Title
Novel methods for investigating human adipose tissue at the single-cell level

Author
Gupta, Anushka

Publication Date
2021

Novel methods for investigating human adipose tissue at the single-cell level

by

Anushka Gupta

A dissertation submitted in partial satisfaction of the

requirements for the degree of

Joint Doctor of Philosophy
with University of California, San Francisco

in

Bioengineering

in the

Graduate Division

of the

University of California, Berkeley

Committee in charge:

Professor Aaron Streets, Chair
Professor Nir Yosef
Professor Zev Gartner

Fall 2021

Novel methods for investigating human adipose tissue at the single-cell level

Abstract

Novel methods for investigating human adipose tissue at the single-cell level

by

Anushka Gupta

Doctor of Philosophy in Bioengineering

University of California, Berkeley

Professor Aaron Streets, Chair

The most important function of adipose tissue is its ability to store fat during periods of feeding, and release fats during periods of fasting and cold. This energy homeostatic activity of the adipose tissue is made possible by the synergistic metabolic functionality of distinct adipocyte-types residing in the tissue (tissue heterogeneity), as well as their developmental dynamics (tissue development). Although, these aspects of adipose tissue biology are very well understood in mice, there knowledge in humans remains poorly understood. Recently, the combination of next generation sequencing (NGS) and microfluidic platforms has led to a revolution in single-cell genomic studies, enabling measurement of molecular features in thousands of single cells at a time. In this body of work, I present novel assays, platforms, and dataset that enable new investigation into adipose tissue at the single cell level, and provide insight into the heterogeneity and developmental lineages within this important tissue type.

Although, advancements in NGS and microfluidic barcoding platforms have significantly increased the throughput of single-cell RNA-seq (scRNA-seq) measurements, many molecules that are critical to understanding the functional roles of cells in a complex tissue or organs, are not directly encoded in the genome, and therefore cannot be profiled with NGS. Lipids, for example, play a critical role in many metabolic processes, and are critical to characterizing an adipocyte's identity, but cannot be detected by sequencing. Recent developments in quantitative imaging, particularly coherent Raman scattering (CRS) techniques, have produced a suite of tools for studying lipid content in single cells. In Chapter 2, I review CRS imaging and computational image processing techniques for non-destructive profiling of dynamic changes in lipid composition and spatial distribution at the single-cell level.

In Chapter 3, I present a microfluidic platform called microfluidic cell barcoding and sequencing (µCB-seq) for combining scRNA-seq measurements with optical imaging measurements, thereby providing a comprehensive characterization of cellular identity at the single-cell level. µCB-seq is enabled by a novel fabrication method that preloads primers with known barcode sequences inside addressable reaction chambers of a microfluidic device. In addition to

enabling multi-modal single-cell analysis, µCB-seq improves gene detection sensitivity, providing a scalable and accurate method for information-rich characterization of single cells.

In Chapter 4, I characterize transcript enrichment and detection bias in single-nuclei RNA-seq for mapping of distinct human adipocyte lineages. scRNA-seq enables molecular characterization of complex biological tissues at high resolution. The requirement of single-cell extraction, however, makes it challenging for profiling tissues such as adipose tissue where collection of intact single adipocytes is complicated by their fragile nature. For such tissues, single-nuclei extraction is often much more efficient and therefore single-nuclei RNA-sequencing (snRNA-seq) presents an alternative to scRNA-seq. However, nuclear transcripts represent only a fraction of the transcriptome in a single cell, with snRNA-seq marked with inherent transcript enrichment and detection biases. Therefore, snRNA-seq may be inadequate for mapping important transcriptional signatures in adipose tissue. In this study, I compare the transcriptomic landscape of single nuclei isolated from preadipocytes and mature adipocytes across human white and brown adipocyte lineages, with whole-cell transcriptome. I demonstrate that snRNA-seq is capable of identifying the broad cell types present in scRNA-seq at all states of adipogenesis. However, I also explore how and why the nuclear transcriptome is biased and limited, and how it can be advantageous. I robustly characterize the enrichment of nuclear-localized transcripts and adipogenic regulatory lncRNAs in snRNA-seq, while also providing a detailed understanding for the preferential detection of long genes upon using this technique. To remove such technical detection biases, I propose a normalization strategy for a more accurate comparison of nuclear and cellular data. Finally, I demonstrate successful integration of scRNA-seq and snRNA-seq datasets with existing bioinformatic tools. Overall, my results illustrate the applicability of snRNA-seq for characterization of cellular diversity in the adipose tissue.

