1098, doi:10.1038/nmeth.2639 (2013).

30   Trombetta, J. J. *et al.* Preparation of Single-Cell RNA-Seq Libraries for Next Generation Sequencing. *Curr Protoc Mol Biol* **107**, 4 22 21-17, doi:10.1002/0471142727.mb0422s107 (2014).

31   Li, H. & Durbin, R. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* **25**, 1754-1760, doi:10.1093/bioinformatics/btp324 (2009).

32   McKenna, A. *et al.* The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Res* **20**, 1297-1303, doi:10.1101/gr.107524.110 (2010).

33   Berger, M. F. *et al.* The genomic complexity of primary human prostate cancer. *Nature* **470**, 214-220, doi:10.1038/nature09744 (2011).

34   Cibulskis, K. *et al.* Sensitive detection of somatic point mutations in impure and heterogeneous cancer samples. *Nat Biotechnol* **31**, 213-219, doi:10.1038/nbt.2514 (2013).

35   Saunders, C. T. *et al.* Strelka: accurate somatic small-variant calling from sequenced tumor-normal sample pairs. *Bioinformatics* **28**, 1811-1817, doi:10.1093/bioinformatics/bts271 (2012).

36   Ramos, A. H. *et al.* Oncotator: cancer variant annotation tool. *Hum Mutat* **36**, E2423-2429, doi:10.1002/humu.22771 (2015).

37   Venkatraman, E. S. & Olshen, A. B. A faster circular binary segmentation algorithm for the analysis of array CGH data. *Bioinformatics* **23**, 657-663, doi:10.1093/bioinformatics/btl646 (2007).

38   Langmead, B., Trapnell, C., Pop, M. & Salzberg, S. L. Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biol* **10**, R25, doi:10.1186/gb-2009-10-3-r25 (2009).

39   Li, B. & Dewey, C. N. RSEM: accurate transcript quantification from RNA-Seq data with or without a reference genome. *BMC Bioinformatics* **12**, 323, doi:10.1186/1471-2105-12-323 (2011).

40   Shalek, A. K. *et al.* Single-cell RNA-seq reveals dynamic paracrine control of cellular variation. *Nature* **510**, 363-369, doi:10.1038/nature13437 (2014).

41   Campton, D. E. *et al.* High-recovery visual identification and single-cell retrieval of circulating tumor cells for genomic analysis using a dual-technology platform integrated with automated immunofluorescence staining. *BMC Cancer* **15**, 360, doi:10.1186/s12885-015-1383-x (2015).

42   Skaland, I. *et al.* Comparing subjective and digital image analysis HER2/neu expression scores with conventional and modified FISH scores in breast cancer. *J Clin Pathol* **61**, 68-71, doi:10.1136/jcp.2007.046763 (2008).

43   Konsti, J. *et al.* Development and evaluation of a virtual microscopy application for automated assessment of Ki-67 expression in breast cancer. *BMC Clin Pathol* **11**, 3, doi:10.1186/1472-6890-11-3 (2011).

Chapter 4: Transcriptomic and genomic correlates of checkpoint blockade in leptomeningeal disease

This chapter is adapted from the following manuscript in preparation:

Prakadan SM*, Alvarez-Breckenridge CA*, Markson SC*, Klein RH, Nayyar N, Navia AW, Kuter BM, Kolb KE, Bihun I, Mora JL, Solana Bertalan M, Shaw B, White M, Kaplan A, Stocking JH, Wadsworth MH II, Lee EQ, Subramanian M, Rotem D, Cahill DP, Adalsteinsson VA, Miller JW, Sullivan RJ, Carter SL[†], Brastianos PK[†], Shalek AK[†] "Transcriptomic and genomic correlates of checkpoint blockade in leptomeningeal disease."

[*,†] Denote equal contribution

Immune checkpoint molecules have been shown to modulate the behaviors of multiple cell types. Thus, immune checkpoint inhibitors (ICIs) might exert unappreciated, complex effects on the tumor microenvironment (TME). To directly explore treatment-induced shifts within a particular TME, we performed single-cell RNA-Seq (scRNA-Seq) and cell-free DNA-Seq (cfDNA-Seq) on cerebrospinal fluid from patients with leptomeningeal disease (LMD) before and after PD-1 blockade. We recover immune and malignant cell types, and characterize changes in T cell cytotoxicity and exhaustion, innate immune polarization, and transient tumor cell responses, including adaptive selection of a tumor subclone. Overall, our study illuminates multicellular changes in the liquid LMD TME following immunotherapy, and demonstrates the potential for longitudinal cell-free and single-cell genomic measurements to better understand potential putative drivers of clinical phenomena.

SMP co-led the design and implementation of the single-cell RNA-Sequencing study with CAA, performed the single-cell RNA-Sequencing with others, and co-led the analysis of the transcriptomic data with SCM.

Forward

We sought to extend our previous investigations of tumor heterogeneity introduced in Chapter 3 to other cancer systems, profiling their responses to treatment using single-cell genomics. However, limitations – including cost of sample preparation, the rarity of acquiring surgical resections, and difficulty profiling longitudinally – challenged widespread adoption. Clinical liquid biopsies presented an opportunity to circumvent these difficulties, but are often too low in cell number and/or volume for traditional single-cell approaches. Here, we apply a novel, nanowell-based single-cell RNA Sequencing technology to profile cerebrospinal fluid (CSF) liquid biopsies from patients with leptomeningeal carcinomatosis (LMD) treated with immunotherapy in a clinical trial studying its efficacy. We develop the first, to date, single-cell atlas of LMD, and use it to study the response of CSF cells to intravenously administered PD-1 blockade by Pembrolizumab. Additionally, we use paired longitudinal single-cell RNA & cell-free DNA sequencing to study clinical correlates in specific patients, investigating potential resistance mechanisms. This pilot study and subsequent data provide a window into the effect of drug on the TME and suggest how the response to treatment may be coordinated across compartments, with nuanced dynamics in individual patients.

4.1: Introduction

Among the immunotherapies currently used to treat cancer[1-4], immune checkpoint inhibitors (ICIs) have shown particularly striking clinical activity across a range of tumors[1,5,6]. These molecules can restore immune function by binding inhibitory co-receptors (or ligands) on T cells (or tumor cells), such as PD-1 and CTLA-4 (or their cognates), that are often engaged (expressed) by tumor cells and other cells[3,5]. Despite early successes, the broad applicability of PD-1 blockade has been challenged by mounting observations of partial response, acquired resistance, and inconsistent benefit across patients, even for the same tumor type[7-10], particularly in extension of ICI beyond initially promising tumor types. Recent evidence suggests that these obstacles may be partially explained by aspects of the tumor microenvironment (TME), such as immune exclusion or the presence of suppressive extracellular factors, among other processes[8,10-12]. Overcoming these mechanisms of resistance (both inherent and acquired) requires a detailed understanding of ICI-induced changes within the TME. Unfortunately, given the invasiveness of conventional sampling methods (e.g., resections or core biopsies) and the paucity of material recovered via their alternatives (e.g., fine needle aspirates, FNAs), it has been challenging to comprehensively characterize the TME before and after treatment[13].

Here, as part of a unique clinical trial of PD-1 blockade by pembrolizumab (NCT02886585, see **Methods**), we longitudinally collected cerebrospinal fluid (CSF) samples from patients diagnosed with leptomeningeal dissemination of cancer (leptomeningeal disease; LMD). LMD is a devastating late-stage metastatic feature of many solid tumors in the central nervous system associated with a median survival of 4-6 weeks[14-16]. Our trial achieved primary endpoint, with improved overall median survival (**Figure 4-1 a**), providing an unprecedented opportunity to study correlative cellular and molecular features of clinical PD1-blockade efficacy in LMD. By profiling CSF cells before and after treatment with Seq-Well, a high-throughput scRNA-Seq

method specialized for paucicellular patient-derived specimens[17,18] (**Figure 4-1**), as well as cfDNA, we characterize changes to the liquid component of the LMD TME[19] upon PD-1 blockade. Our data explicitly detail malignant (tumor) and non-malignant molecular and cellular ICI response features, and demonstrate the clinical utility of longitudinally sampling the TME.
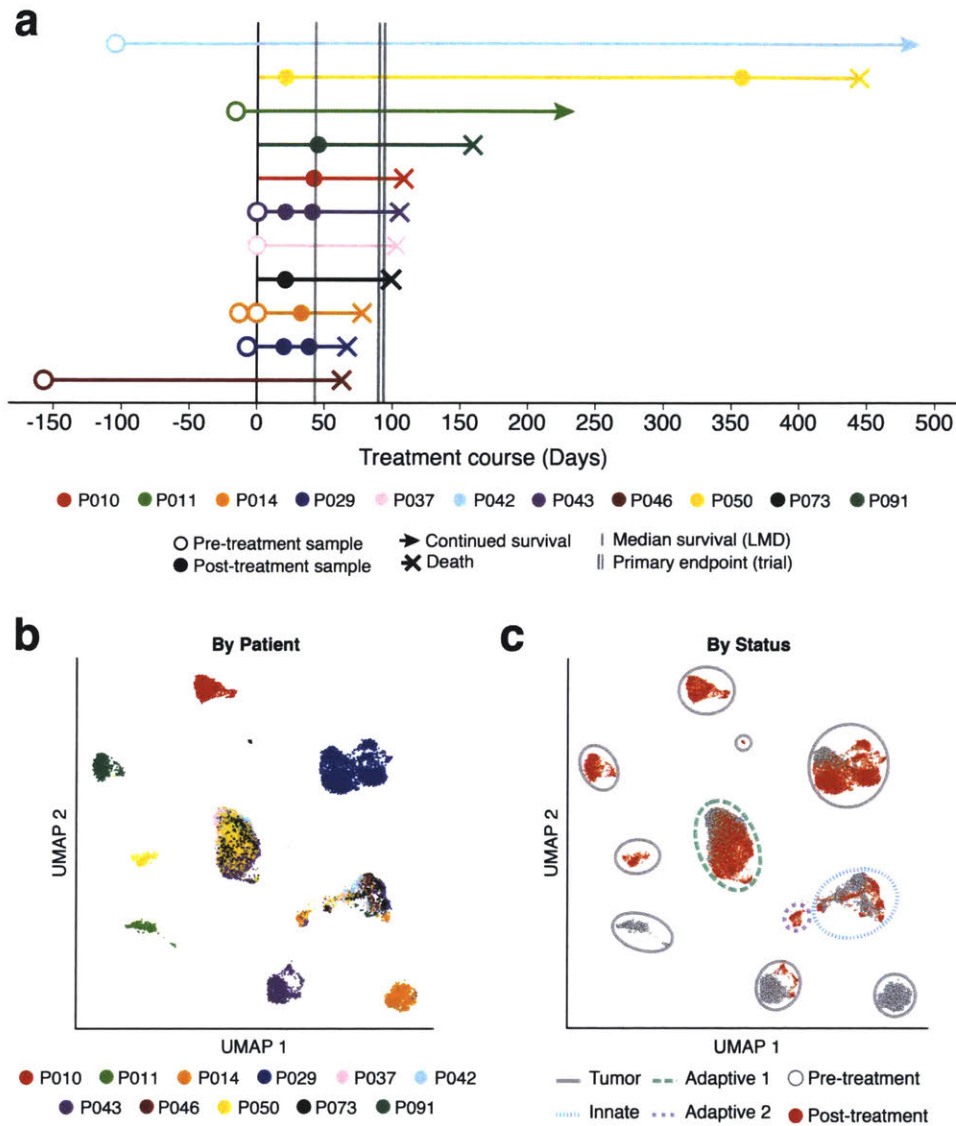

Figure 4-1: A phase-II clinical trial of the efficacy of pembrolizumab for treating LMD analyzed by longitudinal scRNA-Seq: **a**) Schematic representation of the longitudinal sampling performed on patients in this study. **b,c**) UMAP projection calculated from all cells (14,333), colored by patient (**b**) and by treatment status (**c**) with individual clusters identified (see **Figure 4-2**).

Results

4.2: scRNA-Seq of the LMD TME

We performed massively-parallel scRNA-Seq on 18 distinct CSF samples from 11 patients enrolled in a phase-II clinical trial testing the efficacy of pembrolizumab on LMD associated with breast cancer. 8 of these 11 patients surpassed the 3-month survival threshold, and overall, the trial met its primary endpoint[20]. As part of the trial, CSF was drawn every 3 weeks immediately prior to pembrolizumab administration, and a portion of this sample was utilized for single-cell profiling. After filtering for low quality cells and doublets[21,22], we obtained 14,333 single cells which we further identified using standard dimensionality reduction[23] and visualization strategies[24,25] (**Methods**). These analyses initially separated our data into twelve distinct clusters comprised primarily of malignant, T/NK, immunoglobulin-expressing B, and innate immune cells (**Figure 4-1 b,c**, **Figure 4-2 a,c**). The malignant cell clusters (n = 8) exhibited strong patient-specific representation while the immune clusters (n = 4) grouped by phenotype rather than by patient (**Figure 4-2 b**), consistent with previous observations derived through scRNA-Seq of human tumors[21,26].
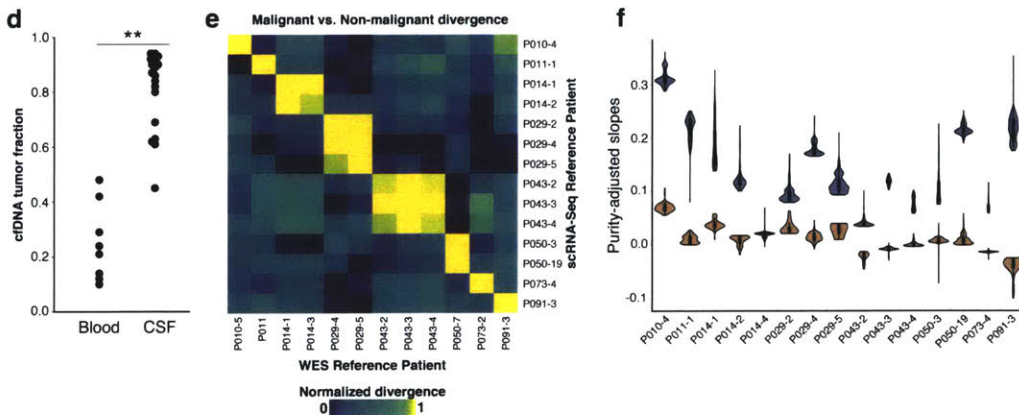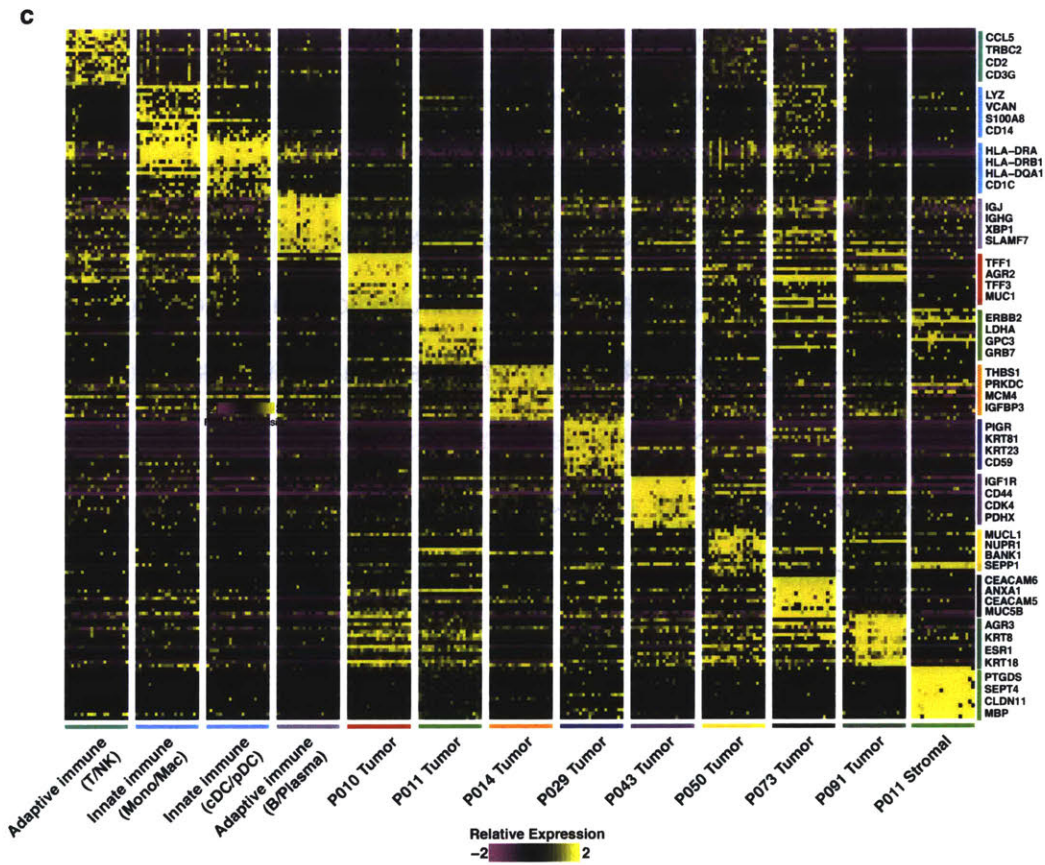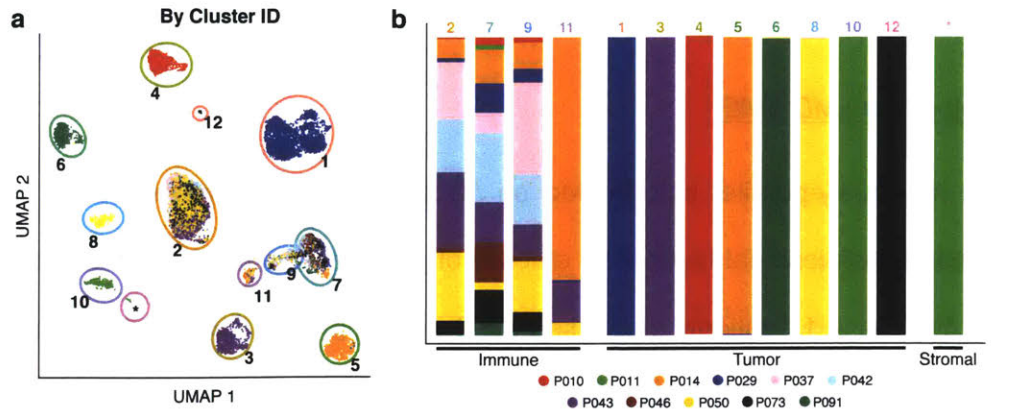
a By Cluster ID

b

Immune — Tumor — Stromal

P010   P011   P014   P029   P037   P042
P043   P046   P050   P073   P091

c

CCL5
TRBC2
CD2
CD3G

LYZ
VCAN
S100A8
CD14

HLA–DRA
HLA–DRB1
HLA–DQA1
CD1C

IGJ
IGHG
XBP1
SLAMF7

TFF1
AGR2
TFF3
MUC1

ERBB2
LDHA
GPC3
GRB7

THBS1
PRKDC
MCM4
IGFBP3

PIGR
KRT81
KRT23
CD59

IGF1R
CD44
CDK4
PDHX

MUCL1
NUPR1
BANK1
SEPP1

CEACAM6
ANXA1
CEACAM5
MUC5B

AGR3
KRT8
ESR1
KRT18

PTGDS
SEPT4
CLDN11
MBP

Adaptive Immune (T/NK) · Innate Immune (Mono/Mac) · Innate Immune (cDC/pDC) · Adeptive Immune (B/Plasma) · P010 Tumor · P011 Tumor · P014 Tumor · P029 Tumor · P043 Tumor · P050 Tumor · P073 Tumor · P091 Tumor · P011 Stromal

Relative Expression −2 ▬ 2

d **e** Malignant vs. Non-malignant divergence **f**

cfDNA tumor fraction — Blood · CSF

Normalized divergence 0 ▬ 1

Purity-adjusted slopes

82

Figure 4-2: Defining cellular identity across patients. a) UMAP of all cells noted by cluster identity calculated by SNN clustering. Gray is to distinguish cluster 9. **b)** Patient representation by cluster, aggregated by cellular identity (immune, tumor, stromal). **c)** Heatmap showing expression of select marker genes for all clusters (50 cells per cluster chosen for representation). **d)** Tumor purity calculated by ABSOLUTE of cfDNA derived from blood and CSF. **e)** Heatmap of all combinations of the mean difference between tumor (T) and non-tumor (NT) Theil-Sen slopes cells of chromosome-level average scRNA-Seq expression vs. chromosome-level WES tCR. **f)** Distributions of purity-adjusted slopes of chromosome-level average vs. WES tCR for each sample's tumor (blue) and non-tumor (orange) cells.

We further confirmed that the identified non-immune clusters were malignant cells by inferring copy number variation (CNV) profiles for each single cell[21,27,28]. As the average complexity of massively-parallel scRNA-Seq data is lower than that of plate-based methods utilized in previous studies[29], we developed an approach that combines whole exome sequencing data (WES)—here, from CSF-derived cfDNA—with patient-matched scRNA-Seq data to detect chromosome-level aberrations (see **Methods**); leveraging our CSF-derived cfDNA data as a genomic CNV reference, we corroborated the malignancy status of our individual cells. We note that we detected significantly higher fractions of genomic material derived from cancer cells in the CSF than in paired peripheral blood samples, suggesting that the CSF is a largely isolated microenvironment (**Figure 4-2 d-f**). Overall, we retained comparable pre- and post-treatment cell numbers for all aforementioned cell types, enabling investigation of the impact of intravenous PD-1 blockade on the heterogeneous cellular subsets present in LMD CSF TME.

4.3: Increased Expression of Genes Associated with Cytotoxicity and Proliferation in T Cells Following ICI

Using a combination of unsupervised and supervised approaches, we first partitioned the T/NK cluster (cluster 2, **Figure 4-2 a**) into CD4+ and CD8+ T cells, and NK cells (which can co-segregate during high-level analyses of scRNA-Seq data based on gene-expression similarity[30]) (**Figure 4-3 a-c**, **Methods**). We then investigated phenotypic shifts within these T cell subsets following PD-1 blockade, beginning with CD8+ T cells, the major target of pembrolizumab[31].
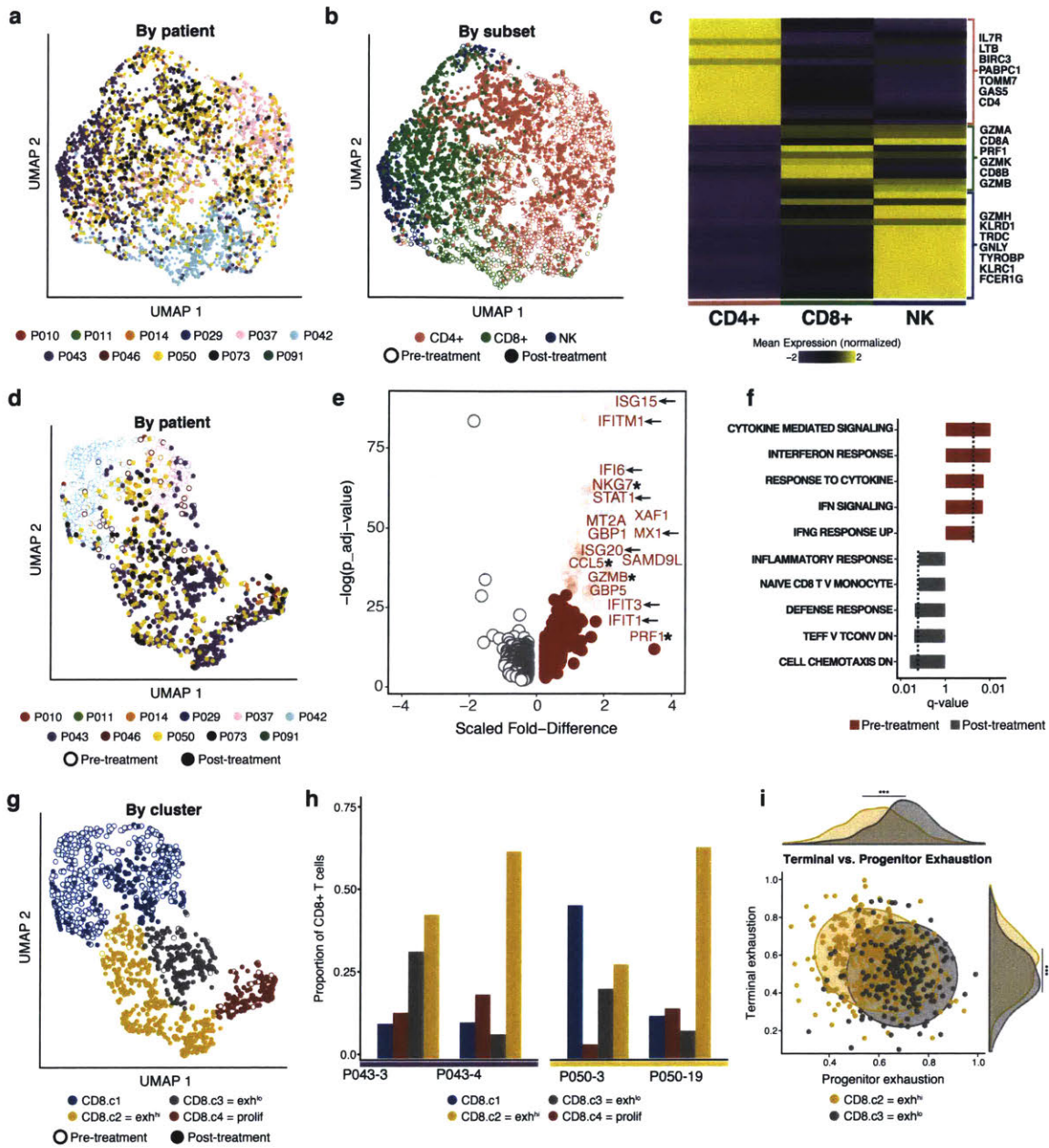
Figure 4-3: T cells exhibit strong differences in the expression of interferon-induced, cytotoxic, and exhaustion genes following PD-1 blockade. a,b) UMAP calculated over all T/NK cells (n = 4,251), colored by patient (**a**) and canonical cell type (**b**) (identified via iterative subclustering; see **Methods**). **c**, Heatmap showing the expression of representative marker genes for each canonical cell type cluster. **NB** CD8+ and NK cells share markers of cytotoxicity. **d**) UMAP calculated over all CD8+ T cells (n = 1,446), colored by patient. **e**) Volcano plot of differentially expressed genes between pre- and post-treatment CD8+ T cells with cytotoxic (stars) and interferon response (arrows) genes highlighted. **f**) Gene set enrichments calculated over genes differentially expressed between pre- and post-treatment CD8+ T cells. **g**) UMAP calculated over all CD8+ T cells colored by SNN-cluster (see **Methods**). **NB** CD8.c2, CD8.c3, and CD8.c4 are additionally referred to as CD8.exh^hi, CD8.exh^lo, and CD8.prolif, respectively. **h**) Proportion of CD8+ T cell cluster representation by post-treatment time points within patients

043 and 050. **i**, Scoring of CD8.exh^{hi} and CD8.exh^{lo} T cells from **e**, according to progenitor and terminally exhausted T cell signatures identified in Miller et al[36] (***, p<0.001; Mann-Whitney U Test).

Unsupervised analysis of expression data from LMD CSF revealed increased CD8+ T cell proliferation following intravenous administration of pembrolizumab, consistent with previous reports[31-33]. CD8+ T cells across all patients (**Figure 4-3 d**) resolved into 4 distinct clusters (**Figure 4-3 g**), largely partitioned by treatment status. One cluster in particular (CD8.c4) expressed high levels of genes associated with proliferation. This cluster was primarily composed of post-treatment cells, and it comprised a larger fraction of total post-treatment cells than pre-treatment (14.9% vs. 1.8% of post-treatment and pre-treatment cells, respectively, **Figure 4-4**).
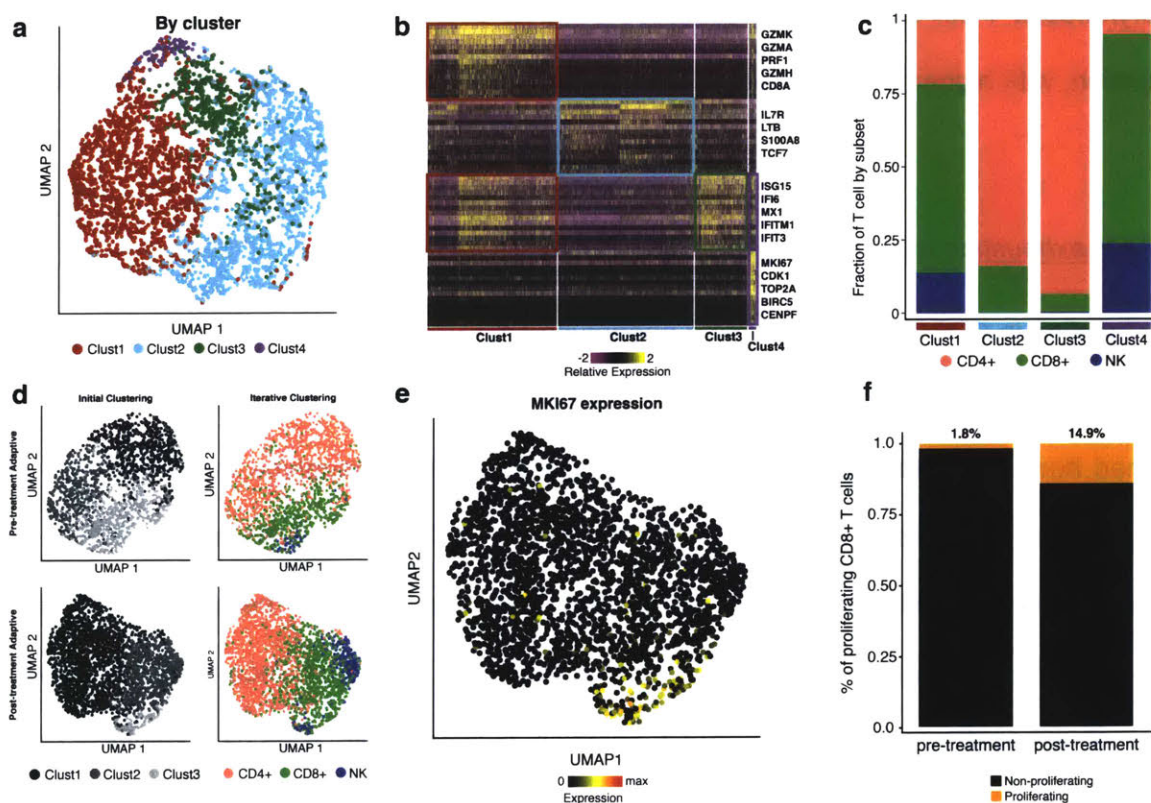

Figure 4-4: Unsupervised analysis of T/NK cluster. a) T/NK cluster UMAP, colored by clusters calculated by SNN. **b**) Heatmap showing the expression of marker genes with select markers per cluster. Markers represented in multiple clusters highlighted. **c**) Fraction of T/NK

cells in each SNN cluster. **d**) SNN clusters and iteratively assigned identities of CD4+ T cells, CD8+ T cells, and NK cells. **e**) MKI67 expression in post-treatment T/NK cells projected over UMAP. **f**) Proportion of cycling cells that comprise all pre- and post-treatment CD8+ T cells.

