A toolset for linking phenotype and gene expression at the single-cell level

by

Robert J. Kimmerling

B.E. Biomedical Engineering
Stony Brook University (2011)

Submitted to the Department of Biological Engineering
in partial fulfillment of the requirements for the degree of:

Doctorate of Philosophy in Biological Engineering

at the

MASSACHUSETTS INSTITUTE OF TECHNOLOGY

February 2017

© Massachusetts Institute of Technology 2017. All rights reserved.
Thesis committee members in favor of defense:

Professor Scott R. Manalis

Professor Darrell J. Irvine

Professor J. Christopher Love
A toolset for linking phenotype and gene expression at the single-cell level

by

Robert J. Kimmerling

Submitted to the Department of Biological Engineering
on December 8, 2016, in partial fulfillment of the requirements for the degree of Doctor of Philosophy in Biological Engineering

Abstract

The development of single-cell RNA-sequencing has led to a new degree of resolution in the characterization of complex, heterogeneous biological systems. However, existing methods are often limited in their ability to link these whole-transcriptome profiles with complimentary measurements of single-cell phenotype. In this thesis we present a microfluidic toolset which allows us to link a panel of single-cell phenotypic measurements—including lineage history, cell cycle stage, cell size, and growth rate— with their corresponding transcriptional profiles.

Using a microfluidic platform that employs an array of hydrodynamic traps to capture and culture single cells for multiple generations we measured single-cell growth kinetics, lineage hierarchies and cell cycle stage. By subsequently releasing individual cells from this device for downstream scRNA-seq we were able to generate whole-transcriptome profiles of primary, activated murine CD8+ T cell and lymphocytic leukemia cell line lineages. For both cell types we found distinct transcriptional patterns associated with single-cell lineage relationships as well as cell cycle progression. In order to link single-cell size and growth rate measurements with gene expression, we have also developed a system that relies on an array of suspended microchannel resonators (SMR)—high resolution single-cell buoyant mass sensors—in combination with an automated method of isolating single cells to conduct scRNA-seq downstream. Using this platform, we were able to collect linked transcriptional and biophysical measurements for a murine leukemia cell line, primary murine CD8+ T cells, and a patient-derived glioblastoma multiforme (GBM) cell line. For all cell models measured, we found that single-cell buoyant mass showed a strong correlation with the expression of cell cycle genes. Furthermore, we found that single-cell growth rate and buoyant mass measurements can be used to characterize the degree to which GBM cells respond to drug treatment as well as determine transcriptional signatures associated with response and resistance.

Taken together, we believe these single-cell phenotypic measurements will offer complementary contextual information to further resolve the heterogeneity of single-cell transcriptional data. As such, we expect these platforms to be broadly useful to fields where heterogeneous populations of cells display distinct clonal trajectories, including immunology, cancer, and developmental biology.

Thesis Supervisor: Scott R. Manalis
Title: Professor of Biological and Mechanical Engineering
Acknowledgements

First and foremost, I would like to thank my advisor Scott Manalis for your guidance and mentorship. Your unwavering support, optimism, and enthusiasm made for an incredible training experience. I genuinely enjoyed the work I was doing and the people I was doing it with, and I consider myself very fortunate for having the opportunity to work in your group.

I would also like to thank the members of my thesis committee, Darrell Irvine and Chris Love. Your feedback and insights into the crossover between engineering and biology did a great deal to further develop the projects I worked on as well as the way I approach science.

Thank you to my colleagues. To Jenny Li for your enthusiasm, optimism and diligence – it was a great pleasure to work with you and I’m excited to see what you do next. To Mark Stevens for lending your ear and critical mind to help bring clarity and vision to countless scientific discussions. Your support and friendship have been invaluable. To Selim Olcum and Nate Cermak, two of the most talented engineers I have had the opportunity to work with, your efforts were instrumental for this work and your patience and willingness to help were much appreciated. To Scott Knudsen, the resident expert on just about everything, for taking the time to answer every question I had from my first day to my last. To Kris Payer, whose fabrication and design expertise were essential to this work. To Nick Calistri, your excitement is matched only by your ability, I have truly enjoyed working with you to keep pushing the limits of what we can achieve. To former and senior lab members in the group including Vivian Hecht, Josephine Bagnall, Nigel Chou, Sungmin Son, Sangwon Byun, Francisco Delgado and Amneet Gulati who, without hesitation, taught me everything I needed to move forward, both scientifically and academically. To current lab members including Bashar Hamza, Joon Kang, Clara Park, Max Stockslager, Davi da Silva, Arif Cetin, Chris Chin, Lucy Yang, and Teemu Miettinen whose abilities and passion have made for an incredible group that I am proud to be a part of.
I would also like to thank the collaborators that I have had the chance to work with. Specifically, Alex Shalek, without whom the gene expression aspects of this story simply wouldn't exist. Your patience and accessibility while teaching me the foundations of single-cell sequencing analysis fundamentally changed the trajectory of my graduate work and for that I am incredibly grateful. Also, members of the Shalek Lab including Jay Prakadan, Alex Genshaft, Sam Kazer and Riley Drake, whose talent and steady hands produced the transcriptional data presented in this thesis. I would also like to thank Greg Szeto, who introduced me to the world of immunology and helped bring a clear biological focus to the hydrodynamic trap work. I greatly appreciated your tenacity and commitment to training younger scientists – myself included.

Thank you to my friends, especially Alec Nielsen, Andrew Warren, Baris Sevinc, James Weiss, Nathan Stebbins, Raja Srinivas, and Vyas Ramanan. You are some of the most intelligent people I have ever met and yet make it impossible to take anything too seriously, and it has made for a great five years.

Thank you to my family. To my parents, who have made incredible sacrifices to give my brother and me every possible opportunity. You have shown me what hard work and decency look like and without your love and support I simply would not be where I am today. To my brother, an easy man to look up to and my lifelong best friend. You, Jess, Kaylee, and Ben are a daily reminder of where I come from and what truly matters. To the Palma family, for welcoming me in to your lives as a sixteen year old and never looking back. Your encouragement and guidance have been a constant source of support that I am incredibly grateful for.

Above all else, thank you to my wife, Erica. You have read every paper, polished every presentation, and helped deal with every crisis of confidence that has come with graduate school. You are the single most important reason I am where I am, and why I have enjoyed every step of the way. In everything that I do, I seek to make you proud, and that is the only motivation I will ever need.
# Table of Contents

1. Chapter 1: Introduction .......................................................................................................................... 15
   1.1 | Single-cell lineage characterization ................................................................................................. 18
   1.2 | Transcriptional characteristics of cell cycle progression ............................................................... 19
   1.3 | Linking biophysical parameters and gene expression ........................................................................... 21

2. Chapter 2: A hydrodynamic trap array for single-cell lineage characterization ............................. 23
   2.1 | Introduction ........................................................................................................................................ 23
   2.2 | Hydrodynamic trap array .................................................................................................................. 24
       2.2.1 | Design and principle of hydrodynamic trapping ............................................................................. 24
       2.2.2 | Hydrodynamic trap array design .................................................................................................... 29
       2.2.3 | Measuring single-cell proliferation kinetics .................................................................................... 34
       2.2.4 | Single-cell lineage tracking ........................................................................................................... 39
       2.2.5 | Single-cell release – Design and implementation ............................................................................ 43
   2.3 | Discussion ......................................................................................................................................... 45
   2.4 | Methods ............................................................................................................................................. 47
       2.4.1 | Device fabrication and system setup ................................................................................................ 47
       2.4.2 | Single-cell culture ............................................................................................................................. 47
       2.4.3 | Loading cells into multiple lanes ...................................................................................................... 48
       2.4.4 | Single-cell release .......................................................................................................................... 48
       2.4.5 | Cell culture and primary cell preparation ......................................................................................... 49
       2.4.6 | Image analysis ............................................................................................................................... 50

3. Chapter 3: Linking scRNA-seq with cell lineage and cell cycle progression ............................. 51
   3.1 | Introduction ........................................................................................................................................ 51
   3.2 | Lineage dependent gene expression patterns .................................................................................... 52
       3.2.1 | Transcriptional similarity amongst related cells ............................................................................. 52
3.2.2 | Clone-dependent transcriptional signatures of CD8+ T cells ........................................ 57
3.3 | Cell cycle gene expression patterns .................................................................................. 63
   3.3.1 | Decoupling cell cycle and lineage dependence ............................................................. 63
   3.3.2 | Latent variable models describing cell cycle transcriptional signatures ................. 66
3.4 | Discussion .......................................................................................................................... 68
3.5 | Methods .............................................................................................................................. 69
   3.5.1 | Single-cell RNA-seq ...................................................................................................... 69
   3.5.2 | Gene expression data pre-processing ............................................................................. 69
   3.5.3 | Transcriptional similarity comparison ............................................................................ 70
   3.5.4 | Constructing single-cell gene expression weight matrices ......................................... 70
   3.5.5 | Unsupervised clustering and lineage reconstruction ...................................................... 72
   3.5.6 | CD8+ T cell subset scoring ............................................................................................. 72
   3.5.7 | Partial least squares regression (PLSR) modeling ....................................................... 72
   3.5.8 | Gene-annotation enrichment analysis ............................................................................ 74
4. Chapter 4: Suspended microchannel resonator platforms for linking single-cell size and growth rate measurements with downstream assays ............................................. 75
   4.1 | Introduction ...................................................................................................................... 75
   4.2 | Array of SMRs with hydrodynamic traps ..................................................................... 77
      4.2.1 | Design and operation .................................................................................................... 77
      4.2.2 | Measuring single-cell growth in the Trap-SMR array ............................................... 81
      4.2.3 | Trap-SMR platform limitations .................................................................................... 81
   4.3 | Serial SMR for single-cell growth and collection measurements ............................ 83
      4.3.1 | System design and implementation ............................................................................. 83
      4.3.2 | Fluidic operation of the serial SMR platform ............................................................. 85
      4.3.3 | Single-cell growth and capture experiments .............................................................. 91
4.3.4 | Potential fluidic improvements ................................................................. 95
4.4 | Discussion ............................................................................................................. 98
4.5 | Methods ................................................................................................................ 100
  4.5.1 | Calibrating the Trap-SMR platform ................................................................. 100
  4.5.2 | Calibrating the serial-SMR platform ............................................................... 100
  4.5.3 | Single-cell growth and release with the Trap-SMR platform ............................. 101
  4.5.4 | Single-cell growth and collection with the serial-SMR platform ....................... 102
  4.5.5 | Cell culture ......................................................................................................... 103
5. Chapter 5: Linking single-cell biophysical measurements with gene expression .... 105
  5.1 | Introduction .......................................................................................................... 105
  5.2 | Transcriptional analysis of hematopoietic cell lines ........................................... 106
  5.3 | CD8+ T cell activation ............................................................................................ 113
    5.3.1 | Distinct transcriptional signatures of different activation time points ............. 113
    5.3.2 | CD8+ T cell growth efficiency during activation .............................................. 115
    5.3.3 | CD8+ T cell mass during activation .................................................................... 117
  5.4 | Glioblastoma multiforme (GBM) drug response ................................................... 121
    5.4.1 | Transcriptional changes in response to drug treatment ................................... 121
    5.4.2 | Mass-dependent gene expression in GBM ...................................................... 123
    5.4.3 | Growth-dependent gene expression in GBM .................................................... 125
  5.5 | Discussion .............................................................................................................. 127
  5.6 | Methods ................................................................................................................ 128
    5.6.1 | Functional enrichment analysis .......................................................................... 128
    5.6.2 | DNA content staining for cell cycle analysis ..................................................... 129
    5.6.3 | Cell cycle scoring .............................................................................................. 129
    5.6.4 | PLSR analysis ..................................................................................................... 130
6. Chapter 6: Appendix ............................................................................................................ 131
  6.1 | Trap system parts list ........................................................................................................ 131
  6.2 | Alternative experimental approaches using the hydrodynamic trap array .......... 133
    6.2.1 | Bead-based immune cell activation ........................................................................... 133
    6.2.2 | Adherent cell growth .................................................................................................. 135
  6.3 | Scaling considerations for growth and collection measurements ................................. 137
7. Bibliography ......................................................................................................................... 139
List of Figures

Figure 2.1 | Hydrodynamic trapping of single cells ................................................................. 25
Figure 2.3 | Optimized hydrodynamic trap design ................................................................. 28
Figure 2.4 | Array of hydrodynamic traps ........................................................................... 30
Figure 2.5 | Immunofluorescence staining on-chip .................................................................. 31
Figure 2.6 | Extended hydrodynamic trap array ...................................................................... 33
Figure 2.7 | Single-cell proliferation kinetics measured in the hydrodynamic trap array .... 35
Figure 2.8 | Cell growth stability in the hydrodynamic trap array ........................................ 36
Figure 2.9 | Proliferation kinetics measured across various hematopoietic cell types ............ 37
Figure 2.10 | Single-cell lineage tracking .................................................................................. 40
Figure 2.11 | Comparing sister cell interdivisionary times ...................................................... 42
Figure 2.12 | Single-cell release schematic .............................................................................. 44
Figure 3.1 | Comparing transcriptional similarity between related and unrelated cell pairs ....................................................................................................................................................... 53
Figure 3.2 | Comparing Spearman distances between related and unrelated cell pairs ...... 54
Figure 3.3 | Unsupervised clustering dendrograms ................................................................. 56
Figure 3.4 | CD8+ T cell clone-dependent transcriptional similarity ....................................... 58
Figure 3.5 | Clone-dependent Granzyme B expression ............................................................ 59
Figure 3.6 | Intraclonal transcriptional comparisons ............................................................... 60
Figure 3.7 | CD8+ T cell subset scoring ..................................................................................... 62
Figure 3.8 | Effect of cell cycle stage proximity on transcriptional similarity ......................... 64
Figure 3.9 | Granzyme B expression level as a function of cell cycle progression .................. 65
Figure 3.10 | Partial least squares regression modeling of single-cell time since division .... 67
Figure 3.11 | Assessing single-cell cDNA library quality ......................................................... 71
Figure 3.12 | Gene selection for partial least squares regression analysis ............................. 73
Figure 4.1 | Suspended microchannel resonator (SMR) for single-cell buoyant mass sensing ....................................................................................................................................................... 76
Figure 4.2 | Suspended microchannel resonator design with a hydrodynamic trap .............. 78
Figure 4.3 | Array of suspended microchannel resonators with hydrodynamic traps .......... 79
Chapter 1

Introduction

The role that single cells play in various biological contexts, either as parts of a complex functional system or as individual actors, has long been recognized. Whether in developmental biology, where a single cell gives rise to a diverse range of progeny, ultimately developing into a complex organism. Or in immunology, where a single T cell can coordinate an effective response to a pathogen or aberrantly growing cell in order to prevent severe infection or tumorigenesis in the host. Or in cancer, where a single cell – if resistant to therapy – is capable of repopulating a tumor and leading to progressive and incurable disease. Although many of the basic mechanisms involved in these processes have been well characterized, studying biological systems at the single-cell level presents distinct technical challenges due to the scarce amount of starting material to conduct traditional biochemical assays as well as the time-sensitive nature of many cell systems which require rapid measurement for accurate characterization. As such, one of the primary focuses in this field – both historically and recently – has been the development of new tools and approaches to enable the collection of more thorough and informative data sets for single cells.

As a result, there are now several broadly accessible approaches for conducting single-cell experimentation. For instance, recent developments in high-content and high-throughput imaging have enabled the routine screening of single-cell morphological and proliferation characteristics
under various conditions. Similarly, tools such as flow cytometry and CyTOF now present the opportunity to characterize tens of markers across thousands of single cells with commercially available systems and well-established experimental protocols. Furthermore, with the development and dissemination of microfabrication approaches – specifically for microfluidic device fabrication – the process of application and system development for single-cell experimentation has itself become more accessible, rapidly accelerating the pace at which new tools are developed. With increased scalability, decreased reagent use and simple customization, microfluidic devices are now widely used for a range of single-cell applications across both academic and industrial settings.

However, while many of these technological advancements have made single-cell analysis a more routine process in biological experimentation, there are a few fundamental limitations incurred by these approaches. For instance, seeing as many of these systems rely on label-based methods for biological readouts – including fluorescent reporter constructs or traditional immunocytochemistry – they have inherent limits to the depth with which they are able to characterize molecular networks within single cells. The recent development of single-cell RNA-sequencing (scRNA-seq) has made significant progress towards addressing these limitations. Measuring single-cell transcription has the unique advantage of being able to amplify the starting material of interest using well-established protocols (i.e. PCR) – a process which is not possible for other molecular readouts such as single-cell proteomics. As such, it is now possible to characterize gene expression at a whole-transcriptome level within single cells. This technology has enabled an unprecedented degree of resolution in characterizing network-level heterogeneity of single cells within complex biological systems.
With the development of scRNA-seq, a primary focus as of late has been to expand the level of information that is attainable with experiments utilizing this technique. One method for achieving this goal is to increase the throughput and decrease the cost of scRNA-seq in order to attain single-cell level characterization of heterogeneity across tissue-level samples. To this end, there have been multiple microfluidics-based approached developed recently which have drastically improved the throughput of scRNA-seq sample preparation to enable the analysis of thousands of single cells as compared to traditional methods which were limited to tens of cells. In fact, there are now commercially available platforms which enable routine collection of these large single-cell data sets at a fraction of the per-cell cost of early sample preparation methods.

Another means of increasing the depth of information achieved with scRNA-seq experiments is to contextualize these transcriptional measurements with other complimentary single-cell data sets. For instance, there have been various methods developed recently to link transcriptional data with tissue-level cellular localization information in order to further analyze gene expression in the context of microenvironmental architecture. Other methods have linked transcriptional measurements with single-cell functional assays – as was done for single neurons using a modified patch-clamping platform. Similarly, various multi-omics approaches have been developed to link measurements such as protein abundance or DNA-sequencing with whole transcriptome characterization in single cells. Overall, these approaches seek to increase biological interpretability of single-cell transcriptional data by contextualizing them with other well-established data sets in order to gain novel biological insights.

In this thesis, we present our efforts to expand the range of single-cell phenotypic measurements that can be linked with gene expression. We demonstrate the implementation of various
platforms which ultimately allowed us to couple of a panel of additional single-cell measurements with scRNA-seq. These phenotypic parameters include single-cell lineage history, cell cycle stage, size, and growth rate. Each of these phenotypic measurements offers a novel means of contextualizing single-cell transcriptional data and as such we envision these systems being broadly applicable to a range of biological questions moving forward.

1.1 | Single-cell lineage characterization

The characterization of single-cell lineage development has long been a widely studied field, particularly in the wake of the development of new tools and approaches which enable high-throughput characterization of single cells over several generations\textsuperscript{19, 20}. These approaches have applications across a broad range of disciplines including immunology, development biology, and cancer where single cells display distinct clonal trajectories over several generations.

One area of research which has been of particular interest as of late in the field of immunology is the mechanisms involved in single T cell development in response to antigenic stimulation. It is well established that a single T cell, in response to activation, can give rise to diverse subsets of cells with distinct functionality ranging from cytotoxic activity to memory cell development\textsuperscript{21}. However, the processes by which these subsets develop from a common ancestor is not well-known and continues to be a point of controversy in the field. On one hand there is a theory of asymmetric cell division in CD8+ T cells which suggests that as early as the first division there is asymmetric partitioning of various functional regulators which gives rise to one memory-like and
one effector-like cell. The linear progression model, on the other hand, suggests that this phenotypic divergence is a more gradual process which takes multiple generations and depends highly on the local microenvironmental cues that single cells experience in their lifetimes.

Recent efforts to characterize the relative contributions of these models of differentiation utilizing either direct observation of cellular division, or computational approaches to reconstruct cell lineage relationships with single-cell gene expression data, have yielded mixed results. In fact, based on the methods used to analyze these data, evidence of either model becomes apparent. However, these data sets do not have the ability to directly link a single cell’s lineage history with its gene expression profile and as such make it difficult to fully resolve the various patterns which may be associated with a cell’s division history. By directly monitoring single-cell lineage histories upstream of scRNA-seq measurements, the hydrodynamic trap array platform presented here offers a means of collecting these linked data sets. We believe this method will prove useful in further characterizing the mechanisms of T cell differentiation.

1.2 | Transcriptional characteristics of cell cycle progression

The mechanisms of cell cycle progression and regulation have been one of the most extensively studied areas in cell biology for several decades. As a result, there are now well-established canonical pathways which have been shown to coordinate various stages of the cell cycle both at a protein and gene expression level. In fact, there are now systems which allow for the direct
observation of these cell cycle regulatory processes at the single-cell level. The development of scRNA-seq has further expanded the depth of these single-cell cell cycle measurements, with access to transcriptome-wide characterization of variation amongst cells. However, it is often difficult to characterize the cell cycle stage of single cells upstream of scRNA-seq. As such, the transcriptional variability introduced by cell cycle variation – without having a sense of the structure of this variability – is often considered a nuisance parameter in the design of scRNA-seq studies. For instance, it is possible that subtle differences in single-cell phenotype are masked by significant differences in transcription associated with cell cycle progression, which may not be the central question of interest.