Finally, in Chapter 5, I utilize snRNA-seq to generate the transcriptional landscape of human white and brown adipogenesis using an *in vitro* model system, derived from a single individual and a single anatomical location. In total, I generate snRNA-seq libraries from $\sim$ 50,000 nuclei isolated from differentiating white and brown preadipocytes at 5 stages of adipogenesis. Using a custom bioinformatic strategy for cellular ordering across a continuum of maturation states, I reveal 5 distinct gene expression modules in both white and brown adipogenesis, each module highlighting the dynamics of biologically relevant functional processes. I identify potentially novel adipogenic as well thermogenic transcription factors, and investigate their involvement in Obesity by analyzing publicly available GWAS, RNA-seq and microarray datasets in lean vs obese humans. Overall, this study, for the first time, provides a comprehensive molecular understanding of both white and brown adipogenesis in humans, thereby serving as an an important resource and a reference to map the future *in vivo* adipogenic studies onto.

To my father, Virendra Kumar Gupta
Hope you would have been proud

Contents

List of Figures

List of Tables

Acknowledgments

I started my PhD in Fall of 2016, and these last 5 years have been the most rewarding experience of my life, both in terms of intellectual as well as personal growth. Of course, none of this would have been minutely possible without the incredible support of my advisor, friends, family, lab-mates, and the broader Bioengineering community at the University of California Berkeley.

I remember joining Prof. Aaron Streets's lab in 2017, because of the immense joy, confidence, and inspiration I felt after talking to him about potential project ideas for my dissertation, and science in general. Without Aaron's constant support, feedback, and mentorship, it would have been simply impossible for me to be at this stage in my career. As a scientist, Aaron has this amazing ability to provide insights, new ideas, and solutions for any challenge you are facing, which has been a key factor in me finishing my PhD. Without all our weekly/biweekly meetings, it would have been impossible to learn everything I have learnt in the last 5 years, and his scientific mindset has definitely guided me in becoming a better scientist.

Through Aaron, I also got the opportunity to meet some of the most brilliant scientists as part of the Streets Lab, who always provided me with the most fun and supportive environment to thrive. When I first joined the lab, Nick was the only grad student working there, and since then, he has been been like a second mentor for me. Stanford is lucky to have you, Nick! Gabriel is an optical wizard, and with some of the best feedback provided in group meetings. From Annie, I hope to learn the art of scientific communication, critical thinking, and record keeping, but more importantly, the art of kindness. Also, thank you so much for inspiring me to pick up Taylor Swift! I think of all the members of the Streets Lab, Zoë has been the most influential force in guiding my PhD thesis. Seeing her swiftly, but critically learn the skills and techniques of scRNA-seq gave my the confidence to pursue my research in this field as well. A huge thanks to Adam, Rodrigo, and Soohong too for always lending an ear for all my research problems, and job-hunting problems! Thanks to Tyler for being an incredible collaborator on the µCB-seq project. I hope your infectious energy rubbed off on me too! A huge shout out to my undergrad mentee Mansi for making some of the most beautiful (and operational!) microfluidic chips ever. I would also like to than Claris Garzon, April Alexander, Martin Witte, Kristin Olson, and Rocio Sanchez for helping us with day-to-day logistical problems in an extremely effective manner.

A huge thank you to my thesis committee members Prof. Zev Gartner & Prof. Nir Yosef, and my qualifying exam committee members Prof. Lydia Sohn, Prof. Bo Huang, Prof. Andreas Stahl. My work was also made possible by our brilliant collaborators Prof. Yu-Hua Tseng, Dr. Mary-Elizabeth Patti, Prof. Farnaz Shamsi!I would also like to acknowledge the contribution of my undergrad mentors Prof. Mina Hoorfar, Prof. Parag Deshpande, Prof. Sunando DasGupta, and Prof. Rabibrata Mukherjee in inspiring me to pursue a PhD after my undergrad.