Additional comparisons suggested that PD-1 blockade increases cytotoxic and interferon response gene expression among CD8+ T cells. Differential gene expression analysis between untreated and treated non-proliferating CD8+ T cells revealed increased expression of a set of genes containing canonical markers of cytotoxicity (*NKG7, PRF1, GZMB*) and interferon-response (*ISG15, STAT1, MX1*) following treatment (**Figure 4-3 e**), as well as overall enrichments for genes associated with leukocyte activation, interferon-γ signaling and response, and cytokine signaling pathways (**Figure 4-3 f**). Conversely, genes upregulated in untreated CD8+ T cells included *IL7R, S100A4, BCL2*, which have previously been associated with memory/effector T cell state[34,35]. We observed a similar bifurcation in CD4+ T cells following ICI introduction, with higher expression of interferon- signaling and response genes post-treatment.

4.4: T cell exhaustion is heterogeneously affected by ICI introduction.

PD-1 signaling in CD8+ T cells induces a phenotype characterized by expression of genes such as *CD39, TIM3, HAVCR2*, and others[36-38], commonly referred to as T cell exhaustion. We examined how PD-1 blockade administration might affect this phenotype in our CD8+ T cells using a series of curated signatures[12,36,38-40] (**Figure 4-6**). Our results suggest that post-treatment CD8+ T cells were more exhausted than pre-treatment cells ($p<2\times10^{-3}$ for all evaluated signatures, Mann-Whitney U Test; see **Table 4-1** for significance tests), and that the degree of exhaustion varied significantly between clusters, with CD8.c2 exhibiting greater exhaustion than CD8.c3 ($p<2\times10^{-6}$ for all significant signatures. Mann-Whitney U Test; see **Table 4-1** for significance tests; **Figure 4-6**). We hereafter refer to CD8.c2 and CD8.c3 as CD8.exh^{hi} and CD8.exh^{lo}, respectively. CD8.exh^{hi} also exhibited significantly higher expression

86

of genes associated with cytotoxicity and interferon response (e.g., *PRF1*, *NKG7*, and *GZMK*) than CD8.exh^{lo}.
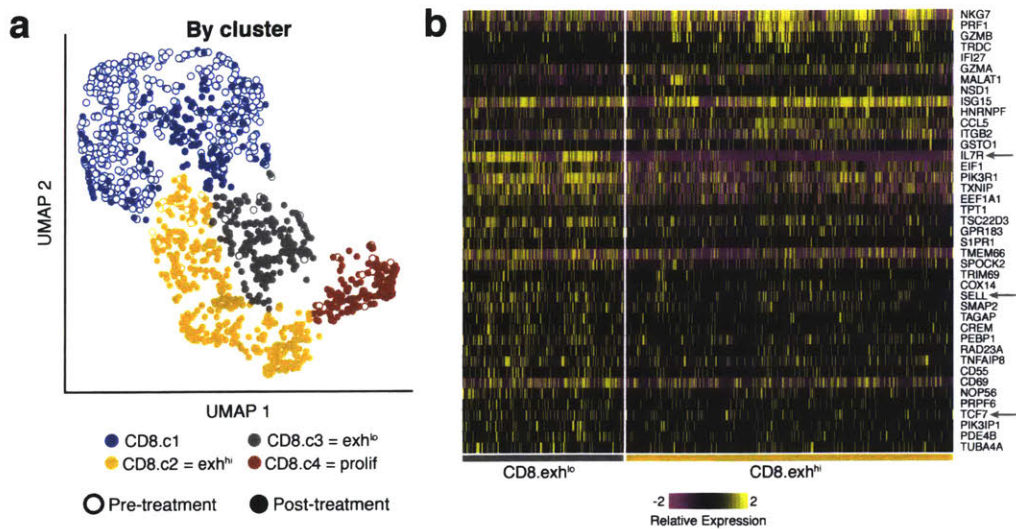

Figure 4-5: Analysis of exhaustion heterogeneity behavior in CD8+ T cells. a) CD8+ T cell clusters represented in UMAP. **b)** Heatmap showing expression of differentially expressed genes between CD8.exh^{hi} and CD8.exh^{lo} CD8+ T cells; arrows point to genes traditionally associated with memory (*IL7R, SELL*) or good prognosis to PD-1 blockade (*TCF7*).
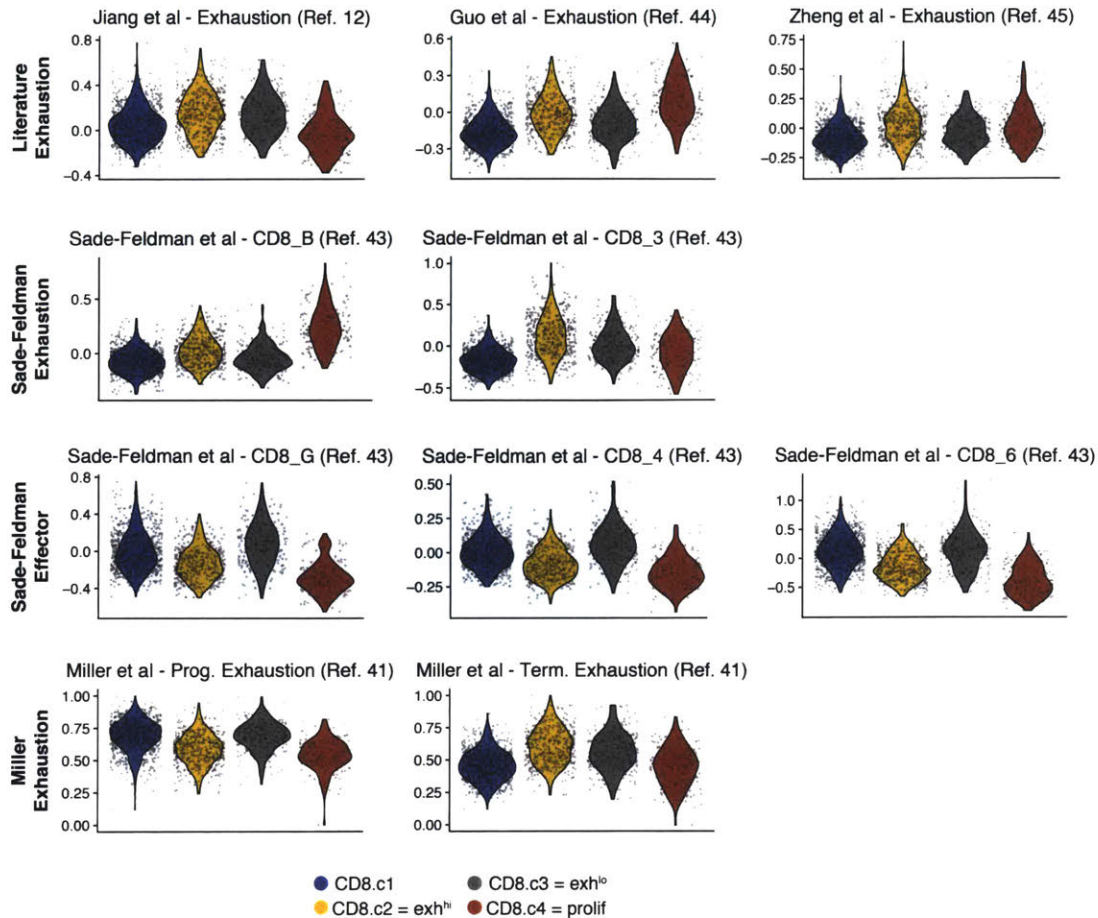
Figure 4-6: Exhaustion heterogeneity over literature-curated signature among CD8+ T cells. Violin plots of all literature-derived scores of T cell exhaustion used in this manuscript stratified by CD8+ T cell clusters.

We next sought to understand how these two exhaustion states varied over the course of PD-1 blockade treatment in patients. For two patients from whom multiple post-treatment CSF draws were available (see **Methods**), we detected a smaller fraction of CD8.exh^{lo} cells at the later post-ICI time point, and a larger proportion of CD8.exh^{hi} (**Figure 4-3 h**). These observations suggest that the CD8.exh^{lo} and CD8.exh^{hi} phenotypes may be generally associated with early and late PD-1 blockade, respectively. We demonstrate here that longitudinal profiling of the same patient's liquid TME reveals congruent phenotypic shifts inferred from the full cohort.

Recent work in transgenic murine models has suggested that PD-1 blockade may transition exhausted CD8+ T cells from a stem-like exhausted state towards a more terminally exhausted one[36,41]. Thus, we scored our CD8+ T cells according to recent work that described exhaustion as a heterogeneous phenomenon defined by at least two distinct phenotypes[36-38,41-44]: progenitor and terminally exhausted CD8+ T cells[36]. We observed high concordance between our clusters and these two states: CD8.exh^{lo} cells more closely resembled progenitor exhausted T cells, while CD8.exh^{hi} cells mirrored terminally exhausted T cells (**Figure 4-3 i**; $p<2x10^{-16}$ for progenitor and $p<2x10^{-4}$ for terminal, Mann-Whitney U Test; see **Table 4-1** for significance tests). While CD8.exh^{hi} cells exhibit higher cytotoxicity, CD8.exh^{lo} cells exhibited increased memory/effector-like function[38] (**Figure 4-6**), as well as increased expression of *TCF7*, a biomarker frequently associated with positive patient prognosis in clinical ICI studies[36,38,45] (**Figure 4-5**). Furthermore, our human data support a model of T cell reactivation whereby CD8.exh^{lo} cells achieve a more terminal exhausted state following treatment with PD-1 blockade, here demonstrated in the same patient's liquid TME over the course of PD-1 blockade.

4.5: Innate Immune Cells Exhibit Pro- and Anti-inflammatory Behavior Following PD-1 Blockade

Although PD-1 blockade directly modulates T cell activity, we also observed clinically relevant behaviors among innate immune cells (clusters 6, 8, **Figure 4-2**). We identified canonical innate immune phenotypes[46] as previously described (**Figure 4-7 a-c**). Of these, plasmacytoid dendritic cells (pDCs) were the only subpopulation comprised of more post-treatment than pre-treatment cells (**Figure 4-8 b**). Since pDCs have been shown to produce large quantities of interferon in response to antigen[47] – also seen in our data – they may drive, in part, the aforementioned interferon responses in the LMD CSF TME.
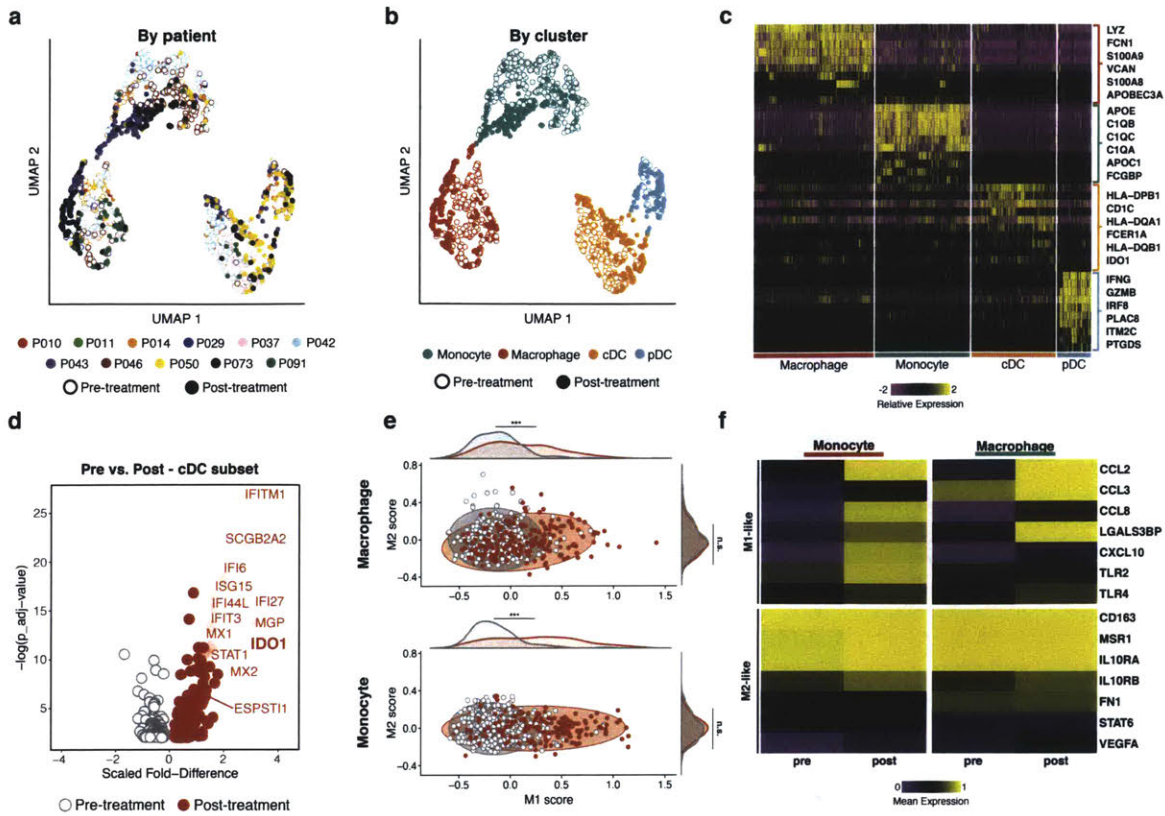
Figure 4-7: Innate immune cells exhibit multifaceted responses to PD-1 blockade. a,b, UMAP calculated over all innate immune cells (n = 1,701), colored by patient (**a**) and canonical cell type (**b**) (identified via SNN; see **Methods**). **c**) Heatmap showing the expression of select markers for the clusters identified by SNN (**b**). **d**) Volcano plot of genes differentially expressed between pre- and post-treatment classical dendritic cells (cDCs). **e**) M1 and M2 signature scores for pre- and post-treatment monocytes and macrophages plotted in 2D (***, p<0.001; Mann-Whitney U Test). **f**) Heatmap showing the mean expression for M1- and M2-associated genes across pre- and post-treatment monocytes and macrophages.

Differential expression between pre- and post-treatment cells within each innate subpopulation revealed increased expression of interferon response and antigen-processing genes post-treatment (**Figure 4-8 c-d**). We found that classical DCs (cDCs), in addition to upregulating interferon response genes post-treatment, also upregulated genes that negatively affect immune response, such as *IDO1*, *EPSTI1*, and *MGP* (**Figure 4-7 d**). *IDO1* production is stimulated by interferon-γ-induced *JAK/STAT* signaling in the tumor-immune microenvironment, and depletes tryptophan from the extracellular matrix, which is required by T cells to proliferate[48]. The observed post-treatment shift among cDCs in our data suggests the possibility

that innate immune mechanisms of ICI resistance may become more active in response to interferon-mediated signaling following PD-1 blockade, potentially impacting clinical benefit in these patients.
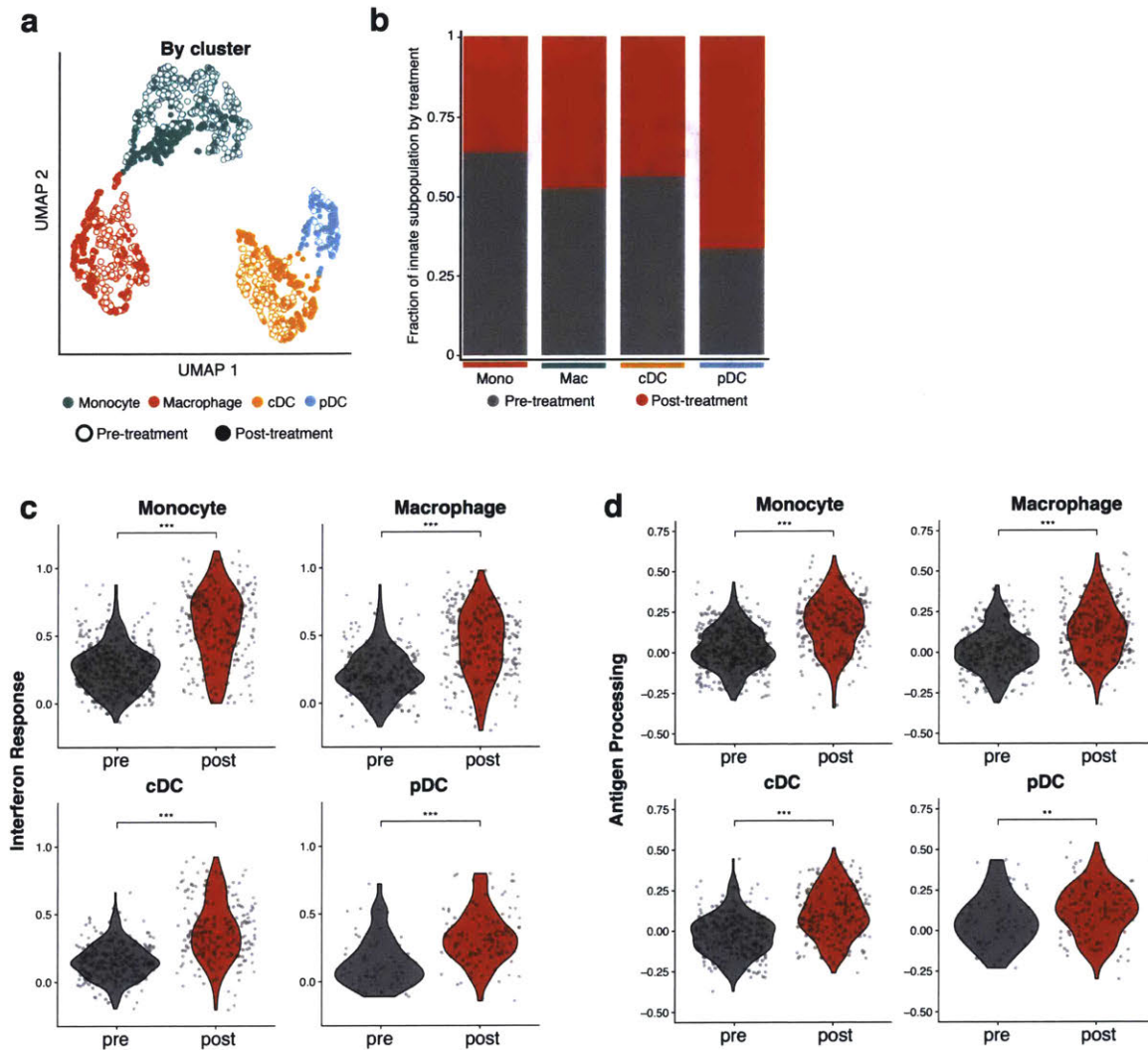

Figure 4-8: Innate immune responses and heterogeneity following PD-1 blockade. a, UMAP of innate immune subpopulations identified from SNN clustering. **b,** Enumeration of the fractional abundance of pre- and post-treatment cells in each subpopulation. **c,** Interferon response and **d,** antigen processing scores for all innate immune subsets stratified by pre- and post-treatment status (***, $p<0.001$; **, $p<0.01$; Mann-Whitney U test).

We further observed that, post-treatment, innate immune subpopulations exhibited independent shifts in the expression of genes associated with the M1-like (proinflammatory) and M2-like (immunomodulatory) phenotypes[12,49,50] (**Figure 4-7 e-f**; $p<3\times10^{-15}$ for all signatures, Mann-

Whitney U Test; see **Table 4-1** for significance tests). In particular, expression of genes such as *CCL2* and *CCL3*, which drive immune recruitment, was elevated in post-treatment monocytes and macrophages (**NB** *CCL2* has also been implicated in M2 reprogramming in solid tumors following its role in recruitment[51]), as were *CCL8*, *LGALS3BP*, *CXCL10* and TLRs, which drive cytotoxicity and defense response and are associated with M1-like behavior in monocytes and macrophages [52,53] (**Figure 4-9**).

Conversely, the M2-like signature and its associated genes – including *CD163*, *MSR1*, *IL10R* – maintained consistent expression levels across treatment in these monocytes and macrophages[52] (**Figure 4-7 c-d**, **Figure 4-9**; $P>0.02$ with effect size<0.1 for all, Mann-Whitney U Test; see **Table 4-1** for significance tests). PD-1 signaling has previously been reported to regulate macrophage activity[54,55], and macrophages have been implicated in resistance to anti-PD-1 therapy[56,57]. Our observation of increased M1-like functionality, despite constant M2-like behavior, suggests that control of pro- and anti-inflammatory programs are contingent on multiple factors, and that PD-1 blockade and subsequent signaling may modify the activity of one, but not necessarily at the expense of the other.
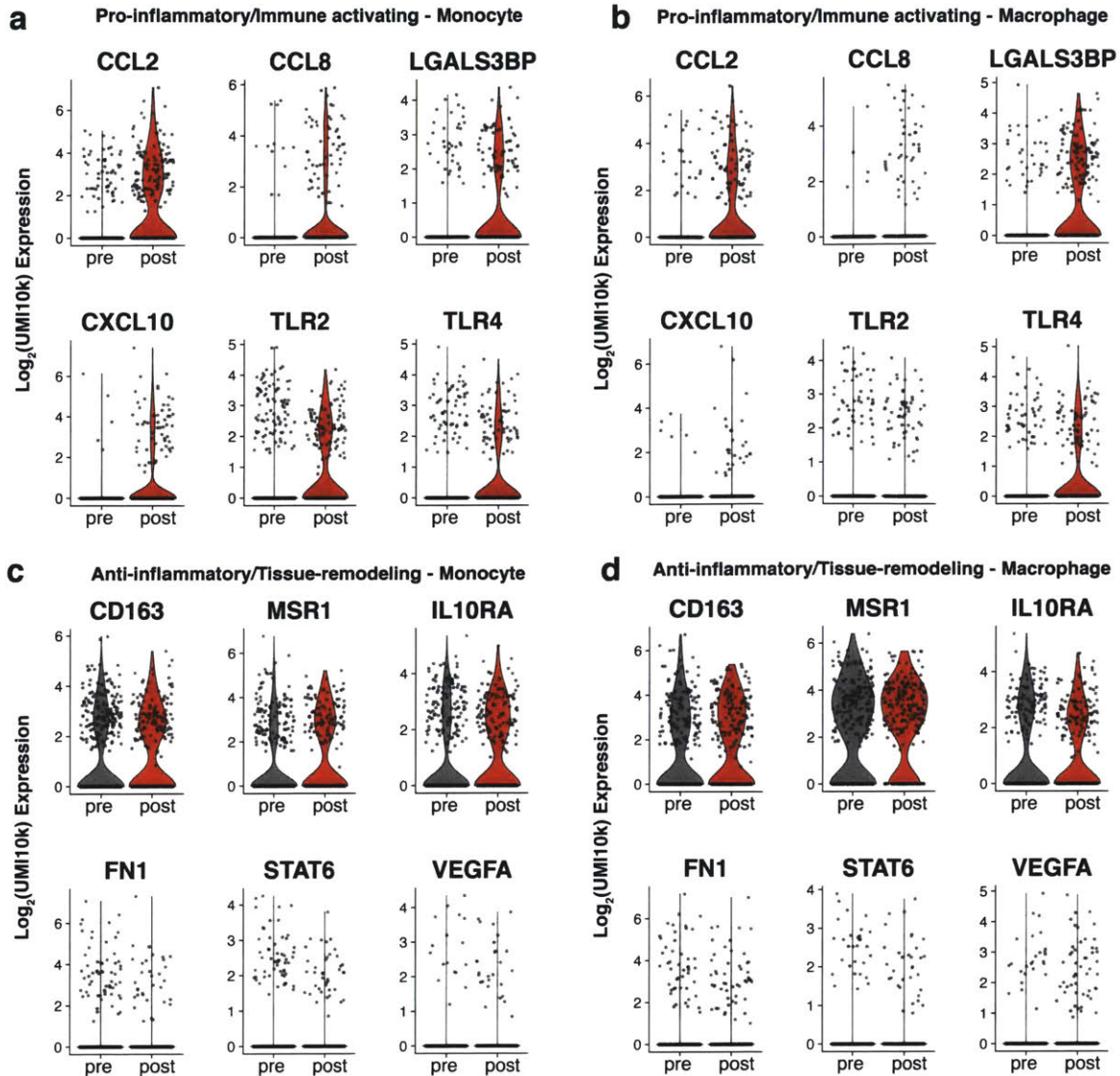
Figure 4-9: Monocyte/macrophage phenotype shifts in polarization following immunotherapy treatment. a,b, Violin plots of pro-inflammatory genes stratified between pre- and post-treatment for (**a**) monocytes and (**b**) macrophages. **c,d**, Violin plots of anti-inflammatory genes stratified between pre- and post-treatment for (**c**) monocytes and (**d**) macrophages.

4.6: Longitudinal Tumor Cell Behavior Following ICI Varies Across Patients

A comparison of pre- and post-treatment tumor cells across patients revealed differences that suggested increased susceptibility to immune activity following treatment, consistent with previous reports[58-60]. We found that both the per-patient average and combined single-cell

scores for interferon-γ response and antigen processing genes were upregulated post-treatment (**Figure 4-10 a**; p<0.03 for per-patient average, $p < 2 \times 10^{-16}$ for combined single-cell, Mann-Whitney U Test; see **Methods**).

Our unsupervised analysis revealed that the primary source of variation across tumor cells was patient identity (**Figure 4-10 b**). We therefore performed a stratified analysis of interferon-γ response and antigen processing on four patients for whom CSF was sampled both before and after PD-1 blockade. This comparison revealed markedly more variable longitudinal activity than was apparent through our inter-patient analyses (**Figure 4-10 c**). In particular, P043 exhibited initial upregulation of interferon response following pembrolizumab administration ($p < 2.2 \times 10^{-16}$, effect size = 1.61, Mann-Whitney U Test; see **Table 4-1** for significance tests), consistent with the general result across patients. However, P029 exhibited only a slight increase in interferon-gamma response ($p < 1.2 \times 10^{-6}$ effect size = 0.2, Mann-Whitney U Test; see **Table 4-1** for significance tests) at the first time point immediately following treatment. Furthermore, following these initial increases, both P029 and P043 showed decreases toward (P043) or below (P029) pre-treatment levels ($p < 4.1 \times 10^{-9}$ for P029 and $p < 4.5 \times 10^{-12}$ for P043, Mann-Whitney U Test; see **Table 4-1** for significance tests). We note that we were unable to evaluate tumor cell behavior in P014, as an insufficient number of tumor cells were detected in the post-treatment time point. These results, taken together, suggest that proper determination of the cellular and molecular features that inform clinical course may be critically impacted by the timing of post-treatment measurements.
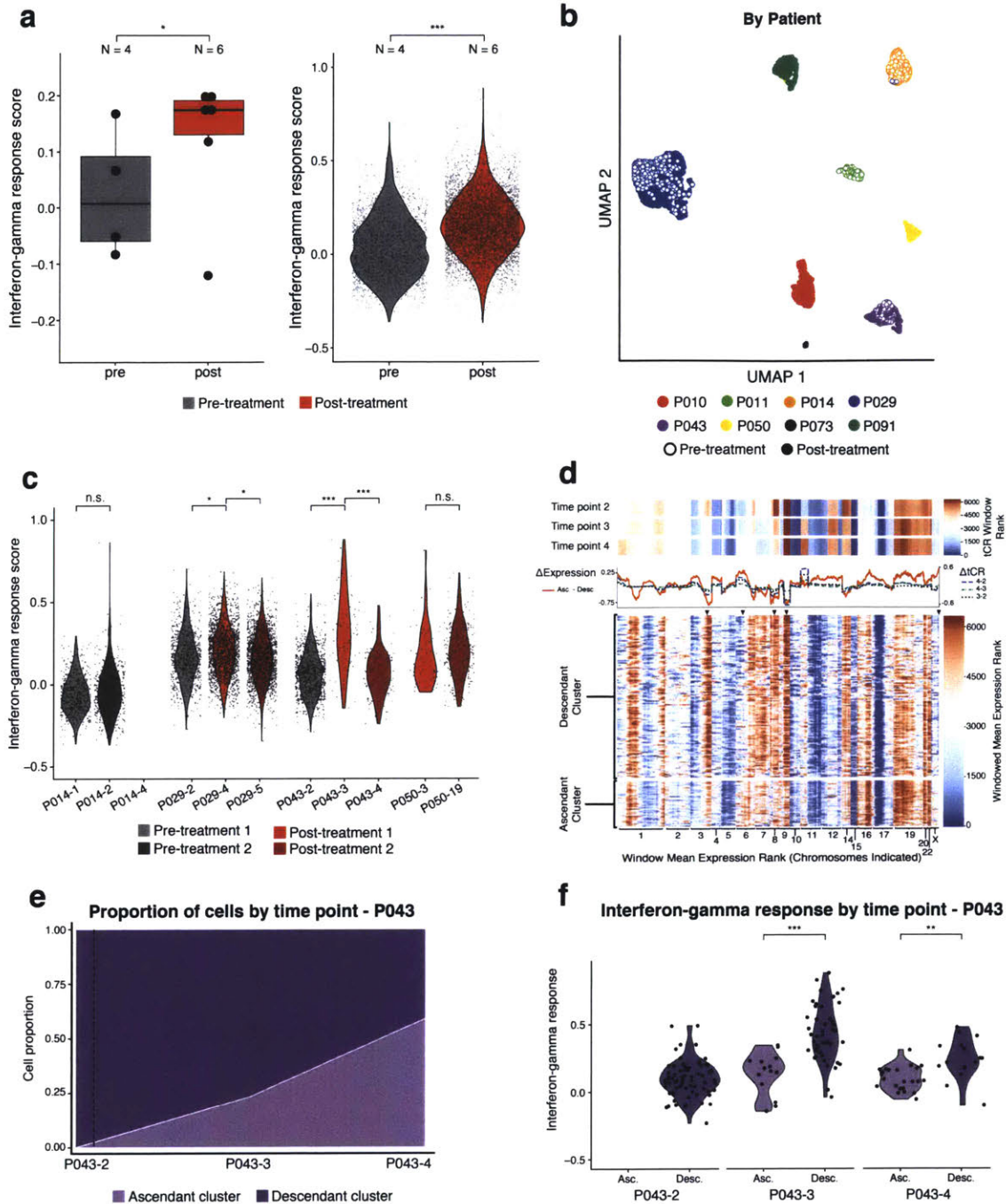
Figure 4-10: Inter-patient vs. intra-patient analysis of tumor cells reveals dynamic, transient response behaviors. **a**) Per-patient averaged (left) and cross-patient single-cell (right) scores for interferon-gamma response genes in tumor cells stratified by pre- and post-treatment status (*, p<0.05; ***, p<0.001; Mann-Whitney U test; see **Methods**). **b**) UMAP calculated over all tumor cells (n = 8,071), colored by patient and denoted by treatment status. **c**) Interferon-gamma response gene scores for individual patients' tumor cells, colored by treatment status (*, p<0.05; ***, p<0.001; Mann-Whitney U test). **d**) Heatmap of ranked windowed mean expression ordered by chromosome and transcriptional start site for all considered tumor cells in P043, separated by agglomerative cluster identity (bottom, arrows

identify peaks of mutual information between clusters). Windows are represented in transcriptomic units (see **Methods**). Plotted above is mean difference in single-cell gene expression between ascendant and descendant clusters overlaid with change in tCR between time points. WES tCR (ranks-by-window) for each of three sampled time points from CSF-derived cfDNA at top. **e**) Proportion of clusters in P043 stratified by time point (Ascendant below, Descendant above). **f**) Interferon-gamma response gene score at each of the three sampled time points in P043 stratified by cluster identity (**, $p < 0.01$; ***, $p < 0.001$; Mann-Whitney U test).