In order to address this issue, multiple groups have established computational methods of determining the structure of cell cycle signatures within scRNA-seq data sets in order to correct for these features in subsequent analyses. Often, these models will use linked measurements of gene expression and cell cycle stage – for instance, by means of DNA content staining upstream of single-cell isolation – in order to construct cell cycle signatures which can be used for future experimentation. These approaches have proven to be effective in reducing the transcriptional variability associated with cell cycle progression in order to reveal other functional differences among cells, such as different immune cell activation states.

Here we present two alternative proxy measurements of cell cycle progression that we can now link with single-cell gene expression profiles: time since division and cell mass. In contrast to phase-dependent markers of cell cycle progression, such as DNA stains or fluorescent reporter models, these phenotypic parameters offer a continuous range of values which can be used to determine the relative stage of cell cycle progression. We can therefore directly determine the level of correlation between gene expression and either of these parameters to develop cell cycle
signatures for various cell types. We have found that for both time since division and cell mass measurements, different cell types have unique transcriptional signatures associated with cell cycle progression. As such, in order to effectively account for the effect of cell cycle heterogeneity in a scRNA-seq study, it may be necessary to characterize a wide range of different cell types in order to determine conserved and cell-specific patterns. In combination with recently developed computational methods, the platforms presented here can further enable the development of such signatures.

1.3 | Linking biophysical parameters and gene expression

Recent technical advancements have enabled the collection of a broad range of single-cell biophysical parameters which have applications in both basic and translational biological studies. Specifically, single-cell buoyant mass and mass accumulation rate (MAR) measurements now offer the ability to measure the regulation of cell size and cell growth rate with an unprecedented resolution\(^7\). This approach has offered novel insights into the relationship between cell size, growth rate, and cell cycle progression as well as the effects and time-scales of nutrient depletion on cell growth\(^8,9\).

In addition to basic cellular biology studies, this approach has also been applied to a range of translational studies including characterizing immune cell activation dynamics and single cancer cell growth rates\(^{40,41}\). The concept of monitoring single-cell growth rates is particularly
amenable to the study of cancer seeing as it is, fundamentally, a disease of dysfunctional cell growth. These growth measurements offer a direct method of determining the heterogeneity of single tumor cells responding to drug therapy simply by comparing the level to which they continue to accumulate mass. This approach has been shown to be effective in determining populations of cells that are sensitive or resistant to a particular therapy as well as determining the relative fraction of these subsets across a population of single-cells\textsuperscript{41}. However, while this approach offers an effective means of ascribing a binary determination of response or non-response to single-cells, these biophysical parameters alone do not offer insight into the mechanisms by which certain cells resist therapy.

On the other hand, scRNA-seq studies which explore the heterogeneity of cells within a tumor offer network level signatures of transcription and cellular function but do not have access to single-cell biophysical measurements and therefore are difficult to contextualize with respect to cell growth characteristics. Recently, various groups have sought to address this limitation computationally by developing signatures for cycling and non-cycling cells which can be used as a proxy for cellular expansion within a tumor\textsuperscript{10,11,42}. Although these signatures may be effective in certain circumstances, we found that in the case of a population of drug treated cells, a cell-cycle signature alone does not sufficiently account for biophysical response heterogeneity.

To address these limitations, here we present a platform which links single-cell biophysical measurements with gene expression. By leveraging the benefits of each of these approaches, we envision these linked data sets offering a novel means of exploring cancer cell growth and single-cell drug response mechanisms.
Chapter 2

A hydrodynamic trap array for single-cell lineage characterization

2.1 | Introduction

Recent developments in microfluidic technology have enabled new methods of capturing and culturing single cells\textsuperscript{44, 45}. When coupled with traditional imaging approaches, these systems offer a robust means of following cellular trajectories over time, but require experimental platforms that can reliably link these measurements to downstream molecular assays\textsuperscript{20, 46}. Alternatively, microfluidic devices which enable the efficient preparation of single cells for such assays – such as the Fluidigm C1 platform for cDNA library preparation – currently lack the long-term culture, progeny capture and time-lapse imaging capabilities necessary to link these transcriptional measurements with lineage information.

In order to address these limitations, we developed a microfluidic platform that allows direct association of these complementary datasets by enabling registered off-chip single-cell isolation after multigenerational lineage tracking. The platform relies on an array of hydrodynamic traps to achieve key operational capabilities including single-cell isolation and culture, progeny capture, and single-cell release for downstream analysis\textsuperscript{43}. In combination with time-lapse imaging, we
were able to characterize the single-cell growth kinetics and lineage hierarchies for a range of cell
types including various cell lines as well as primary murine and human immune cells. Ultimately we
used this platform to generate whole-transcriptome profiles of primary, activated murine CD8+
T cell and lymphocytic leukemia cell line (L1210) lineages in order to validate the process of
single-cell retrieval and downstream scRNA-seq.

2.2 | Hydrodynamic trap array

2.2.1 | Design and principle of hydrodynamic trapping

Hydrodynamic traps are microfluidic structures that rely on a difference in fluidic resistance
between a trapping pocket and a bypassing serpentine channel to deterministically capture single
particles (Figure 2.1)\(^4\). While previous iterations of these fluidic structures have been
implemented to capture particles such as polystyrene beads, single-cell capture requires the
consideration of various additional design constraints. For instance, in comparison to polystyrene
beads – which are rigid and highly monodisperse particles – cells display a wide range of radii
and have varying degrees of deformability. Furthermore, while particle structure may be largely
unaffected by the flow field, cellular physiology can be altered drastically in response to
excessive shear stresses\(^4\). As such, the modified hydrodynamic traps used for single-cell
processing required a delicate balance between high-fidelity single-cell capture and a limited
shear stress in order to ensure long term cellular viability. One of the key design parameters
which determines the fluidic behavior of the hydrodynamic trap structures is the balance of
fluidic resistances between the trapping pocket and the bypassing serpentine channel (Figure 2.1). Tan and colleagues have characterized this relationship numerically by defining the volumetric flow rates directed to either of these two channels (1)44.

![Figure 2.1](image)

**Figure 2.1 | Hydrodynamic trapping of single cells**
Schematic representation of a hydrodynamic trap with a resistor diagram indicating the relative fluidic resistances required for single-cell capture. Prior to loading a single cell or particle, the fluidic resistance of the trapping pocket \((R_T)\) is lower than the resistance of the bypassing channel \((R_B)\) and thus a majority of the flow is directed through the trap leading to a capture event. Once the trap is occupied, however, the relative resistance of the trap increases \((R_T > R_B)\) and the majority of the flow instead diverts to the bypassing channel, thus preventing secondary particle capture. (B) COMSOL simulation of fluid flow within a hydrodynamic trap before (left) and after (right) single particle capture (red sphere).

\[
\frac{Q_T}{Q_B} = \frac{C_B(\alpha_B)}{C_T(\alpha_T)} \cdot \left(\frac{L_B}{L_T}\right) \cdot \left(\frac{P_B}{P_T}\right)^2 \cdot \left(\frac{A_T}{A_B}\right)^3
\] (1)
Here the subscripts $T$ and $B$ refer to the trap and bypass channels respectively. The volumetric flow rate through each portion of the structure is a function of the channel length ($L$), perimeter ($P$), area ($A$) and a constant ($C$) which itself is a function of the aspect ratio of the channel ($a$). In order to effectively capture a particle in the hydrodynamic trap, the volumetric flow rate through an unoccupied trap ($Q_T$) must be greater than the volumetric flow rate directed to the bypassing serpentine channel ($Q_B$). When working with polystyrene beads, high-efficiency capture can be achieved by designing the hydrodynamic trap such that $Q_T \gg Q_B$. Seeing as the radius of the beads can be carefully selected and controlled for, the system can be designed such that once a single bead is captured, it occludes the entire trapping pocket and reduces $Q_T$ to nearly zero. This ensures that once a particle is captured, the efficiency of trapping for any subsequent bead will be negligible and only a single bead will be captured in each trap. For cells, however, this homogenous size constraint is not met. In the limit of $Q_T \gg Q_B$ once a single cell is captured, depending on its size, it may not fully occlude the channel. This leaves sufficient flow through the trapping pocket to potentially capture multiple cells in each structure. Therefore, in order to ensure the capture of one and only one cell, the structures were designed such that the volumetric flow rate through an unoccupied trapping pocket is only slightly greater than the volumetric flow rate through the bypassing serpentine channel. In order to reduce this design to practice, we began by defining a rough parameter set quantitatively using the numerical formula referenced above. From here, we used COMSOL fluidic simulations to further validate a range of geometries and determine device feasibility in silico. Finally, a subset of geometries which demonstrated the most promising fluidic profiles were fabricated and tested experimentally. Upon testing these devices experimentally, additional design constraints became clear. For instance, there appeared to be a spatially-dependent variation in single-cell trapping efficiency.
within and across devices from the same fabricated wafer (2.4.1). These efficiency variations were attributed to minor differences in channel geometries caused by the tolerances of the fabrication process. Seeing as the structures were originally designed to have only minor differences between the fluidic resistances of the trapping and bypassing channels, these small geometric variations were enough to appreciably alter the fluidic behavior of the designs. In order to address this issue, we altered the geometry of the trapping pocket to have a broader, ramp-like entrance as well as a wider fluidic opening within the trap structures – effectively reducing the fluidic resistance through the trap ($R_T$). COMSOL simulations of these new designs demonstrated that they were less susceptible to slight variations in channel geometries caused by fabrication tolerances and had a streamline profile amenable to high-efficiency single cell capture (Figure 2.2).

In addition to high efficiency single-cell capture, these hydrodynamic trap structures also required the ability to release individual cells for downstream analysis – a process that is explored further in section 2.2.5. However, seeing as the fluidic resistance balance between the trapping and bypassing channels in these structures is dictated only by the geometry of the device, and not the flow parameters, the trapping process also occurs when the flow direction is reversed and particles or cells flow past the back side of the trap. In practice this means that when many of these trap structures are arranged in series – as will be described in section 2.2.2 below – a cell that is released from one trap can subsequently be captured to the reverse side of an upstream trap, a phenomenon that will henceforth be referred to as “back-trapping”. In the case of highly efficient back-trapping, single cell release becomes infeasible seeing as the cell
Figure 2.2 | Optimized hydrodynamic trap design
COMSOL simulation demonstrating flow characteristics in hydrodynamic trap structures. (a) Flow field within the hydrodynamic traps before (left) and after (right) single-cell capture. (b) Shear stresses within the flow field and along the surface of a trapped cell. All simulations were calculated based on a mean flow velocity of 100 μm/s, the approximate velocity observed during experimental operation. Scale bar represents 20 microns.

will simply be passed back-and-forth between two adjacent traps without ever exiting the device.

To address this issue it was necessary to ensure high efficiency trapping in the forward-flow direction while minimizing back-trapping efficiency during reverse flow operation. This behavior was achieved by modifying the geometry of the trapping structures to have asymmetric flow profiles in the forward and reverse flow directions. First, a slight ramp structure was added to the entrance of each trapping pocket (Figure 2.2). In addition to increasing trapping efficiency
and making the structures less susceptible to fabrication tolerances— as was described earlier— this ramp structure also ensured that when cells were released from each trap they were focused to the wall opposite the back side of the upstream trap, thus placing the cell along a flow path that made them less susceptible to back-trapping (Figure 2.3). Second, the channel on the back side of each trapping structure was narrowed slightly relatively to the rest of the channel (Figure 2.2). Experimentally, we found that larger cells were less susceptible to back-trapping than smaller cells. We hypothesized that this behavior was due to the fact that larger cells occupy a larger fraction of the flow field and as such when back-trapping does occur there is a larger shear force which helps to clear the cell from the back of the trap. Therefore, by narrowing the channel we were able to ensure that this clearing behavior occurred for a larger range of cell sizes and as such back-trapping frequency decreased substantially. Together, these alterations to the geometry of the trapping structures offered high efficiency single-cell capture in the forward flow direction, minimized shear stresses on trapped cells and reduced the occurrence of back-trapping in order to enable single-cell release from the device.

2.2.2 | Hydrodynamic trap array design

In addition to capturing single cells in this device, the primary objective of the platform was to culture and monitor cells for multiple generations within the system. Therefore, once we were able to optimize the design and operation of a single trapping structure, we sought to implement a large array of these structures in order to conduct these long-term culturing experiments. This array consisted of many of these individual trapping structures arranged in series, with bypass channels flanking either side which were used for fluidic exchange and media perfusion (Figure 2.3). For experimental operation, a single cell is loaded into the first trap of the array. As the cell
Figure 2.3 | Array of hydrodynamic traps
Schematic of the hydrodynamic trap array. Multiple individual trap structures are placed in series with bypass channels flanking either side that enable independent control of upstream (P1) and downstream (P2, P3) pressures. Scale bar represents 50 µm.

Proliferates its progeny are carried downstream and captured in subsequent unoccupied traps. Time-lapse imaging of this process allows for the determination of single-cell proliferation kinetics and identification of lineal relationships between cells – processes which will be explored in full in section 2.2.3. Each lane of traps can accommodate up to 40 cells and thus enables lineage tracking for up to five generations, at which point a single founding cell has given rise to a family of 32 cells.

The application of independent upstream and downstream pressures (P1, P2, P3) drives fluid flow through the device. By establishing unique pressure gradients along (P1-P2 and P1-P3) and across (P2-P3) the bypass channels, this fluidic design decouples the flow through the bypass channels from the flow across each lane of traps. As such, media can be rapidly and continuously
perfused through the bypass channels while maintaining minimal flow across the traps in order to ensure constant nutrient repletion with low and uniform shear stress on the cells. In addition to the fluidic control necessary for single-cell culture, this independent pressure control design also enables rapid and complete buffer exchange within the device without perturbing captured cells. This exchange can be used to conduct basic immunocytochemistry protocols on-chip (Figure 2.4) as well as to change the buffer in which single cells are suspended for compatibility with downstream assays, as will be seen in section 2.2.5 in the context of single-cell release for RNA sequencing.

Figure 2.4 | Immunofluorescence staining on-chip
Brightfield image of OT-I CD8+ T cells captured in the hydrodynamic trap array with overlaid false color image collected for phycoerythrin (PE) fluorescence. Buffer containing PE-conjugated anti-mouse CD8 antibodies was rapidly flushed in to the device via the bypass channels and slowly perfused across the cells for 10 minutes for staining without dislodging any of the trapped cells. Prior to imaging the buffer was exchanged once again to flush the system with phosphate buffered saline in order to reduce auto fluorescence. Scale bar represents 50 μm.
Once we optimized the design of this array of hydrodynamic traps, we sought to increase the throughput of the single-cell experiments we were performing seeing as each lane of traps was able to capture only one cell at a time. To this end, we designed a large array which included many lanes of trap structures. Each lane had common access to both upstream and downstream fluid sources and shared a common reservoir during operation. However, based on the design of the array, each lane maintained independent fluidic control for manipulation and release of single cells – a design and operation that is described in detail in section 2.2.5. This final design allowed us to run experiments in which a single cell is loaded into each of 20 lanes and observed for multiple generations in the device.
Figure 2.5 | Extended hydrodynamic trap array
Schematic demonstrating the final design of the hydrodynamic trap array which included 20 individually controlled lanes of traps with common bypass channels flanking either side. The inset shows an optical micrograph of a single trapped cell in the first field of view of the device – scale bar indicates 50 μm.
2.2.3 Measuring single-cell proliferation kinetics

In order to implement long-term single-cell growth experiments, we incorporated devices with the optimized hydrodynamic trap array design into a system with carefully controlled temperature and pH regulation in order to recapitulate growth conditions typical of in vitro cell culture. After loading a single cell into each lane of the hydrodynamic trap array, we used time lapse imaging to directly observe division events in single cells (472.4.2). In order to collect high magnification images – which offered greater resolution – we implemented an automated multi-axis stage which enabled the periodic collection of images from various fields of view in order to stitch together an image of the entire device footprint (Figure 2.5). Using these time lapse image series, we were able to observe multiple division events from the same single founding cell. As such, we were able to determine the time between these division events – or interdivisionary time – and use this as a direct metric of single-cell division kinetics. As a demonstration of these measurements, we collected single-cell division kinetics measurements for two model cell types: a murine leukemia cell line (L1210) and primary mouse CD8+ T cells which were isolated from C57BL/6J splenocytes and activated in vitro prior to measurement in the device (Figure 2.6). Although these measurements of multigenerational growth suggested that the platform was amenable to long-term single-cell culture, we sought to further validate whether or not the fluidic conditions or confinement in the device altered the proliferation kinetics of these cell types. To determine the long-term stability of proliferation in the device we compared the interdivisionary times across multiple generations of a cell population within the same experiment (Figure 2.7). We found that there was no statistically significant difference in interdivisionary times over multiple generations, suggesting that the system does not induce stress on the cells that causes a drift in growth phenotype.
Figure 2.6 | Single-cell proliferation kinetics measured in the hydrodynamic trap array

Single-cell interdivisionary time measurements for a murine leukemic cell line (L1210, \( n = 526 \)) and primary CD8+ T cells isolated from C57BL/6J mice (CD8+, \( n = 418 \)). Here, the interdivisionary time refers to the amount of time that elapsed (in minutes) between sequential single-cell division events, as determined by time-lapse microscopy.

We further verified that the system does not induce significant changes in the growth characteristics by comparing the average interdivisionary time of single cells with the bulk doubling time of the population in traditional in vitro cell culture conditions (Figure 2.7). These results demonstrated that the proliferation kinetics observed in the device were consistent with bulk culture, suggesting that the system does not significantly perturb the growth of these cells.
Figure 2.7 | Cell growth stability in the hydrodynamic trap array
(a) Comparison of single-cell interdivisionary time measurements for the first and second generations of division collected for CD8+ T cells (left, n = 168 and 131 for generations 1 and 2 respectively) and L1210 cells (right, n = 48 and 92 for generations 1 and 2 respectively) in the device. The groups were compared with a Mann-Whitney U test. Not-significant (n.s.) indicates a p-value greater than 0.05. (b) Proliferation measurements of CD8+ T cells grown in bulk culture collected by periodically counting cell concentration with a Coulter Counter. Points represent triplicate experiments normalized to the initial number of cells for each replicate. The red line represents the expected growth kinetics based on the mean interdivisionary time collected in the trap array (492 minutes) (c) Same analysis as (b) for L1210 cells which had a mean interdivisionary time of 698 minutes in the trap array.
In order to determine if the platform was amenable for cell growth more generally we collected single-cell proliferation kinetics measurements for a panel of hematopoietic cell types (Figure 2.8). These model systems included the constitutively proliferating murine cell lines L1210, FL5.12 and BaF3 as well as primary immune cells which were isolated from murine splenocytes and activated in vitro prior to culture in the device (2.4.5). The platform offered stable, long term growth for all of the model cell systems tested. Taken together, these results suggest that the hydrodynamic trap array offers a reliable means of measuring single-cell proliferation kinetics across a wide range of cell types.

![Figure 2.8: Proliferation kinetics measured across various hematopoietic cell types](image)

**Figure 2.8: Proliferation kinetics measured across various hematopoietic cell types**

Single-cell interdivisionary time measurements collected for a panel of hematopoietic model cell systems including a murine leukemia cell line (L1210, n = 755), a murine pro-B cell line (FL5.12, n=267), a murine pre-B cell line (BaF3, n = 260), CD8+ T cells isolated from a TCR transgenic OT-1 murine model (OT-1 CD8+, n = 75), CD8+ T cells isolated from C57BL/6J mice (WT CD8+, n = 82), and CD4+ T cells isolated from C57BL/6J mice (WT CD4+, n = 117).
In addition to determining bulk growth kinetics of a cell population, these measurements also offer the unique opportunity to compare heterogeneity in single-cell proliferation rates across various populations, a comparison that is not possible with traditional cell-counting methods. For instance, the population of CD8+ T cells isolated from OT-1 mice showed a significantly lower variance in single-cell interdivisionary times when compared with CD8+ T cells isolated from wild-type C57BL/6J mice (p < 0.001, Bonferroni Corrected) (Figure 2.8). The OT-1 mouse model has been genetically modified such that each CD8+ T cell has the same T cell receptor (TCR) specific for a particular OVA peptide (SIINFEKL) \(^48\). Alternatively, CD8+ T cells isolated from wild-type C57BL/6J mice include a wide range of TCR specificities and represent a polyclonal population. As such, these results indicate that a monoclonal population of CD8+ T cells shows reduced variation in interdivisionary times and suggests that clone-specific heterogeneity in proliferation kinetics may play role in T cell expansion. Furthermore, we found that CD8+ T cells isolated from wild-type C57BL/6J mice showed significantly less variation in interdivisionary time measurements as compared with CD4+ T cells isolated from the same mouse strain (p < 0.001, Bonferroni corrected) (Figure 2.8). It is important to note that the activation methods used for CD8+ versus CD4+ T cell proliferation measurements were varied to account for known differences in activation signaling between these cell types. Specifically, CD8+ T cells are known to require only brief TCR engagement in order to begin proliferating and develop functional activity \(^49\). CD4+ T cells, however, require sustained TCR signaling in order to continue proliferating. Therefore, CD8+ T cells were activated in vitro prior to culture on-chip whereas CD4+ T cells were co-cultured in the device with functionalized beads in order to sustain activation (6.2.1). The fact that the resulting proliferation kinetics showed greater
variation for CD4+ T cells suggests that the form and strength of TCR stimulation may affect lymphocyte expansion, a phenomena which may now be further explored with this platform.