I would also like to thank my incredible Berkeley friends here, who always provided me with the most fun environment to chill and relax outside of my PhD. A huge shoutout to

Shruti, Niharika, Anamika, Milind, Koulik, and Stayaki! A huge shout out to Kristine too, who I had the immense luck of meeting as part of my PhD cohort. We have had chats about failed experiments, successful experiments, boba, tasty food, not-so-tasty food, job hunt, offer letters, and life in general. Thank you so much for being there and I look forward to getting more plants from you!

Finally, none of this would have been possible without my family. I feel blessed to have such an incredible support system, and it scares me sometimes to think that I have been so lucky. I can't even begin to thank my mother, Rekha Gupta, for her sacrifices, support, prayers, visits, and love over the last 28 years. She has always pushed me to be the best version of myself, has supported me in going for things that I want to achieve! She knows how impatient and restless I am, and how tensed and anxious I can be, but she has always been there. I wish my father, Virendra Kumar Gupta, was here today to support me as well. I would have loved to get his feedback through some of the worst times of my PhD. But, I am sure he is happy, wherever he is, and would continue to look down upon me. I am also thankful for my sister, Vartika Gupta, and my brother-in-law, Sumit Gahoi, for providing me with a second home, here in the US, for supporting me in living a functional life, and for introducing me to my nephew Ish. I can't believe how much I love him, and how big of a role he has played these last 2 years in keeping me happy. I would also like to thank my in-laws: Vrinda Sakhalkar, Vivek Sakhalkar, Ketaki Sakhalkar, Aditya Samant, and Anika Samant for welcoming me with so much warmth and support. And finally, all my love and thanks to my husband, Siddhesh Sakhalkar. You have supported me through all my tantrums, self-doubts, anxiousness, and restlessness, and I would not have been here today without you. You make me want to be a better person everyday.

Chapter 1

Introduction

In 1665, Robert Hooke published a series of microscopic observations using plants, where he first coined the term 'cell' to describe the micro unit of biological tissues [1]. Since then, the pursuit of characterizing physiological heterogeneity by identifying cells that differ by morphological, functional, or molecular features has occupied a central place in research in the life sciences. Bolstered by the invention [2] and optimization of microscopic techniques [3], early days saw discovery of multiple distinct cell-types based on their distinguished anatomical structure and morphology [4]. With the development of novel optical methods, the features used to define cell types gradually evolved to incorporate physiological properties such as their pH [5], membrane potential [6], and molecular properties such as the presence of specific proteins [7], RNA molecules [8], and epigenetic modifications [9]. Notably, this evolution also inspired a shift in the scale at which cellular investigations were performed, focusing away from bulk, ensemble measurements, towards scalable, single-cell level measurements, thereby providing insights into the underlying heterogeneity within complex biological systems.

Amongst these scalable, single-cell level techniques, Fluorescence-activated cell sorting (FACS) emerged as a gold-standard technique for proteomic characterization of individual cells. FACS allows sorting of up to millions of single cells at a time based on differential expression of key surface proteins. Indeed, FACS has been used to study the cellular heterogeneity of many complex tissues, especially in the field of cellular immunology [10]. FACS, however, requires apriori knowledge of targeted proteins, depends on the availability of antibodies, and is limited in its multiplexing capacity. Notably, last two decades have seen a rapid development of cost-effective, high-throughput DNA-sequencing technologies, led by Illumina's Next-Generation Sequencing (NGS) platforms [11]. NGS enables sequencing of millions of short DNA fragments at a time in a completely unbiased manner. Since RNA molecules can be easily converted to a more stable complementary DNA (cDNA), NGS has also enabled transcriptome-wide gene expression profiling via RNA-sequencing (RNA-seq). This is important because unbiased profiling of protein-encoding mRNA molecules (mRNA-seq) can be used as a proxy for studying the entire proteome, thereby providing a comprehensive methodology to identify different cell-types. However, unlike FACS which has single-cell level resolution, RNA-seq requires micrograms of totalRNA as input, thereby

providing an average gene expression measurement in hundreds to thousands of mammalian cells. Hence, such population-level bulk measurements are unable to take into account the inherent heterogeneity within complex systems. Consequently, last few years have seen further technological innovations resulting in significantly improved gene detection sensitivity of RNA-seq with only picograms of starting material, thereby enabling completely untargeted gene expression profiling at the single-cell level (scRNA-seq) [12–14].