4.7: Tumor Clonal Evolution Following PD-1 Blockade

In one individual – P043 – we observed a dramatic change in the relative abundance of two tumor subclones following PD-1 blockade. In all patients, we assessed the possibility of subclonal tumor heterogeneity[61-65] by inferring single-cell CNV profiles[21,26,27,66] via clustering in windowed mean expression (WME, see **Methods**) space. Visualizing these clusters with reduced dimensions, we found little structured CNV heterogeneity within 6 of the 7 evaluable patients (**Figure 4-11**). For example, in P029, we observed clusters in gene expression space (primarily attributable to cycling status) but not in inferred CNV space; in other patients (P014, P050), inferred CNV variation was limited and restricted to only a few loci.

Integrative longitudinal analysis of cfDNA whole exome sequencing (WES) and scRNA-Seq data from P043 tumor cells revealed evidence of clonal evolution whereby a subclone with low interferon-γ response expression became the majority subclone at the latest post-ICI time point (**Figure 4-10 d**; see **Methods**). CSF-derived cfDNA WES data from P043 was highly consistent with two major subclonal populations, and suggested a monotonic increase in the fractional abundance of one subclone over time (**Figure 4-10 e**). Regions of high divergence in inferred CNV between clusters corresponded to regions of large change in total copy ratio (tCR) between time points. Notably, these CNV-derived clusters were highly concordant with gene expression-based clustering (**Figure 4-10 e**, **Figure 4-11**).
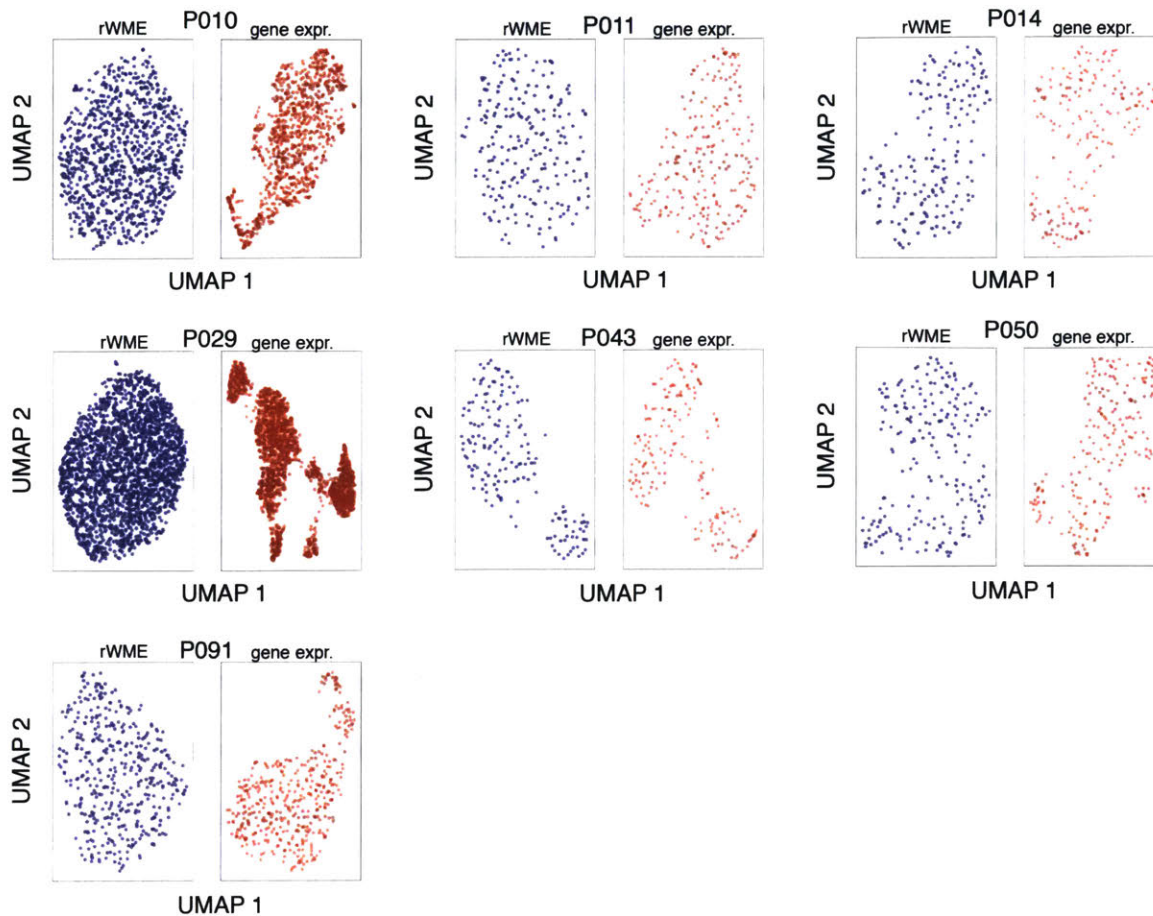
96

Figure 4-11: UMAPs of tumor cell windowed mean expression and expression by patient. UMAPs calculated over rWME (left, blue) and gene expression (right, red) for 7 patients with appreciable numbers of tumor cells. Structured heterogeneity in P014 and P050 is isolated to only a few chromosomes.

We performed differential expression in aggregate and at a single time point between these two CNV-defined subclones in P043 to determine transcriptomic features defining them. This analysis revealed several genes associated with interferon-γ response. We therefore plotted interferon response scores for each subclone at the three time points measured for P043 (**Figure 4-10 f**). These data show that the descendent clone exhibited higher interferon-γ response genes at P043-3 (p<2x10^{-3} for post-treatment time points, Mann-Whitney U Test; see **Table 4-1** for significance tests) – the time point immediately following treatment (seen in **Figure 4-10 f**). In contrast, the ascendant clone exhibited consistently lower interferon response

across this trajectory, eventually predominating at the last time point (P043-4). These observations are potentially consistent with adaptive selection of tumor cells that do not express genes associated with interferon-γ signaling, as the relative abundance of immunogenic cells decreases and these less sensitive cells grow out over the course of treatment.

4.8: Discussion

We present an experimental and computational framework for efficient cell capture and analysis of paucicellular clinical cancer samples. We use these techniques to perform the first scRNA-Seq characterization of the liquid component of the LMD TME, as well as its response to PD-1 blockade. Our data demonstrate shifts in multiple immune programs following ICI introduction, as well as effects on metastatic tumor cells. We find several unique pro- and anti-inflammatory immune features that may have contributed to the clinical benefit observed in our phase-II trial and impacted tumor cell behavior, which we find is highly sensitive to measurement timing. Moreover, we devised a computational workflow which couples cfDNA-Seq with massively-parallel scRNA-Seq to interrogate single tumor cell subclonality and dynamics within individual patients.

Among T cells in the liquid LMD TME, we uncovered shifts toward greater immune-activation following intravenous treatment with pembrolizumab. CD8+ T cells, in particular, exhibited increased proliferation and cytotoxicity[67,68], demonstrating that intravenous biological modulation can result in altered cellular programs in the CNS. Furthermore, samples collected following pembrolizumab administration contained new cell types, including Ig-expressing B cells and γδ T cells, which, to our knowledge, have not been characterized previously in this microenvironment. We additionally determined that post-treatment CD8+ T cells exhibited bifurcating exhaustion profiles[32,36-38,69] – CD8.exh^{hi} and CD8.exh^{lo} – with longitudinal shifts during anti-PD-1 treatment towards CD8.exh^{hi}. We note that we cannot rule out the possibility of

induced immune infiltration to the CNS between treatment time points. Thus, our data does not definitely confirm CD8+ T cell reactivation. Nevertheless, they do support the idea that PD1-blockade leads to CD8+ T cell behaviors that are more proliferative and cytotoxic, but potentially short-lived[32,37]. Future work will explore the clinical implications of accruing this T cell phenotype in LMD and other diseases, as well as the relationship of CNS T cells to those in the periphery.

We additionally profiled the effect of PD-1 blockade on innate immune cells, which, in turn, can modulate the LMD TME to reinforce or abate immune activity. In particular, we identified increased expression of interferon response genes, and, in the case of pDCs, interferon production following pembrolizumab treatment. After ICI introduction, cDCs also exhibited immune regulatory behavior, e.g. expression of *IDO1*. The immunosuppressive effects of *IDO1* upregulation following PD-1 blockade are an important consideration for future work in this and other malignancies[70]. Interestingly, while PD-1 signaling has been reported to affect macrophage polarization[54,55], we observed that PD-1 blockade independently influenced the expression of genes associated with the M1-like and the M2-like phenotypes, suggesting that these axes may be modulated independently in future therapeutic efforts.

Among tumor cells, we observed dynamic behaviors that varied across participants following PD-1 blockade. Notably, we found that elevated expression of interferon-gamma response and antigen processing genes were transient, and that the degree of transience can vary across patients. Further, our data suggest that this behavior may, in select cases, be attributed to adaptive selection of less immunogenic subclones following pembrolizumab treatment. These conclusions were uniquely enabled via multiple post-treatment measurements and the utilization of CNVs inferred through cfDNA-Seq to perform differential gene expression within a single time point in the same patient. Collectively, our results elucidate clinically significant tumor cell behavior in the liquid component of LMD. Future work will be needed to examine how solid meningeal metastases not sampled in the CSF respond to PD-1 blockade treatment.

Analyses of our heterogeneous cohort across all patients exploits the resolution of scRNA-Seq to maximize sampling power, detect conserved behaviors, and provide robustness to technical noise. Our within-patient longitudinal analyses, meanwhile, isolate many experimental and patient-exogenous factors, and crucially reveal the importance of sampling time when attempting to contextualize patient behavior following intervention. While our sample numbers were limited due to clinical considerations (see **Methods**), our results demonstrate the importance of examining a patient's instantaneous state in addition to overall response when deriving putative clinical biomarkers.

Together, our analytical approach empowers hypothesis generation across patients and subsequent testing within patients. Employing this framework, we tracked the dynamics of T cell exhaustion phenotypes and tumor cell interferon responses. These complementary analyses detail a complex, dynamic tumor microenvironment response to ICIs in LMD, with widespread clinical implications in this disease and beyond.

4.9: Methods

Participants and design of study

Study Design and Patients

This study (clinicaltrials.gov identifier NCT02886585) was designed by the principal investigators and the Dana-Farber Harvard Cancer Center (DF/HCC) Institutional review board approved the protocol. The study was conducted in accordance with the provision of the Declaration of Helsinki and Good Clinical Practice guidelines. Eligible patients had histologically confirmed disease from any solid tumor with an ECOG performance status $\leq$ 2, normal organ and marrow function, were on a stable dose of dexamethasone of 2mg or less for 7 days prior to initiation of treatment, and leptomeningeal carcinomatosis as defined by positive cytology. Given the progressive, end-stage nature of LMD, 7/11 patients included in this study were on a steroid regimen at the time of enrollment and 8/11 patients were treated with steroids while receiving pembrolizumab. Written informed consent was obtained for all participants.

Study Design, Treatment, and Endpoints

This was a single arm Phase-II study of patients with solid tumors with leptomeningeal carcinomatosis as defined by positive cytology. Pembrolizumab was administered intravenously at 200mg every 3 weeks until disease progression, death or unacceptable toxicity. A brain MRI and CT chest/abdomen/pelvis CT were obtained every 6 weeks for restaging. The primary endpoint was the rate of overall survival at 3 months (OS3). A total of 11 out of 18 enrolled patients were included for single-cell analysis, 4 of whom were sampled at multiple time points across the course of treatment. All of these 11 were diagnosed with metastatic breast cancer and positive CSF cytology indicative of LMD.

Further details of the subjects', including radiation, surgical history, imaging, disease progression, and survival are provided in **Table 4-2**.

Sample processing. Cerebrospinal fluid (CSF) from patients was extracted via ventriculoperitoneal shunt (VPS), spun at 800G for 5 min to pellet cells, and resuspended in PBS (ThermoFisher 10010023, Ca/Mg-free). Red blood cells (RBCs) were lysed using ACK lysis buffer (ThermoFisher A1049201) for 4 min on ice to remove RBCs. Cells were then washed with sterile PBS and spun down at 800G for 5 min, resuspended as a single-cell suspension in RPMI (Corning) with 10% FBS (ThermoFisher 10082-147) for subsequent scRNA-Seq.

scRNA-Seq with Seq-Well. Resuspended CSF cells were profiled using the Seq-Well platform for massively parallel, high-throughput scRNA-seq for low-input clinical samples. A complete description of methods is available online[17]. Briefly, cells from each CSF sample were manually counted (Bal Supply 808CI) and loaded onto one array preloaded with barcoded mRNA capture beads (ChemGenes). All samples retained fewer than 10,000 cells with the exception of two (CSF029-4 & DFCI010-4; ~100,000 cells). Thus, all available cells were loaded onto a single array, except CSF029-4 and DFCI010-4 where ~10,000 cells were loaded. The loaded arrays containing cells and beads were then sealed using a polycarbonate membrane with a pore size of 0.01 μm, which allows for the exchange of buffers but retains biological molecules confined within each nanowell. Subsequent buffer exchanges facilitate cell lysis, transcript hybridization, and bead recovery before performing reverse transcription *en masse*. Following reverse transcription using Maxima H Minus Reverse Transcriptase (ThermoFisher EP0753) and an Exonuclease I treatment (NewEngland Biolabs M0293L) to remove excess primers, PCR amplification was carried out using KAPA HiFi PCR Mastermix (Kapa Biosystems KK2602) with approximately 2,000 beads per 50 μl reaction volume. Libraries were then pooled into one tube

(with the exception of CSF014-4, CSF029-4, and CSF046-2, which were pooled to two tubes) and purified using Agencourt AMPure XP beads (Beckman Coulter, A63881) by a 0.6X SPRI followed by a 0.8X SPRI and quantified using Qubit hsDNA Assay (Thermo Fisher Q32854). The quality of each WTA product was assessed using the Agilent hsD5000 Screen Tape System (Agilent Genomics) with an expected peak ranging between 800-1500 bp tailing off to beyond 5000 bp, and a small or non-existent primer peak (~100-200 bp). 3' digital gene expression (DGE) libraries were constructed using the Nextera XT DNA tagmentation method (Illumina FC-131-1096) using index primers as described in Gierahn et al[17]. Loaded samples ranged from 600-2,000 pg of pooled cDNA, depending on the peak distribution of the WTA product for the sample. Tagmented and amplified sequences were purified at a 0.6X SPRI ratio followed by a 0.9X SPRI yielding library sizes with an average distribution of 400–750 base pairs in length as determined using the Agilent hsD1000 Screen Tape System (Agilent Genomics). Samples DFCI010-4, CSF011-1, CSF014-1, CSF014-2, and CSF014-4, CSF029-2, CSF029-5, DFCI037-1, CSF046-2, and CSF073-4 were sequenced using an Illumina 75 Cycle NextSeq500/550v2 kit (Illumina 20024906) at a final concentration of 2.2–2.8 pM. Samples CSF029-4, CSF042-1, CSF043-2, CSF043-3, CSF043-4, CSF050-3, CSF050-19, and CSF091-3 were sequenced using an Illumina 100 Cycle NovaSeq6000S kit (Illumina 20027464). The read structure in both cases was paired end with read 1 starting from a custom read 1 primer containing 20 bases with a 12-bp cell barcode and 8-bp unique molecular identifier (UMI) and read 2 containing 50 bases of transcript information.

scRNA-Seq computational pipelines and analysis

Alignment & Pre-processing of scRNA-Seq data

Read alignment was performed as in Macosko et al[71]. In brief, for each Illumina sequencing run, raw sequencing data was converted to demultiplexed FASTQ files using bcl2fastq2 based on

Nextera N700 & N500 indices corresponding to individual samples/arrays. Reads were then aligned to hg19 genome using the Galaxy portal maintained by the Broad Institute for Drop-Seq alignment using standard settings. Individual reads were tagged according to the 12-bp barcode sequenced and the 8-bp UMI contained in Read 1 of each fragment. Following alignment, reads were binned onto 12-bp cell barcodes and collapsed by their 8-bp UMI with a hamming distance correction of 1. DGE matrices (genes-by-barcode) for each sample were obtained from quality filtered and mapped reads, with an automatically determined threshold for barcode count.

DGEs from each sample were individually culled and inspected by unsupervised analysis before inclusion into the full analysis by a combination previously described methods[18,71]. Each barcode was initially scored on: 1) average expression of a list of curated housekeeping genes and 2) complexity, estimated by the total number of genes detected. All sequenced samples were cut to exclude barcodes with low complexity/housekeeping gene expression (400 gene complexity cutoff, housekeeping gene expression cutoff ranged per array to explicitly exclude any outliers). Each sample was then inspected using unsupervised analysis to further identify and curate potential analyzable single cells. Individual arrays were analyzed to determine the extent of cross-cell type gene expression contamination. Minimal cross-cell type gene expression contamination existed between immune subsets, and select barcodes were filtered out according to cross expression of marker genes from other immune subsets. Tumor cells, meanwhile, exhibited some immune cell contamination and thus an immune contamination score was developed to select high-confidence barcodes. Doublets were initially assessed using DoubletFinder[22], and further restrictive analyses incorporating lowered complexity thresholds and count-based downsampling were performed to control for technical confounders wherever relevant. Following curation, all samples were combined and genes expressed in at least 1% of the remaining barcodes were retained. Consecutive samples from the same patient were combined by assigning zeros to all undetected genes per sample and concatenating columns.

104

miRNA and T cell receptor chain genes were subset and saved before cutting genes to ensure information was not lost. This curated, UMI-collapsed data was then normalized to 10,000 UMIs per barcode (tp10k) and $\log_2$-normalized before being input into Seurat[23] v2.3.4 (https://github.com/satijalab/seurat) for further analysis. This yielded a Seurat object of 14,333 single cells and 12,104 genes. The 18 individually sampled time points averaged 815 cells per sample with a range between 212 cells and 1,730 cells (**Table 4-2**).

Unsupervised transcriptomic analysis

Before performing dimensionality reduction, a list of the 2,448 most variable and highly expressed genes was generated by including genes with an average normalized and scaled expression value greater than 0.25 and with a dispersion (variance/mean) greater than 0.25. We then performed principal component analysis (PCA) over the list of variable genes. For both uniform manifold approximation and projection (UMAP) and SNN (shared nearest neighbor) clustering, we used the first 24 principal components. We used FindClusters within Seurat (which utilizes a SNN modularity optimization-based clustering algorithm) with a resolution of 0.5 and UMAP with minimum distance of 0.3 and number of neighbors of 50 to identify 17 clusters across the 18 input samples. 5 of these clusters were collapsed due to gene expression similarity evaluated by prior biological knowledge (2 extraneous divisions in cluster 0, 2 extraneous divisions in cluster 1, and 1 extraneous division in cluster 2, **Figure 4-2**) to arrive at 12 total biological clusters across all samples.

Dimensional reduction on data from the CD8+ T cells and innate immune cells alone was similarly performed using PCA followed by UMAP and SNN clustering, all implemented in Seurat. For CD8+ T cells, principal components 1-7 were used with UMAP parameters of minimum distance 0.3 and number of neighbors 20; a resolution of 0.4 was used to identify clusters. For innate immune cells, principal components 1-5 were used with UMAP parameters

of minimum distance 0.2 and number of neighbors 12; a resolution of 0.4 was used to identify clusters.

Cell type identification and within cell type analysis. To identify genes that defined each cluster, we performed differential expression using the "bimod" test implemented with the FindMarkers function in Seurat based on a likelihood ratio test designed for single-cell differential expression incorporating both a discrete and continuous component. Thresholds were set at an average log-fold difference 0.2 and adjusted p-value (Bonferroni) less than 0.05. Top marker genes with high specificity were used to classify cell clusters into cell types (**Figure 4-2**) based on literature precedent. Two closely related clusters (T/NK clusters) were merged based on biological curation and analysis of hierarchical cluster trees yielding the twelve unique clusters. To investigate further granularity present within the T/NK cell cluster, we subset these cells and re-ran dimensionality reduction and clustering (**Supplementary Figure 4-4**). For T cells, we subclustered first on treatment condition, as we found that the original clusters were dependent on this metadata. The process used for clustering and subset identification was adapted for each iteration of clustering to optimize the parameters of variable genes, principal components, and resolution of clusters desired. Following identification of canonical subsets – CD4+ T cells, CD8+ T cells, and NK cells – these identities were assigned to the main T/NK cluster cells.

NK cell clusters within the pre-treatment and post-treatment T/NK groups were annotated based on expression of *CD2* and *FCRG3A* (*CD16*), lack of expression of *CD3* genes (*CD3D*, *CD3E*, *CD3G*). γ-δ T cells were additionally distinguished from NK cells by comparison of the rate detection of variable chain genes in γ-δ T cells relative to other T cells (27.6% in γ-δ T cells, 17.5% in other CD8+ T cells).

The pDC and cDC clusters were distinguished from the other innate cells as dendritic cells, and then the differentially expressed genes between the two clusters were enriched using GSEA. The top GSEA hits on both gene lists distinguished cDCs and pDCs.

Confirmation of tumor cell malignancy with WES of CSF cfDNA. The mean tumor cell purity[72] measured in the CSF of patients in our study was significantly higher than that of plasma samples ($p<2.3\times10^{-9}$, Mann-Whitney U Test; effect size = 1.55), enabling accurate CNV comparison (**Figure 4-2**).

Malignant cells were distinguished from non-malignant ones by two approaches, similar to previous work[21,26,27,66]:

i) unsupervised clustering on gene expression, with subsequent identification of clusters consistent with breast cancer expressional programs.

ii) the use of inferred copy number variations (CNVs).

Single-cell transcriptomes captured by Seq-Well and other high-throughput methods typically have lower complexity than the SMART-Seq/SMART-Seq2 protocols used in previous, similar work[17]. This, in principle, limits the accuracy of CNV profiles inferred directly from scRNA-Seq gene expression. We measure the concordance of a cell's scRNA-Seq-inferred CNV profile with the patient-matched (and, when possible, time-point matched) 50x coverage whole exome, or, in the case of sample DFCI-010-04, 60x whole genome CNV profile, taking the mean of log(TP10k+1) expressions for each chromosome (the Mean Chromosomal Expression, hereafter MCE). To evaluate the concordance between the single-cell MCE profiles and patient-matched cfDNA obtained from the cerebrospinal fluid (CSF), we took the total copy ratio (tCR) profiles calculated from the WES (or WGS), similar to work done in Brastianos et al[64]. For each

chromosome, the mean tCR was calculated over all contained genes (the Mean Chromosomal tCR, hereafter MCtCR).

On an individual cell basis, we robustly regressed (using a Theil-Sen median of slopes regression) the MCtCR with the MCE to obtain a relation (slope) between these two quantities. Plotting distributions of these slopes for all cells revealed two clean separated peaks – one positive, containing cells previously identified via expressional profiles as having breast cell/cancer cell programs, and one with slopes close to zero, consisting of cells previously identified as being immune associated (in P011, these are likely non-malignant stromal cells).

We note that while the MCE does not capture CNV structure at sub-chromosomal resolution, this maximal grouping of genes is sufficient for distinguishing malignant from non-malignant cells, even at low complexities. We also note that the patients' MCE profiles were highly individual — matching a patient's MCE with the MCtCR from another patient's exome will not successfully distinguish malignant from non-malignant cells.

Differential expression and scoring over gene sets. To identify differentially expressed genes within cell types and subtypes across treated and untreated conditions, we again used the 'bimod' setting in FindMarkers implemented in Seurat. To determine the scores of gene sets and pathways, such as interferon response and antigen processing, we used the 'AddModuleScore' function in Seurat to construct a mean score of supplied genes subtracting a background score constructed from a random selection of genes in bins of average expression across all cells. When comparing scores within a specific subset of cells, AddModuleScore was constructed only over that subset, and recalculated if the subset was further partitioned. Tumor cell scores were calculated both across all patients (to compare pre-treatment and post-treatment time points across patients) and within individual patients (to compare across time points within patients). For specific comparisons of AddModuleScore-derived signatures with

large differences in complexity between groups of cells, an upper complexity threshold and count-based downsampling were used to examine the possibility of complexity-confounded effects. No such effects were observed in comparing between tumor cells across patient and within patient (P029, P043).

Inferred CNV analysis of Patient 043. To more accurately infer CNV patterns in high-complexity tumor cells with sub-chromosomal resolution, we group genes into fixed length windows of 200 genes consecutive along the genome, removing from consideration those genes in the uppermost decile of dropout rate, as well as all immunoglobulin genes. All possible windows were considered where all included genes reside on the same chromosome. We converted the log (TP10k+1) gene expression profiles to TP10k ones. We then took the mean TP10k expression over genes in a window, neglecting the highest 5 gene expressions in that window. This vector of values is hereafter referred to as the unnormalized Windowed Mean Expression (uWME).

Having identified the malignant cells for each patient, we additionally computed a normalized version of the uWME as follows: the uWME from all patients' non-malignant cells were averaged for each window across patients. HLA-* and associated genes on the 6p arm exhibited particularly strong hematopoietic expression; therefore the means of these windows were imputed with the mean (windows) of the mean (patients) WME for all other windows. These values we refer to as the mean non-malignant uWME.

We normalize uWME for malignant cells by dividing the window uWME by the mean non-malignant uWME for each window, hereafter referred to as Windowed Mean Expression (WME).

To reduce possible confounding factors due to experimental or batch effects during subsequent clustering analysis, we converted the WME values in each single cell to ranks, hereafter

referred to as the ranked, normalized WME (rWME). PCA and UMAP were performed on the rWME using the first 50 principal components of all tumor cells. Evidence of consistent structured heterogeneity was detected in only P043 (**Figure 4-11**), with changes in individual chromosomes in P014 and P050 not confidently ascribed to potential subclonality. In this P043, the UMAP-obtained clustering was concordant with that achieved via agglomerative clustering. To perform this clustering, we used as a distance metric $1\text{-}\tau_K$, where τ_K is the Kendall's tau coefficient between the WME of all pairs of cells. Agglomerative clustering was performed with a complete linkage.

To support the subclonal hypothesis without relying on either inferred CNV profiles or unsupervised clustering thereof, we performed a supervised comparison of single-cell expression profiles at each time point to both the early and late cfDNA-derived WES copy ratios (**Figure 4-12**. We calculate the Kendall-Tau correlation for all genes' total copy ratio and single-cell expression, for all single cells. Then we calculate the difference in correlation for all single cells when using total copy ratio from time point 4 (late) vs. time point 2 (early). We observe that CSF043-2 single cells exhibit correlations more similar to WES from time point 2, and that CSF043-4 single cells exhibit correlations more similar to WES from time point 4. At CSF043-3, we observe bimodality in the distribution of the difference of Kendall-Tau correlations. Additionally, we observe that single cells derived from post-treatment time points (CSF043-3 and CSF043-4) exhibit anti-correlation between their interferon-gamma response score and the difference in Kendall-Tau correlations between total copy ratios derived from WES at time point 4 vs. time point 2.
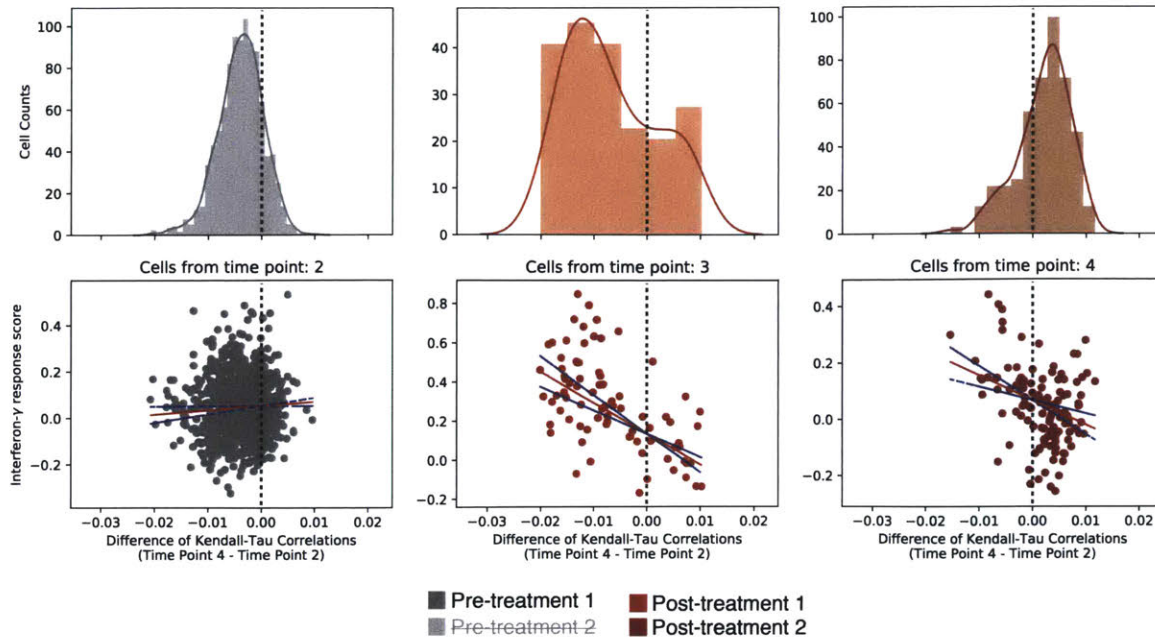
110

Figure 4-12: Density and scatter plots of difference in Kendall-Tau correlations between WES and scRNA-Seq for P043 tumor cells. Lower plots: Difference between Kendall-Tau correlations of gene expressions (scRNA-Seq) and gene copy ratios (WES) at time points 4 and 2 plotted against interferon-gamma response score for each single cell (each point in scatter plot corresponds to one tumor cell). Upper plots: Histogram of difference between Kendall-Tau correlations of gene expressions (scRNA-Seq) and gene copy ratios (WES) at time points 4 and 2. Theil-Sen median slopes plotted in each scatter plot (red line) with 10th and 90th percentile of median slopes (blue, dotted lines).