2.2.4 | Single-cell lineage tracking

Direct observation of single-cell division events in the hydrodynamic trap array allows for the reconstruction of lineage histories after multiple generations of proliferation in the device (Figure 2.9). Specifically, as a single founding cell grows and ultimately divides into two sister cells, eventually one of these sister cells will be carried downstream to the next available trapping structure (Figure 2.3). Using the images collected with time-lapse microscopy, these division events and subsequent downstream re-seeding events can be identified and tracked in order to determine the lineage relationship between all of the cells occupying a lane of the trap array at the end of an experiment.

This ability to track lineage relationships over multiple generations also offers the opportunity to study the proliferation kinetics of single cells in the context of their familial histories. For instance, recent work by Sandler et al. explored whether variations in cell cycling times amongst cells that originated from a common ancestor could be explained by deterministic factors or if these differences were simply the result of stochastic processes. This work demonstrated that for a murine leukemia cell line (L1210) there was little to no correlation between the interdivisionary times of a mother and daughter cell – i.e. there did not appear to be heritability in cell-cycle dynamics. However, there was a fairly strong correlation in interdivisionary times among cells of the same generation – i.e. daughter and even cousin cell pairs demonstrated similar interdivisionary times. The authors suggested that this structure implies an underlying set
of deterministic factors that may explain the heterogeneity in cell cycling times that is seen across a population of single cells. When comparing interdivisionary times collected with the hydrodynamic trap array, we found a similar correlation in division kinetics amongst sister cell pairs of L1210 cells (Figure 2.10). In fact, we found that this correlation between sister cell division kinetics was consistent across various model cell systems. These results are consistent with those collected by Sandler et al. and also suggest that the underlying pattern of division kinetics similarity amongst related cells is a generalized phenomenon common across various cell types. Interestingly, we found that this similarity in proliferation kinetics was even more
pronounced when comparing the interdivisionary times of primary cells – including CD8+ and CD4+ T cells – that were measured in the hydrodynamic trap array. For instance, CD8+ and CD4+ T cells isolated from wild-type C57BL/6J mice showed correlation coefficients of 0.91 and 0.92 when comparing interdivisionary times of sister cell pairs (Figure 2.10). It is important to note that all of the cell lines compared here (L1210, BaF3, and FL5.12) are constitutively proliferating model systems which will continue to proliferate indefinitely in the presence of the proper media and growth factors – such as IL-3 for the BaF3 and FL5.12 cells. The primary T cells, however, require activation via the T cell receptor in order to begin proliferating and do not expand indefinitely, but instead divide for several generations before eventually ceasing division. These divergent modes of proliferation may contribute to the differences in related-cell kinetics similarity that are observed here. The exact mechanisms that give rise to this seemingly deterministic cycling similarity amongst related cells have not yet been fully resolved. However, the measurements enabled by this platform may help to further validate and clarify various models put forth to describe these phenomena as they are developed in the future.
Figure 2.10 | Comparing sister cell interdivisionary times

Plots of interdivisionary times – in minutes – for sister cell pairs of murine cell lines including L1210 (n = 439, $R^2 = 0.51$), FL5.12 (n = 257, $R^2 = 0.42$), and BaF3 (n = 241, $R^2 = 0.57$) as well as primary lymphocytes including OT-I CD8+ T cells (n = 77, $R^2 = 0.76$), wild-type C57BL/6J CD8+ T cells (n = 323, $R^2 = 0.91$), and wild-type C57BL/6J CD4+ T cells (n = 79, $R^2 = 0.92$). For symmetry, each cell is plotted on each axis and therefore each cell pair comparison is represented by two points.
2.2.5 | Single-cell release – Design and implementation

The ability to track multigenerational growth of single-cells within the hydrodynamic trap array along with the fluidic exchange functionality of the platform enables the comparison of proliferation kinetics as well as variable expression of fluorescence markers that can be imaged with standard immunofluorescence techniques (Figure 2.4). However, in order to fully compare the functional phenotypes of single-cells arising from a single common ancestor, it was necessary to implement measurements that offered a greater depth of information, specifically single-cell RNA-sequencing. Seeing as scRNA-seq requires single-cell isolation and lysis prior to cDNA library construction, we needed to implement a cell isolation procedure downstream of the device. To this end, we developed a single-cell release functionality for the hydrodynamic trap array that allowed us to collect one cell at a time directly into lysis buffer for downstream scRNA-seq. Videos documenting this release process allow us to couple transcriptional profiles with lineage relationships and measurements of time since division for each cell.

The single-cell release operation of the device was enabled by a careful design of the fluidic resistances of the hydrodynamic trap array design (Figure 2.11). Specifically, the downstream resistance leading to port P2 (R_{DS1}) is higher than the downstream resistance leading to port P3 (R_{DS2}) due to a longer length of tubing connecting port P2 to its corresponding pressure reservoir. For this reason, a fixed pressure drop (P1-P2 = P1-P3) will generate flow in the direction of cell trapping (left to right in the schematic) even while P2 and P3 are maintained at the same pressure. This allows the pressure reservoir for P2 to be opened and the corresponding connection tube left open to atmospheric pressure. To release cells, P3 is increased until the flow direction changes direction and, as soon as a cell is released into the bypass channel, the tubing leading to port P2 on the device is placed in a PCR tube to collect the volume flushed from the...
Figure 2.11 | Single-cell release schematic
Schematic showing (a) a reduced subset of the hydrodynamic trap array demonstrating the arrangement of individually addressable trap lanes and (b) a resistor diagram representing the fluidic resistances corresponding to the reduced array schematic. The device performance relies on the fluidic resistance across each trap lane ($R_T$) being significantly higher than the resistance along each bypass channel ($R_B$).

system along with the single cell it contains. Once a single-cell is released into the bypass channel, $P_3$ is once again decreased in order to re-establish flow in the direction of cell trapping, thus ensuring only a single-cell is released at a time. By repeating this process multiple times for a lineage of cells we are able to independently isolate each single cell directly into lysis buffer scRNA-seq.

Although this method allows for highly fidelity isolation of single cells from a lineage, if the system were limited to isolating only a single lineage for an experiment, the throughput would be fairly low. Therefore, the system was designed to have several lanes of hydrodynamic traps
which have independent fluidic control and thus allow for the collection of multiple single-cell lineages from a single experiment (Figure 2.11). For releasing cells from each lane of traps, the pressures on the system are adjusted such that $P_1 \gg P_2$ while $P_2$ and $P_3$ are both set to atmospheric pressure. This pressure balance ensures that fluid is constantly flowing through the bypass channels with slight flow directed across the trap lanes to keep the cells captured. Once buffer is flowing through the bypass channels, the pressure $P_3$ is periodically increased such that the flow direction is reversed and cells begin to exit the trap lanes. Each lane of traps is located at a slightly different vertical position along the bypass channels and thus has a different fraction of $P_2$ at the lane entrance seeing as there is a pressure drop along the length of the bypass channel. Therefore each lane has a unique value of $P_3$ at which flow reverses direction and cells begin to exit the traps. Flow originally reverses direction in the lower left-most trap lane and, as $P_3$ is increased, propagates up the array with the top right-most lane reversing direction last. This variability in the pressure required to release cells from each lane, coupled with the difference in proximity to the bypass channels of the left and right columns of lanes, allows for multiple lanes of cells to be released independently with high fidelity.

2.3 Discussion

Through various fabrication iterations and design optimization techniques we have developed a platform that demonstrates reliable and reproducible single-cell manipulation and culture for multiple generations. The proliferation kinetics measurements collected with the hydrodynamic trap array offer the unique ability to monitor both the rate and heterogeneity of cellular growth kinetics within a population of single cells, data sets which we have found to be particularly
useful in characterizing growth characteristics of diverse cell populations such as different primary CD8+ T cell systems. In combination with the rapid buffer exchange enabled by the fluidic design of this system, we envision these measurements being broadly useful when collected in combination with a panel of well-established immunofluorescent markers which may help to contextualize the proliferative characteristics of many diverse cellular subsets.

We have also demonstrated lineage-dependent proliferation characteristics among cells derived from a common ancestor, particularly for primary T cells which show a strong correlation in interdivisionary times between pairs of sister cells. This ability to measure cycling heritability across multiple generations may provide a crucial reference data set for computational models seeking to explain cell intrinsic versus stochastic mechanisms leading to cellular variation, particularly in the context of cell cycle and growth regulation.

Finally, we have implemented a method by which single cells can be released from the hydrodynamic trap array for downstream analysis while maintaining access to the lineage and proliferation data collected with long term time-lapse imaging. As we will demonstrate in chapter 3, this single-cell release process can be used to create high-quality single-cell transcriptomic data with scRNA-seq. In principle, however, we believe that this method will also be applicable to alternative downstream single-cell chemistries focused on whole genome amplification, or epigenomic characterization\textsuperscript{50,51}. In combination with single-cell lineage information these measurements may offer crucial insight into clonal development and stability for fields such as developmental biology and cancer.
2.4 | Methods

2.4.1 | Device fabrication and system setup

All devices were fabricated in 6-inch silicon-on-insulator (SOI) wafers with 17μm deep flow channels created with deep reactive ion etching (DRIE) and an anodically bonded Pyrex lid. Each six inch silicon wafer yields one hundred devices. Fluidic connections were established by securing the devices to a Teflon manifold with PEEK tubing aligned to the access ports. This manifold was secured to a copper clamp maintained at 37°C with a re-circulating water bath (Thermo Scientific). Pressure-driven flow in the device was controlled with electronic pressure regulators (Proportion Air). All fluids were pressurized with 5% CO2 (Airgas) to maintain long-term pH stability of the culture medium for cell growth. Time-lapse imaging was conducted with a custom LabView program (National Instruments) which drove a TTL-triggered white LED light-source (ThorLabs) for illumination as well as two automated stages (Newport) which traversed the X and Y axes to capture multiple fields of view for each frame.

2.4.2 | Single-cell culture

Single cells were manually loaded into the device by introducing a cell sample at a concentration of 2x10^5 cells/mL into the left bypass channel (Figure 2.3) and flowing it into the trap lanes (P2>P3). The fluidic design of the system ensures targeted loading of single cells while avoiding capturing multiple cells in each lane. For long term growth and kinetics measurements, a single cell was loaded in each of the 20 lanes of the device. For the lineage release experiments one cell was loaded in each of the first 8 lanes in order to enable imaging of all lineages during cell release.
Once a single cell is loaded, the bypass channels are flushed to remove any remaining untrapped cells. For continued nutrient repletion, cell growth media is perfused through the bypass channels at a flow rate of 100μL/h with a pressure drop applied along the bypass channels (P1-P2 and P1-P3). A slight pressure drop is concurrently introduced across the traps (P2-P3) to ensure that the cells remain trapped and their progeny flow downstream to unoccupied traps.

2.4.3 | Loading cells into multiple lanes

When loading single cells into the trap array, a plug of cells at a concentration of 2x10^5 is first loaded into the left bypass channel by loading a sample of cells in the vial for the port labeled P2 and pressurizing the system such that P2>P1 and P2=P3 (Figure 2.11). Once cells are in the left bypass channel of the system, the pressures are adjusted such that P2>P1 and P3=P1. Seeing as the total fluidic resistance along the bypass channels of the device (in the resistor schematic: R_ds, 3 R_B, and R_ds connected in series) is significantly lower than the total resistance across the lanes of traps (in the resistor schematic: 4 R_T connected in parallel), for a fixed pressure differential (ΔP = P2-P1 = P2-P3), a larger fraction of flow will be directed along the bypass channel as compared with the flow directed across the trap lanes. For this reason, a single cell can be brought in close proximity to a trap lane entrance and the pressure P1 can periodically be toggled to be higher than P2 to allow the cell to slowly drift in to the trap lane without traveling further down the bypass channel. This design allows for the precise loading of a single cell in to each lane of traps without capturing multiple cells in each lane.

2.4.4 | Single-cell release

After long-term proliferation, the upstream media reservoir is replaced with PBS and the bypass channels are flushed. This exchange is completed within 1 minute. Single cells are then retrieved
from the device by briefly reversing the pressure differential across the lanes of traps (P2-P3) (Figure 2.3). Once a single cell travels to the bypass channel, the original pressure differential is re-established and the cells will once again flow into the traps. During this process, there is a constant pressure drop maintained along the bypass channels (P1-P2) that allows each single cell that enters the bypass channel to be flushed out of the chip. The fluidic design of the system allows for the port P2 to be maintained at atmospheric pressure; this enables single cells to be collected directly from the tubing connected to the device. Furthermore, the design allows for the release of cells from individual lanes to collect multiple single cell lineages with each experiment. A downstream tubing inner diameter of 75μm results in a dead volume of approximately 2 μl. To accommodate this dead volume, each cell is released in 5 μl of PBS. For single-cell RNA-seq measurements, each cell was released directly in to a PCR tube containing 5 μl of 2X TCL lysis buffer (Qiagen) resulting in a total volume of 10 μl of single-cell lysate. These samples were immediately frozen on dry ice and subsequently stored at -80°C prior to library prep and sequencing.

2.4.5 | Cell culture and primary cell preparation

L1210 murine lymphocytic leukemia cells (ATCC CCL219) were cultured in RPMI 1640 (Gibco) with 10% FBS and 1% penicillin-streptomycin solution (Gibco). The L1210 cells used for single-cell RNA-seq were from cultures that had been passaged less than 15 times after the initial thaw of the ATCC aliquot. CD8+ T cells were isolated from mixed-gender, C57BL/6J mice ranging in age from 4-16 weeks (6 mice in total). After splenocyte isolation and red blood cell lysis with ACK buffer (Gibco), CD8+ T cells were purified using a MACS-based CD8a T cell isolation kit (Miltenyi Biotec). These cells were cultured in RPMI 1640 (Gibco) with 10% FBS, 55μM 2-mercaptoethanol (Gibco), 1% penicillin-streptomycin solution (Gibco) and 100
U/mL IL2 (PeproTech). The CD8+ T cells were activated with 5 μg/mL plate-bound anti-mouse CD3 (clone: 145-2C11, BioLegend catalog number 100314) and 2 μg/mL of anti-mouse CD28 (clone: 37.51, BioLegend catalog number 102112) in solution for 30h prior to loading into the device.

Bulk doubling time measurements were determined by fitting an exponential proliferation model to cell count data collected at various time points with a Coulter Counter (Beckman Coulter) for cultures of L1210 and activated CD8+ T cells.

Animals were cared for in accordance with federal, state, and local guidelines following a protocol approved by the Department of Comparative Medicine (DCM) at MIT.

2.4.6 | Image analysis

Lineage relationships and time-since-division measurements were determined by manually tracking division events and subsequent trap locations for single cells throughout time-lapse image stacks (ImageJ). To couple this information to single-cell RNA-seq measurements, the cell release process was recorded and each cell was manually assigned its corresponding sample ID.
Chapter 3

Linking scRNA-seq with cell lineage and cell cycle progression

3.1 | Introduction

The development of single-cell RNA-sequencing (scRNA-seq) has led to a new degree of resolution in the characterization of complex, heterogeneous biological systems. Complimentary technical advances in single-cell isolation using micromanipulation, microfluidics, and fluorescence activated cell sorting (FACS) have further enabled the coupling of traditional measurements of cellular phenotype, such as immunofluorescence staining and optical microscopy, with transcriptional profiles. Together, these approaches have provided crucial insights into the transcriptional heterogeneity of cancer, immune, and pluripotent stem cells.

Because these single-cell isolation platforms rely on single time point measurements, they provide only an instantaneous snapshot of cellular phenotype to link to a transcriptional signature. However, in addition to the transcriptional heterogeneity within a population of cells, the mechanisms involved in generating this heterogeneity over time are often of critical importance. For instance, a cornerstone of adaptive immunity is the ability of single T lymphocytes to generate
diverse progeny that can both acutely respond to a specific antigen and provide long term protection in the event of a future exposure. However, the mechanism by which this diversity is generated from a single founding cell remains a highly controversial topic. Resolving the relative contributions of various models of T cell differentiation – as well as generally defining the mechanisms by which a single cell gives rise to distinctly different progeny in various biological systems – requires a means of directly tracking single cell lineage concurrently with sensitive measurements of cell phenotype. Here we report the results of collecting scRNA-seq measurements downstream of the hydrodynamic trap array, which offer one method of addressing this experimental need.

3.2 | Lineage dependent gene expression patterns

3.2.1 | Transcriptional similarity amongst related cells

After culturing single-cells in the hydrodynamic trap array for multiple generations, single cells were released from the device for scRNA-seq as described earlier. Using time lapse imaging of single-cell proliferation we were able to define sister and cousin cell pairs for each four cell lineage (Figure 2.1). To determine transcriptional patterns associated with lineage relationships in these cell systems, we compared gene expression similarity between related (i.e. sisters and cousins) and unrelated cell pairs. Here, unrelated cell pairs refer to cells that were not derived from a common ancestor as observed in the device (i.e. from different lineages). Transcriptional similarities were determined using Euclidean distance measurements to quantify the distance between two cells in log-transformed transcriptional space with smaller distances signifying more similar gene expression profiles (3.5.3).
Figure 3.1 | Comparing transcriptional similarity between related and unrelated cell pairs

Comparison of Euclidean distances (measured in log-transformed transcripts per million) between sister cells, cousin cells and unrelated cells for CD8+ T cells (n = 43, 73 and 4,544 respectively) and L1210 cells (n = 37, 60 and 3,064 respectively) for the entire gene set (9,997 genes and 10,658 genes for CD8+ T cells and L1210, respectively). Groups were compared with a Mann-Whitney U test. The shaded area overlay on the unrelated cell pair measurements has a width corresponding to differences in time since division for these cells. The widths were constructed using a 250-point moving average of the pairwise differences in time since division for visual clarity. Scale bar indicates 3 hours.

When comparing global expression levels in both cell types, sister cell pairs showed significantly higher transcriptional similarity than unrelated cell pairs (p=0.03 and p<0.001 for L1210 and CD8+ T cells, respectively; Mann-Whitney U test) (Figure 3.1). Seeing as multiple methods exist to quantify distances between paired objects in multidimensional space, we sought to
Figure 3.2 | Comparing Spearman distances between related and unrelated cell pairs
Comparison of Spearman distances between sister cells, cousin cells, and unrelated cells for CD8+ T cells ($n = 43, 73$ and $4,544$ respectively) and L1210 cells ($n = 37, 60$ and $3,064$ respectively) for the entire gene sets for both cell types ($9,997$ genes and $10,658$ genes for CD8+ T cells and L1210, respectively).

determine whether or not this pattern of transcriptional similarity held true for alternative metrics.

We found that conducting these analyses with Spearman distances, a rank-based metric of transcriptional similarity, yielded similar results across all comparisons (Figure 3.2). Seeing as these results suggested a greater degree of global transcriptional similarity between pairs of related versus unrelated cells we sought to determine whether an unsupervised hierarchical clustering analysis of gene expression would faithfully recapitulate the lineage hierarchies we measured using time-lapse imaging (3.5.5).

We found that clustering analysis was in fact reasonably successful in assigning cellular lineage relationships computationally (Figure 3.3). For the CD8+ T cells, 16 out of 97 cells were
correctly paired with their sister cell, while 38 out of 97 cells were paired with a cell from the same lineage (i.e. cousin or sister cell). The probabilities of these sister and lineage pairs occurring by random chance, as modeled with a binomial distribution, are $5.5 \times 10^{-15}$ and $8.9 \times 10^{-32}$, respectively. Similarly, for the L1210 cells, 14 out of 80 cells were correctly paired with their sister cell, while 24 out of 80 cells were paired with a cell from the correct lineage. In this case, the probabilities of these sister and lineage pairs occurring by random chance, once again as modeled with a binomial distribution, are $1.5 \times 10^{-12}$ and $1.1 \times 10^{-15}$, respectively.

These results suggest that unsupervised clustering of single cell transcriptional profiles can effectively reconstruct lineage membership although with a much lower accuracy compared to lineage tracking by means of direct observation of cell division via time-lapse microscopy.

Furthermore, it is important to note that in other single cell experimental designs in which cells are isolated using methods including FACS sorting, the Fluidigm C1 platform, or micromanipulation, there will likely be either larger sample sizes or fewer closely related cells, which will reduce the efficacy of computationally reconstructing lineage relationships. This suggests that direct observation of single cell lineage upstream of transcriptional profiles offer a higher degree of accuracy in analyzing lineage dependent transcriptional signatures. However, this data set can also be used to provide a ground truth reference for lineage reconstruction with more advanced computational methods developed in the future.
Unsupervised hierarchical clustering analysis for scRNA-seq profiles collected for (a) CD8+ T cells and (b) L1210 cells. Branches highlighted in magenta indicate cells that clustered most closely with another cell known to be of the same lineage while asterisks indicate cells that were clustered most closely with the cell known to be their sister. These cases indicate successful reconstruction of lineage relationships with unsupervised clustering.
3.2.2 Clone-dependent transcriptional signatures of CD8+ T cells

Although transcriptome wide comparisons of gene expression similarity amongst related cells provides clear evidence of lineage-dependent gene expression profiles, it is difficult to use these global metrics to further characterize particular biological pathways which show heritability within a single-cell lineage. These effects of lineage and clonality in T cells are of particular interest in the context of effector cell function and differentiation in response to antigenic stimuli. To explore these functional differences we performed a similar transcriptional similarity comparison with a subset of genes with functional relevance to CD8+ T cell activation, differentiation and cytotoxic function. For this gene subset, there was once again significantly greater transcriptional similarity (p<0.001, Mann-Whitney U test) for related cells (sister and cousin cell pairs) as compared with unrelated cell pairs (Figure 3.4). Since each founding CD8+ T cell loaded into the device represents a unique clone, these results suggest lower intra- than interclonal transcriptional variation for these cells.