In 2009, Tang et al. developed the first scRNA-seq protocol, where sequencing library was prepared using mRNA isolated from individual cells of a four-cell stage embryo [12]. In this illustration, individual cells were isolated using mouth pipetting, clearly demonstrating the challenges of manipulating single cells for downstream RNA-seq. Development of microfluidic solutions helped mitigate such challenges by providing a low Reynolds's number environment for predictable, controlled, and programmable cellular manipulation at nanoscale reaction volumes [15, 16]. Streets *et al* first demonstrated the microfluidic implementation of scRNA-seq by utilizing the original Tang protocol on a valve-based, multi-layer, microfluidic device [17]. Similar microfluidic devices were earlier utilized for acquiring genomic measurements from single cells [18–21], and demonstrated exquisite, robust, and precise single-cell manipulation via use of integrated on-chip valves [22]. Such valve-based microfluidic scRNA-seq protocol was first commercialized by Fluidigm as part of the C1 platform [23, 24], providing transcriptomic measurements from 96 cells at a time using a fully automated library preparation protocol. The next wave of microfluidics-enabled development was an exponential increase in the throughput of scRNA-seq protocols, which went from analyzing hundreds of cells to tens of thousands of cells in a single experiment [25, 26]. This advancement was particularly made possible by implementation of automated, molecularly barcoded sequencing protocols on microfluidic technologies utilizing droplets or microwells for single-cell encapsulation. Previous advancements in microfabrication methodologies enabled successful engineering of devices having ∼100,000 microwells [27], or devices generating ∼1000 droplets per second [28, 29], which facilitated a high-throughput isolation of single cells. Molecular barcoding enabled a one-pot, multiplexed library preparation from such vast number of cells, by assigning a unique oligo sequence to cDNA molecules of individual cells and pooling them after reverse-transcription (RT) into a single tube. The first academic demonstration of microwell based scRNA-seq provided gene expression measurements from thousands of cells [30], which, in a few years, was further improved for analysing 100,000 cells in a single experiment [31]. Around 2015, multiple droplet microfluidics based scRNA-seq protocols were developed that enabled gene expression measurements from ∼10,000 cells in a single experiment [32–34]. In such methods, droplets, encapsulating single cells, were formed by precisely combining aqueous and oil flows in a microfluidic device and used as nanoscale reaction chambers to perform molecularly barcoded library preparation. Within a few years, 10x Genomics launched the first, commercial, droplet-microfluidics based scRNA-seq platform called the Chromium, that enabled transcriptomic measurements from 80,000 cells in one experiment, with minimal requirement of user interference during cell encapsulation. Since it's launch, the Chromium platform has been keenly accepted in the life sciences community and has truly turned scRNA-seq into a widely accessible technique. Overall, the development