We note that the relative populations of the two identified clusters in P043 varied significantly across time (**Figure 4-10**). We plotted, for each gene, the mean purity corrected tCR vs change in the WME between all possible pairs of time points. The purity corrected tCR has the following form:

$$tCR_{corrected} = \frac{tCR_{observed} - (1-p) \times tCR_{germline}}{p} = \frac{tCR_{observed} - (1-p)}{p}$$

where p is purity of sample calculated by ABSOLUTE and $tCR_{germline} = 1$[72].

This relationship shows that the windowed expressional change between these clusters is concordant with the change in WES-derived tCR between any two time points. This

concordance is robust to considering only the cells obtained at time point 3 (i.e., the correlated changes in single-cell expression and cfDNA-derived CNV profile cannot be attributed to batch effects confounding the observed scRNA-seq profiles).

Windowed-mean expression results were compared to the InferCNV R package from the Broad Institute, and broad amplifications and deletions were concordant between the two approaches. We note that InferCNV employs a de-noising strategy, which largely reduces intra-sample substructure.

Statistical analyses. Statistical analyses of differential expression were performed using Seurat v2.3.4 implemented in RStudio. All statistical tests of distributions, cluster diversity, and change in representation, etc. were performed using base R packages implemented in RStudio. All statistical tests of gene set enrichment were performed using piano v1.22.0 and implemented in RStudio for all except enrichments of cluster markers for the full dataset, which was implemented in R. All violin plots and boxplots were generated using ggplot2 without modifications to smoothing or density. Boxplot rectangles encompass the 25th to 75th percentile with outliers as individual points above and below the rectangle. All genes comprising signatures tested above and elsewhere are provided in **Table 4-1**. Overlapping genes between IFN response and antigen processing signatures were removed from both before utilization.

As scores followed non-normal distributions as tested for using a Lilliefors normality test, we used a Mann–Whitney U-test where indicated for determining statistical significance. For scores in single-cell data, we report effect sizes in addition to statistical significance as an additional metric to capture the magnitude of the effect observed. The calculation was performed as Cohen's d where: effect size $d = (Mean_1 - Mean_2)/(s.d. pooled)$. All p-values subject to the multiple comparisons problem (such as marker identification by differential expression) were adjusted by Bonferroni correction.

Table 4-1

Jiang et al	p-value	effect size
CD8.c1 vs. CD8.c2	9.00E-15	0.51
CD8.c1 vs. CD8.c3	1.75E-09	0.48
CD8.c2 vs. CD8.c3	0.92	0.05

Guo et al	p-value	effect size
CD8.c1 vs. CD8.c2	2.20E-16	1
CD8.c1 vs. CD8.c3	1.25E-10	0.52
CD8.c2 vs. CD8.c3	9.16E-10	0.54

Zheng et al	p-value	effect size
CD8.c1 vs. CD8.c2	2.20E-16	0.8
CD8.c1 vs. CD8.c3	1.20E-07	0.45
CD8.c2 vs. CD8.c3	1.20E-06	0.41

Sade-Feldman et al - CD8_B	p-value	effect size
CD8.c1 vs. CD8.c2	2.20E-16	0.77
CD8.c1 vs. CD8.c3	2.14E-03	0.29
CD8.c2 vs. CD8.c3	5.32E-09	0.5

Sade-Feldman et al - CD8_3	p-value	effect size
CD8.c1 vs. CD8.c2	2.20E-16	1.26
CD8.c1 vs. CD8.c3	2.20E-16	1.08
CD8.c2 vs. CD8.c3	2.90E-08	0.47

Sade-Feldman et al - CD8_G	p-value	effect size
CD8.c1 vs. CD8.c2	2.20E-16	0.63
CD8.c1 vs. CD8.c3	8.80E-07	0.38
CD8.c2 vs. CD8.c3	2.20E-16	0.97

Sade-Feldman et al - CD8_4	p-value	effect size
CD8.c1 vs. CD8.c2	2.20E-16	0.74
CD8.c1 vs. CD8.c3	3.60E-06	0.38
CD8.c2 vs. CD8.c3	2.20E-16	1.02

Sade-Feldman et al - CD8_6	p-value	effect size
CD8.c1 vs. CD8.c2	2.20E-16	0.9
CD8.c1 vs. CD8.c3	7.50E-01	0.03
CD8.c2 vs. CD8.c3	2.20E-16	0.92

Miller et al - Progenitor

	p-value	effect size
CD8.c1 vs. CD8.c2	2.20E-16	0.81
CD8.c1 vs. CD8.c3	4.20E-01	0.03
CD8.c2 vs. CD8.c3	2.20E-16	0.8

Miller et al - Terminal

	p-value	effect size
CD8.c1 vs. CD8.c2	2.20E-16	1.05
CD8.c1 vs. CD8.c3	2.20E-16	0.8
CD8.c2 vs. CD8.c3	2.00E-04	0.33

IFN response

	p-value	effect size
monocytes	2.20E-16	1.2
macrophages	2.20E-16	0.73
cDCs	2.20E-16	0.54
pDCs	2.00E-09	0.3

Antigen processing

	p-value	effect size
monocytes	2.20E-16	0.93
macrophages	6.30E-15	0.63
cDCs	2.20E-16	0.51
pDCs	9.00E-03	0.59

M1

	p-value	effect size
monocytes	2.20E-16	1.09
macrophages	2.60E-15	0.83

M2

	p-value	effect size
monocytes	7.80E-01	0.01
macrophages	2.00E-02	0.11

IFN response

	p-value	effect size
per-patient - 91 included	1.70E-01	0.85
per-patient - 91 excluded	3.00E-02	1.36
composite - 91 included	2.20E-16	0.5
composite - 91 excluded	2.20E-16	0.35

Antigen processing

	p-value	effect size
per-patient - 91 included	7.00E-02	1.21
per-patient - 91 excluded	1.00E-01	1.34
composite - 91 included	2.20E-16	0.54
composite - 91 excluded	2.20E-16	0.63

IFN response

	p-value	effect size
P014-1 vs. P014-2	1.30E-02	0.22
P029-2 vs. P029-4	2.70E-06	0.36
P029-4 vs. P029-5	4.50E-16	-0.3
P043-2 vs. P043-3	2.20E-16	1.51
P043-3 vs. P043-4	8.36E-14	-1.1

Antigen processing

	p-value	effect size
P014-1 vs. P014-2	7.80E-02	0.22
P029-2 vs. P029-4	2.20E-16	0.36
P029-4 vs. P029-5	2.20E-16	-0.38
P043-2 vs. P043-3	2.20E-16	1.61
P043-3 vs. P043-4	1.10E-11	-1.15

IFN response

	p-value	effect size
P043-2: asc vs. des	3.50E-01	0.33
P043-3: asc vs. des	4.00E-06	-1.33
P043-4: asc vs. des	1.90E-03	-1.17

#	Patient ID	Brastianos ID	Gender (M/F)	Age	Ethnicity	Molecular subtype	Pre-Treatment Steroids	On-Treatment Steroid
1	P010	DFCI-010	F	55	White	ER+/PR+/Her2-	Yes	Yes
2	P011	CSF-011?	F	48	White	ER-/PR+/HER2+	No	Yes
3	P014	CSF-014	F	52	White	BRCA1+, ER-/PR-/HER2-	Yes	Yes
4	P029	CSF-029	F	36	White	BRCA1+, ER-/PR-/HER2-	Yes	Yes
5	P037	DFCI-037	F	50	African American	ER+/PRlow/HER2+	No	No
6	P042	CSF-042	F	51	White	ER+/PR+/HER2-	Yes	Yes
7	P043	CSF-043	F	45	White	ER+/PR+/Her2-	Yes	Yes
8	P046	CSF-046	F	44	Unavailable	ER-/PR-/HER2-	Yes	Yes
9	P050	CSF-050	F	52	White	ER+/PR+/HER2-	Yes	No
10	P073	CSF-073	F	51	White	ER+/PR+/HER2-	No	No
11	P091	CSF-091	F	40	White	BRCA2+, ER+/PR+/HER2-	No	Yes

Table 4-2

Radiation	Pre-Treatment Imaging	Number of Pembrolizumab Doses	Biopsy I (days from treatment initiation)	Biopsy II (days from treatment initiation)
N/A	Leptomeningeal enhancement; Systemic disease	4	42	N/A
WBRT	Brain metastases, Leptomeningeal Enhancement; Systemic disease	3	-16	N/A
WBRT, SRS	Brain metastases, Leptomeningeal Enhancement; Systemic disease	2	-13	0
WBRT	Brain metastases, Leptomeningeal Enhancement; No systemic disease	3	-7	20
WBRT, SRS	Brain metastases, Leptomeningeal Enhancment; Systemic disease	2	0	N/A
WBRT, SRS	Brain metastases, Leptomeningeal Enhancement; No systemic disease	8	-105	N/A
N/A	Leptomeningeal enhancement; Systemic disease	5	0	21
SRS	Brain metastases, Leptomeningeal Enhancement; Systemic disease	3	-157	N/A
WBRT, SRS	Leptomeningeal enhancement; Systemic disease	17	21	357
WBRT	Leptomeningeal enhancement; Systemic disease	5	21	N/A
WBRT	Brain metastases, Leptomeningeal enhancement; Systemic disease	3	45	N/A

Biopsy III (days from treatment initiation)	Progression	Overall Survival (months)
N/A	CNS	2.2
N/A	CNS	8.7
33	CNS and systemic	2.6
39	CNS	2.2
N/A	CNS	3.7
N/A	CNS	15.8
41	CNS and systemic	3.5
N/A	Systemic	2.1
N/A	CNS and systemic	14.8
N/A	Systemic	3.3
N/A	CNS	5.2

4.10: References

1   Hodi, F. S. *et al.* Improved survival with ipilimumab in patients with metastatic melanoma. *N Engl J Med* **363**, 711-723, doi:10.1056/NEJMoa1003466 (2010).

2   Pardoll, D. M. The blockade of immune checkpoints in cancer immunotherapy. *Nat Rev Cancer* **12**, 252-264, doi:10.1038/nrc3239 (2012).

3   Sharma, P., Wagner, K., Wolchok, J. D. & Allison, J. P. Novel cancer immunotherapy agents with survival benefit: recent successes and next steps. *Nat Rev Cancer* **11**, 805-812, doi:10.1038/nrc3153 (2011).

4   Kalos, M. *et al.* T cells with chimeric antigen receptors have potent antitumor effects and can establish memory in patients with advanced leukemia. *Sci Transl Med* **3**, 95ra73, doi:10.1126/scitranslmed.3002842 (2011).

5   Callahan, M. K., Postow, M. A. & Wolchok, J. D. Targeting T Cell Co-receptors for Cancer Therapy. *Immunity* **44**, 1069-1078, doi:10.1016/j.immuni.2016.04.023 (2016).

6   Zou, W., Wolchok, J. D. & Chen, L. PD-L1 (B7-H1) and PD-1 pathway blockade for cancer therapy: Mechanisms, response biomarkers, and combinations. *Sci Transl Med* **8**, 328rv324, doi:10.1126/scitranslmed.aad7118 (2016).

7   Brahmer, J. R. *et al.* Safety and activity of anti-PD-L1 antibody in patients with advanced cancer. *N Engl J Med* **366**, 2455-2465, doi:10.1056/NEJMoa1200694 (2012).

8   Pitt, J. M. *et al.* Resistance Mechanisms to Immune-Checkpoint Blockade in Cancer: Tumor-Intrinsic and -Extrinsic Factors. *Immunity* **44**, 1255-1269, doi:10.1016/j.immuni.2016.06.001 (2016).

9   Postow, M. A., Sidlow, R. & Hellmann, M. D. Immune-Related Adverse Events Associated with Immune Checkpoint Blockade. *N Engl J Med* **378**, 158-168, doi:10.1056/NEJMra1703481 (2018).

10   Sharma, P., Hu-Lieskovan, S., Wargo, J. A. & Ribas, A. Primary, Adaptive, and Acquired Resistance to Cancer Immunotherapy. *Cell* **168**, 707-723, doi:10.1016/j.cell.2017.01.017 (2017).

11   Joyce, J. A. & Fearon, D. T. T cell exclusion, immune privilege, and the tumor microenvironment. *Science* **348**, 74-80, doi:10.1126/science.aaa6204 (2015).

12   Jiang, P. *et al.* Signatures of T cell dysfunction and exclusion predict cancer immunotherapy response. *Nat Med* **24**, 1550-1558, doi:10.1038/s41591-018-0136-1 (2018).

13   Heitzer, E., Haque, I. S., Roberts, C. E. S. & Speicher, M. R. Current and future perspectives of liquid biopsies in genomics-driven oncology. *Nat Rev Genet* **20**, 71-88, doi:10.1038/s41576-018-0071-5 (2019).

14   Grossman, S. A. & Krabak, M. J. Leptomeningeal carcinomatosis. *Cancer Treat Rev* **25**, 103-119, doi:10.1053/ctrv.1999.0119 (1999).

15   Morikawa, A. *et al.* Characteristics and Outcomes of Patients With Breast Cancer With Leptomeningeal Metastasis. *Clin Breast Cancer* **17**, 23-28, doi:10.1016/j.clbc.2016.07.002 (2017).

16   Scott, B. J. & Kesari, S. Leptomeningeal metastases in breast cancer. *Am J Cancer Res* **3**, 117-126 (2013).

17   Gierahn, T. M. *et al.* Seq-Well: portable, low-cost RNA sequencing of single cells at high throughput. *Nat Methods* **14**, 395-398, doi:10.1038/nmeth.4179 (2017).

18   Ordovas-Montanes, J. *et al.* Allergic inflammatory memory in human respiratory epithelial progenitor cells. *Nature* **560**, 649-654, doi:10.1038/s41586-018-0449-8 (2018).

19   Boire, A. *et al.* Complement Component 3 Adapts the Cerebrospinal Fluid for Leptomeningeal Metastasis. *Cell* **168**, 1101-1113 e1113, doi:10.1016/j.cell.2017.02.025 (2017).

20   Brastianos, P. K. *et al.* Phase II study of pembrolizumab in leptomeningeal carcinomatosis. *Journal of Clinical Oncology* **36**, 2007-2007, doi:10.1200/JCO.2018.36.15_suppl.2007 (2018).

21   Tirosh, I. *et al.* Dissecting the multicellular ecosystem of metastatic melanoma by single-cell RNA-seq. *Science* **352**, 189-196, doi:10.1126/science.aad0501 (2016).

22   McGinnis, C. S., Murrow, L. M. & Gartner, Z. J. DoubletFinder: Doublet Detection in Single-Cell RNA Sequencing Data Using Artificial Nearest Neighbors. *Cell Syst* **8**, 329-337 e324, doi:10.1016/j.cels.2019.03.003 (2019).

23   Satija, R., Farrell, J. A., Gennert, D., Schier, A. F. & Regev, A. Spatial reconstruction of single-cell gene expression data. *Nat Biotechnol* **33**, 495-502, doi:10.1038/nbt.3192 (2015).

24   Becht, E. *et al.* Dimensionality reduction for visualizing single-cell data using UMAP. *Nat Biotechnol*, doi:10.1038/nbt.4314 (2018).

25   McInnes, L., Healy, J. & Melville, J. UMAP: Uniform Manifold Approximation and Projection for Dimension Reduction. *arXiv*, doi:arXiv:1802.03426v2 (2018).

26   Puram, S. V. *et al.* Single-Cell Transcriptomic Analysis of Primary and Metastatic Tumor Ecosystems in Head and Neck Cancer. *Cell* **171**, 1611-1624 e1624, doi:10.1016/j.cell.2017.10.044 (2017).

27   Patel, A. P. *et al.* Single-cell RNA-seq highlights intratumoral heterogeneity in primary glioblastoma. *Science* **344**, 1396-1401, doi:10.1126/science.1254257 (2014).

28   Chung, W. *et al.* Single-cell RNA-seq enables comprehensive tumour and immune cell profiling in primary breast cancer. *Nat Commun* **8**, 15081, doi:10.1038/ncomms15081 (2017).

29   Ziegenhain, C. *et al.* Comparative Analysis of Single-Cell RNA Sequencing Methods. *Mol Cell* **65**, 631-643 e634, doi:10.1016/j.molcel.2017.01.023 (2017).

30   Gutierrez-Arcelus, M. *et al.* Lymphocyte innateness defined by transcriptional states reflects a balance between proliferation and effector functions. *Nat Commun* **10**, 687, doi:10.1038/s41467-019-08604-4 (2019).

31   Ribas, A. *et al.* PD-1 Blockade Expands Intratumoral Memory T Cells. *Cancer Immunol Res* **4**, 194-203, doi:10.1158/2326-6066.CIR-15-0210 (2016).

32   Im, S. J. *et al.* Defining CD8+ T cells that provide the proliferative burst after PD-1 therapy. *Nature* **537**, 417-421, doi:10.1038/nature19330 (2016).

33   Kamphorst, A. O. *et al.* Proliferation of PD-1+ CD8 T cells in peripheral blood after PD-1-targeted therapy in lung cancer patients. *Proc Natl Acad Sci U S A* **114**, 4993-4998, doi:10.1073/pnas.1705327114 (2017).

34   Benita, Y. *et al.* Gene enrichment profiles reveal T-cell development, differentiation, and lineage-specific transcription factors including ZBTB25 as a novel NF-AT repressor. *Blood* **115**, 5376-5384, doi:10.1182/blood-2010-01-263855 (2010).

35   Nirmal, A. J. *et al.* Immune Cell Gene Signatures for Profiling the Microenvironment of Solid Tumors. *Cancer Immunol Res* **6**, 1388-1400, doi:10.1158/2326-6066.CIR-18-0342 (2018).

36   Miller, B. C. *et al.* Subsets of exhausted CD8(+) T cells differentially mediate tumor control and respond to checkpoint blockade. *Nat Immunol* **20**, 326-336, doi:10.1038/s41590-019-0312-6 (2019).

37   Huang, A. C. *et al.* T-cell invigoration to tumour burden ratio associated with anti-PD-1 response. *Nature* **545**, 60-65, doi:10.1038/nature22079 (2017).

38   Sade-Feldman, M. *et al.* Defining T Cell States Associated with Response to Checkpoint Immunotherapy in Melanoma. *Cell* **175**, 998-1013 e1020, doi:10.1016/j.cell.2018.10.038 (2018).

39   Guo, X. *et al.* Global characterization of T cells in non-small-cell lung cancer by single-cell sequencing. *Nat Med* **24**, 978-985, doi:10.1038/s41591-018-0045-3 (2018).

40   Zheng, C. *et al.* Landscape of Infiltrating T Cells in Liver Cancer Revealed by Single-Cell Sequencing. *Cell* **169**, 1342-1356 e1316, doi:10.1016/j.cell.2017.05.035 (2017).

41   Siddiqui, I. *et al.* Intratumoral Tcf1(+)PD-1(+)CD8(+) T Cells with Stem-like Properties Promote Tumor Control in Response to Vaccination and Checkpoint Blockade Immunotherapy. *Immunity* **50**, 195-211 e110, doi:10.1016/j.immuni.2018.12.021 (2019).

42   Wherry, E. J. & Kurachi, M. Molecular and cellular insights into T cell exhaustion. *Nat Rev Immunol* **15**, 486-499, doi:10.1038/nri3862 (2015).

43   Blackburn, S. D., Shin, H., Freeman, G. J. & Wherry, E. J. Selective expansion of a subset of exhausted CD8 T cells by alphaPD-L1 blockade. *Proc Natl Acad Sci U S A* **105**, 15016-15021, doi:10.1073/pnas.0801497105 (2008).

44   Huang, A. C. *et al.* A single dose of neoadjuvant PD-1 blockade predicts clinical outcomes in resectable melanoma. *Nat Med* **25**, 454-461, doi:10.1038/s41591-019-0357-y (2019).

45   Wu, T. *et al.* The TCF1-Bcl6 axis counteracts type I interferon to repress exhaustion and maintain T cell stemness. *Sci Immunol* **1**, doi:10.1126/sciimmunol.aai8593 (2016).

46   Villani, A. C. *et al.* Single-cell RNA-seq reveals new types of human blood dendritic cells, monocytes, and progenitors. *Science* **356**, doi:10.1126/science.aah4573 (2017).

47   Swiecki, M. & Colonna, M. The multifaceted biology of plasmacytoid dendritic cells. *Nat Rev Immunol* **15**, 471-485, doi:10.1038/nri3865 (2015).

48   Munn, D. H. & Mellor, A. L. IDO in the Tumor Microenvironment: Inflammation, Counter-Regulation, and Tolerance. *Trends Immunol* **37**, 193-207, doi:10.1016/j.it.2016.01.002 (2016).

49   Beyer, M. *et al.* High-resolution transcriptome of human macrophages. *PLoS One* **7**, e45466, doi:10.1371/journal.pone.0045466 (2012).

50   Xue, J. *et al.* Transcriptome-based network analysis reveals a spectrum model of human macrophage activation. *Immunity* **40**, 274-288, doi:10.1016/j.immuni.2014.01.006 (2014).

51   Ruytinx, P., Proost, P., Van Damme, J. & Struyf, S. Chemokine-Induced Macrophage Polarization in Inflammatory Conditions. *Front Immunol* **9**, 1930, doi:10.3389/fimmu.2018.01930 (2018).

52   Biswas, S. K. & Mantovani, A. Macrophage plasticity and interaction with lymphocyte subsets: cancer as a paradigm. *Nat Immunol* **11**, 889-896, doi:10.1038/ni.1937 (2010).

53   Cassetta, L. *et al.* Human Tumor-Associated Macrophage and Monocyte Transcriptional Landscapes Reveal Cancer-Specific Reprogramming, Biomarkers, and Therapeutic Targets. *Cancer Cell* **35**, 588-602 e510, doi:10.1016/j.ccell.2019.02.009 (2019).

54   Dhupkar, P., Gordon, N., Stewart, J. & Kleinerman, E. S. Anti-PD-1 therapy redirects macrophages from an M2 to an M1 phenotype inducing regression of OS lung metastases. *Cancer Med* **7**, 2654-2664, doi:10.1002/cam4.1518 (2018).

55   Hartley, G. P., Chow, L., Ammons, D. T., Wheat, W. H. & Dow, S. W. Programmed Cell Death Ligand 1 (PD-L1) Signaling Regulates Macrophage Proliferation and Activation. *Cancer Immunol Res* **6**, 1260-1273, doi:10.1158/2326-6066.CIR-17-0537 (2018).

56   Arlauckas, S. P. *et al.* In vivo imaging reveals a tumor-associated macrophage-mediated resistance pathway in anti-PD-1 therapy. *Sci Transl Med* **9**, doi:10.1126/scitranslmed.aal3604 (2017).

57   Rodell, C. B. *et al.* TLR7/8-agonist-loaded nanoparticles promote the polarization of tumour-associated macrophages to enhance cancer immunotherapy. *Nat Biomed Eng* **2**, 578-588 (2018).

58   Rodig, S. J. *et al.* MHC proteins confer differential sensitivity to CTLA-4 and PD-1 blockade in untreated metastatic melanoma. *Sci Transl Med* **10**, doi:10.1126/scitranslmed.aar3342 (2018).

59   Keenan, T. E., Burke, K. P. & Van Allen, E. M. Genomic correlates of response to immune checkpoint blockade. *Nat Med* **25**, 389-402, doi:10.1038/s41591-019-0382-x (2019).

60   Zaretsky, J. M. *et al.* Mutations Associated with Acquired Resistance to PD-1 Blockade in Melanoma. *N Engl J Med* **375**, 819-829, doi:10.1056/NEJMoa1604958 (2016).

61   Kim, C. *et al.* Chemoresistance Evolution in Triple-Negative Breast Cancer Delineated by Single-Cell Sequencing. *Cell* **173**, 879-893 e813, doi:10.1016/j.cell.2018.03.041 (2018).

62   Landau, D. A. *et al.* Evolution and impact of subclonal mutations in chronic lymphocytic leukemia. *Cell* **152**, 714-726, doi:10.1016/j.cell.2013.01.019 (2013).

63   Stachler, M. D. *et al.* Paired exome analysis of Barrett's esophagus and adenocarcinoma. *Nat Genet* **47**, 1047-1055, doi:10.1038/ng.3343 (2015).

64   Brastianos, P. K. *et al.* Genomic Characterization of Brain Metastases Reveals Branched Evolution and Potential Therapeutic Targets. *Cancer Discov* **5**, 1164-1177, doi:10.1158/2159-8290.CD-15-0369 (2015).

65   Campbell, P. J. *et al.* The patterns and dynamics of genomic instability in metastatic pancreatic cancer. *Nature* **467**, 1109-1113, doi:10.1038/nature09460 (2010).

66   Tirosh, I. *et al.* Single-cell RNA-seq supports a developmental hierarchy in human oligodendroglioma. *Nature* **539**, 309-313, doi:10.1038/nature20123 (2016).

67   Riaz, N. *et al.* Tumor and Microenvironment Evolution during Immunotherapy with Nivolumab. *Cell* **171**, 934-949 e916, doi:10.1016/j.cell.2017.09.028 (2017).

68   Krieg, C. *et al.* High-dimensional single-cell analysis predicts response to anti-PD-1 immunotherapy. *Nat Med* **24**, 144-153, doi:10.1038/nm.4466 (2018).

69   He, R. *et al.* Follicular CXCR5- expressing CD8(+) T cells curtail chronic viral infection. *Nature* **537**, 412-428, doi:10.1038/nature19317 (2016).

70   Li, A. *et al.* Indoleamine 2,3-dioxygenase 1 inhibition targets anti-PD1-resistant lung tumors by blocking myeloid-derived suppressor cells. *Cancer Lett* **431**, 54-63, doi:10.1016/j.canlet.2018.05.005 (2018).

71   Macosko, E. Z. *et al.* Highly Parallel Genome-wide Expression Profiling of Individual Cells Using Nanoliter Droplets. *Cell* **161**, 1202-1214, doi:10.1016/j.cell.2015.05.002 (2015).

72   Carter, S. L. *et al.* Absolute quantification of somatic DNA alterations in human cancer. *Nat Biotechnol* **30**, 413-421, doi:10.1038/nbt.2203 (2012).

Chapter 5: Longitudinal analysis of circulating tumor cell response to whole mouse drug treatment

Hamza B*, Ng SR*, Prakadan SM*, Delgado FF, Chin CR, King EM, Yang LF, Davidson SM, DeGouveia KL, Cermak N, Navia AW, Winter PS, Drake RS, Tammela T, Li CM, Papagian-nakopoulos T, Gupta AJ, Bagnall JS, Knudsen SM, Vander Heiden MG, Wasserman SC, Jacks T[†], Shalek AK[†], and Manalis SR[†]. "Optofluidic real-time cell sorter for longitudinal CTC studies in mouse models of cancer," *Proc. Natl. Acad. Sci. U.S.A.* **116** (6): 2232-2236 (2019).

[*,†] Denote equal contribution

Circulating tumor cells (CTCs) play a fundamental role in cancer progression. However, in mice, limited blood volume and the rarity of CTCs in the bloodstream preclude longitudinal, in-depth studies of these cells using existing liquid biopsy techniques. Here, we present an optofluidic system that continuously collects fluorescently-labeled CTCs from a genetically-engineered mouse model (GEMM) for several hours per day over multiple days or weeks. The system is based on a microfluidic cell-sorting chip connected serially to an un-anesthetized mouse via an implanted arteriovenous shunt. Pneumatically-controlled microfluidic valves capture CTCs as they flow through the device and CTC-depleted blood is returned back to the mouse via the shunt. To demonstrate the utility of our system, we profile CTCs isolated longitudinally from animals over a four-day treatment with the BET inhibitor JQ1 using single-cell RNA-Seq (scRNA-Seq) and show that our approach eliminates potential biases driven by inter-mouse heterogeneity that can occur when CTCs are collected across different mice. The CTC isolation and sorting technology presented here provides a research tool to help reveal details of how

CTCs change over time, allowing studies to credential changes in CTCs as biomarkers of drug response and facilitating future studies to understand the role of CTCs in metastasis.

Key Contributions

BH led the design and manufacturing of the optofluidic device as well as mouse CTC collection experiments. SRN led efforts to generate mice, perform mouse viral infection and IVIS imaging as well as processing of terminal mouse blood, primary and metastatic tumor tissue, and immunofluorescence staining. SMP led the single-cell RNA sequencing efforts and performed statistical analysis on the RNA sequencing data.

Forward

While studies of patient-derived samples are powerful, they can be difficult to implement and their results difficult to interpret, due to unavailability of patient samples, incomplete patient histories, variable treatment regimens, and lack of control groups. Model organisms enable more organized experimental control and design, and murine models are particularly effective for enriching and studying rare cell types and phenomena, such as circulating tumor cells (CTCs) in blood, which are typically difficult to detect, capture, enrich, and profile. CTCs could serve an important purpose as peripheral proxy cells for the behavior of primary tumors, but this relationship is still under investigation. Furthermore, the volume of blood required to capture sufficient CTCs over the acute window of treatment cannot be collected while maintaining the physiology of a mouse. Here, we develop and implement a low-input platform to study CTCs from a genetically engineered mouse model (GEMM) that was specifically designed to release CTCs for detection/capture. We investigate phenotypic CTC heterogeneity across genetically

identical mice, study among the first responses of CTCs to treatment acute time points, and compare their responses to matched primary tumors.

5.1: Introduction

Circulating tumor cells (CTCs) are an intermediate in the hematogenous spread of tumors during metastasis[1]. Given their accessibility and potential prognostic and diagnostic value, CTCs have been the focus of significant clinical research efforts monitoring response to therapy and predicting risk of relapse[2-5]. Over the last decade, novel microfluidic liquid biopsy-based techniques as well as *in vivo*, vein-catheter-based methods have been developed to detect and collect CTCs directly from the blood of human patients[2-4,6-10]. Combined with recently developed single-cell profiling methods, such as single-cell RNA-Seq[11-14] (scRNA-Seq), in-depth examination of CTCs is now possible. Such studies can provide new insights into the genomic properties of CTCs, as well as their relationship to matched primary and metastatic tumors[3,4,15-18].

Genetically-engineered mouse models (GEMMs) of cancer, which mimic the natural multistage evolution of their human counterparts, facilitate characterization of both acute perturbations (e.g. drug treatment) and long-term phenotypic changes (e.g. tumor evolution) not possible in human subjects. However, despite the usefulness of GEMMs in cancer research, the combination of the small total murine blood volume (~1.5 mL) and the rarity of CTCs in circulating blood (fewer than 100 cells per mL)[3,19] precludes the use of existing liquid biopsy techniques for longitudinal CTC studies in mice. When repeated blood samples are required at short intervals, a maximum of 1.0% of an animal's total blood volume can be removed every 24 hours (~16.5μL for a 25g mouse)[20] – a miniscule volume that does not yield a sufficient sample of CTCs for analysis.