In addition to these aggregate measurements, we compared intra- and interclonal similarity at the single gene level. The genes that showed significantly more intra- than interclonal similarity were highly enriched for gene ontology annotations relating to T cell activation and immune cell function (p<0.001, Mann-Whitney U test) (3.5.8). Interestingly, gene expression for Granzyme B (Gzmb) – whose protein product plays a key role in cytotoxic T cell mediated target cell killing – showed one of the highest levels of clonal similarity, with strong correlations in Gzmb expression levels between sisters (R^2=0.524, n = 43) and cousins (R^2=0.517, n = 73) as compared to unrelated cells (R^2=0.002, n = 4,544) (Figure 3.5). These clonal signatures appeared to manifest as differences in intraclonal transcriptional similarities as well. For instance, at the intraclonal level for CD8+ T cells, genes which showed stronger correlation
between sister cells than between cousin cells were also enriched for gene ontology terms relating to T cell activation and immune cell function (p<0.05; , Mann-Whitney U test) (Figure 3.6). A similar analysis in L1210 cells, which do not require activation and are not actively differentiating, demonstrated that genes with expression levels more correlated in sister cells were instead enriched for basic biological functional annotations including cell metabolism and biosynthetic processes. Both cell types showed very few genes (<10) that were more similarly expressed in cousins as compared to sisters and these genes did not reveal any functional enrichment. This result is consistent with the positive skew observed for the $\rho_{\text{diff}}$ distributions for both L1210 and CD8+ T cells ($\gamma = 0.294$ and 0.318, respectively), which suggests that, for both
Figure 3.5 | Clone-dependent Granzyme B expression
Plot of Granzyme B expression levels (measured in log-transformed transcripts per million) in sister cell pairs (blue circles), cousin cell pairs (red circles), and unrelated cell pairs (grayscale density plot with darkest regions corresponding to highest relative occupancy). The correlation coefficients of related and unrelated cell pairs were compared with a Fisher's z transformation (p = 0.005).

...cell types, there are more genes with a higher correlation between sister cell expression levels than between cousin cell expression levels. To our knowledge, these measurements offer the first direct comparison of inter- and intraclonal variability in activated CD8+ T cells with a priori knowledge of lineage relationships and, when taken together, suggest lineage-dependent transcriptional signatures corresponding to unique functional phenotypes.

Although these data demonstrate clone-specific transcriptional patterns in CD8+ T cells they do not necessarily explain the source of this variability. For instance, it is possible that these clone specific variations in gene expression are due to the existence of pre-existing subsets of CD8+ T cells prior to activation as opposed to variations in differentiation progression in response to antigen stimulation. In order to account for the possibility of pre-existing subsets, we assigned a
Figure 3.6 | Intraclonal transcriptional comparisons
The difference in Spearman correlation coefficients for expression levels in sister cell pairs and cousin cell pairs ($\rho_{\text{diff}} = \rho_{\text{sisters}} - \rho_{\text{cousins}}$) was determined for each gene in (a) CD8+ T cells (9,997 genes total) and (b) L1210 cells (10,658 genes total). To determine genes that were expressed more similarly in sister or cousin cells, we defined an expected null distribution of $\rho_{\text{diff}}$ with a mean of zero and a standard deviation approximated as the average standard deviations of $\rho_{\text{diff}}$ we observed in these two cell types (0.15). We then defined the values of $\rho_{\text{diff}}$ corresponding to the top and bottom 1% of this distribution ($\pm 0.349$, vertical dashed lines) as the thresholds for the highest and lowest values of $\rho_{\text{diff}}$ to determine the genes that are more similarly expressed in sister cells or cousin cells, respectively.

score to each single CD8+ T cells based on gene expression patterns related to either activated memory or activated naïve CD8+ T cells (3.5.6). The subset scores do not appear to demonstrate lineage-dependent values that partition into clear groups of high and low scores (Figure 3.7). This does not fully exclude the effects of pre-existing CD8+ T cell subsets but suggests that the lineages studied here do not display diverging transcriptional profiles driven primarily by the phenotype of the founding cell. To quantify this effect further, unrelated pairs of CD8+ T cells were ranked by the absolute value of the difference in their subset scores ($\Delta$ subset score) and the dataset was split in thirds to produce groups with high (top third, $n = 1,514$) and low (bottom third, $n = 1,514$) differences in subset scores. The Euclidean distances between cell pairs in each of these groups was then compared with a Mann-Whitney U test for the entire gene list (9,997
genes) and a subset of genes relating to T cell activation and function (142 genes). Both of these comparisons revealed that cells with a smaller difference in subset scores were more transcriptionally similar, with the effect appearing to be more pronounced in the subset of genes relating to T cell function (Figure 3.7).

To reduce the effects of pre-existing CD8+ T cell subsets when comparing transcriptional similarity, Euclidean distance measurements of CD8+ sister (n = 43) and cousin (n = 73) cell pairs were compared to unrelated cell pairs with the smallest differences in subset scores (low Δ subset score, n = 1,514) for the complete gene list (9,997 genes) and for a subset of genes relating to T cell activation and function (142 genes). These results indicate that unrelated cell pairs with similar subset scores still show less transcriptional similarity than related cell pairs which suggests that, although pre-existing CD8+ T cell subsets may partially contribute to inter-lineage transcriptional variability, they are not the main drivers of differences in transcriptional similarity between related and unrelated cell pairs (Figure 3.7).
Figure 3.7 | CD8+ T cell subset scoring
A score describing the relative expression of genes related to activated-memory versus activated-naïve cells was determined for each single cell as described in 3.5.6. Each vertical line in (a) indicates a single CD8+ T cell lineage with points indicating individual cells. The Euclidean distance between cells with high or low differences in subset scores were then computed for (b) the complete gene list as well as (c) a subset of genes relating to T cell activation and function. Finally, the Euclidean distances between related cell pairs were compared only to unrelated cell pairs that had similar subset scores for (d) the complete gene list as well as (e) a subset of genes related to T cell activation and function. After Bonferroni correction: * p<0.05, ** p<0.01, ***p<0.001.
3.3 | Cell cycle gene expression patterns

3.3.1 | Decoupling cell cycle and lineage dependence

Recent work has demonstrated that cell cycle-dependent transcriptional profiles in single-cell RNA-seq measurements may obfuscate underlying phenotypic relationships between cells\(^{35}\). Therefore, we sought to determine whether the observed intraclonal transcriptional similarities were primarily the result of cell cycle stage proximity or lineage relationship. When released from the device, cells derived from a single clone are inherently at similar cell cycle stages by virtue of originating from common division events. Unrelated cell pairs, however, are drawn from various lineages and have, on average, a greater difference in cell cycle proximity. Direct observation of proliferation in the device allowed us to account for this confounding effect by independently determining the time since division, an approximate measurement of cell cycle progression, for each cell analyzed. Differences in time since division for unrelated cell pairs were then used as a proxy for the extent to which cell cycle stage differed for each of the transcriptional similarity measurements.

Comparison of unrelated L1210 and CD8\(^+\) T cells demonstrated that cell pairs with smaller differences in times since division were significantly more transcriptionally similar (p<0.001, Mann-Whitney U test) (Figure 3.8). For CD8\(^+\) T cells, this effect was even more pronounced when considering a subset of genes with cell cycle related functional annotation (p<0.001, Mann-Whitney U test). Similarly, in L1210 cells, genes which showed greater expression level similarity among related cells as compared to unrelated cells were enriched for cell cycle associated gene ontology terms (Figure 3.6). These results suggest that global transcriptional similarity for related L1210 and CD8\(^+\) T cells is at least partially due to cell cycle stage.
Figure 3.8 | Effect of cell cycle stage proximity on transcriptional similarity

(a) Global gene expression-based Euclidean distances between unrelated pairs of L1210 and CD8+ T cells were ranked by their corresponding difference in times since division (Δt) and the dataset was split in thirds to produce groups with low (bottom third, n = 1,514 and 1,021 for CD8+ and L1210 cells, respectively) and high (top third, n = 1,514 and 1,021 for CD8+ and L1210 cells, respectively) Δt values corresponding to cells with more and less similar cell cycle stages, respectively. These groups were then compared with a Mann-Whitney U test. This same analysis was applied for (b) a subset of genes with cell cycle related gene annotations (688 genes total) for CD8+ T cells and (c) a subset of genes with gene annotations related to T cell activation and function (142 genes total) for CD8+ T cells. Euclidean distance measurements of sister (n = 43 and 37 for CD8+ and L1210 respectively) and cousin cell pairs (n = 73 and 60 for CD8+ and L1210 respectively) were compared to unrelated cell pairs that had a difference in times since division (Δt) of less than 2 hours (n = 1,006 and 495 for CD8+ and L1210 respectively) for (d) the entire gene list (9,997 genes and 10,658 genes for CD8+ T cells and L1210 respectively), (e) genes relating to cell cycle progression in CD8+ T cells (688 genes total) and (f) genes relating to T cell activation and function (142 genes total)\textsuperscript{43}. 

---

a. Complete gene list  

b. Cell cycle gene list  
c. T cell function gene list  
d. Complete gene list  
e. Cell cycle gene list  
f. T cell function gene list
proximity and not entirely governed by lineage relationships. However, this dependence on cell cycle stage was less pronounced for a subset of genes in CD8+ T cells with functional annotations relating to T cell activation and function (Figure 3.8). Expression levels of these genes – including Gzmb – showed a strong correlation between related cells but no correlation with time since division (p=0.005, Fisher's z transformation) (Figure 3.9). Furthermore, when limiting the transcriptional similarity analysis to include only unrelated cell pairs that have similar cell cycle stages, sister and cousin cell pairs still demonstrate greater transcriptional similarity (Figure 3.8). These results suggest that lineage relationships, and not cell cycle stage proximity, are the dominant factor accounting for the similar expression patterns of genes relating to T cell function observed in single CD8+ T cell clones.

Figure 3.9 | Granzyme B expression level as a function of cell cycle progression
Plot of single-cell Granzyme B expression levels (log-transformed transcriptions per million) versus time since division measurements collected for each cell. As seen in Figure 3.5, related cells show similar expression levels of Granzyme B. Here we see that this similarity is not due to cell cycle proximity between related cells seeing as there is no correlation between cell cycle stage and Granzyme B expression level.
3.3.2 | Latent variable models describing cell cycle transcriptional signatures

To further examine cell cycle-dependent transcriptional patterns, we used partial least squares regression (PLSR) models to find the multivariate, weighted combination of genes — referred to as the latent variable — that covaried most significantly with single-cell measurements of time since division (3.5.7). After initial model construction with the entire gene set, we limited our analysis to subsets including the 300 genes that accounted for the most covariance with time since division for each cell type (Figure 3.12). Reconstructing the models with these reduced gene lists resulted in strong correlations between the latent variable scores and experimental measurements of time since division for both the CD8+ T cells ($R^2 = 0.77 \pm 0.03$ s.d., $n = 10$ iterations) and L1210 cells ($R^2 = 0.84 \pm 0.02$ s.d., $n = 10$ iterations) (Figure 3.10). Cross validation of these models demonstrates that these transcriptional patterns are not artifactual signatures associated with variations amongst cells. Furthermore, random permutation analysis verifies that the results of each model were not due to over fitting ($p<0.001$, pairwise Wilcoxon signed rank test). Therefore, these transcriptional signatures are able to explain a majority of the gene expression variability associated with cell cycle progression. As such, these models could be used to predict cell cycle stage without prior knowledge of time since division for these cell types.

For each model, the optimized subsets of 300 genes were highly enriched for genes with cell cycle-related functional annotations. However, there were only 28 genes that were common to both subsets, indicating that cell cycle progression transcriptional signatures are different between the two cell types. This cell-type specificity may be in part due to dissimilarities in cell cycle regulation between L1210, a constitutively proliferating lymphocytic leukemia, and CD8+ T cells, which require activation and growth factor stimulation to induce proliferation. Furthermore, these results
are consistent with previous work that demonstrated distinct cell cycle signatures associated with various cell types of hematopoietic origin\textsuperscript{56}. Altogether, these observations suggest that independent measurements of cell cycle-associated transcriptional patterns may be necessary to describe different cell types that have distinct mechanisms driving cell growth.

Figure 3.10 | Partial least squares regression modeling of single-cell time since division
Plot of scores on the first latent variable of partial least squares regression models constructed with expression measurements as predictor variables and the time since division for each single cell as the response variable for CD8\(^+\) T cells (left) and L1210 cells (right). The final models were constructed with genes corresponding to the top 300 VIP scores for each cell type (3.5.7).
3.4 | Discussion

We have demonstrated the utility of linked gene expression and lineage measurements with results indicating both lineage and cell cycle dependent gene expression profiles in two different cell types. While the system is limited to in vitro studies, it allows for the analysis of single cell development and lineage progression under highly-controlled, user-defined culture conditions. Such an approach enables the study of cell intrinsic developmental patterns that can serve as a benchmark for measurements of differentiation in vivo, where it may be difficult to deconvolve the effects of the microenvironment on cellular development. Specifically, this method may shed light on to the relative contributions of various CD8+ T cell differentiation models, which have been difficult to reconcile with in vivo single cell measurements alone \cite{28, 30, 54}.

The ability to characterize transcriptional profiles associated with cell cycle progression also offers a novel means of exploring various biological questions. For instance, these linked measurements of gene expression and time since division can be used to further explore the mechanisms of cell cycle regulation and progression in various cell types. Here we found that cells with similar developmental trajectories – i.e. various cells of hematopoietic origin – still demonstrated unique transcriptional signatures associated with cell cycle progression. We therefore envision divergent cell models, for instance healthy versus oncogenically transformed cells, will display distinct transcriptional signatures which may be further defined using the approaches described here.

In addition to exploring the mechanisms of cell cycle regulation, these linked data sets may also offer the opportunity to account for the confounding effects of cell cycle progression which are often considered nuisance parameters in scRNA-seq analyses. In combination with recently
developed computational approaches\textsuperscript{35} and high-throughput single-cell sequencing methods\textsuperscript{12, 13}, this multiplexing of data may allow for more sensitive analyses of single-cell RNA-seq profiles and help to distinguish subtle, but meaningful, functional signatures from cell cycle in various biological contexts.

3.5 | Methods

3.5.1 | Single-cell RNA-seq

Single-cell RNA isolation, cDNA library synthesis, next generation sequencing, read alignment, and gene expression estimation were performed as described previously\textsuperscript{8}. Briefly, Smart-Seq2 whole transcriptome amplification and library preparation was performed on harvested single cells\textsuperscript{5, 7}. Single-cell libraries were then sequenced on a NextSeq500 using 30-bp paired end reads to an average depth of $1,229,637 \pm 60,907$ reads (s.e.m.), and expression estimates (transcripts per million; TPM) for all UCSC-annotated mouse genes (mm10) were calculated in RSEM. The average transcriptomic alignment percentage was $64.3 \pm 0.63\%$ (s.e.m) and the average number of detected genes was $6925 \pm 170$ (s.e.m).

3.5.2 | Gene expression data pre-processing

All analysis was performed on log-transformed expression level measurements ($\ln(\text{TPM}+1)$). Any cell doublets (as observed during single cell release), as well as single-cell libraries with less than 1,000 mapped genes, were excluded from further analysis. All genes which were expressed at a level of $\ln(\text{TPM}+1)>1$ for greater than 10% of cells were included in the final analysis. These
constraints yielded 97 single CD8+ T cells (out of 106 total) and 80 single L1210 cells (out of 88 total) with 9,997 and 10,658 genes for further analysis, respectively.

3.5.3 | Transcriptional similarity comparison

For the analysis of the full gene lists, as well as cell cycle and T cell specific gene lists, weighted Euclidean distance was used as a metric for transcriptional similarity between cell pairs. These aggregate measurements of pairwise Euclidean distances for sister cells, cousin cells and unrelated cells were compared with a Mann-Whitney U test. P values for these tests were Bonferroni adjusted to account for multiple hypothesis testing. These analyses were robust to the use of different distance metrics, such as Spearman distance, which yielded the same results (Figure 3.2). All statistical analysis was performed in MATLAB (Mathworks).

3.5.4 | Constructing single-cell gene expression weight matrices

Because scRNA-seq relies on a very low RNA input for cDNA library construction and amplification, there is a fair amount of dropout that occurs particularly for low copy number transcripts. This dropout results in concentration dependent false negative rate that can vary from cell to cell based on slight differences in the quality and fidelity of library construction. Shalek et al. formalized this behavior mathematically by developing an equation that assigns a weight to each cell/gene pair based on the quality of the cDNA library7. Briefly, genes are first binned in to groups by their average expression levels across all single cells. Then, the number of genes detected in each bin for each single cell is used to construct a sigmoidal fit representing the fraction of genes detected versus the average expression level (Figure 3.11). These curves can be
Figure 3.11 | Assessing single-cell cDNA library quality
Plots of fraction of genes detected versus average gene expression values (in log-transformed transcripts per million) for L1210 and CD8+ T cells. Each separate line demonstrates a logistic regression fit for each single-cell collected for sequencing constructed using the method developed by Shalek et al. used as a measure of cDNA library quality. For instance, cells which show a high fraction of genes detected at a low average expression level indicate a library that is able to detect even lowly expressed genes and thus suggests a high quality library. Alternatively, a library which shows a high fraction of genes expressed only for genes with the highest average expression levels indicates poor library construction. This sigmoidal curve is used to determine the weight of cell/gene pairs by finding the value along this trajectory for a gene with a given average expression level in a particular cell. This ensures that a higher weight is placed on genes that have a higher detection accuracy. These weights were used for all subsequent analyses.
3.5.5 | Unsupervised clustering and lineage reconstruction

To perform unsupervised clustering of single CD8+ T cells and L1210 cells, we first constructed lists of highly variable genes using the Seurat package in R (Jackstraw method). This resulted in subsets of 777 and 647 highly variable genes for CD8+ T cells and L1210 cells, respectively.

We then performed agglomerative clustering for each cell type in MATLAB using Ward’s method after mean centering and unit variance adjustment (Figure 3.3). We compared the results of hierarchical clustering for these two cell types to the known lineage information gathered with time-lapse microscopy.

3.5.6 | CD8+ T cell subset scoring

In previously published work, Hinrichs et al. sorted naïve and memory CD8+ T cells from murine splenocytes prior to activation with anti-CD3 and anti-CD28—the same activation scheme used in this study. Microarray analysis of these two populations revealed sets of genes that were differentially regulated in naïve cells after activation (activated-naïve cells) relative to memory cells after activation (activated-memory cells; MSigDB systematic names M3660 and M3662 for up-regulated and down-regulated gene lists, respectively). To determine a subset score for each cell, we calculated the weighted mean of expression levels for up-regulated genes and subtracted the weighted mean of expression levels for down-regulated genes.

3.5.7 | Partial least squares regression (PLSR) modeling

All PLSR modeling was performed with the commercially available PLS Toolbox for MATLAB (EigenVector Research). Single latent variable models were computed with single-cell time since division measurements used to construct the response variable and corresponding single-cell gene expression measurements used as predictor variables. For variable selection, the models were
Figure 3.12 | Gene selection for partial least squares regression analysis
Plot of the coefficients of determination ($R^2$) between the first latent variable scores and single-cell measurements of time since division as a function of the number of VIP ranked genes included in the model for (a) CD8+ T cells and (b) L1210 cells. These $R^2$ values indicate the fraction of gene expression variance due to time since division that is explained by these models. The variance explained by the full model (black) and cross-validated model (red) are both included for comparison. The amount of variance explained by the model appears to plateau at around 300 genes for each cell type. For this reason, we used a subset of genes with the top 300 VIP scores for the final model construction presented in Figure 3.10.

Iteratively constructed with 90% of the data and genes were sorted based on the average variable importance to the projection (VIP) scores across ten iterations. To determine the optimal number of top-ranked genes to include, the model was iteratively re-run with an increasing number of genes (Figure 3.12). For the final model, we chose to use the genes with the top 300 VIP scores; at this point, the addition of more genes did not appear to increase the amount of variance captured by the cross validated model. These genes were subsequently used as predictor variables for construction of the final models. The final models were cross-validated by iteratively constructing models with 90% of cells and using them to find the latent variable scores for the remaining 10%.

The mean coefficients of determination ($\pm$ s.d.) across all ten iterations are reported in Figure 3.10.
The final models were tested for over-fitting by randomly permuting the response variables, recalculating the model and determining the permuted model residuals. A Mann-Whitney U test demonstrated significantly (p<0.001) smaller residuals for the original models as compared with the permuted models, indicating that the model is not over-fit for either cell type.