of micro-scale, high-throughput, and low-cost scRNA-seq techniques has revolutionized the way we study complex biological systems and has paved the way for undertakings such as constructing the human cell atlas [35]. Recent work has already composed cellular catalogues of both mouse and human organs, including the brain [36, 37], the thymus [38], the pancreas [39], and two recent reports of a comprehensive mouse atlas [31, 40]. However, while scRNA-seq has proven to be a robust tool for quantifying cellular identity, there are many molecules, which play critical roles in cellular function, that are not directly encoded in the genome and therefore cannot be detected with measurements that are based on sequencing. Metabolites and lipids are examples of such molecules that cannot be profiled using NGS but are important for regulation of cellular function. Lipids participate in providing structural integrity to biological membranes [41, 42], signaling pathways [41, 43], and interact with proteins to regulate their functions [44, 45]. The most important role of lipids, however, is serving as a reservoir for energy storage as part of the adipose tissue. Resident adipocytes in the tissue store lipids in organelles called lipid droplets (LDs), that undergo lipolysis or lipogenesis to provide/store energy for a variety of physiological conditions, thereby maintaining a system-wide energy balance. Hence, in order to comprehensively characterize the cell-types within adipose tissue, scRNA-seq measurements need to be paired with measurements interrogating inter-cellular LDs. Adipose tissue also serves as a case-study for highlighting the second drawback of scRNA-seq. One requirement of scRNA-seq is the extraction of intact single cells, which is challenging from primary adipose tissue samples, where collection of intact single adipocytes is complicated by their fragile nature. For such tissues, single-nuclei extraction is often much more efficient and therefore single-nuclei RNA-sequencing (snRNA-seq) presents an alternative to scRNA-seq. However, snRNA-seq has low transcript complexity, and is marked by transcript enrichment and detection biases, which distort the biological signal of interest. Therefore, there is a need to understand the transcriptomic similarities and differences between single-cell and single-nucleus profiles in the context of the human adipose tissue, for which there is growing need to rely on snRNA-seq.

In this dissertation, I first worked on developing a novel platform that enables simultaneous interrogation of molecular and LD features in single cells, by pairing scRNA-seq and quantitative imaging measurements for each cell. Furthermore, I also developed novel approaches to systematically compare the transcriptomic profiles derived using scRNA-seq and snRNA-seq in matched adipose tissue cell-types, and finally implement snRNA-seq to uncover new biology regarding adipose tissue development and function. Overall, this body of work provides novel tools and strategies to investigate the adipose tissue at the single-cell level, thereby fulfilling key outstanding needs currently faced by the scientific community in this field.

1.1 Adipose tissue and its role in maintaining energy homeostasis

Fat or adipose tissue, in the popular mind, is a way to store excess energy during conditions of nutritional affluence. And while that is true, it is an incomplete picture of the functionality of adipose tissue in maintaining system-wide energy balance. Almost all animals have two functionally distinct types of fat: the white adipose tissue (WAT) and the brown adipose tissue (BAT). WAT stores excess nutritional energy in the form of lipids during fed state. However, during fasting state, lipids stored in WAT are processed to provide energy to peripheral organs and tissues. BAT, on the other hand, regulates non-shivering thermogenesis, an energy-spending process where stored lipids are metabolized to generate heat, resulting in the maintenance of core body temperature. This is achieved through the actions of uncoupling protein-1 (UCP1) [46], a BAT-specific protein located within the mitochondria. Therefore, WAT and BAT function together to maintain a balance between lipid accumulation and energy expenditure.

There are two key aspects about adipose tissue biology that are critical to maintaining this balance discussed above. First is the development of adipocytes, the metabolically-active cell-types in the tissue. Adipocytes are generated via a cellular developmental process called adipogenesis, where resident preadipocytes (adipocyte precursors) differentiate into lipid-laden adipocytes. Notably, excess of fat manifests itself in the development of syndromes such as Obesity, whereas deficiency of fat is associated with disorders such as lipodystrophy. Therefore, a firm understanding of the molecular underpinnings of fat development, a.k.a., adipogenesis, is key to elucidating the pathophysiology and potential treatment modalities of such pathological cases. Second key aspect about adipose tissue biology is its cellular makeup, which is critical for determining its metabolic capacity. Rodent models have been the most prevalent model-type in the field of adipose tissue biology, and hence most of our current understanding about the tissue's cellular makeup comes from this species. Traditionally, adipocytes within WAT and BAT are classified as white or brown types, respectively, in rodents. White adipocytes within WAT are marked by the presence of a single large lipid droplet (unilocular), whereas brown adipocytes within BAT are histologically identified as having multiple small lipid droplets (multilocular) [47]. Unlike white adipocytes, brown adipocytes are UCP1+ (thereby enabling the thermogenic function of BAT) and descend from a PAX7+/MYF5+ smooth-muscle cell lineage [48]. Because of the energy expenditure thermogenic capacity of brown adipocytes, promotion of BAT function has emerged as a promising anti-Obesity therapeutic target [49]. More recently, multiple investigations have reported identification of a third kind of beige adipocyte, which are inter-dispersed within WAT of rodents [50]. Notably, like brown adipocytes, beige adipocytes also exhibit UCP1-dependent thermogenic capacity and multilocular morphology [51]. However, beige adipocytes are not from PAX7+/MYF5+ lineage, thereby suggesting an overlapping but distinct gene expression pattern compared to brown adipocytes [51]. Naturally, the thermogenic capacity of beige adipocytes have also raised important questions about its therapeutic impli-