GEMMs have been developed that combine genetic perturbations (manipulation of oncogenes and tumor suppressor genes) with genetically-encoded fluorescent markers, enabling the unbiased detection and isolation of CTCs from the bloodstream. *In vivo* flow cytometry (IVFC) techniques have been used to enumerate CTCs in ear capillaries or tail veins longitudinally without sacrificing animals[21-23], but these techniques do not permit isolation and downstream molecular characterization of CTCs. On the other hand, *in vivo*, vein-catheter-based techniques allow for the direct capture and isolation of CTCs from much larger blood volumes in real time[6,10]. However, these techniques are currently limited to detecting EpCAM-expressing CTCs, which may result in only a subpopulation of CTCs being detected and isolated[24].

Results

5.2: Optofluidic Platform Design and Characterization

To enable longitudinal, in-depth studies of CTC biology in GEMMs and other murine cancer models, we have developed an optofluidic system capable of detecting and capturing fluorescent CTCs in living mice over several hours, days, or weeks. Key components of the system include a polydimethylsiloxane (PDMS)-based microfluidic CTC sorter chip, a fluorescence detector, and computer-controlled pneumatic valves (**Figure 5-1 a-b**). A cannulated mouse with two permanent catheters easily accessible on its back allows for continuous blood withdrawal from the left carotid artery and return through the right jugular vein. Blood flows at a rate of 30 µL/min into the CTC sorter chip. Two closely-spaced laser beam lines illuminate the main flow channel of the chip. As such, each fluorescent cell that passes through the device emits two pulses of light, which are detected by a photomultiplier tube (**Figure 5-1 c**). The second laser line allows the controller to compute the velocity of the cells, which is essential to ensure reliable CTC capture. Similar to the ensemble-decision aliquot ranking (eDAR) technique for sorting CTC-containing aliquots of blood[5,9], upon fluorescent cell

detection, the controller instantly operates pneumatic valves[25] to redirect a small blood volume that includes the CTC toward a collection tube (mean±s.d. = 127±47 nL/sort event, **Figure 5-1 d, Methods**). Blood from the collection tube can then be further enriched for CTCs and run through a secondary single-CTC sorting chip for downstream characterization using techniques such as scRNA-Seq (**Figure 5-1 e**).
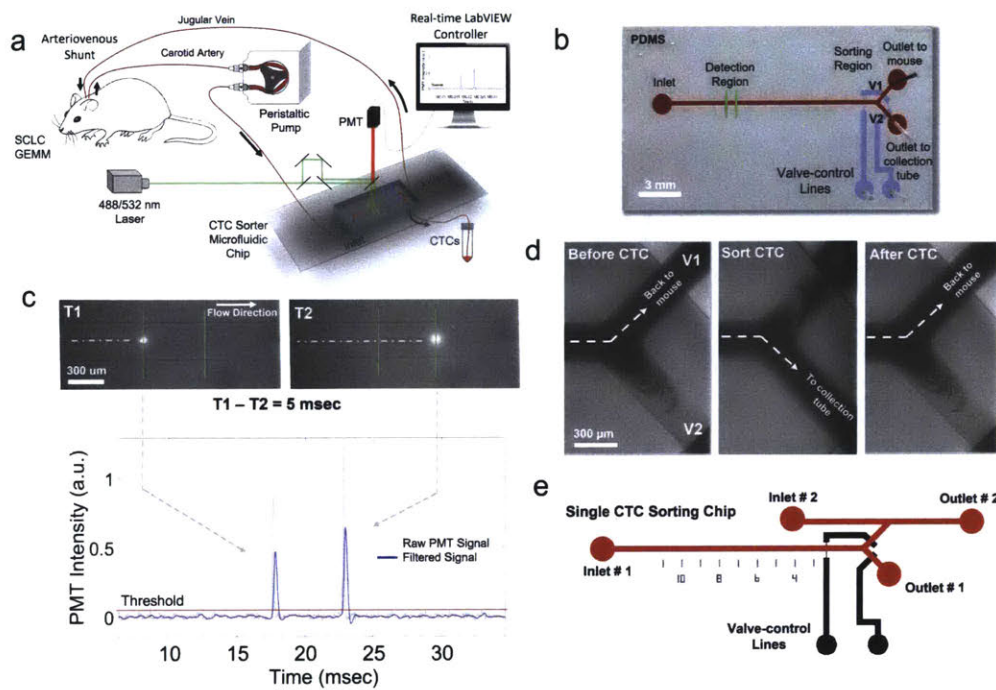

Figure 5-1: A microfluidic sorter for longitudinal CTC studies in GEMMs. **a)** A peristaltic pump withdraws blood from a surgically-implanted cannula in the carotid artery of a mouse at a flow rate of 30 μL/min. The blood is directed into the main flow channel of the CTC sorter chip. For tdTomato-positive cells, a green (532 nm) laser illuminates two points along the main flow channel of the CTC chip separated by a known distance. Thus, fluorescent CTCs emit two red-shifted pulses of light, which are detected by a photomultiplier tube (PMT). Based on the timing of the pulses, a LabVIEW program computes the velocity of the cells and operates computer-controlled pneumatic valves to redirect fluorescent CTCs toward a collection tube. After exiting the chip, CTC-depleted blood returns to the jugular vein of the mouse via a second surgically-implanted cannula. **b)** Top-view image of the CTC sorter microfluidic chip showing the inlet, outlets, and the valve actuation lines (V1 and V2, scale bar = 3 mm). **c)** Illustration of the CTC detection mechanism using the two excitation laser lines. A low-pass filter is applied to the raw data for determining true peaks. **d)** The outlet by which blood is returned to the mouse is briefly sealed while the opposite outlet is opened to allow for CTC isolation in real time. **e)** After collection, CTCs are further enriched by a secondary CTC-sorting chip designed with a parallel channel to flush CTCs into wells containing cell-lysis buffer for downstream scRNA-Seq.

128

5.3: Validation of Platform's Detection Limits with Fluorescent Microbeads and Cells

To ascertain the detection limit of our CTC sorter, we passed a sample of healthy mouse blood spiked with flow-cytometry calibration beads through the system. The reference beads comprised 5 fluorescence intensity groups, including one with zero fluorescence. The system consistently detected the two brightest fluorescence levels (peaks 4 and 5) and approximately the brightest 30% of level 3 (peak 3, **Figure 5-2 a-c**). This sensitivity was sufficient to detect nearly the entire population of tdTomato-expressing murine small cell lung cancer (SCLC) cells spiked in healthy mouse blood. We also tested blood isolated from autochthonous SCLC tumor-bearing mice that exhibit metastasis to distant organs in a pattern similar to metastatic spread in human patients[26]. Tumors in these mice were initiated by Cre-mediated deletion of tumor suppressor genes *Trp53*, *Rb1* and *Pten* in the murine lung epithelium[26]. This GEMM also includes a Cre-activated tdTomato allele[27] that engenders fluorescence in all tumor cells after tumor initiation, including CTCs. The majority of CTCs from the blood of SCLC tumor-bearing mice were above the detection threshold (**Figure 5-2 d, Methods**).
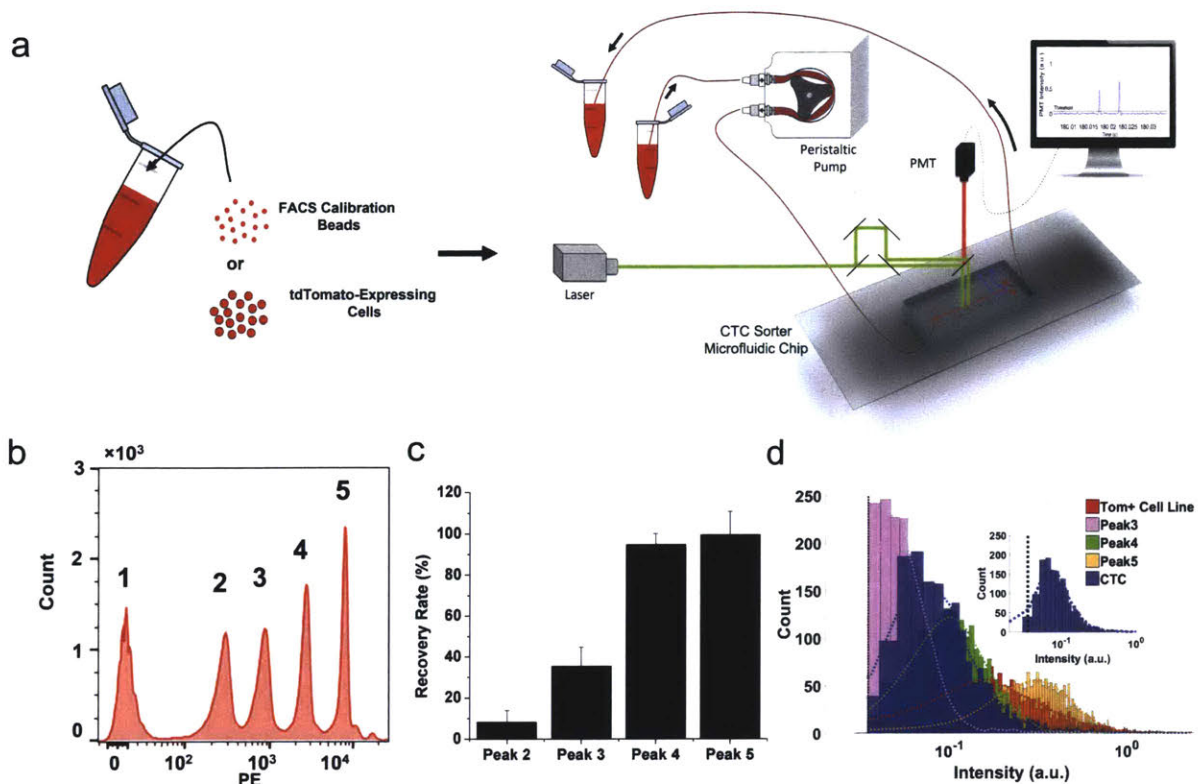
Figure 5-2: CTC sorter system characterization. a) Healthy mouse blood spiked with either fluorescent beads or tdTomato-positive cells is used to determine the detection limits of the system. Each sample is flowed through the sorter at a flow rate of 30 μL/min. **b)** Five separate peaks representing the five different intensity levels of the SpheroTech PE Calibration beads as detected by FACS. **c)** Beads representing fluorescent peaks 2 through 5 were spiked separately into four separate 500 μL samples of healthy mouse blood. Each sample was run through the CTC sorter to measure the recovery rate (i.e. Recovery Rate for Peak 2 sample = total detected/1,000×100). N = 3, results demonstrated as mean±s.d. **d)** Comprehensive histogram plot with Kernel smoothing function fit (dotted envelope) demonstrating the result of all spiked-blood experiments using beads, a tdTomato-positive fluorescent cell line, or terminal bleed samples from SCLC tumor-bearing mice containing tdTomato-positive CTCs (N=1,900 events per population. For CTC population, six terminal bleed samples accounted for the 2,000 events). Inset represents the CTC population that lies above the detection threshold (vertical dotted line).

5.4: Validation of Platform's Sorting Functionality

After establishing that the sensitivity of the optical detector was sufficient, we characterized and optimized the sorting efficacy using blood samples from healthy mice spiked with low concentrations of tdTomato-expressing murine SCLC cells. In samples containing 100 cells/mL

or more, over 80% of detected tdTomato-positive cells were successfully captured. For samples with only 10 cells spiked into 500 µL of healthy mouse blood, the sorted sample contained 6.0 ± 0.7 cells (mean±s.d., N=3 repeats). Applying a slight delay in actuating the pneumatic valves until the cell has moved closer to the sorting region decreased the collected blood volume per CTC to 76 ± 28 nL (mean±s.d.) without compromising the capture efficiency. At this volume, on the order of 700 neighboring white blood cells (WBCs) and over 70,000 red blood cells (RBCs) and platelets in the bloodstream are collected in addition to the target CTC on each valve actuation. These experiments demonstrate that the CTC sorter is capable of isolating fluorescent CTCs from blood even at very low concentrations.

5.5: Longitudinal CTC Collection from SCLC Tumor-Bearing Mice

Next, we conducted a longitudinal study of CTCs collected from autochthonous SCLC tumor-bearing mice treated with the BET bromodomain inhibitor JQ1, which has been demonstrated to have anti-proliferative effects in SCLC[28-30]. CTCs were isolated from mice over a two-hour period prior to treatment (0-hour) and at 24-hour intervals following treatment initiation, continuing over 96 hours (**Figure 5-3, Methods**). CTCs were enriched from the samples by RBC lysis, followed by WBC depletion using magnetic-activated cell sorting (MACS), and finally by passing through a secondary, single-cell CTC sorting chip. Enriched CTCs were then processed using Smart-Seq2[31] (**Figure 5-1 e, Methods**). Cells with insufficient gene complexity for downstream analysis post-scRNA-Seq were eliminated computationally, in addition to cells with high expression of immune and platelet signature genes[3,4,16] (**Methods**). The overall yield of the process (from blood to successful scRNA-Seq library) was 11.5% and 5.3% for samples from treated and untreated mice, respectively (median values with a range of 7.4-31% for treated samples and 3.3-6.7% for untreated samples).
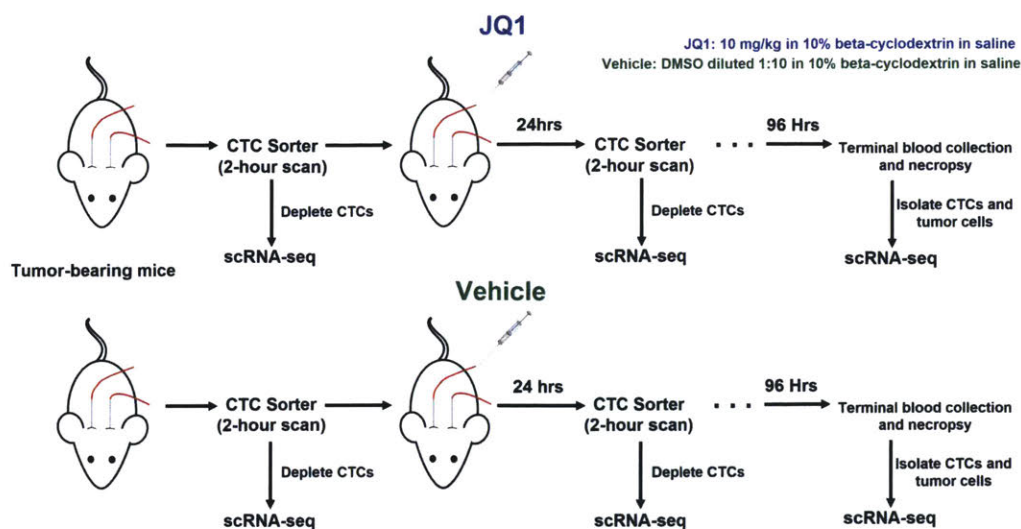
Figure 5-3: Longitudinal JQ1 study timeline. Autochthonous SCLC tumor-bearing mice undergo arteriovenous surgery, followed by treatment with either the BET bromodomain inhibitor JQ1 (top row), or vehicle (bottom row). Each mouse is scanned to collect an initial CTC population before treatment. Intravenous injection of either JQ1 or vehicle then follows and continues daily following a two-hour scan for collection of CTCs. Collected blood samples from each mouse then undergo enrichment processes to sort single CTCs for downstream scRNA-Seq.

5.6: Analysis of Single-CTC Transcriptomes Across Different Mice & Within Each Mouse.

We then examined our data to determine how the information collected longitudinally from the same mouse with our system compared to the common approach of capturing CTCs across different mice using asynchronous terminal bleeds[16,19]. To analyze our longitudinal CTC data, we pooled our collected CTC transcriptomes across all mice, performed a principal component analysis (PCA), visualized by t-distributed stochastic neighbor embedding (tSNE)[13,32,33], and identified clusters (using k-means clustering) over the significant principal components (PCs)[14,34] (**Figure 5-4 a-c, Methods**). This unsupervised analysis revealed that mouse of origin contributed significantly to the variation observed in our dataset, with cluster representation driven primarily by individual mice[3,4,8] (**Figure 5-4 c**). We next performed PCA on CTCs collected from each mouse individually. Here, we found that PC1 significantly correlated

(Spearman correlation) with time since treatment (p<0.05; Student's *t*-test following a Lilliefors test for normality) when independently calculated for each of the treated mice but not for either control (**Figure 5-4 d-e**). This suggests that by isolating CTCs from the same animal longitudinally, we are able to eliminate potentially confounding differences between animals that could otherwise mask biologically-relevant gene expression changes that occur over time.

In comparison, the conventional approach for performing a longitudinal CTC analysis would be to begin the experiment with a cohort of mice and obtain terminal bleeds from a subset at each time point. We simulated this approach from our measurements by selecting a different treated mouse to represent each of the 0, 48, and 96hr time points (**Figure 5-5**); here, regardless of which mouse was chosen to represent which time point, we found that the mean PC1 coordinate of treated mouse 1 existed outside the interquartile range of the other mice, suggesting a consistent mouse-specific effect that dominates the first PC (**Figure 5-4 f**). As such, conclusions drawn from analysis of CTCs from terminal bleeds at different time points across mice would be confounded by organism-specific features from the different mice.
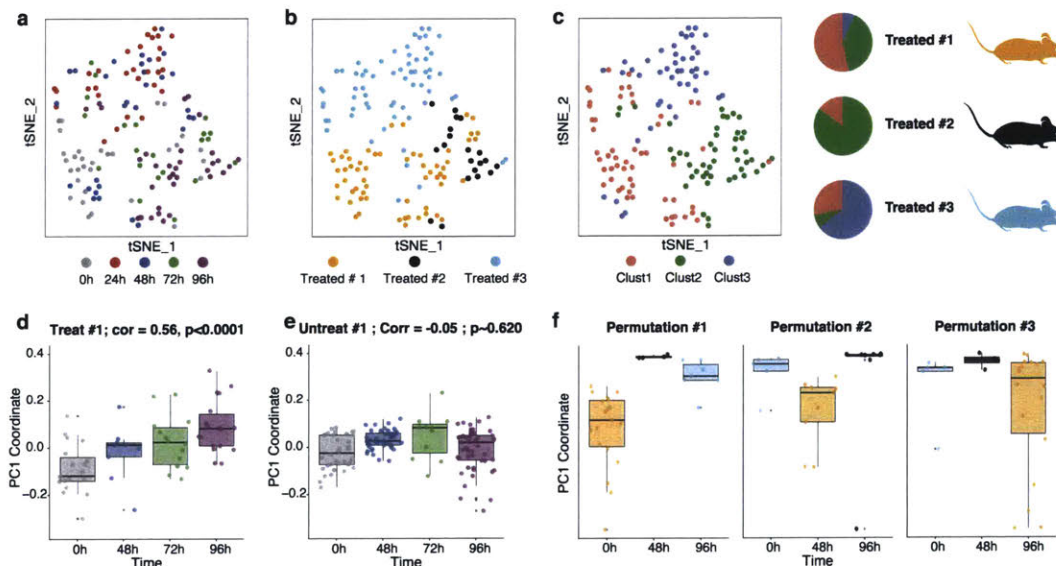

Figure 5-4. scRNA-Seq of captured CTCs demonstrates utility of intra-mouse CTC profiling. a-c) tSNE of all CTCs collected across three JQ1-treated mice colored by time point post-treatment **(a)**, mouse **(b)** and cluster of assignment based on kNN clustering **(c)**. Pie charts on the top right show the fractional representation of each cluster in each treated mouse. **d-e)**

Boxplots of the first principal component of CTC transcriptomes from PCAs obtained from longitudinally following the same treated mouse (Corr = 0.56) **(d)** or untreated mouse (Corr = -0.05) **(e)**. Each point represents a CTC. **f)** Boxplots of the first principal component from three different "mock terminal bleed" permutations across three treated mice (**Figure 6-5, Methods**).
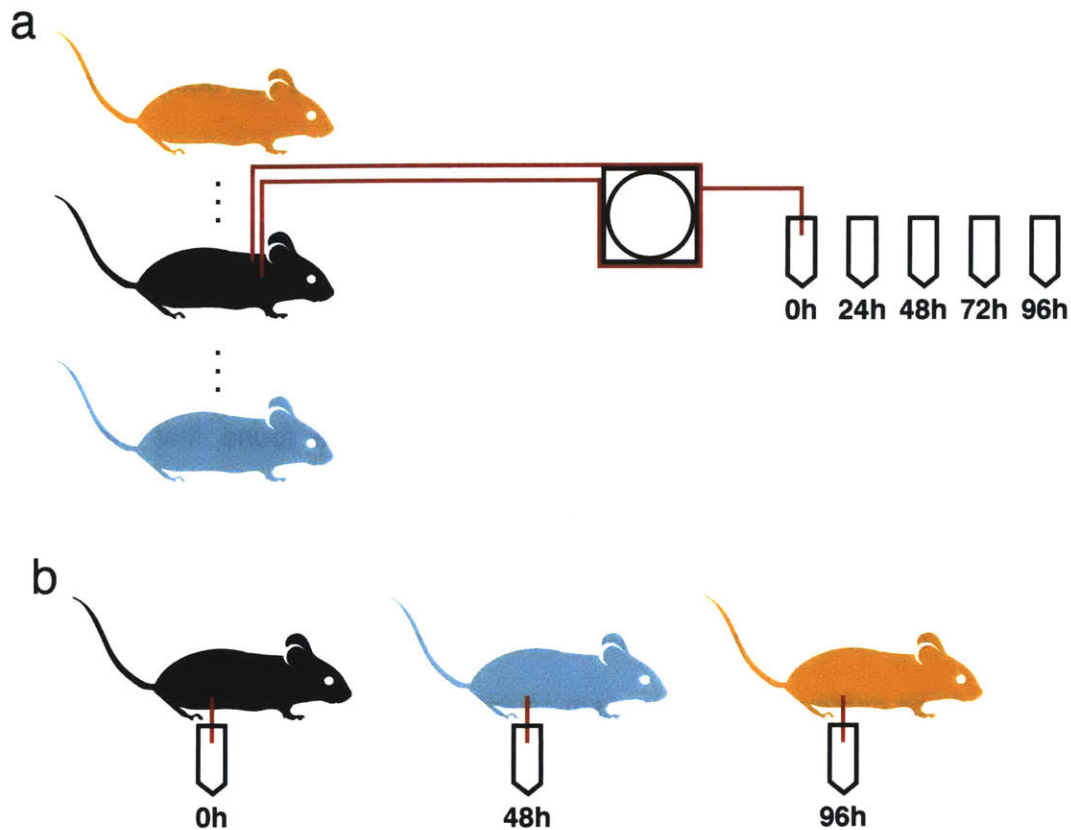

Figure 5-5: Schematic of true **(a)** and mock **(b)** data for simulations of experiments. **a)** shows the process by which single CTCs are collected using our CTC sorter. **b)** shows one of the combinations of mice that could lead to a time course collection of CTCs, which simulates a terminal bleed experiment of CTCs. NB: we analyzed all six possible permutation combinations of **(b)**.

5.7: Supervised Analysis of Single-CTC Transcriptomes

To more formally examine treatment-induced shifts in gene expression, we calculated differential expression across all pairs of time points within each longitudinally-profiled mouse[14,35] and once again simulated terminal bleed data (**Figure 5-5, Methods**). Our analyses showed that the majority of the differentially expressed genes within each mouse (per-mouse) were unique (**Figure 5-6, Methods**). Furthermore, each per-mouse differentially expressed

gene set shared little overlap (p<0.05; Hypergeometric test) with those calculated from mock

terminal bleed datasets, regardless of the chosen mouse for the different time points. At each

time point, differentially expressed genes in the mock terminal bleed data were enriched for

several functional processes, such as mitochondrial function, cellular organization, and

metabolism[36,37]; however, upon further inspection of the different mock terminal bleed

permutations, we found that these enrichments were linked primarily to mouse rather than time

point (**Figure 5-4 f**), suggesting confounding mouse-to-mouse heterogeneity. This is evocative

of the marked inter- and intra-patient heterogeneity observed in CTCs longitudinally collected

from human patients[6-8], and suggests the importance of examining the same mouse over time.
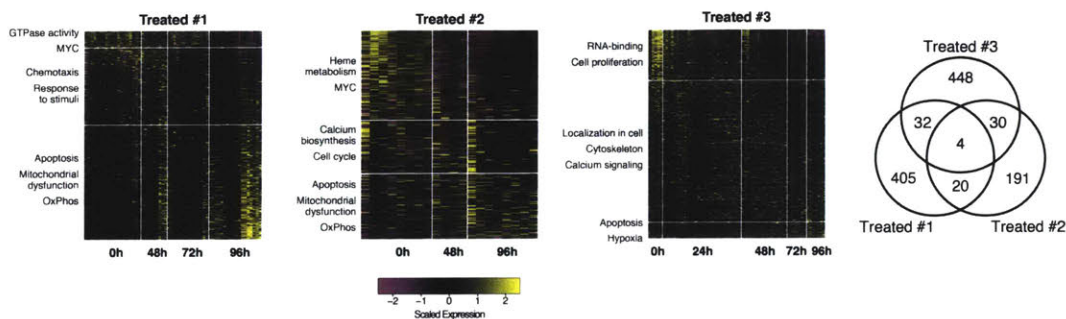

Figure 5-6: Differentially expressed genes across time course for each mouse. Left:
Heatmaps of differentially expressed genes between all pairs of time points within each treated
mouse. Rows represent genes, columns represent cells. Time points sampled are provided on
the x-axis and selected gene set enrichments for each differentially expressed gene module
(identified by hierarchical clustering; see **Methods**) are displayed to the left of the relevant
cluster. The majority of these gene set enrichments are unique to each pair of mouse and time
point. **Right:** Venn diagram shows overlap of differentially expressed genes across all treated
mice.

5.8: Discussion

The platform outlined here represents an important advancement in the detection and

continuous capture of single CTCs from the same mouse over time. Our method enables CTCs

to be isolated in low blood volumes and prepares them for downstream characterization. Here,

we used scRNA-Seq to show that continuous CTC capture from the same mouse eliminates

biases driven by inter-mouse heterogeneity that can occur when CTCs are collected across different mice. Although future work will be needed to elucidate the underlying drivers of this variability, given the baseline genetic homogeneity of the animals used to generate our GEMM, one potential explanation could be underlying differences in the cellular composition of the primary tumors across different animals. scRNA-Seq results from the primary tumor samples harvested from each animal after terminal CTC collection (96hr) are consistent with this hypothesis (**Figure 5-7 a-b, Methods**). These data suggest that primary tumors from each mouse may contain multiple malignant gene expression states[32,33], which appear to be shared across mice. Although some of these differences could be attributed to the presence of multiple independently evolving primary tumors within each mouse, the fact that each gene expression state is comprised of cells from multiple mice suggests that these states may be a shared feature of our SCLC GEMM, though future experiments will be needed to robustly validate this finding. Intriguingly, when we computationally assigned terminally collected CTCs (96hr, N=92 cells from 5 mice) to these shared states based on gene expression similarity (**Methods**), ~67% (62 cells) of the CTCs matched. Furthermore, we observed comparable state frequencies between a mouse's CTCs and its corresponding primary tumor sample (p>0.5, Fisher's Exact test, **Figure 5-7 c**). For mice with fewer than 10 CTCs, the statistical power in our comparison was in some cases limited. While further experimentation will be needed to corroborate this preliminary finding, our data demonstrate the potential value of having matched primary tumor samples as a reference in mouse models of cancer, and that analysis of CTCs in our murine model of SCLC may reveal similar biology to primary tumors from the same mouse, suggesting their utility as a surrogate for matched tumors under specific circumstances.
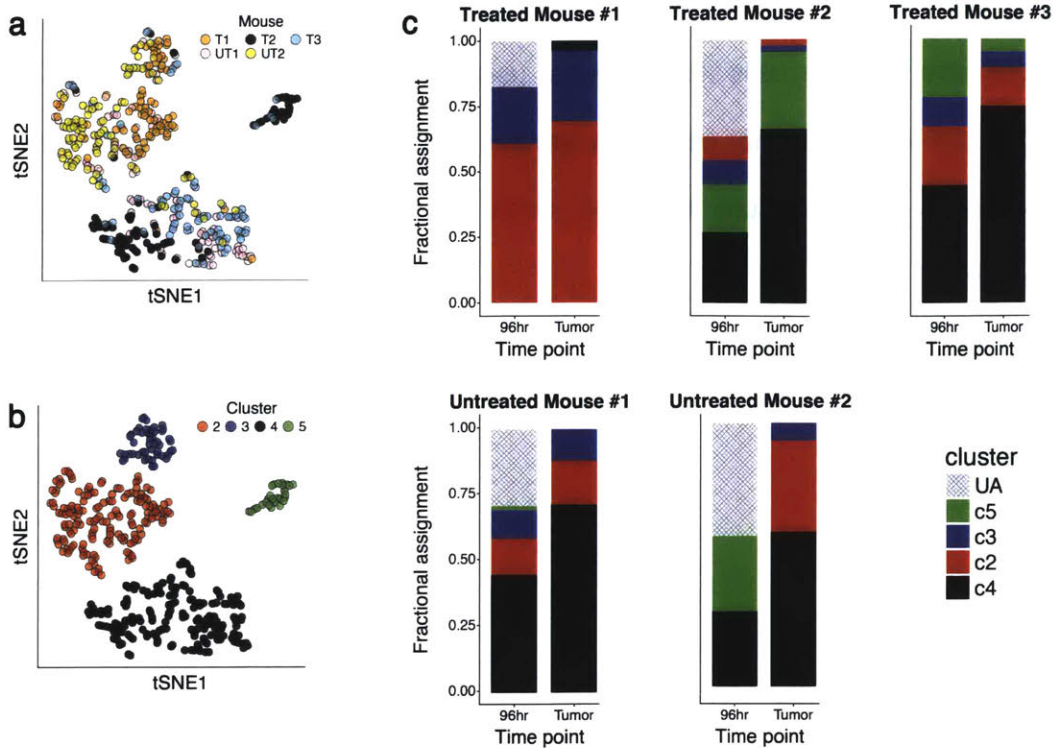
Figure 5-7: scRNA-Seq of endpoint primary tumors demonstrates heterogeneity in phenotype. a) tSNE of primary tumor cells across treated and untreated mice, colored by clusters called from kNN clustering. **b)** tSNE of primary tumor cells across treated and untreated mice, colored by mouse. C) Computational cluster assignments (**Methods**) for 96-hr CTCs next to their matched primary for a representative treated and untreated mouse plotted as bar plots (N=18 and 82 cells for treated mouse #1 96hr CTCs and tumor cells, respectively, and N=52 and 84 cells for untreated mouse #1 96hr CTCs and tumor cells, respectively). Neither pairing is significantly different (p=0.99 and 0.66 for Treated Mouse # 1 and Untreated Mouse #1, respectively, Fisher's Exact Test).