3.5.8 | Gene-annotation enrichment analysis

Functional enrichment analysis was performed with DAVID v6.7. The gene lists used for inter- and intraclonal transcription comparison include genes that demonstrated significantly greater intra- than interclonal expression similarity (FDR<0.05). The lists used to analyze genes which showed greater transcriptional similarity between sister cells than between cousin cells include genes with the top 1% of \( \rho_{\text{diff}} \) values for both cell types (Figure 3.6). The lists used to analyze cell cycle stage-dependent transcription include genes with the top 300 VIP scores as determined by PLSR modeling for the full gene lists of L1210 and CD8+ T cells (3.5.7). The full lists of detected genes in each cell type were used as the background gene lists for all enrichment measurements.
Chapter 4

Suspended microchannel resonator platforms for linking single-cell size and growth rate measurements with downstream assays

4.1 | Introduction

In the last several years, the Manalis Laboratory has designed and implemented systems to collect a wide range of high-resolution single-cell biophysical measurements. The core technology used to implement these platforms is the suspended microchannel resonator (SMR) (Figure 4.1), which has been described in depth elsewhere\(^{37, 38, 60}\). Briefly, a silicon cantilever with an embedded microfluidic channel is constantly actuated at its resonant frequency. As a cell passes through the embedded fluidic channel in this sensor, the total mass of the system changes and thus there is a measurable shift in the resonant frequency. These shifts in resonant frequency are directly proportional to the buoyant mass of the cell or particle passing through the SMR and as such can be used to determine high-resolution mass measurements (Figure 4.1). Previous implementations of these devices have required the traversal of a single cell across a cantilever in order to effectively collect mass measurements. This process can be repeated periodically over...
Figure 4.1 | Suspended microchannel resonator (SMR) for single-cell buoyant mass sensing
(a) Schematic demonstration of a suspended microchannel resonator (SMR) – a microfabricated cantilever with an embedded microfluidic channel which allows single cells to pass through – and the flow path a single cell takes as it is measured. (b) As the cell traverses the cantilever, the change in the mass of the system is measured by means of a resonant frequency shift. This frequency shift ($\Delta f$) is linearly proportional to the buoyant mass of the cell being measured ($m_b$). This buoyant mass is a product of the volume of fluid displaced by the cell ($V_{cell}$) and the difference in density between the cell ($\rho_{cell}$) and the fluid it is suspended in ($\rho_{fluid}$).

$$\Delta f \propto m_b$$

$$m_b = V_{cell}(\rho_{cell} - \rho_{fluid})$$

an extended period of time by passing a cell back-and-forth across a single mass sensor in order to measure the accumulation of buoyant mass over time as a single cell is growing$^{38}$.

Although previously developed platforms offer robust means of measuring single-cell size and growth rate, they are limited in their ability to efficiently isolate single cells downstream, thus precluding additional molecular assays to be linked to these biophysical measurements. Here, we have developed alternative platforms which utilize the same mass sensing technology while also enabling single-cell isolation in order to further expand the depth of information attainable with these systems.
4.2 Array of SMRs with hydrodynamic traps

4.2.1 Design and operation

In order to enable downstream collection of single cells after biophysical measurements, we developed a platform that relies on hydrodynamic trapping of single-cells within a suspended microchannel resonator (Figure 4.2). This design utilizes the same mass sensing mechanism as a traditional SMR, however instead of flowing a cell across a cantilever and monitoring the magnitude of the corresponding peak, a single-cell is captured in a hydrodynamic trap within the cantilever and step changes in the resonant frequency are used to determine the buoyant mass of a single cell. In order to measure the mass accumulation rate – or growth rate – of a single cell, the pressure can periodically be altered to reverse the flow direction within the channel and cause a cell to be returned to the base of the cantilever. This enables a new measurement of the baseline resonant frequency before the cell is once again flown into the device and another frequency shift is measured. Repeating this process periodically allows for many single-cell mass measurements to be collected over time in order to observe single cell growth rates.

Although this method enables high-fidelity measurements of single-cell growth rates, it is limited in that it only allows a single cell to be measured at a time. In order to address this issue, we implemented a device that contains an array of cantilever mass sensors, each with their own hydrodynamic trap to capture many single-cells simultaneously (Figure 4.3). This design consists of eight separate resonating mass sensors with common fluidic bypass channels in order to manipulate a set of eight cells simultaneously. Once a set of cells is loaded into the device,
Figure 4.2 | Suspended microchannel resonator design with a hydrodynamic trap

The device (a) consists of a resonating cantilever (SMR – outlined in red) with an embedded microfluidic channel that contains a single hydrodynamic trap. As a cell enters the device it will flow into the cantilever and eventually become trapped in the hydrodynamic trap (position C in the diagram). The additional buoyant mass of the cell entering the cantilever and becoming captured in the hydrodynamic trap leads to a shift in the resonant frequency of the device (b). The magnitude of this resonant frequency shift is directly proportional to the buoyant mass of the cell being captured. By periodically reversing the flow to release the cell from the trap and return it to the base of the cantilever (position A in the diagram) the cell’s buoyant mass can once again be flown into the device to determine changes in buoyant mass – or mass accumulation rate – over time.

the cells are periodically passed into and out of the hydrodynamic traps in order to determine the corresponding buoyant masses, and thus mass accumulation rate, of each single cell.

For this device, each cantilever has a slightly different length in order to ensure that each mass sensor has a different baseline resonant frequency and thus all eight sensors can be monitored simultaneously, as described by Olcum et al. To address the differences in sensitivity arising from these different physical dimensions, each cantilever was calibrated with polystyrene beads in order to convert from a resonant frequency shift measurement to an absolute buoyant mass measurement.
Figure 4.3 | Array of suspended microchannel resonators with hydrodynamic traps
(a) An array of eight suspended microchannel resonators, each with a hydrodynamic trap, are connected in parallel between two bypass fluidic channels with independent control of upstream (P1) and downstream (P2, P3) pressures in order to allow for independent fluidic control of flow across the traps and along the bypass channels. Each cantilever is designed to be a slightly different length in order to have different baseline resonant frequencies to distinguish each sensor. (b) By trapping 8 cells and periodically releasing and re-capturing them within the hydrodynamic trap structures, 8 mass accumulation rate measurements can be collected simultaneously prior to releasing the cells one-at-a-time for downstream analysis.
Because the fluidic design of the trap-SMR devices shares the same basic structure as the hydrodynamic trap array, we were able to apply the same fluidic operation principles in order to manipulate single cells within the array. Specifically, the ability to independently control the upstream and downstream pressures for either bypass channel allowed us to decouple the flow rate along the bypass channels with the flow rate across the array of traps (Figure 4.3). This fluidic design allowed us to rapidly perfuse cell growth media through the bypass channels for the duration of a cell growth measurement in order to ensure a continuous nutrient supply to the cells within the array. At the same time, we were able to maintain a high degree of resolution when controlling the direction and magnitude of flow in the array. For the majority of the experiment, the flow is directed into the trapping structures—i.e. from left to right in Figure 4.3 with a pressure regime of $P_2 > P_3$. Periodically, however, the flow direction was briefly reversed ($P_3 > P_2$) in order to release the cells from the traps and out of the cantilevers to re-measure the baseline resonant frequency of the system. We found that for collecting high-quality single-cell growth trajectory measurements without applying excessive stress on the cells, a duty cycle of roughly 60 seconds for this release and re-capture protocol was optimal (4.5.3).

Utilizing this fluidic design also enabled the collection of single-cells downstream of the device after short term growth rate measurements were collected for each cell. Specifically, these cells were collected directly into lysis in order to perform single-cell RNA-seq as described in 2.4.4. After collecting single-cell MAR measurements for roughly 45 minutes, the bypass channels of the device were rapidly flushed with RNAse-free PBS in order to ensure that the cells were collected in a buffer optimized for high-quality cDNA library construction downstream. Subsequently, single cells were released from the trap-SMR by altering the relative pressures applied to the device as described in 2.2.5. By recording this process we were able to ultimately
link each single-cell buoyant mass and MAR measurement with single-cell transcriptome signatures collected downstream.

4.2.2 | Measuring single-cell growth in the Trap-SMR array

As described in Figure 2.2, capturing single-cells within a hydrodynamic trap structure produces an altered sheer stress environment that a cell would not experience under traditional operation of an SMR device where a cell is perpetually flowing through a rectangular channel. In order to determine whether or not this altered environment had an effect on single-cell growth rate, we measured the mass accumulation rates of single L1210 and FL5.12 cells, two model systems we had measured previously with the traditional SMR platforms (Figure 4.4). We found that the buoyant mass and MAR measurements overlaid well with previous measurements, suggesting that the platform does not induce a level a stress on the cells that manifests as alterations in growth properties.

4.2.3 | Trap-SMR platform limitations

The trap-SMR platform offers a high degree of fluidic stability consistent with the operation of the hydrodynamic trap array. This stability allows for the consistent collection of growth trajectories followed by single-cell isolation off-chip for scRNA-seq. However, there are a few operating constraints we discovered while performing these growth and release measurements. For instance, while the platform works quite well for monodisperse cells with a roughly spherical morphology – as was the case for the model cell lines L1210 and FL5.12 – the trapping efficiency decreases for cells with a wider range of cell sizes and a more irregular cell shape. This issue was particularly apparent for murine CD8+ T cells which were activated in vitro prior to attempting growth and release measurements with the trap-SMR platform. Within the first 24
Figure 4.4 | Single-cell growth measurements collected with the trap-SMR designs
Plot of buoyant mass and mass accumulation rate (MAR) measurements collected with the trap-SMR platform (black points) and traditional SMR platform (grey points) for (a) a murine leukemia cell line (L1210, left) and (b) a murine pro-B cell line (FL5.12, right).

hours of activation – a time point which is of particular interest when studying early T cell activation dynamics – these cells display a wide range of cell sizes and begin to develop a non-spherical cellular morphology as they begin to increase in size. We therefore found that these cells were difficult to continuously capture and release within the trap platform without the cells occasionally failing to re-capture or smaller cells squeezing through the hydrodynamic traps.

Another drawback of the trap-SMR platform is that the throughput is limited to roughly 8 single cell growth and release measurements per hour. While this throughput is sufficient for conducting studies of continuously proliferative cell lines where cohorts of cells can be collected over several experiments across multiple days, it is not amenable to measurements of time-sensitive samples such as clinical isolates or differentiating cells which may change behavior significantly over the course of days. Furthermore, these single-cell growth and collection
experiments are labor intensive and require extensive user training to conduct successfully. Therefore, it is difficult to envision a significant increase in the scale or efficiency of these collection experiments with the implementation of the trap platform presented here.

These limitations suggest that the trap-SMR platform is most well-suited for measurements of model cell systems which demonstrate stable phenotypes over the course of several days and have morphological characteristics which are amenable to long term hydrodynamic trapping within the device. To address these limitations, we sought to implement a platform which enabled higher-throughput single-cell growth and release experiments, required less user training and intervention, and could be used for a wider range of cell types. These improvements are discussed below.

4.3 | Serial SMR for single-cell growth and collection measurements

4.3.1 | System design and implementation

In order to increase the throughput and ease of use of the single-cell growth and collection experiments, we implemented a modified version of a serial-SMR device, a platform developed by Cermak et al.\textsuperscript{40}. Briefly, the system relies on an array of SMR mass sensors placed periodically along the length of a long microfluidic channel (Figure 4.5). Along the length of this channel, a cell’s mass is measured roughly every two minutes as the cell traverses the array. Therefore, at the end of twenty minutes, there is a series of sequential mass measurements collected for each single cell. This series of mass measurements can subsequently be used to
Figure 4.5 | Serial SMR platform with downstream single-cell collection
Schematic showing the design of a serial SMR device (left) as described by Cermak et al.40. As a cell traverses the microfluidic channel, its mass is measured periodically over the course of roughly twenty minutes in order to determine the mass accumulation rate – or growth rate – for each single cell, data which is represented schematically here (right). As a cell reaches the final mass sensor in the array (blue outline) a three-dimensional motorized stage is triggered to position a PCR tube to capture the cell directly into lysis buffer. The final processed data allows each single-cell growth trajectory to be linked with a corresponding single-cell lysate which is subsequently prepared for scRNA-seq.

determine the mass accumulation rate – or growth rate – for a single cell. Furthermore, seeing as this device relies on unidirectional flow through the array of sensors, many single cells can be passed through device in series, thus drastically increasing the throughput of single-cell size and growth rate measurements as compared with single SMR devices which have been used previously38.

This original serial SMR design was modified in order to enable single-cell collection downstream of size and growth rate measurements. With real time access to the data collected from each single-cell mass sensor, we are able to use the mass measurements collected at the final sensor as an indicator of when a cell is exiting the array and entering the bypass channel of
the device (Figure 4.5). We use these cell exiting events to trigger the motion of a three-dimensional motorized stage, which switches between a waste collection and cell collection configuration. Therefore, each cell that exits the array is collected directly in to lysis buffer in a separate PCR tube to prepare single-cell lysates for downstream scRNA-seq (4.5.4). By continuously recording the position of the three-dimensional stage as well as the frequency data collected from each mass sensor, we are able to link each cell single cell growth rate measurement with its corresponding transcriptional profile collected downstream. In order to ensure the stability of RNA released from single lysed cells, these experiments are performed using 8-tube PCR strips in order to limit the amount of time that single-cell lysate samples are held at room temperature prior to long-term storage at -80°C.

4.3.2 | Fluidic operation of the serial SMR platform

The serial SMR devices employ a similar H-channel geometry used for the hydrodynamic trap array devices discussed in 2.2.2. For instance, the system has independent control of both upstream and downstream pressures applied to the bypass channels. This control enables the establishment of different volumetric flow rates along the bypass channels as compared to flow rate across the mass sensor array (Figure 4.6). In order to measure single-cell growth rates, a constant flow rate is maintained across the array of mass sensors – i.e. Pin-Pout is maintained at a constant value for the entirety of a growth measurement experiment. For the majority of the experiment, the upstream and downstream pressures applied to the bypass channel on the cell-loading side of the array will be held constant in order to flush cells in to the mass sensor array (P1 = P3). However, this fluidic balance leads to very low volumetric flow rate – on the order of 1 µl per hour – and therefore is not suitable for flushing the dead volume of the bypass channel.
Figure 4.6 | Fluidic operation of the serial SMR platform
(a) Schematic of the serial SMR device implementation used to capture single-cells downstream of growth rate and mass measurements. COMSOL fluidic modeling of the highlighted area at the entrance of the array is displayed below for the fluidic regimes used for cell loading and cell flushing (left and right respectively). (b) Resistor diagram describing the fluidic components of the serial SMR design. Based on the symmetry of the bypass channel designs and the tubing used to run fluid through the device all bypass resistances (R1) are equivalent with the exception of the tubing used to collect cells from the device (R2). Seeing as this tubing has a smaller inner diameter, it leads to a higher fluidic resistance (R2>R1). This schematic also includes the key pressure values which determine the fluidic operation of the system including the upstream pressures (P1 and P2), downstream pressures (P3 and P4) and the pressures at the entrance and exit of the mass sensor array (Pin and Pout).
and loading a sample of cells into the platform for measurement. Therefore, in order to load a sample, a significantly higher flow rate is generated along the cell-loading bypass channel \((P_1 \gg P_3)\). During this flushing period, \(P_{in}\) is maintained at a constant value by increasing \(P_1\) and decreasing \(P_3\) by the same value. This ensures consistent flow speed across the mass sensor regardless of whether the cell-loading bypass channel is in a state of cell loading or cell flushing as depicted in Figure 4.6. Depending on the type of cell sample being measured, this flushing regime may also be implemented periodically in order to deliver a fresh plug of cells for measurement or clear any debris that may aggregate in the bypass channel.

When running the platform for single-cell collection experiments downstream of growth measurements, it is necessary to flush the cell from the device as quickly as possible and in the smallest possible volume in order to maintain sufficiently high-throughput as well as maintain compatibility with downstream scRNA-seq protocols. In order to minimize the volume each cell is collected in, we use a smaller inner diameter tube (75 \(\mu\)m) connected to the port from which cells are collected (P4). This increases the resistance \((R_2)\) on the cell-collection side of the device, however this can be compensated for with an increase pressure applied upstream (P2). For the duration of a cell collection experiment, a fixed pressure drop is maintained along the cell collection bypass channel – i.e. P2-P4 is held constant. By extension this results in a constant pressure at the exit of the array of mass sensors (Pout), thus ensuring constant flow across the array as well. The pressure drop along the cell collection bypass channel is typically adjusted such that the volumetric flow rate is roughly 15 \(\mu\)l per minute. With each cell being collected in 5 \(\mu\)l of buffer, this leads to a collection time of roughly 20 seconds per cell.

This time required to flush single-cells from the device imposes a restriction on the frequency at which cells can be passed across the array of mass sensors. For instance, if cells are spaced on
average by ten seconds, there will multiple events where a second cell is released in to the bypass channel while another cell is still being flushed from the system. These cells will subsequently be captured in the same tube containing lysis buffer and the matched transcriptional and physical measurements will be lost. It is therefore necessary to ensure adequate spacing between cells such that these co-collection events happen as infrequently as possible. One approach to controlling the frequency of cell loading is to adjust the concentration of cells that are loaded in to the array of mass sensors. Based on Poisson statistics, the volumetric flow rate and cell concentration per unit volume can be used to determine the average expected time between cells entering the array (Figure 4.7). Although this approach is effective for limiting the number of co-collection events, it imposes inherent throughput limits on the platform. For instance, if the minimum time between cells required for collection is 20 seconds, in order to reduce the co-collection frequency to less than ten percent, the cell concentration would have to be adjusted to yield an average time between cells of roughly 60 seconds. This time is increased further when attempting to achieve a higher success rate for single-cell capture. As such, although the maximum throughput of the system as determined by the time required to flush cells is on the order of 180 cells per hour, a dilution approach alone limits the throughput to just 60 cells per hour.

In order to address the limitations of concentration-based cell loading, we implemented an active loading regime for the serial SMR devices (Figure 4.7). This fluidic process relies on active switching between the flushing and loading configurations presented in Figure 4.6. With real-time access to the data generated by the first mass sensor, we are able to determine when a cell has entered the array based on the corresponding shift in resonant frequency. We use this frequency shift to trigger a switch from the cell loading configuration to the cell flushing
Figure 4.7 | Fluidic approaches for loading single-cells into the serial SMR array
Schematic of the entrance to the mass sensor array (left) which shows the cell-loading bypass channel as well as the first mass sensor that each cell flows through once it enters the array. A shift in resonant frequency is measured each time a single-cell traverses this first mass sensor (right), each vertical spike indicates a single-cell measurement. In the case of passive loading (top), single-cells enter the mass sensor array in a concentration-dependent manner following a Poisson distribution, leading to variability in the time spacing between sequential cell events. For active loading (bottom), the resonant frequency shift associated with a single-cell entering the array is used to trigger a switch from the loading to flushing fluidic regimes presented in Figure 4.6 until a desired time has elapsed and the system reverts back to a cell-loading configuration. This active switching ensures that an equally spaced stream of cells enters the array of mass sensors.

configuration. Although the volumetric sampling from the cell solution is equivalent between these two modes – based on the consistent flow maintained across the array of mass sensors – the volumetric flow fraction directed along the bypass channel while flushing is significantly greater than the flow directed to the array. Therefore, during the flushing configuration, the majority of the streamlines continue along the bypass channel. Because cells are of finite size and occupy multiple streamlines, they are directed along the bypass channel and are not drawn into the array. Therefore, as soon as the system is switched to a flushing regime, cells will not be loaded
in to the array. Once a cell loading event is triggered and flushing begins we instead wait for a set amount of time – for instance 20 seconds, the time required to flush a cell from the system – before switching back to a cell loading regime. In order to capture a cell quickly after switching back to loading, we use a significantly higher concentration cell sample than is typical of experiments relying on Poisson based loading. Using this approach we are able to ensure a continuous stream of cells entering the array with a fixed separation in time.

Although a consistently-spaced stream of cells entering the device significantly increases the maximum achievable throughput of single-cell collection, there is still a limitation imposed by the fluidic behavior of the cells once they have entered the array. Because each cell has a slightly different size and is located at a different location within the cross section of the microfluidic channel, the natural velocity gradient within the channel causes each cell to traverse the mass sensor array at a different speed. Therefore, cells with an optimal spacing of 20 seconds at the first cantilever will have a range of times between cells that averages 20 seconds but with some amount of variability. This variability leads to occasional co-collection events and thus a decreased yield and throughput of the system. In order to address this limitation, we implemented a stop-flow functionality based on multiple cell events registered at the final cantilever of the device. For instance, once a single-cell is measured at the final mass sensor it triggers the motion of the three dimensional motorized stage for collection. If another cell passes the final cantilever while the system is still flushing the first cell – an event that would typically lead to a failed co-collection event – we increase the upstream pressure on the cell collection bypass channel (P2) such that the pressure at the exit of the mass sensor array (Pout) also increases and temporarily stops flow within the array. This stop-flow configuration is maintained
until a sufficient amount of time has elapsed to capture the first cell released from the array at which point normal forward flow is re-established.

This combination of consistent cell spacing at the entrance of the mass sensor array as well as secondary control of the frequency at which cells exit the device ensures that the system can operate at its maximum achievable throughput. Currently, this throughput is limited by the amount of dead volume leading from the chip interface to the tubes used to collect single cells and the amount of time required to clear this volume for each sample.