cations [50]. Currently, we are in a period in which new information regarding heterogeneity within white, brown, and beige adipocytes is being accumulated rapidly, and therefore, an atlas of all the residing adipocyte-types in the tissue is fundamental to designing better therapies for metabolic diseases.

Focusing on adipocyte heterogeneity, in the following sections, I will first review how the applicability of scRNA-seq has already fueled discovery of multiple adipocyte sub-types in the adipose tissue. I will then discuss how we can further advance identification of adipocyte sub-types by combining scRNA-seq with other cellular phenotypic measurements. Next, focusing on adipogenesis and its molecular regulation, I will first discuss our current understanding of adipogenic regulation in rodents, and a lack thereof in humans. Finally, I will discuss how we can bridge this gap by utilizing snRNA-seq for building a comprehensive transcriptional landscape for white and brown fat development in humans.

1.2 scRNA-seq enables investigation into adipose tissue heterogeneity

Over the last few year, scRNA-seq has identified extensive heterogeneity within white and brown adipocytes. For example, Ramirez et al. performed scRNA-seq on abdominal human preadipocytes undergoing adipogenesis and identified at least two distinct classes of subcutaneous white adipocytes [52]. Song et al. performed scRNA-seq of BAT and identified two populations of brown adipocytes, with one cell-type similar to the classical, well-studied, highly thermogenic BAT, and the other cell-type with substantially lower thermogenic activity [53]. These low-thermogenic BAs were functionally assessed and determined to be enriched in genes for UCP1-independent thermogenesis. Spaethling et al. performed scRNA-seq of primary brown adipocytes and observed significant heterogeneity in UCP1 expression, as well as heterogeneity in numerous additional genes associated with the brown thermogenic phenotype [54]. Overall, scRNA-seq has provided a deeper understanding of adipocyte sub-types within fat depots. However, other phenotypic LD-associated features play critical role in adipocyte diversity, that are not directly encoded in the genome and therefore cannot be detected with measurements that are based on sequencing DNA.

1.3 Paired single-cell imaging and sequencing for improved adipocyte sub-type discovery

Although scRNA-seq has popularized the notion of single-cell level measurements, for the longest time, optical microscopy was at the forefront of this endeavor, enabling phenotypical, functional, and even compositional measurements of single cells at nanoscale resolution. In the context of adipose tissue, early applications of optical imaging revealed heterogeneity in lipid droplet morphology in broad classes of adipocytes, with white cells having a single

large lipid droplet (unilocular), and brown & beige cells having small, but numerous lipid droplets (multilocular) [47]. Further applications of optical imaging at the single-cell level revealed heterogeneity in lipid droplet morphology and adipocyte size within both cultured cell lines [55, 56] and in primary fat cells, such as in murine white adipocytes [57, 58]. Development of new methods to utilize imaging for functional characterization of single cells further revealed distinct adipocyte function based on cell-size. For example, optical analysis of large and small white adipocytes revealed decreased insulin-mediated glucose uptake [59] and insulin-sensitivity [60] in larger adipocytes. Notably, advancements over the last decade in quantitative optical imaging techniques enabled lipid droplet compositional characterization, revealing extensive cell to cell heterogeneity in the lipidomic profile of mature adipocytes [61]. Therefore, while scRNA-seq is effective for measuring mRNA in large number of single cells, adipocyte's identity is also described by its size, lipid droplet morphology, and its lipidomic profile, features that are not directly encoded in the genome and hence cannot be detected with sequencing alone. Since isolation of single cells is a requirement for scRNA-seq, single-cell transcriptomic measurements could be combined with upstream single-cell optical imaging measurements to provide a truly comprehensive view of adipocyte identity. Such paired measurements, however, would require imaging techniques to be non-destructive, since cells need to be preserved for downstream sequencing measurements. In chapter 2, I review existing label-free, non-destructive, and quantitative optical imaging techniques that enable comprehensive characterization of cell-size, lipid droplet morphology and its composition [62]. I also review multiple computational image processing techniques and finally propose strategies for pairing such imaging techniques with scRNA-seq, using either microfluidic platforms or *in situ* platforms. In chapter 3, I then develop microfluidic cell barcoding and sequencing (μCB-seq), a microfluidic platform that combines high-resolution imaging and sequencing of single cells [63].