Future work of this kind has the potential to shed new light on the relationship between CTCs, primary tumors, and metastases, allowing for the exploration of their utility as biomarkers and facilitating examination of how individual CTCs contribute to metastasis. Moreover, they may help elucidate the features that inform shifts observed upon perturbation, such as drug treatment. Ultimately, these data show that our platform opens the door for novel CTC experimentation, such as examining longitudinal drug responses and comparing CTCs to primary tumors (shown here), characterizing their relationship to metastases, and measuring the rate of CTC production in an acute window. With additional development, our device could

enable longitudinal studies in mice to find associations between individual CTCs and clusters of CTCs, profile rare immune cells (for example, using genetic reporters or based on tetramer-staining), monitor mesenchymal cells in a variety of contexts (including wound healing and tumor formation), and measure induction rates of drugs or nanoparticles in circulating mononuclear cells.

Materials and Methods. All RNA-seq data generated in this study (raw data and processed data matrices) have been deposited into the Gene Expression Omnibus database hosted at the National Center for Biotechnology Information under the accession code GSE122233. For information on mouse models, cell culture, shunt surgery, optofluidic platform design and fabrication, real-time data processing and analysis, CTC and tumor cell processing and enrichment, single-cell RNA-sequencing sample preparation and data analysis, please refer to SI Appendix.

Mouse model and cell lines. The $Trp53^{fl/fl}$; $Rb1^{fl/fl}$; $Pten^{fl/fl}$; $Rosa26^{LSL-Luciferase/LSL-Luciferase}$ (PRP-L/L) mouse model of SCLC has been described previously[26]. $Rosa26^{LSL-tdTomato/LSL-tdTomato}$ mice were obtained from Jackson Laboratories ($Gt(ROSA)26Sor^{tm14(CAG-tdTomato)Hze}$) and crossed into the PRP-L/L model to obtain $Trp53^{fl/fl}$; $Rb1^{fl/fl}$; $Pten^{fl/fl}$; $Rosa26^{LSL-tdTomato/LSL-Luciferase}$ mice. Tumors were initiated by intratracheal delivery of $2×10^8$ plaque-forming units (p.f.u.) of adenovirus expressing Cre recombinase under the control of a CGRP promoter (Ad5-CGRP-Cre[38]), as previously described[39]. Adenoviral stocks were purchased from the Viral Vector Core Facility at the University of Iowa Carver College of Medicine. Candidates for arteriovenous shunt surgery were identified by *in vivo* bioluminescence imaging using the IVIS Spectrum In Vivo Imaging System (PerkinElmer).

Murine SCLC cell lines (AF1281-M1, AF3291LN) were generated from mSCLC tumors isolated from $Trp53^{fl/fl}$; $Rb1^{fl/fl}$; $Rosa26^{LSL-tdTomato/+}$ or $Trp53^{fl/fl}$; $Rb1^{fl/fl}$; $Pten^{fl/fl}$; $Rosa26^{LSL-tdTomato/LSL-Luciferase}$ mice as previously described[40].

Shunt Surgery. All animal-based procedures were approved by the Massachusetts Institute of Technology Committee on Animal Care (CAC), Division of Comparative Medicine (DCM). The arteriovenous shunt method was validated as an approach for real-time manipulation of nutrient levels in the serum of anesthetized mice[41-43]. We adopted the same technique for continuous sampling of mouse blood for CTCs. Briefly, catheters are inserted into the right jugular vein and the left carotid artery and are externalized using standard cannulation surgical techniques. A peristaltic pump (Instech Laboratories Inc., Plymoth Meeting, PA, USA) is then connected to the catheters for blood sampling and return through the carotid artery and jugular vein, respectively, in the conscious mouse (**Figure 5-1a**). During the four-day longitudinal studies, the total collected blood volume is monitored. If depleted blood volume exceeds 1% of the animal's body weight (for example 260 µL for a 25g mouse), per MIT DCM guidelines, healthy-mouse donor blood (of same strain and sex) is infused directly into the mouse using the jugular vein catheter, equivalent to the amount removed.

Optical Detection Platform. The optical system is comprised of two optical trains, making up two compact vertical microscopes. The top optical train divides the laser beam (OBIS 532 LS, Coherent Inc) into two separate beams that are focused along one axis to produce two illumination lines projected at the sample focal plane (perpendicular to the blood flow channel) for precise velocity measurements of the flowing CTCs. The laser passes through a line filter and polarizing beam splitters to generate the two beams with minimal losses. Next, the two laser lines pass through a cylindrical lens to focus the two beams into lines. The focused lines are then projected onto the microfluidic channel with a 4F optical system. The dichroic mirror and longpass filters, placed directly above the detection region, pass a filtered fluorescence signal to the PMT (Hamamatsu H10722-20) by blocking the 532 nm laser line signal with a

140

notch filter (532 nm StopLine single-notch filter, Semrock). A 90:10 beam splitter is added before the PMT to allow for imaging of the illumination region for device alignment purposes.

The bottom optical train is similar in configuration to the top train and uses a second green laser to illuminate a circular region covering the valves, enabling fluorescent and bright field imaging for confirming successful CTC sorting and valve functionality throughout the experiment.

Microfluidic Device Design and Fabrication. The microfluidic device for sorting minute amounts of blood containing single CTCs was designed to have one inlet to a 1000µm-long microfluidic channel that bifurcates into two channel outlets (90° apart, **Figure 5-1 b-c**); one for returning the blood to the mouse and the other for collecting the sorted CTC-containing blood sample. The fabrication was performed using standard soft lithographic techniques on two four-inch wafers. A single layer of photoresist (SU8 2050, Microchem, Newton, MA) was patterned to create the pneumatic channels on the valve control wafer. For the blood flow channel, AZ9260 positive resist (MicroChemicals) was exposed, developed, and then reflowed at 120°C for 10 minutes to create the half ellipsoid channel profile necessary for a complete valve seal[25].

A mixture of PDMS (Polydimethylsiloxane) and its curing agent (SYLGARD 184 A/B, Dowcorning, Midland, MI, USA) at a 10:1 ratio was spun on top of the actuation wafer to a thickness of 50 µm and baked in an oven set to 65 °C for at least 3 hours. For the flow channel layer, the mixture was poured to a thickness of ~5 mm and cured at 65 °C for 3 hours. Afterwards, the flow channel layer was peeled off and punched with a 0.75 mm puncher (Harris Uni-Core, Ted Pella Inc., Reading, CA) to define the inlet and outlets to and from the channel, respectively, and diced to prepare for bonding. The flow channel devices and the actuation layers were then treated with oxygen plasma (100 watt, 1 ccm, 140 torr, 10 sec). Next, the flow layer was aligned to the actuation layer and transferred to a hot plate set to 85 °C. After 15 minutes, the assembled PDMS layers were peeled off and punched with a 0.75 mm puncher to

define inlets to the actuation channels. The assembled PDMS layers were treated with oxygen plasma (100 watt, 1 ccm, 140 torr, 10 sec) for irreversible bonding to a glass slide (Fisherbrand 1x3", Fisher Scientific, Pittsburgh, PA).

Prior to flow experiments, the device was aligned to project the two laser lines across the flow channel approximately 8 mm away from the valve actuation region. The device was then primed with Heparinized-Saline (diluted to 100 USP units per mL, NDC 25021-400-30) to prevent any clotting within the microfluidic channel.

Real-Time Data Processing and System Control. The PMT module generates an output voltage that is sampled by a NI USB-6009 (National Instruments) 14 bit analog-to-digital converter. This analog-to-digital converter is also used to output the control voltage signals for the peristaltic pump and the PMT gain. The PMT voltage is acquired and displayed in real time in LabVIEW (National Instruments) at 20,000 samples per second. The raw data is filtered with a low pass filter to remove spurious noise and the output of a running median filter is subtracted to remove low frequency drift or any DC offset. Upon transit of a fluorescent cell or particle, the PMT output signal passes a specific threshold, triggering the NI USB-9472 device to actuate the SMC Solenoid Valves (S070A-6DC, 8 ports) that are connected to an external 6-volt power supply.

During non-sorting operation of the system (i.e. cell counting experiments), the first valve (V1) region is kept sealed to force the blood to flow directly through to the opposite outlet and back into the jugular vein catheter of the mouse (Top-view image, **Figure 5-1 b**). When the sorting functionality is activated, and upon detection of a cell, its velocity is calculated and used to estimate the time of its arrival at the valve region of the channel, all using a virtual instrument (VI) program in LabVIEW. V2 then closes immediately and V1 opens for an amount of time equal to twice the estimated time of arrival of the cell at the valve to deflect a 127±47 nL (or

142

76±28 nL with a valve actuation delay) blood bolus containing the cell. This amount of blood per single sort provides an upper limit for the number of events we can collect from a mouse over the four-day study. For a 25g mouse, a maximum number of ~2,600 total events can be collected without the need for fluid replacement, per MIT DCM guidelines.

After CTC enrichment with RBC lysis and MACS (described below), the cells were sorted using our secondary single-CTC sorting chip. Upon detection of a single fluorescent CTC in the secondary single-cell sorting chip (**Figure 5-1 e**), micro-valves actuate to push the cell into the parallel channel. The micro-valves then seal the primary channel and sample flow is stopped to introduce fresh buffer into the parallel channel, releasing the cell out of the chip into the tubing and then into a collection well containing TCL buffer (QIAGEN) with 1% 2-mercaptoethanol for downstream scRNA-seq.

CTC enrichment by RBC lysis and MACS. Following CTC isolation, the collected blood sample is first treated with ACK Lysing Buffer (Gibco A10492-01) to remove RBCs, then rinsed, filtered using a 30μm Pre-Separation Filter (Miltenyi Biotec, #130-041-407), and processed through magnetic activated cell sorting (MACS) using mouse CD45 MicroBeads (Miltenyi #130-052-301) and MS columns (Miltenyi #130-042-201) to remove CD45-positive cells, according to the manufacturers' protocols. The final product is then diluted to a total volume of 2 mL of MACS buffer and flowed through the secondary single-CTC sorting chip.

Characterization with Cell Line and Beads. To validate the sensitivity of the system to detect fluorescent cells in blood, 5-peak FACS calibration beads (Sphero PE 5-peak, Spherotech Inc) were spiked into mouse blood and flowed through the device at 30 μL/min (**Figure 5-2**). To validate the sorting functionality of the system, a tumor cell line (AF1281-M1) was established

from a mouse with autochthonous small cell lung cancer expressing a bright and stable cytoplasmic tdTomato fluorescent protein. Cells were then counted using Coulter Counter (Multisizer 4, Beckman-Coulter) and re-suspended in 500 µL of mouse blood at different densities (10, 100, 500, and 1000 cells).

JQ1 Treatment of Tumor-Bearing Mice. Tumor-bearing SCLC mice were treated with 10 mg/kg JQ1 (Cayman Chemical) by intravenous injection daily for the duration of the study. JQ1 was dissolved in DMSO to make a 20 mg/mL stock, then diluted 1:10 with 10% beta-cyclodextrin in 0.9% saline to obtain a working concentration of 2 mg/mL. Vehicle-treated control mice received an equivalent dose of DMSO diluted 1:10 with 10% beta-cyclodextrin in 0.9% saline (**Figure 5-3**).

Dissociation of Tumor Samples for Single-Cell RNA-Sequencing Analysis. Primary tumors from tumor-bearing animals were dissected, dissociated into single cells using a lung dissociation kit according to the manufacturer's protocol (Miltenyi Biotec #130-095-927), then stained with APC-conjugated antibodies against CD11b (eBioscience #17-0112-82), CD31 (BioLegend #102510), CD45 (eBioscience #17-0451-83) and TER-119 (BD Biosciences #557909). tdTomato-positive, APC-negative cells were single-cell sorted by FACS into TCL buffer (QIAGEN) containing 1% 2-mercaptoethanol, then frozen at -80°C for downstream processing for RNA-seq.

Single-Cell RNA-Sequencing Sample Preparation. Both CTC samples and primary tumor samples collected in TCL buffer were processed through Smart-Seq2 as follows. Total nucleotide material from lysed single cells was extracted with RNA-clean AMPure nucleotide

144

extraction beads (Beckman-Coulter) and washed with 80% ethanol before undergoing reverse transcription with Maxima enzyme (Thermoscientific), followed by PCR with a KAPA Hotstart Readymix 2x kit (KAPA biosystems). Following quantification and quality control analysis by Qubit DNA quantification (Thermoscientific) and tape station (Agilent), whole transcriptome amplifications (WTAs) of each single cell were transformed into sequencing libraries with a Nextera XT kit (Illumina) and barcoded with unique 8-bp DNA barcodes. cDNA libraries were pooled, quantified, and sequenced on an Illumina NextSeq 500 to an average depth of 1.2M reads/CTC. Raw sequencing data

Analysis of Raw Sequencing Data. Following sequencing, BCL files were converted to merged, demultiplexed FASTQs. Paired-end reads were mapped to mm10 mouse transcriptome (UCSC) with Bowtie. Expression levels of genes were log-transformed transcript-per-million (TPM[i,j]) for gene i in sample j, estimated by RSEM in paired-end mode. For each cell, we enumerated genes for which at least one read was mapped, and the average expression level of a curated list of housekeeping genes. We excluded from analysis profiles with fewer than 500 detected genes or an average housekeeping expression below 0.5 $\log_2(\text{TPM})$.

Identification of leukocytes and correction of platelet effect. A matrix of TPM estimates for all genes across all cells resulted from preprocessing of RNA-Seq data. A raw principal component analysis was run for each sample, and the first PC separated cells of immune lineage from cells expressing epithelial genes. We selected from overlaps between this first PC and a curated set of immune-related genes to score CTCs according to their immune score. Cells with an average immune score below 0.5 $\log_2(\text{TPM})$ were taken forward for analysis of their transcriptomes as CTCs.

Likewise, a curated set of platelet-related genes was used to develop a platelet signature for each CTC transcriptome. The signal of this platelet signature was then regressed out using the RegressOut function in Seurat.

Principal Component Analysis of CTCs. We performed a gene selection based on a binning strategy across expression using Seurat to define a unique set of variable genes for each treated and untreated mouse, as well as combined sets of treated and untreated mice. Principal component analysis was performed over these variable gene sets for each case (per-mouse or combined). PCs were correlated to time using Spearman correlation. To test for significance of PC1 correlation with time, we permuted assignments of time (N = 1000) and calculated statistics for each set of CTCs per mouse. The sets of correlations for each mouse were tested with Lilliefors test for normality and used to determine PCs with significant correlations with time.

Generation of Mock Data. To simulate current methods of terminal bleed assays for CTC collection across an acute time scale, we generated a series of mock datasets using our true, continuously collected data. At each time point, the CTCs from a single treated mouse were selected without replacement. This process was exhaustively repeated five times to generate all unique treated mock data over which further differential expression analysis was performed.

Differential Expression Analysis. For each of the treated mice, as well as the mock data, we performed differential expression using Seurat's built-in single-cell differential expression tool, with a bimodal distribution model. Differential expression was performed between all pairs of time points available for each dataset. Genes with avg_diff > 1 and p-value < 0.01 were selected and visualized using the DoHeatmap function in Seurat. Furthermore, these genes

146

were enriched for upstream regulators using gene set enrichment analysis (GSEA) through the Broad Institute.

Analysis of Primary Tumor Data. Variable genes across primary tumor cells of all mice were calculated and principal component analysis was performed as described above. Relevant PCs were determined by visualizing percent variance explained in an elbow plot, significance by Jackstraw[44], and manual inspection of loadings and coordinates. Following inspection, PCs 1-5 were selected for downstream visualization of the primary tumor cells by tSNE in Seurat, with perplexity set to 15, and 2500 iterations run.

Assignment of CTCs to Primary Tumor Cluster. Differential expression, as reported above, was used to identify marker genes that describe the resultant clusters in the primary tumor. These marker genes were used to develop a signature score for each cluster for each of the 96-hr CTCs, using weighted averaging of the genes for each signature. Next, we permuted random sets of genes with similarly binned expression distribution and size to create cluster-independent background scores for each CTC. CTCs with cluster-specific signatures above their cluster-independent background were "assigned" to that cluster. CTCs were assigned to all clusters which were above background – if a CTC scored above background for more than one cluster, it was assigned to all those clusters; if a CTC scored above background for no clusters, it was classified as "unassigned." The total number of CTCs per mouse for each cluster was visualized by a stacked bar plot, generated through ggplot2 in R. Distributions of assignment between 96-hr CTCs and primary tumors from mice were tested by Fisher's Exact Test, with a cutoff for statistically significant difference set at $p < 0.05$.

5.10: References

1   Pantel, K. & Speicher, M. R. The biology of circulating tumor cells. *Oncogene* **35**, 1216-1224, doi:10.1038/onc.2015.192 (2016).

2   Yu, M., Stott, S., Toner, M., Maheswaran, S. & Haber, D. A. Circulating tumor cells: approaches to isolation and characterization. *J Cell Biol* **192**, 373-382, doi:10.1083/jcb.201010021 (2011).

3   Miyamoto, D. T. *et al.* RNA-Seq of single prostate CTCs implicates noncanonical Wnt signaling in antiandrogen resistance. *Science* **349**, 1351-1356, doi:10.1126/science.aab0917 (2015).

4   Ting, D. T. *et al.* Single-cell RNA sequencing identifies extracellular matrix gene expression by pancreatic circulating tumor cells. *Cell Rep* **8**, 1905-1918, doi:10.1016/j.celrep.2014.08.029 (2014).

5   Johnson, E. S., Anand, R. K. & Chiu, D. T. Improved detection by ensemble-decision aliquot ranking of circulating tumor cells with low numbers of a targeted surface antigen. *Anal Chem* **87**, 9389-9395, doi:10.1021/acs.analchem.5b02241 (2015).

6   Gorges, T. M. *et al.* Enumeration and Molecular Characterization of Tumor Cells in Lung Cancer Patients Using a Novel In Vivo Device for Capturing Circulating Tumor Cells. *Clin Cancer Res* **22**, 2197-2206, doi:10.1158/1078-0432.CCR-15-1416 (2016).

7   Ozkumur, E. *et al.* Inertial focusing for tumor antigen-dependent and -independent sorting of rare circulating tumor cells. *Sci Transl Med* **5**, 179ra147, doi:10.1126/scitranslmed.3005616 (2013).

8   Sarioglu, A. F. *et al.* A microfluidic device for label-free, physical capture of circulating tumor cell clusters. *Nat Methods* **12**, 685-691, doi:10.1038/nmeth.3404 (2015).

9   Schiro, P. G. *et al.* Sensitive and high-throughput isolation of rare cells from peripheral blood with ensemble-decision aliquot ranking. *Angew Chem Int Ed Engl* **51**, 4618-4622, doi:10.1002/anie.201108695 (2012).

10   Vermesh, O. *et al.* An intravascular magnetic wire for the high-throughput retrieval of circulating tumour cells in vivo. *Nat Biomed Eng* **2**, 696-705, doi:10.1038/s41551-018-0257-3 (2018).

11   Zeisel, A. *et al.* Brain structure. Cell types in the mouse cortex and hippocampus revealed by single-cell RNA-seq. *Science* **347**, 1138-1142, doi:10.1126/science.aaa1934 (2015).

12   Macosko, E. Z. *et al.* Highly Parallel Genome-wide Expression Profiling of Individual Cells Using Nanoliter Droplets. *Cell* **161**, 1202-1214, doi:10.1016/j.cell.2015.05.002 (2015).

13   Shalek, A. K. *et al.* Single-cell transcriptomics reveals bimodality in expression and splicing in immune cells. *Nature* **498**, 236-240, doi:10.1038/nature12172 (2013).

14   Satija, R., Farrell, J. A., Gennert, D., Schier, A. F. & Regev, A. Spatial reconstruction of single-cell gene expression data. *Nat Biotechnol* **33**, 495-502, doi:10.1038/nbt.3192 (2015).

15   Lohr, J. G. *et al.* Whole-exome sequencing of circulating tumor cells provides a window into metastatic prostate cancer. *Nat Biotechnol* **32**, 479-484, doi:10.1038/nbt.2892 (2014).

16   Aceto, N. *et al.* Circulating tumor cell clusters are oligoclonal precursors of breast cancer metastasis. *Cell* **158**, 1110-1122, doi:10.1016/j.cell.2014.07.013 (2014).

17   Vishnoi, M. *et al.* The isolation and characterization of CTC subsets related to breast cancer dormancy. *Sci Rep* **5**, 17533, doi:10.1038/srep17533 (2015).

18   Alix-Panabieres, C. & Pantel, K. Clinical Applications of Circulating Tumor Cells and Circulating Tumor DNA as Liquid Biopsy. *Cancer Discov* **6**, 479-491, doi:10.1158/2159-8290.CD-15-1483 (2016).

19   Rhim, A. D. *et al.* EMT and dissemination precede pancreatic tumor formation. *Cell* **148**, 349-361, doi:10.1016/j.cell.2011.11.025 (2012).

20   Parasuraman, S., Raveendran, R. & Kesavan, R. Blood sample collection in small laboratory animals. *J Pharmacol Pharmacother* **1**, 87-93, doi:10.4103/0976-500X.72350 (2010).

21   Georgakoudi, I. *et al.* In vivo flow cytometry: a new method for enumerating circulating cancer cells. *Cancer Res* **64**, 5044-5047, doi:10.1158/0008-5472.CAN-04-1058 (2004).

22   Zettergren, E. *et al.* Instrument for fluorescence sensing of circulating cells with diffuse light in mice in vivo. *J Biomed Opt* **17**, 037001, doi:10.1117/1.JBO.17.3.037001 (2012).

23   Nedosekin, D. A., Verkhusha, V. V., Melerzanov, A. V., Zharov, V. P. & Galanzha, E. I. In vivo photoswitchable flow cytometry for direct tracking of single circulating tumor cells. *Chem Biol* **21**, 792-801, doi:10.1016/j.chembiol.2014.03.012 (2014).

24   Chudziak, J. *et al.* Clinical evaluation of a novel microfluidic device for epitope-independent enrichment of circulating tumour cells in patients with small cell lung cancer. *Analyst* **141**, 669-678, doi:10.1039/c5an02156a (2016).

25   Unger, M. A., Chou, H. P., Thorsen, T., Scherer, A. & Quake, S. R. Monolithic microfabricated valves and pumps by multilayer soft lithography. *Science* **288**, 113-116, doi:10.1126/science.288.5463.113 (2000).

26   McFadden, D. G. *et al.* Genetic and clonal dissection of murine small cell lung carcinoma progression by genome sequencing. *Cell* **156**, 1298-1311, doi:10.1016/j.cell.2014.02.031 (2014).

27   Madisen, L. *et al.* A robust and high-throughput Cre reporting and characterization system for the whole mouse brain. *Nat Neurosci* **13**, 133-140, doi:10.1038/nn.2467 (2010).

28   Lenhart, R. *et al.* Sensitivity of Small Cell Lung Cancer to BET Inhibition Is Mediated by Regulation of ASCL1 Gene Expression. *Mol Cancer Ther* **14**, 2167-2174, doi:10.1158/1535-7163.MCT-15-0037 (2015).

29   Kato, F. *et al.* MYCL is a target of a BET bromodomain inhibitor, JQ1, on growth suppression efficacy in small cell lung cancer cells. *Oncotarget* **7**, 77378-77388, doi:10.18632/oncotarget.12671 (2016).

30   Jahchan, N. S. *et al.* Identification and Targeting of Long-Term Tumor-Propagating Cells in Small Cell Lung Cancer. *Cell Rep* **16**, 644-656, doi:10.1016/j.celrep.2016.06.021 (2016).

31   Picelli, S. *et al.* Full-length RNA-seq from single cells using Smart-seq2. *Nat Protoc* **9**, 171-181, doi:10.1038/nprot.2014.006 (2014).

32   Patel, A. P. *et al.* Single-cell RNA-seq highlights intratumoral heterogeneity in primary glioblastoma. *Science* **344**, 1396-1401, doi:10.1126/science.1254257 (2014).

33   Tirosh, I. *et al.* Dissecting the multicellular ecosystem of metastatic melanoma by single-cell RNA-seq. *Science* **352**, 189-196, doi:10.1126/science.aad0501 (2016).

34   Chung, N. C. & Storey, J. D. Statistical significance of variables driving systematic variation in high-dimensional data. *Bioinformatics* **31**, 545-554, doi:10.1093/bioinformatics/btu674 (2015).

35   McDavid, A. *et al.* Data exploration, quality control and testing in single-cell qPCR-based gene expression experiments. *Bioinformatics* **29**, 461-467, doi:10.1093/bioinformatics/bts714 (2013).

36   Huang da, W., Sherman, B. T. & Lempicki, R. A. Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nat Protoc* **4**, 44-57, doi:10.1038/nprot.2008.211 (2009).

37   Subramanian, A. *et al.* Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proc Natl Acad Sci U S A* **102**, 15545-15550, doi:10.1073/pnas.0506580102 (2005).

38   Sutherland, K. D. *et al.* Cell of origin of small cell lung cancer: inactivation of Trp53 and Rb1 in distinct cell types of adult mouse lung. *Cancer Cell* **19**, 754-764, doi:10.1016/j.ccr.2011.04.019 (2011).

39   DuPage, M., Dooley, A. L. & Jacks, T. Conditional mouse lung cancer models using adenoviral or lentiviral delivery of Cre recombinase. *Nat Protoc* **4**, 1064-1072, doi:10.1038/nprot.2009.95 (2009).

40   Dooley, A. L. *et al.* Nuclear factor I/B is an oncogene in small cell lung cancer. *Genes Dev* **25**, 1470-1475, doi:10.1101/gad.2046711 (2011).

41   Ayala, J. E. *et al.* Hyperinsulinemic-euglycemic clamps in conscious, unrestrained mice. *J Vis Exp*, doi:10.3791/3188 (2011).

42   Ayala, J. E., Bracy, D. P., McGuinness, O. P. & Wasserman, D. H. Considerations in the design of hyperinsulinemic-euglycemic clamps in the conscious mouse. *Diabetes* **55**, 390-397, doi:10.2337/diabetes.55.02.06.db05-0686 (2006).

43   Ayala, J. E. *et al.* Standard operating procedures for describing and performing metabolic tests of glucose homeostasis in mice. *Dis Model Mech* **3**, 525-534, doi:10.1242/dmm.006239 (2010).

44   Zheng, G. X. *et al.* Massively parallel digital transcriptional profiling of single cells. *Nat Commun* **8**, 14049, doi:10.1038/ncomms14049 (2017).

Chapter 6: Linking single-cell measurements of mass, growth rate, and gene expression

Mass and growth rate are highly integrative measures of cell physiology not discernable via genomic measurements. Here, we introduce a microfluidic platform that enables direct measurement of single-cell mass and growth rate upstream of highly-multiplexed single-cell profiling — e.g., single-cell RNA-sequencing. Applying our approach, we resolve transcriptional signatures associated with single-cell mass and growth rate in L1210 and FL5.12 cell lines. Further, we demonstrate a framework using these linked measurements to characterize biophysical heterogeneity of a patient-derived glioblastoma cell line in the presence or absence of drug. Our results highlight the value of coupled phenotypic metrics in guiding single-cell genomic analyses.

Key Contributions

RJK led the design and implementation of the platform as well as the experimental biophysical workflow. SMP led the single-cell RNA-Sequencing efforts. RJK and SMP analyzed the transcriptomic data.

Forward

The studies described thus far primarily utilize single-cell transcriptomics alone to profile cellular phenotype. However, other cellular measurements also provide important information about phenotype, such as genetics, epigenetics, protein expression, spatial orientation, intercellular communication, and biophysical measurements. Mass and growth rate of a cell are particularly powerful proxies for overall cellular health, which are useful when studying perturbations that deplete a population before gene expression measurements can be made. As discussed in Chapter 2, pairing these other measurements with transcriptomics could provide more accurate depictions of cellular behavior, especially following perturbations such as drug treatment. In this chapter, we describe the extension of a new microfluidic tool that can profile biophysical changes (mass/growth) to additionally profile gene expression of the same single cell. These multiplexed measurements enable interpretation of the presumed trajectory of a single cell during its differentiation, activation, or response to treatment (either towards sensitivity or resistance), and investigation of gene expression that underlies variation in these properties.

6.1: Introduction

Recent experimental advancements have dramatically improved the throughput and cost-efficiency of single-cell RNA-sequencing (scRNA-seq)[1-3]. However, gene expression measurements alone cannot fully describe many complex cellular processes[4]. Thus, parallel

efforts have focused on linking single-cell transcriptomics with complementary data that can provide further information to help guide analyses and contextualize distinct cellular states. For instance, various multi-omic methods have been developed to link measurements such as protein abundance, DNA sequence or methylation, with gene expression from the same single cell[5-8]. Gene expression measurements have also been linked to single-cell location within a tissue to enable the study of cellular development and differentiation with unprecedented detail[9-11]. Moreover, single-cell functional assays have been coupled with mRNA expression to obtain novel insights into the relationships among cellular electrophysiology, morphology, and transcription[12]. Taken together, these approaches demonstrate the value of linked single-cell data sets to afford a deep understanding of various cellular phenotypic states that may be difficult to glean through transcriptomic measurements alone.

Linked gene expression data sets are of particular interest when considering recent technological developments that have enabled the precise measurement of various single-cell biophysical properties, such as mass and growth rate[13,14]. As highly integrative metrics of cellular state, these parameters offer unique insights into a wide range of biological phenomena, including: **i)** basic patterns of single-cell mass and growth regulation; **ii)** biophysical changes associated with immune cell activation; and, **iii)** cancer cell heterogeneity in the presence or absence of drug[15-17]. However, the approaches and devices previously used to collect these biophysical measurements have precluded linking these properties with molecular profiling of the same cell.

To our knowledge, there have been no methods reported to date that allow for linked measurements of cellular mass, growth rate, and transcriptome-wide gene expression from the same cell. It has therefore been challenging to characterize the underlying transcriptional programs associated with cellular mass and growth rate variability observed in a range of normal and dysfunctional biological contexts.

Here, we describe and characterize a microfluidic platform that enables the measurement of single-cell mass and growth rate immediately upstream of a range of highly multiplexed single-cell endpoint assays. We leverage this approach in combination with scRNA-Seq to examine linked single-cell biophysical and transcriptomic properties in cell lines and primary cells. Finally, we further apply this method to examine biophysical heterogeneity in a patient-derived glioblastoma (GBM) cancer cell line in the presence or absence of drug, highlighting the potential utility of guiding single-cell genomic measurements with biophysical metadata.