4.3.3 | Single-cell growth and capture experiments

As discussed in section 4.2.3, the higher throughout growth and collection measurements enabled by the serial SMR platform offer the ability to study cell systems which may show phenotypic divergence over time and are thus not well-suited for lower throughput alternatives such as the trap-SMR platform. In order to demonstrate this utility we sought to measure two different model cell systems that display variations in phenotype over time. First, we conducted single-cell growth and collection measurements for activated murine CD8+ T cells (4.5.5). In response to T cell receptor stimulation, these cells are known to progressively change their metabolic, biophysical and transcriptional characteristics over time. As such, studying this process is not amenable to lower throughput methods which would require multiple days in order to collect a sufficient number of measurements for comparison. In order to characterize the transcriptional and biophysical properties of single-CD8+ T cells during activation, we conducted growth and collection measurements after 24 and 48 hours of stimulation in vitro from the same mouse. For each time point, we were able to collect a total of 96 single-cell lysate samples – each with a linked size and growth rate measurement – in approximately four hours.
Figure 4.8 | Growth characteristics of activated CD8+ T cells
(a) Plot of mass and mass accumulation rate measurements collected for single murine CD8+ T cells that were activated in vitro for either 24 (blue points, n=49) or 48 (red points, n=59) hours. Each single-cell displayed here was captured downstream for scRNA-seq. (b) Plot of mass-normalized mass accumulation rate for each single-cell measurement depicted in (a). Here, the normalized MAR is used as a proxy for the growth efficiency of each single-cell.

Both time points displayed a range of mass and mass accumulation rates that was consistent with previous measurements collected under similar conditions (Figure 4.8)$^{40,43}$. However, we found that although the mass range of cells was similar between the 24 and 48 hour time points, the CD8+ T cells measured at 48 hours demonstrated a higher mass-normalized mass accumulation rate—a proxy for cellular growth efficiency per unit mass. By collecting each of these populations of cells downstream, we were able to explore the transcriptional characteristics associated with these biophysical differences observed between time points, analysis which is described in depth in Chapter 5.
As a second demonstration of the higher-throughput growth and collection measurements, we sought to characterize the biophysical response of a patient-derived glioblastoma multiforme (GBM) cell line to drug treatment. We utilized BT159 cells, a patient-derived GBM cell line which has been studied previously with measurements of single-cell size and growth rate. In the absence of drug treatment we observed biophysical characteristics that were in agreement with these previously collected data sets (Figure 4.9). In order to measure the effect of drug therapy on these biophysical properties we treated the cells with a second-generation MDM2 inhibitor (RG7388) for 16 hours prior to measurement. MDM2 acts as a negative regulator of p53—a broadly functional tumor suppressor protein—by targeting it for proteasomal degradation. For many neurological malignancies, MDM2 activity is often deregulated and thus leads to p53 suppression and a phenotype that is consistent with genetically dysfunctional p53, even in the case of wild-type expression of this protein. As such, treatment of these tumors with an MDM2 inhibitor has been found to stabilize the p53 axis and re-establish its functionality as a tumor suppressor. However, the response of these model cell systems to MDM2 inhibition has been found to show a strong time dependence, with an increase in toxicity with extended exposure to a fixed concentration of the drug. As such, in order to compare single-cell drug response from the same culture of treated cells it is important to collect measurements at a high enough throughput to avoid significant differences in the times each cell was exposed to therapy, a constraint which is achievable with the serial collection platform.

For the BT159 cells we found that there was a clear biophysical response to MDM2 inhibition therapy (Figure 4.9). A majority of the cells that were measured showed a reduced mass accumulation rate as compared to untreated cells. However, there was a sub-population of cells which appeared to have size and growth characteristics that were indistinguishable from the
untreated population of single-cells. Because we captured each of these cells downstream for scRNA-seq, we were able to further explore any transcriptional characteristics that may be associated with this lack of drug response for a subpopulation of cells, analysis which is explored fully in Chapter 5.

**Figure 4.9 | Growth measurements of patient-derived glioblastoma multiforme (GBM) cells**
Plot of mass versus mass accumulation rate (MAR) for a patient-derived GBM cell line BT159 in the absence (black points, n=90) or presence (red triangles, n=91) of drug treatment. For drug treatment, the cells were exposed to a second-generation MDM2 inhibitor (RG7388) for 16 hours prior to measurement. Each cell represented here was captured downstream for scRNA-seq.
4.3.4 | Potential fluidic improvements

As discussed in 4.3.2, one of the main parameters that constrains the maximum achievable throughput of the cell growth and collection platform is the time required to flush a single-cell from the system once it has exited the mass sensor array. One of the most straight-forward means of approaching this limitation is to further optimize the system in order minimize the length of tubing running from the cell collection port (P4) to the tubes used to capture each sample. This reduction provides a benefit on two fronts seeing as it decreases the amount of the dead-volume leading to the sample collection tube and also reduces the hydrodynamic resistance of the collection path, thus increasing the volumetric flow rate of flushing and decreasing the time required to capture single cells. Along the same lines, the system can be operated at a higher running pressure (increased P2) in order to increase the volumetric flow rate along the cell collection bypass channel.

Regardless of the improvements, however, the fluidic design presented in Figure 4.6 has inherent geometric constraints that ultimately limit the rate at which cells can be released from the platform. Namely, as a single-cell exiting the mass sensor array it enters the cell-collection bypass channel which has a significantly greater volumetric flow rate – a similar streamline distribution as seen in the flushing model presented in Figure 4.6. As such, the cell is focused very closely to the wall of the bypass channel as it is being flushed from the device. Based on the parabolic flow profile within the bypass channel, the cell follows a fluidic path which causes it to move slower than the average velocity across the entire channel. Therefore, rather than flushing the system for only long enough to clear the dead volume – based on the volumetric flow rate – it instead has to be flushed for significantly longer in order to ensure that the cell is cleared from the system. For instance, during normal operation of the platform we typically have a fluidic dead volume on the
order of 2 µl, however we flush the system for a total of 5 µl for each cell in order to ensure that the cell exits the device.

There are multiple potential means of addressing this issue of cellular location within the flow field as single-cells are being released. The first involves a modified fluidic design that enables pressure-driven flow focusing in the cell collection bypass channel (Figure 4.10). Here, instead of utilizing an H channel geometry – as presented in Figure 4.6 – the cell collection bypass channel is designed as a T junction. The flow coming from the array of mass sensors acts as the center stream of fluid flow while a pressurized sheath buffer flows on either side of this stream. This sheath flow acts to focus the stream of fluid containing the cell in the center of the channel, where the average flow velocity is highest. Seeing as these devices are operated in a Stokes flow regime, this streamline focusing will be maintained for the length of the channel until the cell is released into a collection tube. As such, each cell that is released from the array travels along the fastest possible flow path out of the device and therefore in order to ensure each cell is released it is only necessary to flush for a time sufficient to clear the dead volume of the collection tubing.

Another means of achieving cell focusing at the center of the release channel is with a slightly varied fabrication approach (Figure 4.10). The original serial mass sensor device utilizes a bypass channel that is etched in the same plane as the channel used to form the array. In this case, once a cell reaches the bypass channel it enters along the side wall as described previously. However, if the bypass channel were etched in a separate layer – for instance in the top glass layer alone – the exit of the array channel could be placed at the center of the cell collection bypass channel. This placement would ensure that cells exiting the mass sensor array were centered in the cell-collection bypass channel, thus ensuring the same increased flushing speed described for the previous design strategy.
Either of these design approaches can theoretically reduce the minimum time required to flush a single cell from the platform and therefore increase the maximum achievable throughput of the growth and collection experiments.

Figure 4.10 | Future improvements to the serial-SMR cell collection device
Centering a cell within the flow field of the cell-collection bypass channel as it exits the mass sensor array will ensure that the cell follows the fastest flow path out of the device, thus minimizing the time required to flush each single-cell from the platform. This can be achieved with (a) a pressure-driven flow focusing design which using sheath fluid to focus the stream of fluid containing a single-cell to the center of the channel or (b) a geometric flow-focusing design which uses an alternative cell-release bypass channel etched in to the glass layer of the device and places the exit of the mass sensor array channel in the center of this bypass channel – as can be viewed by the cross-sectional view of the design.
4.4 | Discussion

We have demonstrated the efficacy of two unique platforms capable of collecting single cells downstream of cellular size and growth rate measurements – a functionality we envision being broadly useful to a range of scientific disciplines moving forward.

As discussed, the limited throughput of the trap-SMR platform makes it most well-suited for studies of stable cell lines with a highly monodisperse morphology. This platform is particularly advantageous for slow growing cell model systems for which brief measurements of cellular growth rate – as in the case of measurements collected with a serial-SMR system – would not be sufficient to accurately characterize cellular growth. For these models, the cells can simply be measured for a longer period of time within the trap-SMR system. Similarly, the trap-SMR platform enables multigenerational single-cell growth rate measurements which are unachievable with the serial-SMR system. In an experimental design similar to that presented in 2.4.2, a single founding cell can be loaded in the first hydrodynamic trapping structure of the trap-SMR platform. As this cell grows and ultimately divides, one of the progeny will travel downstream and be captured by the next available trapping structure. Using this method, it may be possible to further explore how single-cell size and growth characteristics are inherited across multiple generations of cells derived from a common ancestor. Furthermore, using the single-cell release functionality presented above will enable further characterization cellular biophysical properties in the context of differential transcriptional profiles measured downstream.

As a simple flow-through fluidic design, the serial-SMR platform also offers the ability to incorporate well-established optical measurement modalities to complement single-cell biophysical measurements. For instance, photomultiplier tube-based measurements of
fluorescent intensity can be used to characterize the relative levels of markers across multiple fluorophore spectra, as was described by Son et al. In combination with fluorescent reporter model cells systems or standard immunocytochemical methods, these measurements enable a wide range of linked single-cell biophysical and protein level data sets. Similarly, simple bright-field image capture can also be implemented with the serial-SMR thus allowing for cellular morphological characteristics to be linked with single-cell biophysical measurements.

Furthermore, either of these optical measurement approaches – fluorescent or bright-field – can be linked with transcriptional signatures characterized downstream of the platform. The system can therefore ultimately be used to collect comprehensive linked measurements that incorporate morphological, biophysical, protein and transcriptional information from a single cell.

The ability to isolate single-cells downstream of biophysical measurements – offered by both of the platforms presented here – offers a broad range of different linked data types that may be accessible in the future. Although here we focus primarily on the application of scRNA-seq, the ability to isolate individual cells presents an opportunity to conduct various other biochemical single-cell methods that have been developed recently. For instance, in addition to single-cell transcriptional measurements, various groups have recently reported methods for single-cell DNA sequencing, epigenetic profiling, and proteomic profiling. As these biochemical methods become more robust and easier to implement we expect that the platforms presented here will offer new complimentary linked data sets that will help to further explore the regulation of cellular biophysical parameters across a range of physiologic contexts.
4.5 Methods

4.5.1 Calibrating the Trap-SMR platform

In comparison to single-SMR devices and the serial array of SMRs, the Trap-SMR system required an altered method of calibration in order to avoid the complications associated with capturing polystyrene beads in the hydrodynamic traps. For instance, if the calibration were attempted by flowing polystyrene beads directly into the traps they would simply be captured and it would be difficult to load many individual beads without manually reloading the system. We therefore altered the system to run beads in the reverse direction across the array (right to left in Figure 4.3). It is important to note here that if the beads are too small (in the range of 4-7 µm) they will become trapped to the back side of the trap and thus trains of beads build up and prevent accurate measurement of peaks that can be used to calibrate each sensor. We therefore utilized 10 µm polystyrene beads in order to avoid back-trapping and establish accurate relative (cantilever to cantilever ratio) and absolute (Hz to pg) calibration for the platform.

4.5.2 Calibrating the serial-SMR platform

Seeing as each single-cell growth rate measurements relies on a compilation of single-cell mass measurements collected across a series of SMRs it is important to have accurate and stable mass calibration measurements for each individual mass sensor. This calibration is typically achieved by flowing a population of highly monodisperse polystyrene beads through the SMR array and measuring the corresponding buoyant mass shifts. Because these beads have a well-defined size and material density, the mode of the frequency shift measurements can subsequently be used to determine the mass sensitivity for each individual cantilever. For experiments involving
only single-cell growth rate measurements – without collection downstream – the population of beads can be run concurrently with the cell population, as long as the size ranges are non-overlapping. In the case of the collection experiments, however, we avoided running particles concurrently with cells in order to avoid collecting polystyrene particles along with the cell lysate used for scRNA-seq and potentially interfering with downstream cDNA library preparation steps. Therefore, for these experiments, we ran a population of polystyrene beads for calibration both before and after each collection experiment in order to determine optimal calibration factors for each mass sensor in the case of any time-dependent drift throughout the course of the experiment.

4.5.3 | Single-cell growth and release with the Trap-SMR platform

For single-cell growth measurements in the trap-SMR platform, single-cells were loaded in a manner similar to that described in 2.4.3. First, a population of cells at a concentration of approximately $2 \times 10^5$ cells/ml is loaded in to the loading bypass channel (left channel in Figure 4.3). Once a cell is observed, the pressure difference along this bypass channel ($P_1 - P_2$) is periodically switched from positive to negative to pass the cell back-and-forth near the entrance of the bypass channel until it drifts in to the array of SMRs. This process is repeated for a total of 8 cells to fill the array prior to flushing the bypass channel in order to ensure that any free cells are cleared from the device before single cells are released for capture after the growth rate measurements are conducted.

Once a group of 8 cells is captured in the array, the pressure balance across the array is periodically changed from $P_2 > P_3$ (flow directed in to the traps) to $P_3 > P_2$ (flow directed out of the traps). This change in flow direction causes cells to periodically be released from each hydrodynamic trap. As soon as the cell exits and the cantilever, and before it flows in to the path
of the upstream mass sensor, the flow direction is re-established and the cell is recaptured in the hydrodynamic trap within the cantilever mass sensor. This process allows for measurements of resonant frequency shifts to be collected periodically, as is depicted in Figure 4.2.

The single-cell collection procedure is functionally identical to that described in 432.2.5 wherein single-cells are released from the hydrodynamic traps one-at-a-time and flushed from the bypass channel in PBS directly into lysis buffer.

4.5.4 | Single-cell growth and collection with the serial-SMR platform

Single-cell growth measurements were collected in the serial-SMR platform as described previously\(^4\). Briefly, a population of cells at a concentration of approximately \(2 \times 10^5\) was loaded into the cell-loading bypass channel of the device (left channel in Figure 4.5). For cell loading, the upstream and downstream pressures applied to the cell loading bypass channel were set to the same value (\(P_1 = P_3\)). One distinction from the original method of cell loading was the addition of active cell loading, as is described in Figure 4.7. Here, once a single cell enters the array of cantilevers, the fluidics are switched to a cell flushing regime where \(P_1 > P_3\) and cells are flushed past the entrance to the array of mass sensors. This ensures that an evenly spaced sequence of cells is loaded into the array.

For downstream collection of cells from the device, the P4 port – from which single cells are released – is maintained at atmospheric pressure. A high pressure applied to port P2 establishes a significant volumetric flow rate of PBS along the release bypass channel – on the order of 15 μl per minute. Therefore, as soon as a cell exits the array of mass sensors it is quickly flushed from the system and captured directly into lysis buffer. The flow rate through the array of mass
sensors is on the order of 1 μl per hour during typical growth rate measurements. Therefore, although the cells are in cell growth media as they are entering the cell release bypass channel, this media is diluted significantly by the PBS used to flush the cell, effectively leading to a complete buffer exchange prior to capturing the cells for scRNA-seq.

As described earlier, the mass signal generated at the final mass sensor triggers the motion of a three-dimensional motorized stage which positions a PCR tube to capture each cell as it is being flushed from the system. The collection manifold includes mounts for two 8-tube PCR tube-strips, one which is used for waste collection between cell events and one that is used for collecting single cells. The waste tubes are prefilled with 75 μl of PBS in order to rinse the exterior of the collection tubing between each sample. The sample tubes are prefilled with 5 μl of 2x Buffer TCL (Qiagen) with 2% 2-mercaptoethanol. Once a cell exits the array, the collection manifold switches from a waste tube to a sample tube and flushes for 30 seconds before returning to the next available waste tube and waiting for the next cell to be released.

### 4.5.5 | Cell culture

L1210 murine lymphocytic leukemia cells (ATCC CCL219) were cultured in RPMI 1640 (Gibco) with 10% FBS and 1% penicillin-streptomycin solution (Gibco). The L1210 cells used for single-cell RNA-seq were from cultures that had been passaged less than 15 times after the initial thaw of the ATCC aliquot. CD8+ T cells were isolated from a 13 week-old male, C57BL/6J mouse. After splenocyte isolation and red blood cell lysis with ACK buffer (Gibco), naïve CD8+ T cells were purified using a MACS-based naïve CD8a T cell isolation kit (Miltenyi Biotec). These cells were cultured in RPMI 1640 (Gibco) with 10% FBS, 55μM 2-mercaptoethanol (Gibco), 1% penicillin-streptomycin solution (Gibco) and 100 U/mL IL2.
(PeproTech). The CD8+ T cells were activated with 5 μg/mL plate-bound anti-mouse CD3 (clone: 145-2C11, BioLegend catalog number 100314), 2 μg/mL plate-bound murine ICAM (R&D Systems), and 2 μg/mL of anti-mouse CD28 (clone: 37.51, BioLegend catalog number 102112) in solution for either 24 h or 48 h prior to loading into the device43. GBM PDCLs were generated from patient tissue collected under an informed consent protocol (Dana Farber Harvard Cancer Center protocol #10-417) and two waived consent protocols (Dana Farber Harvard Cancer Center protocol #10-043 and Partner's Human Research Center protocol #2002 P000995). All protocols mentioned have been approved by Dana Farber Harvard Cancer Center and Partner's Human Research Center institutional review boards. Cells were grown as tumorspheres in NeuroCult NS-A proliferation media (Miltenyi) 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, Cat#3814), which were kept in a 37 °C, 5% CO₂, and humidified incubator. Prior to loading in the SMR, the PDCLs were dissociated with Accutase (Sigma-Aldrich, Cat#A6964) at 37 °C for 7 min and plated as a single-cell suspension44. For drug treatment measurements, the tumorsphere cultures used the same media with the addition of 250nM RG7388 (Roche), a second-generation MDM2 inhibitor. The cells were treated for 16 h prior to dissociation and measurement in the system.
Chapter 5

Linking single-cell biophysical measurements with gene expression

5.1 | Introduction

The mechanisms of cell size and growth regulation have historically been a key focus of systems biology. Such studies have been of particular interest as of late with the recent development of various molecular, optical and technological tools which have enabled an unprecedented degree of resolution in characterizing single cells. While these studies have provided novel insight into the biophysical and protein-level regulation of cell size and growth in mammalian cells, to date it has been difficult to further contextualize these parameters with complementary single-cell transcriptional profiles. Recently, Padovan-Merhar et al. implemented an imaging approach which enabled linked measurements of cell size and transcript abundance that sought to address this limitation. In conducting this work, these authors found previously unobserved phenomena wherein transcript abundance is maintained across a range of cell sizes by means of differential transcriptional burst size and frequency. However, based on the hybridization and imaging approaches utilized, this method is limited to a single time point, with a small panel of transcriptional targets, and with a fairly limited throughput. As such, technical improvements are required to further explore transcriptome-wide signatures associated with either cell size or growth rate with a throughput that enables the study of a broader range of model systems. Here
we present the results of scRNA-seq with linked cell size and growth measurements for a range of cell systems collected using the platforms described earlier. We believe this approach offers a means of addressing these aforementioned technical limitations and will ultimately present complimentary data to existing approaches for characterizing transcriptional regulation of cell size and growth.

5.2 | Transcriptional analysis of hematopoietic cell lines

In order to validate the quality and sensitivity of the transcriptional measurements collected downstream of the SMR platform, we performed scRNA-seq on populations of single L1210 and FL5.12 cells (3.5.1) for which we had collected the biophysical parameters presented in Figure 4.4. The biophysical properties of these cell types – namely single-cell size and growth rate – have been studied extensively\textsuperscript{38, 62, 72, 75}. Therefore, these linked biophysical and transcriptional measurements offer a unique opportunity to explore the gene expression patterns associated with previously observed physical characteristics of these models.

In order to characterize transcriptional patterns associated with cell size, we first determined the degree to which each gene had a gene expression level that correlated with single-cell buoyant mass measurements. Using these correlation coefficients, we were subsequently able to generate
Table 5.1 | Functional enrichment of gene expression correlated with cell mass

List of top twenty enriched gene ontology terms and the corresponding false discovery rate (FDR) values resulting from gene set enrichment analysis of genes ranked by the correlation coefficient between single-cell mass and expression level for L1210 (left) and FL5.12 (right) cells. Green coloration indicates enrichment within genes that showed a positive correlation with cell mass whereas red coloration indicates an enriched term which showed a negative correlation with cell mass. FDR entries of zero indicate that a value less than 1x10^-6 was returned from GSEA.

For both the L1210 and FL5.12 cells, we found that gene ontology terms relating to cell cycle progression and regulation were highly enriched within these ranked lists (Table 5.1). Furthermore, the correlation directionality for both cell types was consistent with various mechanisms of cell cycle progression. For instance, genes relating to the initiation of DNA replication and the G1-S transition showed a higher expression level in smaller...
cells whereas genes relating to processes such as cytokinesis showed a higher expression level in larger cells. In addition to demonstrating the quality of the data generated from the aforementioned isolation method, these results also indicate that single-cell transcriptional profiles have enough sensitivity to resolve subtle biophysical variations amongst cells— in particular, size variability associated with cell cycle progression.