1.4 Molecular regulation of adipogenesis in rodents and its applicability to humans

Historically, adipogenic transcriptional cascades were extensively studied using a variety of murine cell culture systems, most common of them being the 3T3-L1 model system. Early investigations led to the identification of the core adipogenic transcriptional network, including principal transcription factors (TFs) C/EBPβ and C/EBPδ overseeing early adipogenic commitment, and PPARγ and C/EBPα overseeing the terminal differentiation process [64, 65]. Typically, hormonal induction of *in vitro* adipogenic differentiation is rapidly followed by the expression of C/EBPβ and C/EBPδ, which reach peak expression levels within the next few days and then begin to drift downward. Initial activity of C/EBPβ and C/EBPδ induces low levels of PPARγ and C/EBPα, which are then able to induce each other's expression in a positive feedback loop that promotes gene expression changes characteristic of mature adipocytes. PPARγ and C/EBPα then remain elevated for the life of the cell. Notably,

this core transcriptional hierarchy remains preserved during brown adipogenesis as well, although with an accompanying exclusive activity of thermogenic TFs such as PRDM16, EBFs, and PPARγ co-activators (PGC1A and PGC1B)[66]. With the molecular core of white and brown adipogenesis well understood, modern transcriptomic investigations have instead focused on identifying auxiliary transcription factors, that serve as either positive or negative regulators of adipogenic/thermogenic response in rodents. This has resulted in the identification of novel adipogenic TFs such as KLFs [67], SREBP1C [68], CREB [69], ZFPs [70], GATA2 & GATA3 [71], and FOXA1 & FOXA2 [72]. Similarly, many transcriptional regulators have been identified that regulate brown fat-specific gene expression in rodents. This includes activators such as ZFP516, KLF11, IRF4, TAF7L, ZBTB16, EWS, PLAC8, and repressors such as FOXO1, TWIST1, p107, LXRα, pRB, RIP140, TLE3, REV-ERBα, and ZFP423 [73].

The growing number of transcriptomic and epigenomic studies continues to strengthen our understanding of how brown and white adipogenesis is transcriptionally regulated in rodents. However, molecular regulation of human adipogenesis remains poorly understood. Studies have confirmed the applicability of principal TFs PPARγ and C/EBPs in human adipogenesis [74]. However, other studies have also reported dramatic differences in the features, locations, and transcriptomic properties of fat depots across rodents and humans. For example, BAT, which is abundantly present in the interscapular depot in mice, was only found to be present in adult humans over the last decade [75]. Furthermore, rodent BAT lies in a well-defined anatomical location and is homogeneously composed of brown adipocytes, whereas human BAT is widely dispersed and occurs as a mixture of white, and brown adipocytes [76]. Focusing on locations of fat, a large percentage of visceral fat in humans is contained in the omentum, which is barely present in rodents [77]. Conversely, the large epididymal fat pads of male mice, which are frequently sampled as representative of visceral fat, do not exist in men [78]. Notably, recent studies have also highlighted BAT metabolic functions that do not translate from the rodents to the human [79]. Therefore, although studies in rodent models of adipogenesis offer significant insights, their applicability to humans is actually limited by the existing differences in their metabolism and physiology. Consequently, there is an immediate need to comprehensively understand the transcriptional control of adipocyte formation in humans. Recent rapid development of human adipogenic model systems further brings us one step closer to achieving this undertaking [80].