Results

6.2: Serial SMR platform with downstream collection for scRNA-seq

Our system relies on a modified version of a previously described serial suspended microchannel resonator (sSMR) device (**Figure 6-1**) that utilizes an array of high-resolution single-cell buoyant mass sensors placed periodically along the length of a long microfluidic channel to allow a single cell's mass to be measured periodically as it traverses the channel[15]. In addition to providing mass information, this series of measurements can also be used to determine the mass accumulation rate (MAR), or growth rate, of each cell. Here, taking advantage of real-time access to the data generated by each SMR mass sensor, we have modified the system to use peak detection in the final cantilever. Detection at this cantilever indicates a cell exiting the mass sensor array and triggers the motion of a three-dimensional motorized stage which positions a PCR tube containing lysis buffer to capture each single cell as it is flushed from the system. This enables, for the first time, measurements of the biophysical properties of mass and growth rate to be linked to genomic profiles — here RNA-seq — at the single-cell level (**Methods**).
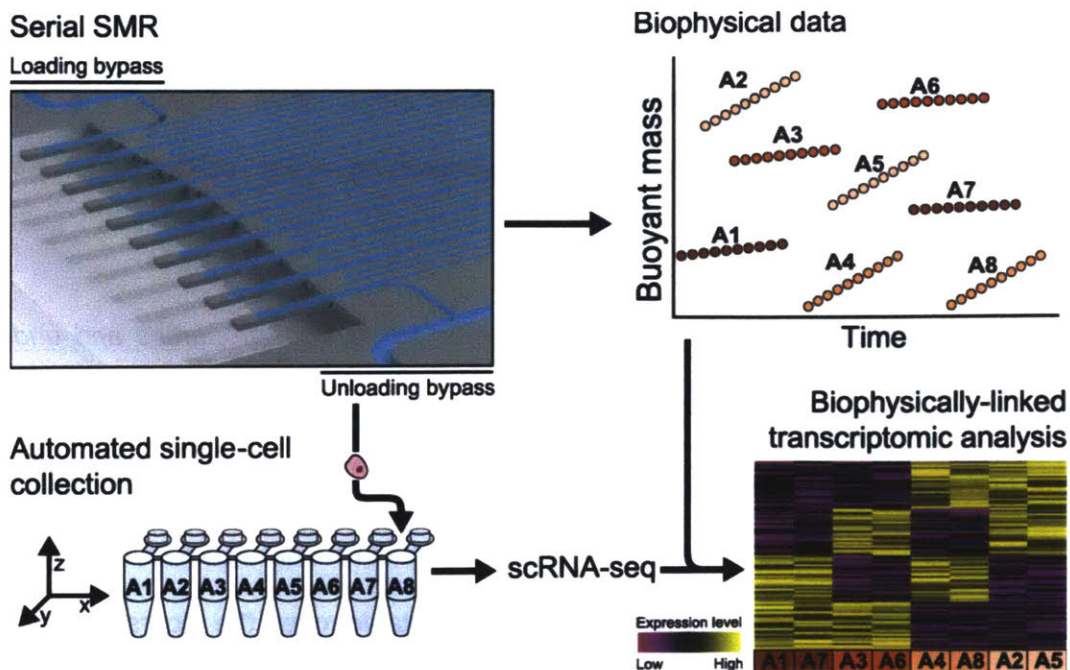
154

Figure 6-1: Serial SMR platform with downstream collection for scRNA-seq.
Schematic representation of the serial SMR platform, which includes an array of SMR mass sensors, separated by a serpentine delay channel to periodically measure the buoyant mass of a single cell. Independent control of the upstream and downstream pressures applied to two bypass channels allows for single-cell spacing at the loading entrance of the array (top left of sSMR image) and single-cell isolation at the unloading exit (bottom right of sSMR image) (**Supplementary Note 1**). Using real-time peak detection at the final mass sensor a three-dimensional motorized stage is triggered to capture each individual cell directly in to lysis buffer for downstream single-cell RNA-sequencing. Based on well location each cell is subsequently matched to its corresponding biophysical data collected from the sSMR including mass and MAR, as schematized in the top-right panel. These linked single-cell data sets can then be used to determine gene expression signatures associated with mass and growth rate variability, as schematized in the bottom-right panel.

We sought to endow our platform with sufficiently high throughput to enable measurements on populations of cells that may change over time. The total time required to flush the system's dead volume and release each single cell (20 seconds for the system implementation described here) sets a theoretical maximum throughput for the platform to avoid the collection of multiplets. Crucially, to minimize the frequency of failed capture events, we implemented a new fluidic scheme whereby single cells are loaded into the array of mass sensors at fixed intervals (**Note 6-1**)[18]. Ultimately, this fluidic scheme allows us to achieve a throughput of one cell approximately every thirty seconds (for a throughput of up to 120 cells per hour) with minimal

failed collection events due to co-release. This offers a two-fold throughput improvement over previous implementations of biophysical measurements alone, while affording the additional ability to capture each individual cell downstream for processing — e.g., scRNA-seq.

6.3: Unique gene expression profiles related to specific biophysical properties and underlying cell biology

To validate our method for collecting linked single-cell biophysical and gene expression data, we first measured two murine lymphoblast cell lines (L1210 and FL5.12) that have well-characterized mass and growth properties which are stable over the course of long-term propagation in bulk culture (**Figure 6-2**)[14-16,19]. Single cells collected downstream of the sSMR for scRNA-seq consistently yielded high-quality cDNA libraries, with 85 out of 87 individual L1210 cells and 124 out 144 individual FL5.12 cells with paired biophysical data passing initial quality controls (e.g., number of genes detected greater than 4,000, **Methods**).

In order to determine the transcriptional signatures associated with the spectrum of biophysical states in these cells, we ranked genes by how strongly their expression levels correlated with single-cell biophysical data (Spearman's correlation coefficients; **NB** Both Spearman and Pearson correlation methods yielded similar results for all comparisons considered). We then utilized the GSEA Preranked tool to determine which gene sets showed significant enrichment at either end of these ranked lists (FDR<0.05, **Methods**)[20]. For both cell lines, genes ranked by correlation strength with single-cell mass (final mass measurement collected before cell lysis) were highly enriched for functional annotations relating to cell cycle progression (FDR<0.05, **Figure 6-2**). Specifically, genes related to early cell cycle events immediately following cell division — such as DNA replication initiation — were more highly expressed in cells with lower masses, whereas genes related to late cell cycle events that occur just prior to division — such as chromosome segregation — were more highly expressed in cells with higher masses.

156

Interestingly, both cell lines revealed a larger number of genes that showed a significant positive correlation with mass relative to the number of genes with a significant negative correlation, though this may be impacted, in part, by the transcript capture inefficiencies inherent in scRNA-Seq protocols[21].
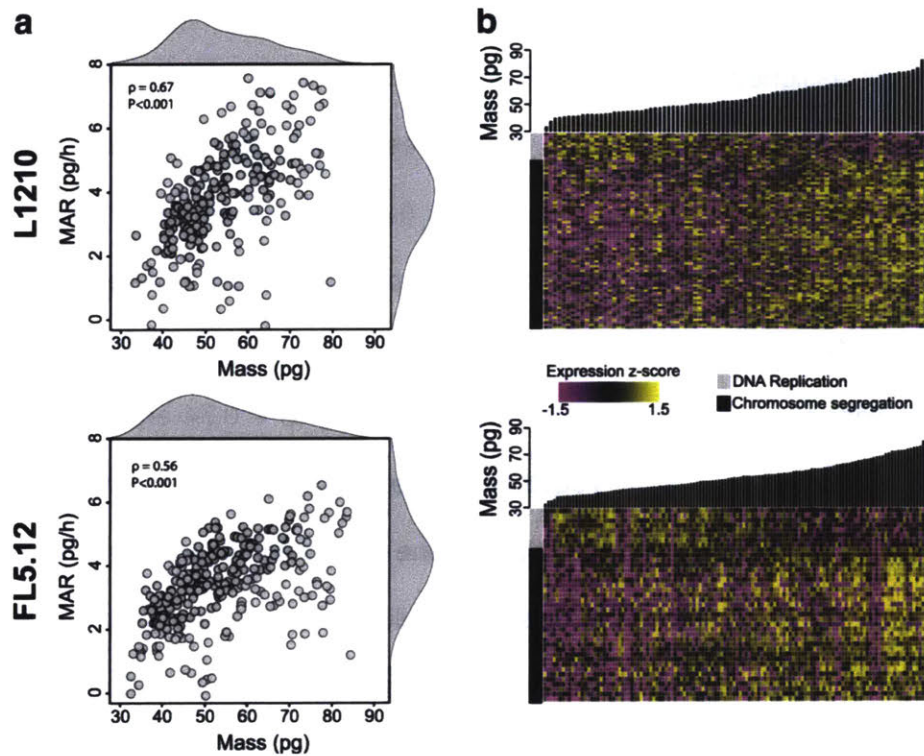

Figure 6-2: Linked biophysical & gene expression measurements of L1210 and FL5.12 cells. (a) Plot of mass accumulation rate versus buoyant mass for single L1210 cells (top, n = 234) and single FL5.12 cells (bottom, n=296) measured in the sSMR. Kernel density plots are included on both axes. **(b)** Heat maps showing the relative expression of various cell cycle-related genes for subsets of the L1210 (top, n=85) and FL5.12 (bottom, n=124) cells depicted in **(a)** that were captured downstream for scRNA-seq. Cells are ordered by buoyant mass (bar plots above heat maps). Entries are colored by row-wise expression level rank where the cell with the highest expression level for a particular gene corresponds to yellow and the cell with the lowest expression level corresponds to magenta. As a demonstration, the heat map includes genes with expression levels that showed a significant correlation with buoyant mass from the chromosome segregation (black bar, n=58 and n=31 for the L1210 and FL5.12, respectively) and DNA replication (gray bar, n=11 and n=8 for the L1210 and FL5.12, respectively) gene ontology subsets (FDR<0.05, **Methods**).

The manifestation of cell cycle-related gene expression in scRNA-seq data has been of particular interest to both further characterize mechanisms of cell cycle progression and regress

out the contributions of cell cycle variability that may act as a nuisance parameter, dominating gene expression heterogeneity amongst single cells and masking more subtle biological signals[3,22,23]. We therefore sought to determine how previously annotated cell cycle signatures corresponded to the single-cell mass measurements collected here. To do so, we established cell cycle phase specific (G1/S and G2/M) scores using an approach inspired by Macosko et al[3] (**Figure 6-3**). Cells assigned to either the G1/S or G2/M phases of the cell cycle using gene expression data alone showed significant differences in cell mass for both the L1210 and FL5.12 cells that were consistent with expectations (i.e., lower mass for G1/S cells; P<0.001, Mann-Whitney U Test). Furthermore, for both cell types, cell mass showed a clear negative correlation with G1/S scoring (ρ = -0.46 and ρ = -0.25 for L1210 and FL5.12, respectively; P<0.005) and a clear positive correlation with G2/M scoring (ρ = 0.74 and ρ = 0.54 for L1210 and FL5.12, respectively; P<0.001). Together, these results provide additional evidence of coordination between mass and cell cycle related gene expression in actively proliferating cells.

To further confirm the consistency and reproducibility of the linked biophysical and gene expression data sets collected with this platform, we compared the L1210 and FL5.12 results with scRNA-seq data from additional independent experiments. For L1210 cells, we found that genes that showed significant correlations with cell mass were also significantly enriched amongst those previously shown to correlate with time since cell division, a proxy for cell cycle progression (FDR<0.05)[24]. In FL5.12 cells, meanwhile, we observed that the genes which showed significant correlations between their expression levels and biophysical properties were highly reproducible across two independent linked biophysical and gene expression experiments (FDR<0.05). These results demonstrate the quality and reproducibility of transcriptional measurements collected downstream of the sSMR.
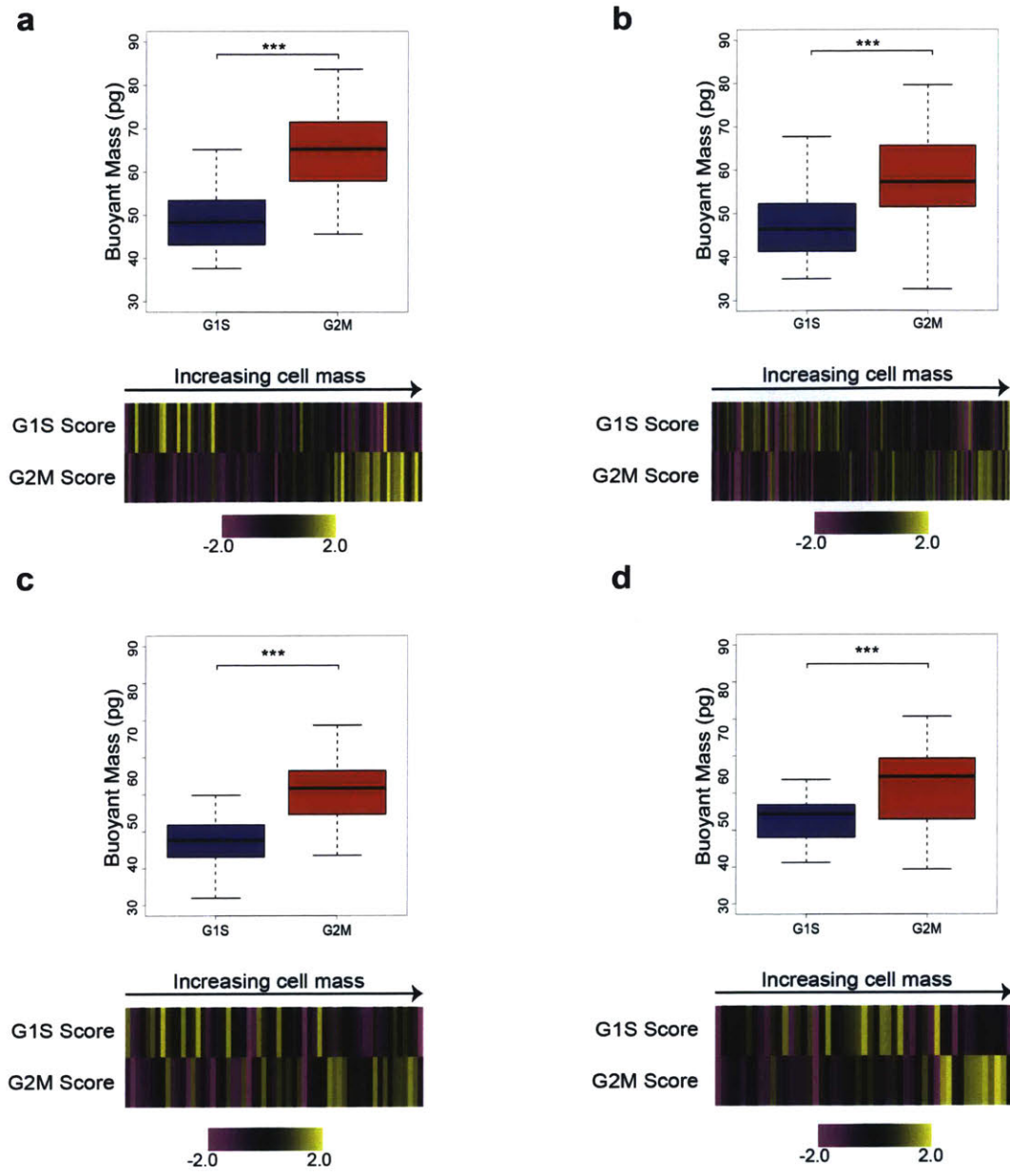
Figure 6-3: Comparison of cell cycle gene expression and cell mass. Boxplots (top) showing the mass distribution of single-cells classified as being in the G1/S or G2/M phases of the cell cycle for **(a)** L1210 cells (n = 48 and 37 for G1/S and G2/M, respectively), **(b)** FL5.12 (n = 63 and 61 for G1/S and G2/M, respectively), **(c)** CD8+ T cells after 24h of activation (n = 25 and 34 for G1/S and G2/M, respectively), and **(d)** CD8+ T cells after 48h of activation (n = 25 and 24 for G1/S and G2/M, respectively) (*** indicates P<0.001, Mann-Whitney U Test). Below each boxplot is a heatmap showing the G1/S and G2/M expression scores for cells ranked by buoyant mass.

Given that we identified a linear relationship between mass and MAR in these cell types (ρ = 0.67 and ρ = 0.56 for L1210 and FL5.12, respectively; P<0.001, **Figure 6-2**), we focused our

analysis on mass-normalized MAR, determined by dividing each cell's MAR by its corresponding mass. We used this parameter, which measures a single cell's growth efficiency decoupled from mass-related confounders, to resolve growth-related transcriptional signatures in these two cell lines[25]. For L1210 cells, genes ranked by strength of correlation between expression level and growth efficiency did not reveal any statistically significant enrichment of functional annotations (FDR>0.05). The FL5.12 cells, however, showed significant positive enrichment for functional annotations related to cell cycle progression amongst genes ranked by correlation strength with growth efficiency (FDR<0.05). Specifically, subsets of genes implicated in the G1-S transition showed a higher level of expression in cells of intermediate mass with the highest growth efficiencies (**Methods**)[26]. These results are consistent with previous FL5.12 single-cell growth measurements, which revealed an increase in growth efficiency approaching the G1-S transition followed by a decrease later in the cell cycle[14].

6.5: Characterizing single-cell biophysical heterogeneity of patient derived cancer cell lines

Cancer cell drug responses are known to be highly heterogeneous at the single-cell level[17,25], and it is now well established that the presence of even a small fraction of cells that are unresponsive to a therapy can lead to resistance and recurrence of cancers[33]. Single-cell transcriptional profiling has been shown to provide a powerful means of characterizing such heterogeneity in clinically relevant tissue samples[34,35], yet the direct interrogation of drug response is still most commonly measured in clinical trials and the laboratory using bulk viability assays[36]. Although effective in quantifying the relative fraction of resistant cells within a heterogeneous population, these assays rely on endpoint measurements. Taken too late, they may miss responding cells (which are lost to cell death) and/or the preceding molecular events that impact survival; taken too early, bulk measurements can muddle the features of responding and non-responding cell subsets. (**Figure 6-4 a**). However, we have previously shown that, prior

160

to viability loss, single-cell biophysical changes of mass and MAR collected with the SMR can predict response to drug treatment[17]. Therefore, we reasoned that downstream molecular characterization could be used to further contextualize single-cell mass and growth rate heterogeneity both at baseline and in response to perturbation with drug treatment.

To demonstrate a framework for the characterization of single-cell biophysical heterogeneity in the presence or absence of drug, we decided to measure the effect of an MDM2 inhibitor (RG7388, Roche) on BT159 cells, a patient derived cell line (PDCL) generated from a primary glioblastoma (GBM) (**Methods**). GBM PDCLs are known to be particularly heterogeneous with respect to cell lineage and have a cancer stem cell like hierarchy proposed to contribute to profound treatment resistance of these tumors[37]. MDM2, meanwhile, typically binds to p53 inhibiting its transcriptional activity and leading to proteasome-mediated degradation[38]. In prior work from our group, pharmacologic inhibition of MDM2 has proven to be a promising therapeutic avenue in GBM patients with wild-type TP53, due to the fact that, in preclinical patient derived models, the drug leads to increased expression and stability of p53, significant responses, and even tumor regression via induction of apoptotic cell death[39]. However, *in vivo* testing revealed that, upon withdrawal of MDM2 inhibition, tumors consistently relapsed, suggesting variable response to treatment[39].
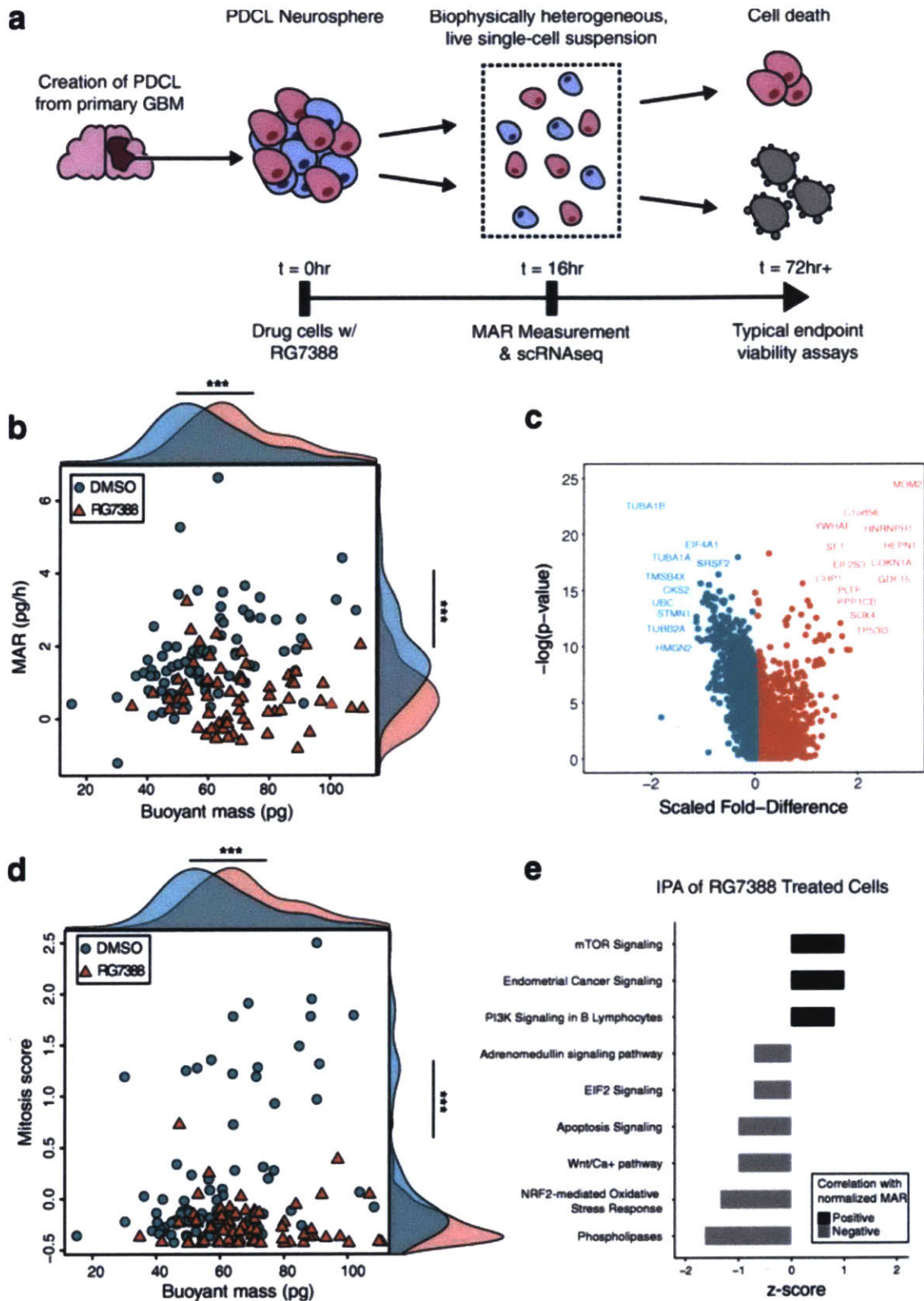
Figure 6-4: Characterizing single-cell drug response in BT159 GBM cells. (a) Schematic representation of GBM PDCL generation, drug treatment *in vitro*, and subsequent

characterization therapeutic response using the sSMR collection platform. Mass and growth measurements are collected after 16h of treatment, prior to loss of cell viability, which enables downstream molecular characterization with scRNA-seq (**Methods**). **(b)** Plot of single-cell MAR versus mass for BT159 GBM cells treated with either DMSO (blue circles, n = 83) or RG7388 (an MDM2 inhibitor, red triangles, n = 66) for 16h. Kernel density plot, using the same color scheme, are include in the margins for both populations. *** Indicates P<0.001, Mann-Whitney U Test. **(c)** Volcano plot showing log-transformed average expression fold change and log-transformed P-values (Bonferroni corrected) for genes upregulated (red) or downregulated (black) in BT159 cells treated with RG7388 as compared with DMSO treatment. **(d)** Plot of mitosis scores versus buoyant mass for BT159 cells treated with DMSO (blue circles, n = 83) or RG7388 (red triangles, n = 66) for 16h. Mitosis scores were calculated by taking the average z-score adjusted gene expression values of a panel of mitosis-related genes (n = 29, **Methods**). Kernel density plot, using the same color scheme, are include in the margins for both populations. *** Indicates P<0.001, Mann-Whitney U Test. **(e)** Plot of significantly enriched canonical pathways (FDR<0.05) in RG7388 treated BT159 cells (n = 66), as determined by Ingenuity Pathway Analysis, amongst genes with significant positive (black) or negative (gray) correlations with normalized MAR. (**Methods**).

To characterize biophysical heterogeneity at the single-cell level we collected linked mass, MAR and gene expression measurements for single BT159 cells that had either been treated for 16h with RG7388 or DMSO (control) (**Methods**). Overall, the drug treated population of cells showed a marked reduction in average MAR and an increase in average mass as compared to the control population of cells, as expected due to cell cycle exit and apoptosis (P<0.001, Mann-Whitney U Test, **Figure 6-4 b**). However, there was also considerable heterogeneity in biophysical response to drug, with some cells continuing to show a positive MAR at the time of measurement (**Figure 6-5**). Since these measurements were collected at a single time point, it is difficult to assess whether the cells that continue to grow in the presence of drug are, in fact, resistant to therapy or simply display a delayed response to treatment. Nonetheless, the biophysical heterogeneity found in these results affords the opportunity to determine transcriptional signatures that correlate with this variability at this particular time point.

We next considered only the transcriptional data. As expected, an unbiased analysis (dimensionality reduction by principal components analysis (PCA) and visualization using a t-stochastic neighbor-embedding (tSNE) plot, **Methods**) revealed distinct transcriptional profiles for drug-treated and control cell populations (**Figure 6-6 a**). Relative to DMSO-treated cells,

drug-treated cells displayed gene expression signatures consistent with the mechanism of MDM2 inhibition, with genes positively regulated by p53, such as *CDK1NA* (p21) and *MDM2*, showing significant upregulation, and genes negatively regulated by p53, such as *CDK1* and *CDC20*, showing significant downregulation (Bonferroni-corrected P<0.05, **Figure 6-4 c**)[40]. We then performed dimensionality reduction (PCA) and graph-based clustering (k-nearest neighbors, KNN) on the transcriptional data from the drug-treated cells alone, and visualized our results using a tSNE plot (**Figure 6-6 b**, **Methods**). This clustering analysis did not reveal any clear subsets of drug-treated cells with distinctly different responses to MDM2 inhibition.
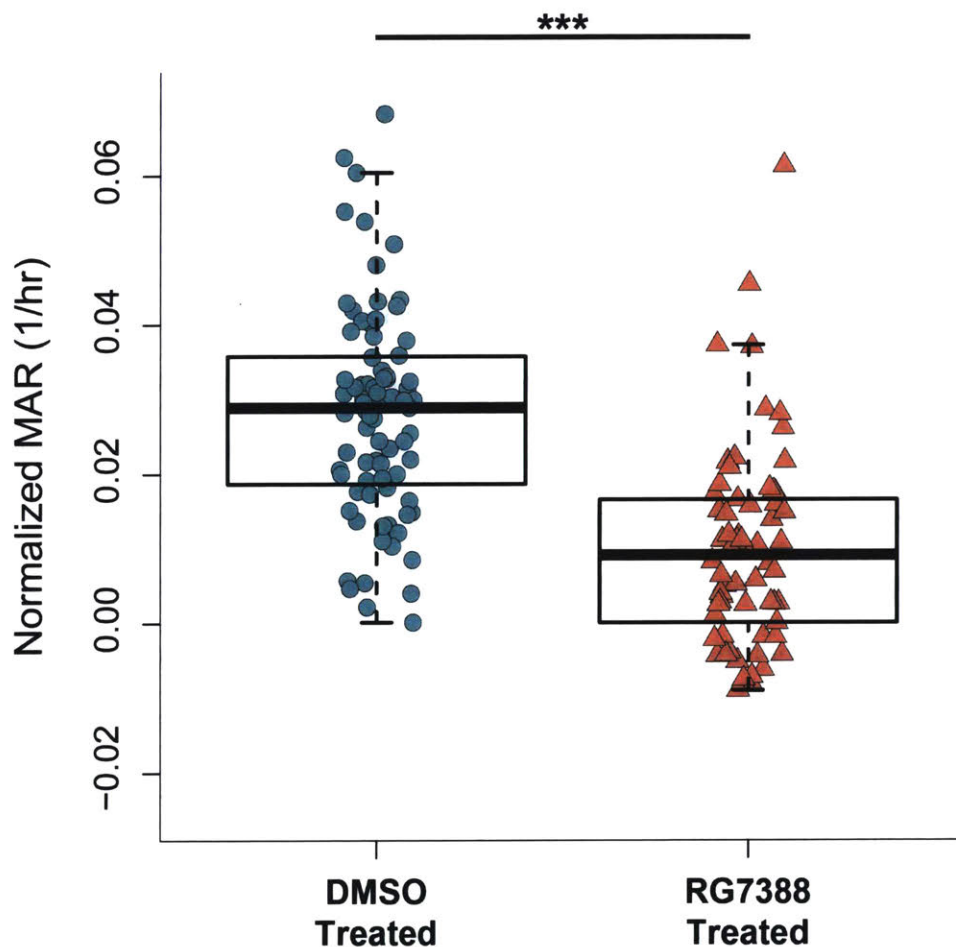

Figure 6-5: Mass-normalized MAR measurements for BT159 cells. Single-cell mass-normalized MAR measurements and corresponding boxplots for BT159 cells treated for 16h with either DMSO (left, n= 83) or RG7388 (right, n = 66). *** indicates P <0.001, Mann-Whitney U-test.
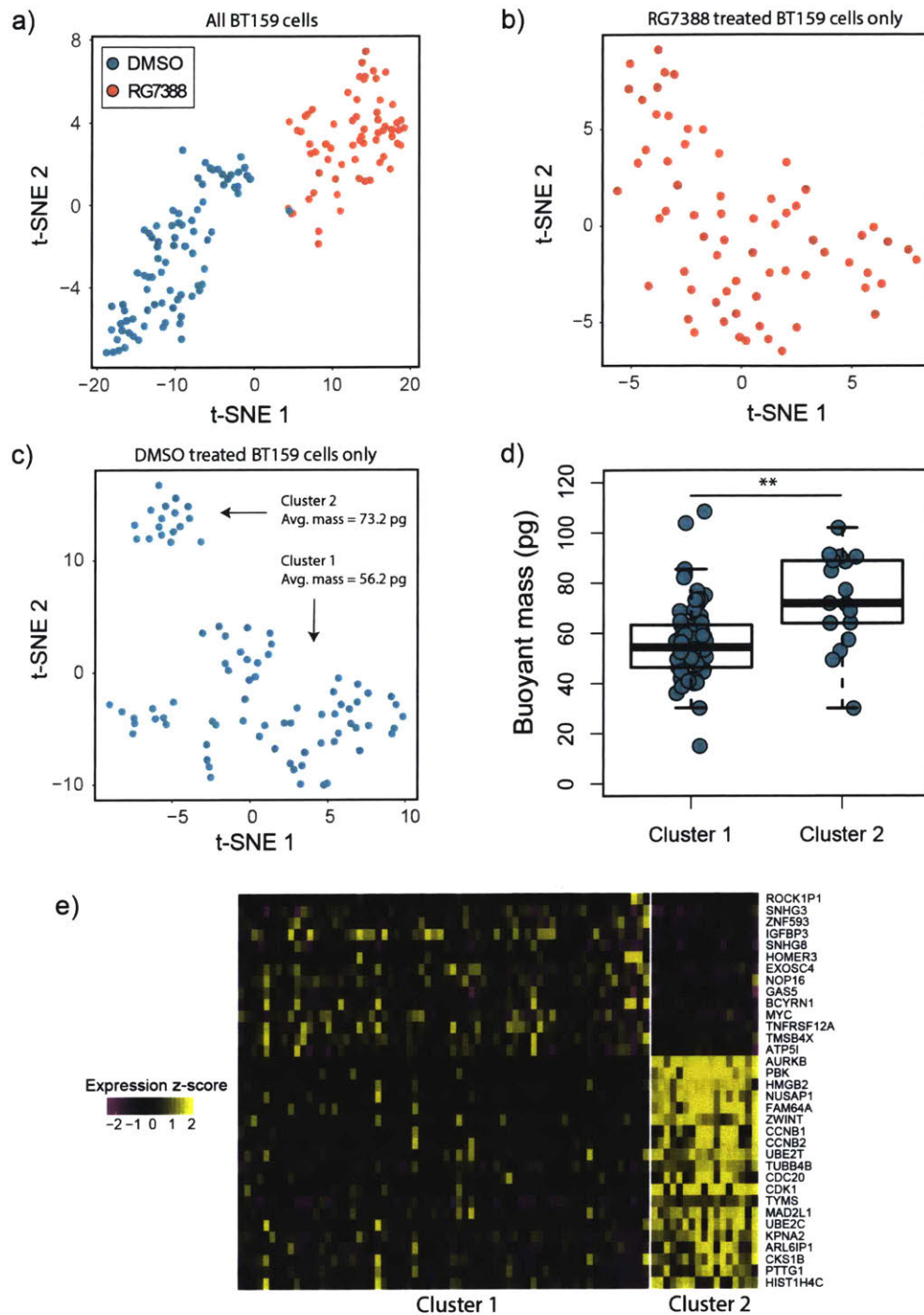
Figure 6-6: *t*-SNE analysis of BT159 cells. (a) t-SNE plot of both DMSO treated (n = 83, red points) and RG7388 treated (n = 66, blue points) BT159 cells analyzed together. **(b)** t-SNE plot of RG7388 treated BT159 cells alone. **(c)** t-SNE plot of DMSO treated BT159 cells alone. Annotations indicate the two distinct subpopulations identified by KNN clustering over significant PCs and the corresponding average masses of cells within these clusters. **(d)** Single-cell mass measurements and corresponding boxplots for two clusters (n = 66 and n = 17 for Cluster 1 and Cluster 2, respectively) of DMSO treated BT159 cells identified in **(c)**. ** Indicates P<0.01, Mann-Whitney U-test. **(e)** Heatmap showing z-score scaled expression values for genes

defining the two clusters depicted in **(c)**. Cluster 2, which is composed of cells with a significantly higher mass (P<0.01, Mann-Whitney U-test), shows increased expression of various genes relating to cell cycle progression.