To further validate the mass-dependent cell cycle signatures obtained from these correlative analyses, we collected a complimentary set of measurements wherein we determined the range of single-cell sizes associated with different cell cycle stages. Specifically, L1210 cells were treated with a fluorescent DNA content stain (Hoechst 33342) prior to collecting buoyant mass measurements in a single-sensor SMR device (5.6.2). This platform is also equipped with a photomultiplier tube along with fluorescence excitation and detection optics in order to collect fluorescent intensity measurements corresponding to each single-cell mass measurement (Figure 5.1). These measurements revealed distinct populations of cells corresponding to n and 2n levels of DNA content, which can be used as a proxy measurement for cells in either the G1 or G2 phase of the cell cycle respectively. Utilizing these differences in fluorescent intensity we were able to approximate the corresponding buoyant mass ranges associated with the G1 (less than 55 pg) and G2 (greater than 60 pg) phases. Subsequently, we were able to use these size bins to determine genes that were differentially expressed between cells at early and late stages of the cell cycle. Once again, we found that genes which were differentially expressed between these groups of single cells were enriched for functional annotations relating to cell cycle progression. Furthermore, the directionality of these differentially expressed genes was also consistent with expectations (Figure 5.2). For instance, Cdc6, a gene associated with the G1-S transition is expressed at a higher level in smaller cells assigned to the G1 phase, whereas Pre1, a gene
Figure 5.1 | Measuring L1210 cell cycle phase with a DNA content stain
Plot of fluorescence intensity of Hoechst 33342—a DNA content stain measured in arbitrary units—versus buoyant mass for L1210 cells, collected using a single SMR device with coupled fluorescence optics. The points are colored by relative density, with areas of higher occupancy indicated with orange and low occupancy with dark blue. The bar chart on the far rate indicates the relative frequency of cells in each fluorescence intensity bin (light blue) and the corresponding relative density fit (red dashed line). The bimodal shape of this density demarcates cells with \( n \) or \( 2n \) levels of DNA, corresponding to cells in the G1 or G2 phase of the cell cycle respectively.

Associated with cytokinesis is expressed at a higher level in larger cells, consistent with its role in late G2.

It is important to note that while multiple scRNA-seq studies have demonstrated distinct transcriptional profiles associated with different cell cycle phases, these studies have focused primarily on either primary cells or adherent cell line models which have doubling times on the order of 24 hours or higher\(^{10-12,36}\). The L1210 and FL5.12 model systems used here, however, have a significantly faster proliferation rate with doubling times in the range of 10-12 hours\(^{38}\).

Despite this condensed time-scale, the single-cell transcriptional data presented here nonetheless display distinct gene expression patterns associated with cell cycle progression. In
addition to demonstrating the resolution of single-cell transcriptomic profiles, these results point to an interesting phenomenon in which conserved cell cycle progression mechanisms can act across distinctly different time scales in different biological systems.

Although both the L1210 and FL5.12 cells displayed cell cycle signatures that correlated with single-cell mass, we sought to further characterize how these signatures differed between these cell types. To this end, we focused on the genes with expression levels that showed the strongest correlation with cell mass in each cell type. From this subset we then isolated all genes which had a cell cycle gene ontology functional annotation associated with them in order to visualize how the expression level of these genes corresponded to single-cell mass measurements (Figure 5.3). We found that L1210 cells showed a stronger correlation between cell cycle gene expression levels and single-cell buoyant mass measurements as compared to FL5.12 cells.
Figure 5.3 | Cell cycle gene expression versus mass
Heat maps showing ranked expression levels (red demonstrates cell with highest expression level, blue the lowest) for genes with functional annotations relating to cell cycle progression for L1210 (left) and FL5.12 (right) cells. The cells are ranked by increasing buoyant mass, represented with the bar plot above each heat map. The labels to the right of each heat map indicate the identities of the cell cycle genes represented for each cell type.

Furthermore, among cell cycle related genes with the strongest correlation with cell mass, FL5.12 cells showed a greater number of genes with a negative correlation with cell size – i.e. genes which show higher expression levels in smaller cells. In order to further compare cell cycle-relate gene expression between these two cell types we determined the correlation coefficients between single-cell mass and the expression levels of a curated list of genes relating to cell cycle progression presented by Buettner et al.\(^\text{35}\) (Figure 5.4). Here we found that, on average, L1210 cells showed a stronger correlation – in terms of absolute correlation coefficients – between single cell mass and the expression levels of these cell cycle genes. Furthermore, FL5.12 cells had a larger fraction of cell cycle genes with a negative correlation with single-cell mass.
mass. However, in addition to cell cycle regulation, genes with expression levels correlating with cell mass in FL5.12 cells also showed significant enrichment for other functional annotations including mitochondrial activity and oxidative phosphorylation (Table 5.1). Taken together, these results suggest that FL5.12 and L1210 cells display distinct transcriptional signatures associated cell size, ranging in biological function from mitochondrial activity to cell cycle regulation. Although further work is required to fully characterize the mechanisms responsible for these diverging transcriptional signatures, the platform and analyses presented here offer a means of generating and testing hypotheses relating to these mechanisms.

Figure 5.4 | Cell cycle gene expression correlation with mass in L1210 and FL5.12 cells
Violin plots and corresponding boxplots showing the Pearson correlation coefficients between the expression levels of a panel of genes related to cell cycle progression (n = 495) and buoyant mass for L1210 (left) and FL5.12 (right) cells.
5.3 | CD8+ T cell activation

In response to antigenic stimulation, CD8+ T cells undergo a well-known set of biophysical, transcriptional and functional changes in the course of developing into proliferative cytotoxic cells. However, the relationship between these dynamic changes—particularly with respect to the interplay between biophysical and transcriptional properties—has been difficult to study and remains largely unexplored. To address this, we sought to characterize the transcriptional profiles of single CD8+ T cells which we collected downstream of the biophysical measurements presented in 4.3.3.

5.3.1 | Distinct transcriptional signatures of different activation time points

As described in Figure 4.8, we observed clear biophysical differences between single CD8+ T cells that had been activated for either 24 or 48 hours in vitro prior to measurement. Specifically, we found that at the 48 hour time point, the cells showed a significantly greater mass accumulation rate per unit mass—used here as a proxy for single-cell growth efficiency—as compared with the 24 hour time point. Utilizing transcriptomic profiles generated with scRNA-seq for each of these time points, we sought to determine whether there were clear differences in the transcriptional profiles of these cells over the course of activation as well. After conducting principal component analysis on these populations, we found that the two activation time points clustered into distinct populations (Figure 5.5). This suggests that are unique transcriptional signatures associated with the 24 and 48 hour activation populations. To determine the functional
**Figure 5.5 | Transcriptional differences over the course of CD8+ T cell activation**

Principle component analysis (PCA) plot showing the scores on the first two principle components for scRNA-seq profiles collected for single CD8+ T cells after 24 (blue) or 48 (right) hours of activation *in vitro*. Each point represents a single cell.

Differences between these populations, we ranked the gene set by the principal component loadings and conducted gene set enrichment analysis. Here, we found that genes which were expressed at a higher level in CD8+ T cells after 48 hours of activation – as indicated by strong negative loadings – showed functional enrichment for T cell effector function and immune activation, consistent with what is expected for cells further along in the differentiation process (Table 5.2). We also found that cells from the 48 hour time point had higher expression levels of genes relating to various metabolic pathways such as glucose metabolism, hexose metabolism and lipid biosynthesis. This increased metabolic activity may account for the higher single-cell growth efficiencies measured for the 48 hour time point.
Table 5.2 | Functional enrichment of genes differentially expressed during CD8+ T cells activation

List of the top twenty gene ontology terms and their corresponding false discovery rates (FDR) that are enriched amongst genes that are differentially expressed between CD8+ T cells activated for either 24 or 48 hours in vitro. To determine enrichment, genes were ranked by their loadings on the first principal component before conducting gene set enrichment analysis. Red coloration indicates enrichment amongst genes that were more highly expressed in the 48 hour time point, which is true for all of the top twenty hits.

<table>
<thead>
<tr>
<th>Enriched Gene Ontology Terms</th>
<th>FDR (q value)</th>
</tr>
</thead>
<tbody>
<tr>
<td>SMALL MOLECULE BIOSYNTHETIC PROCESS</td>
<td>0.008522995</td>
</tr>
<tr>
<td>PROTEIN TARGETING TO MEMBRANE</td>
<td>0.008519964</td>
</tr>
<tr>
<td>MONOSACCHARIDE METABOLIC PROCESS</td>
<td>0.007296835</td>
</tr>
<tr>
<td>HEXOSE METABOLIC PROCESS</td>
<td>0.006690088</td>
</tr>
<tr>
<td>GLUCOSE METABOLIC PROCESS</td>
<td>0.005877795</td>
</tr>
<tr>
<td>STEROID BIOSYNTHETIC PROCESS</td>
<td>0.009064798</td>
</tr>
<tr>
<td>ALCOHOL BIOSYNTHETIC PROCESS</td>
<td>0.013638102</td>
</tr>
<tr>
<td>EXTERNAL SIDE OF PLASMA MEMBRANE</td>
<td>0.018446493</td>
</tr>
<tr>
<td>STEROL BIOSYNTHETIC PROCESS</td>
<td>0.017457398</td>
</tr>
<tr>
<td>POSITIVE REGULATION OF CELL ACTIVATION</td>
<td>0.016419977</td>
</tr>
<tr>
<td>PROTEIN LOCALIZATION TO ENDOPLASMIC RETICULUM</td>
<td>0.021761729</td>
</tr>
<tr>
<td>CELL SURFACE</td>
<td>0.025183068</td>
</tr>
<tr>
<td>POSITIVE REGULATION OF IMMUNE EFFECCTOR PROCESS</td>
<td>0.024748994</td>
</tr>
<tr>
<td>ORGANIC HYDROXY COMPOUND BIOSYNTHETIC PROCESS</td>
<td>0.029493490</td>
</tr>
<tr>
<td>REGULATION OF IMMUNE EFFECCTOR PROCESS</td>
<td>0.039346456</td>
</tr>
<tr>
<td>LIPID BIOSYNTHETIC PROCESS</td>
<td>0.039573282</td>
</tr>
<tr>
<td>INTRINSIC COMPONENT OF PLASMA MEMBRANE</td>
<td>0.040801425</td>
</tr>
<tr>
<td>POSITIVE REGULATION OF IMMUNE SYSTEM PROCESS</td>
<td>0.041686251</td>
</tr>
<tr>
<td>REGULATION OF PRODUCTION OF MOLECULAR MEDIATOR OF IMMUNE RESPONSE</td>
<td>0.046267263</td>
</tr>
<tr>
<td>POSITIVE REGULATION OF CELL CELL ADHESION</td>
<td>0.049296609</td>
</tr>
</tbody>
</table>

5.3.2 | CD8+ T cell growth efficiency during activation

In addition to the biophysical differences in single CD8+ T cells across the 24 and 48 hour time points, we also found that there was heterogeneity in single-cell phenotypes within each of these populations, particularly with regards to growth efficiency. We therefore sought to determine if there were transcriptional signatures associated with single-cell growth efficiency for either of these time points. Specifically, we ranked the gene lists for both time points by how strongly each gene’s expression level correlated with the normalized mass accumulation rates for each cell. From here, we were able to conduct gene set enrichment analysis on this ranked gene list in
Figure 5.6 | Ribosomal functional enrichment in genes related to CD8+ T cell growth efficiency
Gene set enrichment plot for the ribosome gene ontology term compared to the full set of genes mapped for CD8+ T cells after 48 hours of activation. The gene list is ranked by the correlation strength with single-cell normalized growth rate. The enrichment score (ES) for this term is 0.648 with a false discovery rate value less than 1x10^-5. This is a positive enrichment, meaning that genes relating ribosome activity are expressed at a higher level in cells with a higher growth rate per unit mass.

In order to determine what, if any, biological functions were differentially regulated in cells with different growth efficiencies. Interestingly, we found that there were no statistically significant hits (FDR<0.05) in the 24 hour set of cells for genes that showed either a strong positive or negative correlation with single-cell growth efficiency. For the 48 hour time point, however, we found that genes relating to ribosomal activity and protein translation initiation were significantly positively enriched meaning they showed higher expression levels in cells with a higher mass-normalized growth rate (Figure 5.6). These results suggest a possible mechanism by which cells
which are growing more efficiently are regulating increased biomass production. However, whether this transcriptional signature is causal or the result of other regulatory pathways which themselves control growth efficiency cannot be determined from these data and will require further study. The different enrichment profiles seen at 24 versus 48 hours also seem to suggest a divergence in the transcriptional regulation of single-cell biomass accumulation at these time points. It is interesting to note that these populations correspond to time points before (24 hours) and after (48 hours) the first division event following stimulation and as such suggest that these transcriptional differences arise from unique behavior found in blasting versus expanding T lymphocytes.

5.3.3 | CD8+ T cell mass during activation

When comparing the biophysical properties of CD8+ T cells after 24 or 48 hours of activation in vitro we found that while they displayed different growth properties, their range of single-cell mass values were similar for both time points. However, seeing as these time points were characterized by unique transcriptional profiles, it was important to determine whether the gene expression signatures associated with cell mass were also different between these time points. To address this question, we focused on a set of genes known to be associated with cell cycle progression and division. This list of genes, determined by Best et al., was constructed by identifying clusters of genes which were differentially expressed at various time points after murine CD8+ T cell activation, and showed similar time-dependent expression levels. Seeing as this cluster of genes was differentially expressed between blasting – activated lymphocytes that have not yet divided for the first time – and proliferative CD8+ T cells we hypothesized that single cell mass would show different trends with regards to this gene set between the 24 and 48
hour time points. To characterize the relationship between this gene cluster and single-cell mass, we determined the correlation strength between the expression levels of each of these genes and buoyant mass for both time points (Figure 5.7). As a reference data set, we also constructed null distributions of expected correlation coefficients by randomly assigning single-cell mass values to each single-cell transcriptional profile and iteratively reconstructing the range of correlation coefficients. We found that for the 24 hour time point, the correlation coefficient values were not significantly different than the range of values expected from random mass assignment. At the 48 hour time point, however, there was a statistically significant increase in the range of
correlation coefficients suggesting that there is a measurable correlation between cell cycle and division relate gene expression and cell size for this time point (Figure 5.7). These results are further evidence that the relationship between single-cell biophysical properties and gene expression changes throughout the T cell activation process, particularly when comparing cells before and after the first division event following stimulation. Specifically, these data show that genes relating to cell cycle progression and cell division demonstrate a stronger correlation with cell mass for actively proliferating cells, as is the case for cells at the 48 hour time point.

To further explore this relationship between cell cycle gene expression and buoyant mass for different stages of T cell activation, we utilized a set of genes derived from previous measurements of single CD8+ T cells presented in Chapter 3. Specifically, we focused on the set of genes found to correlate with cell cycle progression, as determined by the strong correlation between their expression values and the time since division measurements collected with the hydrodynamic trap array. It is important to note that for these lineage tracking experiments, we focused exclusively on proliferative CD8+ T cells which were beyond their first division event. We therefore hypothesized that this particular cell cycle gene list would show a stronger correlation with the single-cell mass measurements collected for the 48 hour time point, an actively dividing population of cells. To characterize this relationship we found the subset of genes for each activation time point that correlated most strongly with single-cell mass measurements. We then cross-referenced this list of genes with the cell cycle list generated with time since division measurements to determine a union gene list. Finally, we used this union gene list to construct a cell cycle score for each cell – a weighted average expression level across this gene list – and determined the correlation between this cell cycle score and the buoyant mass for each single cell (Figure 5.8). We found that, as expected, this cell cycle score showed a
Figure 5.8 | Cell cycle scoring for CD8+ T cells
Plots of cell cycle score versus buoyant mass for single CD8+ T cells after (a) 24 or (b) 48 hours of activation in vitro. The scores were constructed by finding the weighted average of a set of genes known to be associated with cell cycle progression in CD8+ T cells that also showed a correlation with buoyant mass in either of these time points (5.6.3).

stronger correlation with mass for CD8+ cells collected at the 48 hour time point as compared to those collected at 24 hours. However, there was still an appreciable correlation between this cell cycle score and mass for the 24 hour time point as well. This suggests that while cell cycle gene expression appears to have a stronger relationship with cell mass in proliferative cells, there appears to be a degree of functional overlap where certain cell cycle gene expression patterns are measurable prior to the first division in CD8+ T cells as well.
5.4 | Glioblastoma multiforme (GBM) drug response

As discussed in Chapter 4, we found that BT159 cells – a patient-derived GBM cell line – showed distinct biophysical changes in response to treatment with a targeted MDM2 inhibitor. Specifically, we found that a majority of the cells displayed a reduced growth rate after 16 hours of treatment with drug. In contrast to the CD8+ T cell activation measurements discussed above, which displayed biophysical changes over time due to natural progression through various differentiation states, the GBM cells were subjected to treatment with a targeted small molecule inhibitor which caused the observed changes in phenotype. We were therefore able to determine whether the biophysical changes we observed in response to drug treatment had associated transcriptional signatures that were consistent with targeting the MDM2/p53 axis.

5.4.1 | Transcriptional changes in response to drug treatment

The changes in growth rate of the BT159 GBM cells in response to drug treatment suggested the persistence of a small population of single cells which continued to grow at a rate consistent with the population of untreated cells (Figure 4.9). Therefore, we sought to determine if there was a similar sub population of cells that showed a transcriptional similarity to untreated cells that would also indicate a reduced drug response. To this end, we performed principal component analysis on the scRNA-seq data for populations of single-cells collected downstream of biophysical measurements either with or without MDM2 inhibition. We found that the two populations of single cells formed distinct clusters, suggesting that there is in fact a
Figure 5.9 | Distinct transcriptional signature of drug-treated GBM cells

Plot of scores on the first two principle components for PCA analysis conducted on BT159 patient-derived GBM cells with (red triangles) or without (black points) treatment with a second-generation MDM2 inhibitor (RG7388) for 16 hours prior to measurement. These cells were isolated downstream of the biophysical measurements presented in Figure 4.9.

distinguishable shift in gene expression in these cells in response to drug treatment (Figure 5.9).

Furthermore, when we analyze the gene loadings associated with this principal component analysis we find that genes which are upregulated by p53, such as MDM2, CDKN1A, and TP53I3 are expressed at a higher level in drug-treated cells\textsuperscript{79}. Similarly, genes which are known to be down-regulated by p53, including CDK1, CDC20, and UBE2C all show reduced expression in the drug-treated population. The transcriptional response observed in these populations is thus consistent with the mechanism of the drug used here – an MDM2 inhibitor which increases the stability of p53 signaling.
Interestingly, however, we found that there did not appear to be a subpopulation of drug-treated cells that had transcriptional properties consistent with the untreated sample—a population which would have appeared as a small number of drug-treated cells which clustered more closely with the untreated population than the majority of the treated population. Therefore, the most significant transcriptional differences between these populations of cells—which were determined with principal component analysis—do not explain the heterogeneous biophysical response to drug treatment that these cells show. This result suggests that there may be more subtle transcriptional signatures associated with biophysical response that are masked when considering single-cell transcriptional profiles alone. It is therefore important to analyze these gene expression profiles further in the context of their corresponding biophysical response measurements.

5.4.2 Mass-dependent gene expression in GBM

In response to drug treatment, the BT159 cells displayed a similar range of single-cell masses as compared with the untreated population of cells (Figure 4.9). However, seeing as these two populations of cells had distinct gene expression profiles, it was important to determine whether they also had different patterns associated with cell size for either population. To determine these patterns, we constructed ranked gene lists based on the correlation between gene expression levels and single-cell mass measurements for each population of cells. We then used these lists to perform gene set enrichment analysis in order to determine if there were any biologically meaningful pathways which correlated with cell size for either treated or untreated cells. For the untreated population of cells, we found that there was a strong enrichment for genes relating to cell cycle progression functions such as sister chromatid segregation, kinetochore activity and initiation of DNA replication (Table 5.3). For drug treated cells, however, there were no
Table 5.3 | Functional enrichment of mass-dependent gene expression in untreated BT159 cells

List of the top twenty gene ontology terms and their corresponding false discovery rate (FDR) values that were enriched amongst genes that demonstrated a strong correlation with single-cell mass in BT159 cells in the absence of drug treatment. Green coloration indicates that the gene set corresponding to this functional term is expressed at a higher level in larger cells, this is true for all of the top hits presented here. FDR entries of zero indicate that a value less than $1\times 10^{-6}$ was returned from GSEA.

Biological pathways that were statistically significantly enriched (FDR<0.05) within genes that showed a correlation with single-cell mass. This result demonstrates that in response to treatment with an MDM2 inhibitor, there is a decoupling of cell cycle related gene expression and single-cell size. Although this disruption in cell cycle regulation is perhaps unsurprising considering the broad activity of p53 as a central regulator of cell cycle related gene expression, it has important implications for the methods used to analyze single-cell gene expression data in the context of drug treatment. For instance, several previous studies which have used scRNA-seq to characterize tumor heterogeneity in various primary cancer samples have relied on cell cycle gene expression signatures to determine the fraction of actively cycling cells – a metric which
can be used as a proxy for tumor aggressiveness\textsuperscript{10,11,42}. However, the measurements collected here suggest that cell cycle gene expression alone may not be sufficient when describing the phenotypic response of cells to drug therapy. As such, an alternative analysis approach may be necessary to fully characterize drug response using single-cell transcriptional profiles.

5.4.3 | Growth-dependent gene expression in GBM

Although the mass ranges of the BT159 cells did not appear to change in response to drug treatment, there was a clear decrease in the mass accumulation rate of the majority of cells treated with drug. However, there was also a subpopulation that continued to grow at a rate similar to cells that were not treated with drug. We therefore sought to determine if there was a gene expression signature associated with this heterogeneous growth response observed in drug-treated cells. To this end, we used partial least squares regression (PLSR) analysis to determine the weighted subset of genes that best predicted the growth rate of BT159 cells after treatment with an MDM2 inhibitor (5.6.4). After ranking the genes by their variable importance to the projection (VIP) scores—a numerical value indicating the predictive value of each gene’s expression level in determining the outcome variable, which in this case was single-cell mass accumulation rate—we determined what, if any, biological pathways were enriched amongst the genes which showed the greatest predictive value. We found that there was significant enrichment for genes relating to mitochondrial activity and specifically, mitochondrial genes which are involved in the initiation and regulation of apoptosis such as \textit{BAX} and \textit{BID}. In order to validate this signature, we reconstructed a PLSR model using only the genes relating to mitochondrial activity amongst the list of genes with high predictive value—a subset of 39 genes
Figure 5.10 | Mitochondrial gene expression predicts growth rate in drug-treated BT159 cells
Plot of single-cell mass accumulation rates predicted with a partial least squares regression (PLSR) model — using a subset of genes relating to mitochondrial function — versus the actual mass accumulation measurements presented in Figure 4.9 for BT159 cells which either had been treated with and MDM2 inhibitor for 16 hours prior to measurement (right, red triangles) or left untreated prior to measurement (left, gray points).

in total (Figure 5.10). We found that the mass accumulation rates predicted with this mitochondrial gene subset showed a strong correlation with the actual measurements collected upstream for each single cell ($R^2 = 0.547$). This correlation is the result of a leave-one-out cross-validated method wherein the PLSR model is iteratively reconstructed with all of the cells except for one and subsequently used to predict the value for the remaining cell. As such, this correlation strength is not the result of overfitting underlying noise structures within the data but instead appears to reveal an actual phenotypic correlation. Interestingly, the expression level of this set of mitochondrial genes does not show any correlation with the mass accumulation rates measured for the untreated population of BT159 cells ($R^2 = 0.017$). This result suggests that this subset of mitochondrial genes does not necessarily play a universal role in cell growth regulation.
in these cells but instead is associated with differential response to drug treatment. The fact that
the subset of cells that continue to grow in the presence of drug also show a reduced expression
of genes relating to apoptotic activity appears to be biologically consistent and suggests that
there are in fact transcriptional signatures that can resolve differences in biophysical response to
drug at the single-cell level.

5.5 | Discussion

Here we have demonstrated the ability to collect high-quality transcriptional profiles downstream
of biophysical measurements for a range of model cell systems. Furthermore, we have found that
these transcriptional profiles have sufficient resolution to distinguish subtle phenotypic variations
amongst single cells – for instance, the variations in cell mass associated with cell cycle
progression.

When considering the transcriptional signatures associated with cell mass in two different
hematopoietic cell lines we found that while there were similar cell cycle related patterns, there
were also unique and divergent signatures for each cell type. In combination with existing
molecular tools such as fluorescent reporters of cell cycle progression or protein synthesis, these
transcriptional measurements may provide unique insight into the basic mechanisms of cell size
and growth regulation that are broadly conserved as well as cell-type specific.

Beyond studies of basic biological mechanisms, we have also demonstrated that this method is
amenable to studying cell systems which are not stable for extended periods of time – as was the
case for activating CD8+ T cells and drug-treated GBM cells. This platform is therefore
compatible with clinical sample exploration. For instance, in the case of a primary tumor resection, a set of linked phenotypic and transcriptional single-cell measurements could be collected for hundreds of cells in a matter of hours following surgery. This approach will offer the unique opportunity to explore the transcriptional and biophysical heterogeneity of single tumor cells with minimal time \textit{ex vivo}. Furthermore, these baseline measurements can ultimately be compared with a similar set of data collected for patient-derived cell lines or xenografts developed from the same sample. Together, these results can help to resolve the limitations of model cell systems derived from clinical samples and further develop methods of therapeutic selection and cancer drug development.

5.6 | Methods

5.6.1 | Functional enrichment analysis

Functional enrichment analysis was performed using two main resources: the DAVID Bioinformatics Resource (v6.8) and the Broad Institute’s Gene Set Enrichment Analysis (GSEA) toolset \textsuperscript{59,76}. For analyses in which we were determining functional enrichment within a small subset of genes – as was the case for the gene set relating to mass accumulation rate in GBM cells – we used the DAVID bioinformatics resource as described in 3.5.8. In this case, we used the full list of genes that were detected for the BT159 cells as the background gene list. For analyses in which we utilized the entire gene set and ranked the genes prior to analysis, we used the GSEA Pre-ranked tool. When analyzing the relationship between single-cell biophysical properties and gene expression – including mass, mass accumulation rate, or growth efficiency comparisons for the L1210, FL5.12, CD8+ T or BT159 cells – the genes were ranked by the value of the Spearman coefficient after finding the correlation between gene expression and the
biophysical property of interest. Alternatively, when looking for enrichment in genes that were differentially expressed between the 24 and 48 hour activation time points in the CD8+ T cells or the treated and untreated populations of BT159 cells, we used the gene loadings values on the first principle component in order to rank the gene lists.

5.6.2 | DNA content staining for cell cycle analysis

In order to measure linked cell cycle and buoyant mass measurements in a single SMR device, L1210 cells were stained with Hoechst 33342 (Thermo Fisher Scientific). Cells were stained in complete growth medium (RPMI 1640 + 10%FBS + 1% Antibiotic/Antimycotic + 25 mM HEPES) with 1 μg/ml of Hoechst 33342 for 30 minutes at 37 degrees Celsius. After incubation, cells were spun at 300g for 10 minutes and re-suspended in cold PBS at 2x10^5 cells/ml and stored on ice for the duration of the experiment.

5.6.3 | Cell cycle scoring

Gene subset scores were constructed using a modified method that was described previously by Shalek at al.7. Briefly, the genes used to construct the cell cycle score were determined by comparing the top 300 genes that showed a correlation with cell mass in both the 24 and 48 hour CD8+ T cell time points with the top 300 genes associated with time since division measurements in CD8+ T cells collected using the hydrodynamic trap array 43. Using the union of these gene sets for each time point, we first mean centered the expression data before finding the weighted average of all genes that showed a positive correlation with cell mass and subtracting the weighted average of all genes that showed a negative correlation with cell mass. The weights used for each gene were calculated using the Seurat package which implements the weighting method described earlier (3.5.4)14.
5.6.4 | PLSR analysis

All PLSR modeling was performed with the PLS package in R. Single latent variable models were computed with single-cell mass accumulation rate measurements used to construct the response variable and corresponding single-cell gene expression measurements used as predictor variables. In order to determine the variable importance to the projection (VIP) scores for each gene, we iteratively constructed with 90% of the data and the average scores across ten iterations were used to rank the gene list. After determining the functional enrichment of the genes with the top three hundred VIP scores, we selected a subset of genes relating to mitochondrial function for further PLSR modeling. Here, we once again constructed PLSR models using the single-cell biophysical measurements as the outcome variable but only the subset of interest as the prediction data. For this final model we used a leave-one-out cross validation method wherein the model was constructed using all but one cell and the value of this remaining cell was predicted with the model. These cross validated results are presented in Figure 5.10.
Chapter 6

Appendix

6.1 Trap system parts list

Below we have included a list of all of the necessary components to construct a hydrodynamic trap array platform. The list includes the supplier, part number, part description, and pricing for each component. The only elements of the system that are not commercially available are the metal clamp and face plate which are used to mount the microfluidic chips as these are custom designed parts which are fabricated by an external vendor (ProtoLabs). CAD files for these custom components are available upon request.
<table>
<thead>
<tr>
<th>Supplier</th>
<th>Part #</th>
<th>Description</th>
<th>Quantity</th>
<th>Price</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Thorlabs</td>
<td>TRG-PS</td>
<td>1/2&quot; optical post, 6&quot; length</td>
<td>1</td>
<td>$30.00</td>
<td>$30.00</td>
</tr>
<tr>
<td>Thorlabs</td>
<td>TR2-PS</td>
<td>1/2&quot; optical post, 2&quot; length</td>
<td>2</td>
<td>$25.00</td>
<td>$50.00</td>
</tr>
<tr>
<td>Thorlabs</td>
<td>BAB20-PS</td>
<td>Right-angle clamp for 1/2&quot; optical posts</td>
<td>1</td>
<td>$50.00</td>
<td>$50.00</td>
</tr>
<tr>
<td>Thorlabs</td>
<td>SH250SO5</td>
<td>1/4&quot;-20 Stainless steel Cap Screw, 1/2&quot; long, Pack of 25</td>
<td>1</td>
<td>$10.00</td>
<td>$10.00</td>
</tr>
<tr>
<td>Thorlabs</td>
<td>APIE25E</td>
<td>8-32 to 1/4&quot;-20 adapter</td>
<td>4</td>
<td>$2.00</td>
<td>$8.00</td>
</tr>
<tr>
<td>Thorlabs</td>
<td>S5250SO5</td>
<td>1/4&quot;-20 setscrew, 1/2&quot; long, pack of 25</td>
<td>1</td>
<td>$5.00</td>
<td>$5.00</td>
</tr>
<tr>
<td>Thorlabs</td>
<td>SH85038</td>
<td>8-32 cap screw, 1/8&quot; long, pack of 50</td>
<td>1</td>
<td>$6.00</td>
<td>$6.00</td>
</tr>
<tr>
<td>Thorlabs</td>
<td>WS8038</td>
<td>#8 Washer, pack of 100</td>
<td>1</td>
<td>$1.00</td>
<td>$1.00</td>
</tr>
<tr>
<td>Thorlabs</td>
<td>W2550SO5</td>
<td>1/4&quot; Washer, pack of 100</td>
<td>1</td>
<td>$5.00</td>
<td>$5.00</td>
</tr>
<tr>
<td>Thorlabs</td>
<td>MB1218</td>
<td>Aluminum Breadboard, 12&quot; x 18&quot; x 1/2&quot;, 1/4&quot;-20 Taps</td>
<td>1</td>
<td>$184.00</td>
<td>$184.00</td>
</tr>
<tr>
<td>Thorlabs</td>
<td>RDF1</td>
<td>Rubber damping feet</td>
<td>2</td>
<td>$12.00</td>
<td>$24.00</td>
</tr>
</tbody>
</table>

**Estimated total (without automation):** $458.00

<table>
<thead>
<tr>
<th>Supplier</th>
<th>Part #</th>
<th>Description</th>
<th>Quantity</th>
<th>Price</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cole-Parmer</td>
<td>EW-13042-21</td>
<td>Poly-soft recirculating water bath</td>
<td>1</td>
<td>$2,400.00</td>
<td>$2,400.00</td>
</tr>
<tr>
<td>McMaster</td>
<td>5346K13</td>
<td>1/4&quot;-NPT to barb (1/4&quot; OD) Adapters (pack of 10)</td>
<td>1</td>
<td>$20.00</td>
<td>$20.00</td>
</tr>
<tr>
<td>McMaster</td>
<td>5346K14</td>
<td>1/4&quot;-NPT to barb (1/4&quot; OD) Adapters (pack of 10)</td>
<td>1</td>
<td>$20.00</td>
<td>$20.00</td>
</tr>
<tr>
<td>McMaster</td>
<td>S0785K18</td>
<td>1/8&quot; NPT plug</td>
<td>4</td>
<td>$2.00</td>
<td>$8.00</td>
</tr>
<tr>
<td>McMaster</td>
<td>6516T33</td>
<td>1/2&quot; OD x 1/4&quot; ID I.D. Teflon tubing (25 ft)</td>
<td>1</td>
<td>$75.00</td>
<td>$75.00</td>
</tr>
<tr>
<td>McMaster</td>
<td>S074K32</td>
<td>Worm-drive hose clamps for soft hose and tubing</td>
<td>2</td>
<td>$9.00</td>
<td>$18.00</td>
</tr>
</tbody>
</table>

**Estimated total (with automation):** $14,041.00
6.2 | Alternative experimental approaches using the hydrodynamic trap array

6.2.1 | Bead-based immune cell activation

As discussed in 2.2.3, the hydrodynamic trap array can be used to measure single-cell proliferation kinetics for a wide range of model cell types, particularly immune cells such as CD8+ and CD4+ T lymphocytes. In the case of CD8+ T cells, the experimental approach for collecting these growth kinetics measurements is simplified by the fact that after brief T cell receptor stimulation in vitro – as described in 2.4.5 – the cells will continue to proliferate and differentiate for several generation without the need for sustained stimulation. CD4+ T cells, however, are known to require sustained activation to proliferate and differentiation effectively. We therefore had to adapt an alternative experimental protocol in order to sustain activation throughout the course of our measurement in order to ensure continuous proliferation of the CD4+ T cells. To this end, we implemented a co-trapping strategy whereby we first captured 7.4 μm functionalized polystyrene beads within the hydrodynamic traps before loading in naïve, murine CD4+ T cells (Figure 6.1). The beads were functionalized with anti-murine CD3 and anti-murine CD28 in order to recapitulate the TCR and co-stimulation approach used for traditional in vitro T cell activation assays.

For these experiments, a small subset of naïve CD4+ T cells was loaded in to each lane of the hydrodynamic trap array. Seeing as the beads did not fully occlude the trapping pockets, there was enough flow remaining to direct single-cells in to each trap, however with a lower efficiency than single-cell loading alone. Once these bead-cell conjugates are captured the CD4+ T cells activated and ultimately begin to proliferate. Upon division, one of the two sister cells will
eventually be released and travel downstream to the next available trap that already contains a functionalized bead, thus ensuring stimulation is maintained across multiple generations. The fluidic approach and cell tracking methodology is otherwise the same as presented in 2.4.2.

The beads were functionalized using standard EDAC-mediated protein coupling to carboxylated polystyrene beads (Bangs Laboratory) as per manufacturer’s instructions (Poly Sciences, Inc.). The antibodies used for functionalization were anti-mouse CD3 (clone: 145-2C11, BioLegend catalog number 100314) and anti-mouse CD28 (clone: 37.51, BioLegend catalog number 102112).

Figure 6.1 | Bead-based co-culture in the hydrodynamic trap array
Schematic depicting the experimental approach used to activate single CD4+ T lymphocytes within the hydrodynamic trap array (a). The blue circles represent functionalized polystyrene beads which are loaded in to the array of traps prior to loading single-cells for culture (indicated in orange). The area highlighted in red is depicted as an optical micrograph in (b) where a single CD4+ T cell is captured along with a bead functionalized with anti-CD3 and anti-CD28 antibodies. The cell is bound to the bead and has begun to show an elongated shape consistent with T cell activation via TCR engagement.
6.2.2 | Adherent cell growth

Although section 2.2.3 focuses primarily on proliferation measurements collected for hematopoietic cell systems which grow in suspension, the hydrodynamic trap array can also be utilized to conduct similar growth kinetics and lineage tracking experiments with cell models that growth adherently. However, adherent cell types require a slightly modified experimental protocol and system preparation to measure. For instance, when studying suspension cell models the channels are typically coated with PLL-g-PEG in order to reduce cell sticking over the course of an experiment. However, for adherent cells the system is used without any pre-coating such that cells are exposed to native silicon and can adhere effectively.

Single-cell loading in to the array also requires a slight modification in sample preparation seeing as cells must be detached from a cell culture surface in order to transfer them in to the device. Therefore, prior to loading, cells are exposed to trypsin-EDTA in a manner consistent with adherent cell passaging in order to achieve a single-cell suspension. This cell suspension is then used in the same fashion as described in 2.4.2. After cells are loaded in to the array and recover from trypsinization they once again adhere to the surface and begin to proliferate (Figure 6.2). This proliferation process can be tracked in the same fashion as described for the single cells grown in suspension. However, upon a division event, adherent cells do not travel downstream and in to the next available trapping structure as is the case for cell models grown in suspension. Therefore, in order to ensure adequate separation between single cells and high-fidelity lineage tracking, adherent cell experiments require the periodic introduction of trypsin in to the system. This buffer exchange relies on the same fluidic approach used to introduce a collection buffer as described in 2.4.4. After the brief introduction of trypsin-EDTA, the cells detach from the channel surface and are thus able to be re-seeded in new hydrodynamic trap structures before
switching back to growth media perfusion for continued lineage tracking. By recording this re-seeding process we are able to maintain the lineage information collected for each duration of time-lapse imaging and thus complete this process several times in order to construct multigenerational lineage trees for single adherent cells.

Figure 6.2 | Hydrodynamic trapping of single adherent cells
Schematic representations (a) and corresponding optical micrographs (b) for single HeLa cells loaded in to the hydrodynamic trap array following trypsinization. Immediately following trypsinization, these cells display a spherical morphology consistent with suspension cell models (left panels). However, after switching back to cell growth media the cells re-adhere to the channel surface after roughly one hour (right panels).
6.3 | Scaling considerations for growth and collection measurements

The throughput of the serial SMR platform is amenable to conducting measurements of time-sensitive samples – including clinical isolates or differentiating cell models – which are not stable enough to study over the course of multiple days. However, in order to implement this technology as a centralized resource capable of running various clinical samples with multiple users, it is important to consider how to scale these systems accordingly.

Below we present the cost and throughput considerations for implementing a cluster of systems that can operate as a core resource. The projections for a cluster of systems that run cell growth measurements alone or a cluster that implements downstream cell collection are both included for comparison. The footprint required for either of these system clusters is assumed to be approximately 100 square feet. This includes a 3’ x 6’ optical table which can accommodate either 8 cell growth platforms or 4 cell growth and collection platforms as well as the peripheral space required for technician operation.

The projections presented here include the current implementation of these platforms and do not assume any engineering developments which may reduce the footprint or decrease the cost of the systems moving forward. Similarly, the throughput estimates for each system are based on what is currently achievable. It is important to note that, if implemented, the fluidic and operational improvements presented in 954.3.4 may lead to significant improvements in the per-system throughput, which appears to offer the best opportunity for reducing the per-cell cost and increasing the overall throughput of a cluster of systems.
### Collection platform | Growth platform
--- | ---
**Equipment costs**
Hardware and software$^1$ | $15k per system | $14k per system
Environmental control$^2$ | $6k per cluster | $6k per cluster
Troubleshooting and maintenance$^3$ | $2k per cluster | $2k per cluster
**Total cluster cost$^4$ | $68k | $120k

### Operational costs
<table>
<thead>
<tr>
<th></th>
<th>Collection platform</th>
<th>Growth platform</th>
</tr>
</thead>
</table>
Labor$^5$ | $480 per day per cluster | $960 per day per cluster
Consumables$^6$ | $120 per day per cluster | $20 per day per cluster
**Total throughput$^7$ | 1,920 cells per day per cluster | 7,680 cells per day per cluster
**Total operational cost** | $0.31 per cell$ | $0.13 per cell$ 

1. All of the mechanical, optical, pneumatic, fluidic, and electronic components required to construct the core functional unit of each platform. This cost also includes the computer, software and peripheral electronics required to run each system. The cost of the collection platform also includes the three-dimensional motorized stages required for single-cell collection downstream of the serial SMR.

2. The components necessary for temperature and atmospheric regulation of the systems including a recirculating water bath and biological atmosphere gas supply. These components can be shared amongst multiple systems and thus distribute cost across a full cluster.

3. Electronic instruments such as oscilloscopes and multi-meters to monitor and optimize system performance as well as basic mechanical tool sets required to assemble and maintain the platforms.

4. Assuming a total of either 4 collection platforms or 8 growth platforms per cluster.

5. Approximate technician cost of $30 per hour for an eight hour work day. This figure assumes a rate of $20 per hour with 50% overhead, figures which may vary by institution. Assuming each technician can effectively operate either 2 collection platforms or 2 growth platforms.

6. Includes reagents necessary to clean and passivate device channels as well as the vials and tubing required to run samples. For the collection system, the cost also includes the 96-well format plates used for single-cell collection. This cost does not include the growth media required to run each sample – a cost which may vary substantially based on the cell type of interest.

7. Assuming a throughput of 60 cells per hour per system (4 systems per cluster) for the collection platforms and 120 cells per hour per system (8 systems per cluster) for the cell growth platforms operating for 8 hours per day.

8. Only includes the cost of isolating single cells downstream of growth measurements, does not include the cost of additional single-cell assays. The cost of these assays may vary substantially, ranging from cents per cell for simple outgrowth measurements to roughly ten dollars per cell for scRNA-seq.
Bibliography

2. Kelso, A. & Groves, P. A single peripheral CD8+ T cell can give rise to progeny expressing type 1 and/or type 2 cytokine genes and can retain its multipotentiality through many cell divisions. *Proceedings of the National Academy of Sciences of the United States of America* **94**, 8070-8075 (1997).