Since our transcriptional measurements suggested that all MDM2-inhibitor treated cells were actively experiencing drug but our biophysical measurements revealed mass and MAR heterogeneity, we decided to explicitly examine whether the linked nature of our measurements could be used to shed light on the drivers of biophysical variability at this time point after treatment with DMSO or RG7388. When examining linked measurements of gene expression and cell mass in DMSO treated cells, we found that genes ranked by correlation strength with mass were highly enriched for functional annotations relating to cell cycle progression. Also, as with the other cell types presented here, larger cells in the control population expressed a higher level of genes associated with late cell cycle events, specifically mitosis **(Figure 6-4 d)**. Interestingly, an unsupervised clustering analysis (PCA followed by KNN clustering, **Methods**) of the DMSO-treated cells alone revealed two distinct subsets which had significantly different average masses (P<0.01, Mann-Whitney U-test, **Figure 6-6 d**), and an upregulation of genes relating to cell cycle progression in the subset with a larger average mass **(Figure 6-6 e)**.

MDM2 inhibitor-treated cells, meanwhile, showed significantly reduced expression of mitosis-specific genes (P<0.001, Mann-Whitney U Test, **Figure 6-4 d**). Moreover, in these cells, we did not observe any significant cell cycle-related functional enrichments among those genes correlated with cell mass (FDR>0.05). These results demonstrate that upon MDM2 inhibition and stabilization of p53 signaling in these cells, cell cycle arrest is achieved as expected but there is no longer a correlation between cell mass and cell cycle-related gene expression (ρ = 0.47, P<0.001 for DMSO-treated cells; ρ = -0.07, P=0.54 for drug-treated cells). Furthermore, since a subset of cells within the drug-treated population displayed a positive MAR despite ablated cell cycle gene expression **(Figure 6-4 b)**, our data suggest that cell cycle gene expression alone does not fully account for variability in the single-cell biophysical response. In

166

fact, we did not observe a significant correlation between PCs computed for the drug-treated single-cell transcriptomes and any biophysical properties measured (P>0.05; **Methods**).

To determine transcriptional signatures that may underlie this biophysical heterogeneity, we utilized the corresponding single-cell MAR data to further contextualize gene expression. Genes ranked by correlation strength with mass-normalized MAR in the MDM2-inhibitor treated population of cells showed a significant negative enrichment (i.e. higher expression in cells accumulating less mass over time) for functional annotations related to apoptosis regulation, specifically related to p53 signaling (FDR<0.05). The DMSO treated population of cells, meanwhile, did not show any significant functional enrichments amongst genes ranked by correlation with normalized-MAR (FDR>0.05). Similarly, Ingenuity Pathway Analysis (IPA, Qiagen) performed on drug treated cells revealed significant enrichment of canonical apoptosis signaling amongst genes showing significant negative correlations with normalized MAR (FDR<0.05, **Figure 6-4 e**) while the same analysis on DMSO treated cells did not reveal any apoptosis-related signaling significantly correlated with MAR (FDR>0.05, **Figure 6-7**). Together, these results suggest that cells with a higher normalized MAR had a lower expression of genes related to apoptotic signaling orchestrated by p53, but only in drug treated cells, consistent with the mechanisms of MDM2. IPA of drug treated cells further revealed significant enrichment for PTEN signaling (a negative regulator of AKT) and mTOR signaling (a positive regulator of AKT) amongst genes significantly negatively and positively correlated with normalized MAR, respectively (FDR<0.05)[41,42]. IPA of DMSO treated cells, however, revealed significant enrichment for PTEN signaling (FDR<0.05) in genes negatively correlated with normalized MAR but did not show significant enrichment for mTOR signaling in genes positively correlated with normalized MAR (**Figure 6-7**). Together, these results suggest that cells which continue to grow in the presence of MDM2 inhibition may exhibit more stable AKT signaling, which itself drives MDM2 expression, as compared with cells with decreased normalized MAR, pointing to a

potential mechanism of cell survival in the presence of treatment[43,44]. Though preliminary, these results demonstrate the unique insight offered by linked measurements of biophysical phenotype and gene expression when examining cancer cell drug response at the single-cell level.
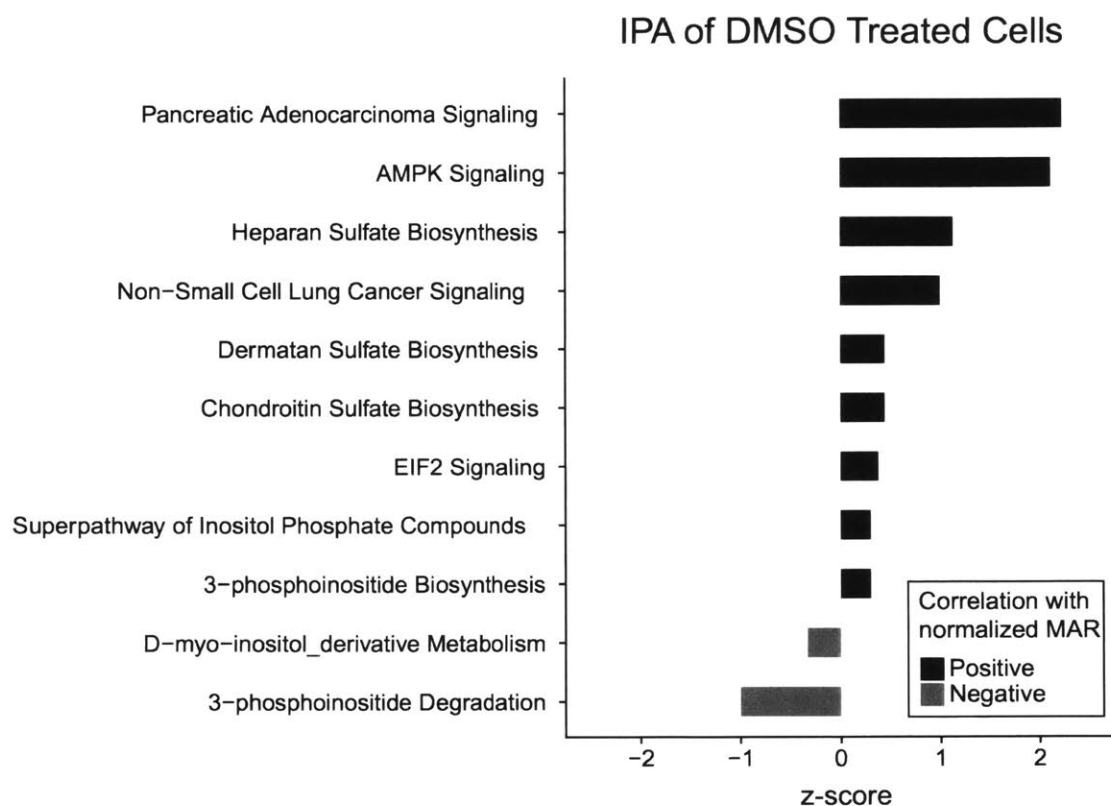
IPA of DMSO Treated Cells

Figure 6-7: IPA of DMSO treated BT159 cells. Plot of significantly enriched canonical pathways (FDR<0.05) in DMSO treated BT159 cells (n = 83), as determined by Ingenuity Pathway Analysis, amongst genes with significant positive (black) or negative (gray) correlations with normalized MAR. (**Methods**).

6.6: Conclusion

The platform presented here enables linked measurements of single-cell biophysical properties and gene expression. Having demonstrated the resolution and reproducibility of these linked data sets with measurements of stable cell lines (L1210 and FL5.12 cells), we present

frameworks for two key applications of these linked data sets: **i)** characterizing immune cell activation and differentiation, and **ii)** examining cancer cell drug response at the single-cell level.

While the primary focus of this work was on conducting scRNA-seq downstream of the sSMR, we also envision this platform being a useful tool for linking biophysical data with other recently developed approaches that enable DNA sequencing, epigenomic characterization, or multi-omic measurements of single cells[5,6,45].

We believe that these linked measurements will offer a novel means of exploring a range of biological questions. For instance, when paired with recently developed computational approaches, these linked biophysical and transcriptional measurements may offer insights into cell cycle regulation as well as provide an additional approach for addressing the potentially confounding effects of cell cycle in scRNA-seq analyses[22]. Clinically, mass and MAR have proven to be effective biomarkers for characterizing cancer cell drug susceptibility at the single-cell level[17,25]. The ability to link these biophysical measurements with gene expression or genetic profiling offers the exciting opportunity to move beyond the simple classification of responding and non-responding cells and begin to explore the molecular mechanisms that may drive such behaviors. We envision that this and related approaches may one day inform more effective precision medicine pipelines[46].

Cell culture and primary cell preparation. L1210 murine lymphocytic leukemia cells (ECACC) were cultured in RPMI 1640 (Gibco) with 10% fetal bovine serum and 1% antibiotic-antimycotic (Gibco). FL5.12 murine pre-B cells (gift from the Vander Heiden Lab, MIT) were cultured in the same media with the addition of 10 ng/ml IL-3 (R&D Systems). For all growth and collection experiments, cells were passaged to a concentration of 5×10^5 cells/ml the night before to ensure consistent culture confluence at time of measurement.

The GBM PDCL BT159 was generated using patient tissue collected under an informed consent protocol (Dana Farber Harvard Cancer Center protocol #10-417) approved by Dana Farber Harvard Cancer Center and Partner's Human Research Center institutional review boards. Cells were harvested from excess tissue resection specimens through cycles of enzymatic (neural tissue dissociation kit with papain, Miltenyi Biotec) and mechanical dissociation in a tissue grinder (gentleMACS dissociator, Miltenyi Biotec). Cells were grown as tumorspheres in NeuroCult NS-A proliferation media (Miltenyi Biotec) supplemented with 2 µg/ml Heparin, 20 ng/ml human epidermal growth factor (EGF), 10 ng/ml human bFGF in ultra-low attachment coated flasks (Corning). Prior to measurement, the BT159 cells were dissociated with Accutase (Sigma-Aldrich) at 37 °C for 7 min. For drug experiments, cells were treated with 250 nM of the MDM2 inhibitor RG7388 (Roche) or DMSO for 16h prior to dissociation for measurement.

Single-cell growth measurements and collection. For all experiments, cells were adjusted to a final concentration of 2.5×10^5 cells/ml to load single cells into the mass sensor array as described in **Supplementary Note 1**. Single-cell growth measurements were conducted as described previously [15]. In order to exchange buffer and flush individual cells from the system, the release side of the device was constantly flushed with PBS at a rate of 15 µL per minute (P2 to P4). Upon detection of a single cell at the final cantilever of the sSMR, as indicated by a

170

supra-threshold shift in resonant frequency, a set of 3-dimensional motorized stages (ThorLabs) was triggered to move a custom PCR-tube strip mount from a waste collection position to a sample collection position. The location of these motors was written to a file for the duration of the experiment in order to annotate single-cell mass and MAR measurements with well position, and thus transcriptional profiles, downstream. Each cell was collected in 5 µl of PBS directly in to a PCR tube containing 5 µl of 2X TCL lysis buffer (Qiagen) with 2% v/v 2-mercaptoethanol (Sigma) for a total final reaction volume of 10 µl. After each 8-tube PCR strip was filled with cells, the strip was spun down at 1000g for 30 seconds and placed immediately on dry ice. Following collection, samples were stored at -80 C prior to library preparation and sequencing.

scRNA-Seq. Single-cell RNA isolation, cDNA library synthesis, next generation sequencing, read alignment and gene expression estimation were performed as described previously[47]. Briefly, Smart-Seq2 whole transcriptome amplification and library preparation were performed on single-cell lysates collected with the sSMR[48]. Single-cell libraries were then sequenced on an Illumina NextSeq 500 using 30-bp paired end reads. Data was filtered to exclude cell doublets or cells with failed matching of masses for growth rate measurement. This step left 87 out of 96 L1210 cells, 144 out of 192 FL5.12 cells, and 181 out of 192 BT159 GBM cells. Next, cells that exceeded a preliminary complexity threshold (4,000 genes for L1210 and FL5.12 cells or 1,000 genes for BT159 cells) were selected for further analysis. Overall, this yielded 85 out of 87 L1210 cells, 124 out of 144 FL5.12 cells, and 149 out of 192 BT159 cells. These cells selected for analysis were sequenced to an average depth of $1,698,879 \pm 106,027$ (s.e.m.) reads for L1210 cells, $760,919 \pm 36,679$ (s.e.m.) reads for FL5.12 cells, and $993,629 \pm 75,796$ (s.e.m.) reads for BT159 cells respectively. Reads were aligned using TopHat2 and expression estimates (transcripts per million; TPM) for all UCSC-annotated mouse genes (mm10, for L1210, FL5.12) or human genes (hg19, for BT159 cells) were calculated using RNA-seq by

expectation maximization (RSEM)[49,50]. The average transcriptome alignments were 67.4 $\pm$ 0.38

% (s.e.m.) for L1210 cells, 64.8$\pm$ 0.51 % (s.e.m.) for FL5.12 cells, and 35.2 $\pm$ 0.84% (s.e.m.) for

BT159 cells. The average number of genes detected was 7,207 $\pm$ 94 (s.e.m.) for L1210 cells,

6,891 $\pm$ 81 (s.e.m.) for FL5.12 cells, and 5,347 $\pm$ 173 (s.e.m.) for BT159 cells.

Gene expression analysis. All analysis was performed on log-transformed expression level

measurements (ln(TPM+1)). Data pre-processing was conducted with the Seurat package for

R[9]. All genes that were detected in >5% of cells were included in the final analysis for each

group of cells (L1210, FL5.12, and BT159 GBM cells).

Significance-testing. To define the null distribution of correlation coefficients, we determined

the Spearman correlation between cell cycle gene expression levels and mass for randomly

shuffled data sampled from experimental values (i.e., mismatching single-cell mass and gene

expression data). After 10,000 iterations, we used the average mean and standard deviation

values of these correlation coefficient distributions to define the null distributions presented.

We computed the null distributions for the correlation coefficients between either mass, MAR, or

normalized MAR and the principal components for either the DMSO-treated, drug-treated, or

combined transcriptomic data sets using a similar random shuffling of PC coordinates across

single-cells. Following 10,000 iterations, the mean and standard deviation of these distributions

was compared to the correlation of each biophysical parameter with all significant principal

components (PCs). For each data set, the PCElbow plot and jackstraw functions in Seurat were

used to select significant PCs whose explained variation preceded a precipitous drop in

cumulative explained variation (elbow). In each data set, for consistency, the top 10 PCs were

investigated, although in some cases fewer than 10 PCs preceded the elbow. Correlation

172

coefficients were deemed insignificant if they were within two standard deviations of the mean determined from random shuffling.

Gene set enrichment analysis. Ranked gene lists were created for each cell population by determining the gene-wise correlation coefficient (Spearman) between log-transformed gene expression levels and either single-cell mass or growth efficiency (MAR/mass). Spearman and Pearson correlation coefficients yielded similar results for all conditions measured. Gene set enrichment was computed for these ranked lists using the GSEA Preranked tool, implemented with the fgsea package in R[20,51].

Differential expression. Differential expression analysis for the DMSO versus RG7388 treated BT159 cells, was performed using FindMarkers in Seurat with the Wilcoxon rank sum test. All P values presented are Bonferroni corrected, as per Seurat documentation recommendation.

Dimensionality reduction. Variable genes for DMSO-treated, drug-treated, and combined data sets were identified using Seurat's FindVaribleGenes. Principal components analysis (PCA) was performed over these genes for each of the three sets of cells, followed by non-linear dimensionality reduction by t-stochastic neighbor embedding (tSNE). Clusters were identified in the linear PC space using K-nearest neighbor (KNN) clustering, and cluster assignments were visualized on the non-linear tSNE space. For the DMSO-treated cells, we detected two distinct clusters (**Figure 6-6 c**); for the RG7388 treated cells, we only detected one (**Figure 6-6 b**).

Ingenuity pathway analysis. Ingenuity pathway analysis (IPA, Qiagen) was performed on canonical pathways using genes which significantly correlated positively and negatively with

normalized MAR. Briefly, correlation and p-values for significant genes were uploaded into IPA and analyzed using the "Core Analysis" function. Correlations were input as "Expression: Other" measurements with range from -INF to INF. Significant canonical pathways and upstream regulators (determined by hypergeometric test) with positive and negative z-scores are plotted in **Figure 6-4 e**.

Note 6-1 | Maintaining minimum cell spacing in mass sensor array

Loading single cells into the mass sensor array at fixed, minimum spacing requires the implementation of active switching between two distinct fluidic states. Initially, equivalent pressures are applied to upstream and downstream ports on the bypass channel leading in to the array (ports P1 and P3). In this "loading" configuration, all streamlines are directed into the array and therefore cells in the bypass channel will enter the array. An imaging region at the entrance to the mass sensor array is used as an indication of when a cell has been successfully loaded. Real-time optical peak detection is used to switch from this loading fluidic state to a "flushing" regime wherein the upstream pressures (P1) is increased and the downstream pressure (P3) is decreased such that a vast majority of streamlines continue along the bypass channel with a small fraction entering the array. Because cells are of finite size and occupy several streamlines, they are directed along the bypass channel and not drawn in to the array. Importantly, during this process the pressure at the entrance to the mass sensor array is maintained at a fixed value, therefore any cells that have entered the array continue to flow at a constant speed. Therefore, although the volumetric flow rate is maintained across the array while flushing, no additional cells are loaded. After a desired amount of time has elapsed the system is automatically returned to the loading configuration to obtain the next cell for measurement.

6.8: References

1   Gierahn, T. M. *et al.* Seq-Well: portable, low-cost RNA sequencing of single cells at high throughput. *Nat Methods* **14**, 395-398, doi:10.1038/nmeth.4179 (2017).

2   Klein, A. M. *et al.* Droplet barcoding for single-cell transcriptomics applied to embryonic stem cells. *Cell* **161**, 1187-1201, doi:10.1016/j.cell.2015.04.044 (2015).

3   Macosko, E. Z. *et al.* Highly Parallel Genome-wide Expression Profiling of Individual Cells Using Nanoliter Droplets. *Cell* **161**, 1202-1214, doi:10.1016/j.cell.2015.05.002 (2015).

4   Easwaran, H., Tsai, H. C. & Baylin, S. B. Cancer epigenetics: tumor heterogeneity, plasticity of stem-like states, and drug resistance. *Mol Cell* **54**, 716-727, doi:10.1016/j.molcel.2014.05.015 (2014).

5   Genshaft, A. S. *et al.* Multiplexed, targeted profiling of single-cell proteomes and transcriptomes in a single reaction. *Genome Biol* **17**, 188, doi:10.1186/s13059-016-1045-6 (2016).

6   Dey, S. S., Kester, L., Spanjaard, B., Bienko, M. & van Oudenaarden, A. Integrated genome and transcriptome sequencing of the same cell. *Nat Biotechnol* **33**, 285-289, doi:10.1038/nbt.3129 (2015).

7   Angermueller, C. *et al.* Parallel single-cell sequencing links transcriptional and epigenetic heterogeneity. *Nat Methods* **13**, 229-232, doi:10.1038/nmeth.3728 (2016).

8   Stoeckius, M. *et al.* Simultaneous epitope and transcriptome measurement in single cells. *Nat Methods* **14**, 865-868, doi:10.1038/nmeth.4380 (2017).

9   Satija, R., Farrell, J. A., Gennert, D., Schier, A. F. & Regev, A. Spatial reconstruction of single-cell gene expression data. *Nat Biotechnol* **33**, 495-502, doi:10.1038/nbt.3192 (2015).

10   Achim, K. *et al.* High-throughput spatial mapping of single-cell RNA-seq data to tissue of origin. *Nat Biotechnol* **33**, 503-509, doi:10.1038/nbt.3209 (2015).

11   Stahl, P. L. *et al.* Visualization and analysis of gene expression in tissue sections by spatial transcriptomics. *Science* **353**, 78-82, doi:10.1126/science.aaf2403 (2016).

12   Cadwell, C. R. *et al.* Electrophysiological, transcriptomic and morphologic profiling of single neurons using Patch-seq. *Nat Biotechnol* **34**, 199-203, doi:10.1038/nbt.3445 (2016).

13   Godin, M. *et al.* Using buoyant mass to measure the growth of single cells. *Nat Methods* **7**, 387-390, doi:10.1038/nmeth.1452 (2010).

14   Son, S. *et al.* Direct observation of mammalian cell growth and size regulation. *Nat Methods* **9**, 910-912, doi:10.1038/nmeth.2133 (2012).

15   Cermak, N. *et al.* High-throughput measurement of single-cell growth rates using serial microfluidic mass sensor arrays. *Nat Biotechnol* **34**, 1052-1059, doi:10.1038/nbt.3666 (2016).

16   Son, S. *et al.* Cooperative nutrient accumulation sustains growth of mammalian cells. *Sci Rep* **5**, 17401, doi:10.1038/srep17401 (2015).

17   Stevens, M. M. *et al.* Drug sensitivity of single cancer cells is predicted by changes in mass accumulation rate. *Nat Biotechnol* **34**, 1161-1167, doi:10.1038/nbt.3697 (2016).

18   Calistri, N. L. *et al.* Microfluidic active loading of single cells enables analysis of complex clinical specimens. *Nat Commun* **9**, 4784, doi:10.1038/s41467-018-07283-x (2018).

19   Hecht, V. C. *et al.* Biophysical changes reduce energetic demand in growth factor-deprived lymphocytes. *J Cell Biol* **212**, 439-447, doi:10.1083/jcb.201506118 (2016).

20   Subramanian, A. *et al.* Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proc Natl Acad Sci U S A* **102**, 15545-15550, doi:10.1073/pnas.0506580102 (2005).

21   Prakadan, S. M., Shalek, A. K. & Weitz, D. A. Scaling by shrinking: empowering single-cell 'omics' with microfluidic devices. *Nat Rev Genet* **18**, 345-361, doi:10.1038/nrg.2017.15 (2017).

22   Buettner, F. *et al.* Computational analysis of cell-to-cell heterogeneity in single-cell RNA-sequencing data reveals hidden subpopulations of cells. *Nat Biotechnol* **33**, 155-160, doi:10.1038/nbt.3102 (2015).

23   Kowalczyk, M. S. *et al.* Single-cell RNA-seq reveals changes in cell cycle and differentiation programs upon aging of hematopoietic stem cells. *Genome Res* **25**, 1860-1872, doi:10.1101/gr.192237.115 (2015).

24   Kimmerling, R. J. *et al.* A microfluidic platform enabling single-cell RNA-seq of multigenerational lineages. *Nat Commun* **7**, 10220, doi:10.1038/ncomms10220 (2016).

25   Cetin, A. E. *et al.* Determining therapeutic susceptibility in multiple myeloma by single-cell mass accumulation. *Nat Commun* **8**, 1613, doi:10.1038/s41467-017-01593-2 (2017).

26   Ashburner, M. *et al.* Gene ontology: tool for the unification of biology. The Gene Ontology Consortium. *Nat Genet* **25**, 25-29, doi:10.1038/75556 (2000).

27   Best, J. A. *et al.* Transcriptional insights into the CD8(+) T cell response to infection and memory T cell formation. *Nat Immunol* **14**, 404-412, doi:10.1038/ni.2536 (2013).

28   Fox, C. J., Hammerman, P. S. & Thompson, C. B. Fuel feeds function: energy metabolism and the T-cell response. *Nat Rev Immunol* **5**, 844-852, doi:10.1038/nri1710 (2005).

29   Verbist, K. C. *et al.* Metabolic maintenance of cell asymmetry following division in activated T lymphocytes. *Nature* **532**, 389-393, doi:10.1038/nature17442 (2016).

30   Wang, R. & Green, D. R. Metabolic checkpoints in activated T cells. *Nat Immunol* **13**, 907-915, doi:10.1038/ni.2386 (2012).

31   Araki, K. *et al.* Translation is actively regulated during the differentiation of CD8(+) effector T cells. *Nat Immunol* **18**, 1046-1057, doi:10.1038/ni.3795 (2017).

32   Obst, R. The Timing of T Cell Priming and Cycling. *Front Immunol* **6**, 563, doi:10.3389/fimmu.2015.00563 (2015).

33   Batlle, E. & Clevers, H. Cancer stem cells revisited. *Nat Med* **23**, 1124-1134, doi:10.1038/nm.4409 (2017).

34   Patel, A. P. *et al.* Single-cell RNA-seq highlights intratumoral heterogeneity in primary glioblastoma. *Science* **344**, 1396-1401, doi:10.1126/science.1254257 (2014).

35   Tirosh, I. *et al.* Dissecting the multicellular ecosystem of metastatic melanoma by single-cell RNA-seq. *Science* **352**, 189-196, doi:10.1126/science.aad0501 (2016).

36   Barretina, J. *et al.* The Cancer Cell Line Encyclopedia enables predictive modelling of anticancer drug sensitivity. *Nature* **483**, 603-607, doi:10.1038/nature11003 (2012).

37   Suva, M. L. *et al.* Reconstructing and reprogramming the tumor-propagating potential of glioblastoma stem-like cells. *Cell* **157**, 580-594, doi:10.1016/j.cell.2014.02.030 (2014).

38   Chene, P. Inhibiting the p53-MDM2 interaction: an important target for cancer therapy. *Nat Rev Cancer* **3**, 102-109, doi:10.1038/nrc991 (2003).

39   Verreault, M. *et al.* Preclinical Efficacy of the MDM2 Inhibitor RG7112 in MDM2-Amplified and TP53 Wild-type Glioblastomas. *Clin Cancer Res* **22**, 1185-1196, doi:10.1158/1078-0432.CCR-15-1015 (2016).

40   Fischer, M. Census and evaluation of p53 target genes. *Oncogene* **36**, 3943-3956, doi:10.1038/onc.2016.502 (2017).

41   Song, M. S., Salmena, L. & Pandolfi, P. P. The functions and regulation of the PTEN tumour suppressor. *Nat Rev Mol Cell Biol* **13**, 283-296, doi:10.1038/nrm3330 (2012).

42   Saxton, R. A. & Sabatini, D. M. mTOR Signaling in Growth, Metabolism, and Disease. *Cell* **168**, 960-976, doi:10.1016/j.cell.2017.02.004 (2017).

43   Wee, K. B., Surana, U. & Aguda, B. D. Oscillations of the p53-Akt network: implications on cell survival and death. *PLoS One* **4**, e4407, doi:10.1371/journal.pone.0004407 (2009).

44   Daniele, S. *et al.* Combined inhibition of AKT/mTOR and MDM2 enhances Glioblastoma Multiforme cell apoptosis and differentiation of cancer stem cells. *Sci Rep* **5**, 9956, doi:10.1038/srep09956 (2015).

45   Buenrostro, J. D. *et al.* Single-cell chromatin accessibility reveals principles of regulatory variation. *Nature* **523**, 486-490, doi:10.1038/nature14590 (2015).

46   Shalek, A. K. & Benson, M. Single-cell analyses to tailor treatments. *Sci Transl Med* **9**, doi:10.1126/scitranslmed.aan4730 (2017).

47   Trombetta, J. J. *et al.* Preparation of Single-Cell RNA-Seq Libraries for Next Generation Sequencing. *Curr Protoc Mol Biol* **107**, 4 22 21-17, doi:10.1002/0471142727.mb0422s107 (2014).

48   Picelli, S. *et al.* Smart-seq2 for sensitive full-length transcriptome profiling in single cells. *Nat Methods* **10**, 1096-1098, doi:10.1038/nmeth.2639 (2013).

49   Kim, D. *et al.* TopHat2: accurate alignment of transcriptomes in the presence of insertions, deletions and gene fusions. *Genome Biol* **14**, R36, doi:10.1186/gb-2013-14-4-r36 (2013).

50   Li, B. & Dewey, C. N. RSEM: accurate transcript quantification from RNA-Seq data with or without a reference genome. *BMC Bioinformatics* **12**, 323, doi:10.1186/1471-2105-12-323 (2011).

51   Sergushichev, A. An algorithm for fast preranked gene set enrichment analysis using cumulative statistic calculation. *bioRxiv* (2016).

52   Whitfield, M. L. *et al.* Identification of genes periodically expressed in the human cell cycle and their expression in tumors. *Mol Biol Cell* **13**, 1977-2000, doi:10.1091/mbc.02-02-0030 (2002).