1.5 single-nuclei RNA-sequencing for mapping human adipogenesis

Besides discovery of distinct cell-types, scRNA-seq has also enabled assessing the molecular progression of individual cells along a continuously changing biological process of interest, by providing a snapshot of the transcriptome of thousands of single cells in a cell population, which are each at distinct points of the dynamic process under study. Recent bioinformatic

advancements have further made it possible to analyze this wealth of transcriptional information for computationally inferring lineage developmental trajectories and gaining detailed insights into the underlying molecular programmes executed, by ordering cells along a variable called the "pseudotime" [81, 82]. Pseudotime can be computationally inferred based on the cells' gene expression profiles measured by scRNA-seq, and can be thought of as a time-like variable indicating the relative position a cell takes in a developmental lineage. Based on this pseudotemporal ordering, expression dynamics for individual genes as well as regulatory TFs can be investigated at an extremely high cellular as well as temporal resolution. Indeed, pseudotemporal analysis has been implemented to study developmental dynamics in multiple tissue types such as mouse pancreas [83], zebrafish embryos [84], and human lungs [85], thymus [86], and embryos [87] etc . In the context of adipose tissue, recent studies are beginning to utilize scRNA-seq and appropriate bioinformatic techniques, for investigating the molecular regulation of adipogenesis in rodents [88, 89]. However, recovery of mature adipocytes is severely limited in such studies, in part because of the technical barriers associated with isolating intact, single adipocytes from *primary* tissue samples. Primary adipocytes can be difficult to work with due to their fragile nature, high buoyancy, and large size [90]. Existing protocols for tissue digestion and single-cell suspension preparation often result in complete or partial adipocyte lysis and therefore are not compatible with scRNA-seq. To address the challenge of working with tissues that are difficult to dissociate into single cells, recent studies have turned to single-nucleus RNA-sequencing (snRNA-seq) as an alternative approach for transcriptomic profiling of cellular heterogeneity within primary tissue [91–97]. These studies rely on nuclear mRNA to serve as a proxy for the single-cell transcriptome, and take advantage of protocols which enable efficient extraction of intact nuclei. However, a single nucleus contains 10-100-fold less mRNA than whole-cells, with snRNA-seq marked with inherent transcript enrichment and detection biases. Therefore, it is not clear whether snRNA-seq has enough sensitivity for mapping adipogenic transcriptional signatures in the human adipose tissue. Consequently, in Chapter 4, I systematically compare the transcriptomic landscape of single nuclei isolated from preadipocytes and mature adipocytes across human white and brown adipocyte lineages, with whole-cell transcriptome [98]. I demonstrate that snRNA-seq is capable of identifying the broad cell types present in scRNA-seq at all states of adipogenesis. However, I also explore how and why the nuclear transcriptome is biased and limited, and how it can be advantageous. I robustly characterize the enrichment of nuclear-localized transcripts and adipogenic regulatory lncRNAs in snRNA-seq, while also providing a detailed understanding for the preferential detection of long genes upon using this technique. To remove such technical detection biases, I propose a normalization strategy for a more accurate comparison of nuclear and cellular data. Overall, my results illustrate the applicability of snRNA-seq for characterization of transcriptomic diversity in the adipose tissue. Hence, in Chapter 5, I generate a high-resolution temporal transcriptional landscape of adipogenesis in humans by performing large-scale snRNA-seq experiments on differentiating white and brown preadipocytes.

1.6 Scope of the dissertation

In Chapter 2, I review quantitative imaging and computational image processing techniques for non-destructive profiling of lipid droplet morphology, composition and spatial distribution at the single-cell level. I also discuss experimental strategies for combining lipidomic and transcriptomic analysis at the single-cell level, enabling a more comprehensive profiling of adipocyte identity. In chapter 3, I follow up on one of the experimental strategies discussed in Chapter 2 and develop (μCB-seq), a microfluidic platform that combines high-resolution imaging and sequencing of single cells, thereby enabling a comprehensive adipocyte identity characterization. A requirement of such paired measurements, however, is extraction of intact single adipocytes with preserved morphology, which is currently possible for only *in vitro* adipogenic model systems (primary samples undergo adipocyte lysis during tissue digestion).

In chapter 4 and 5, I shift my focus away from adipose tissue heterogeneity onto adipose tissue development, where snRNA-seq is emerging as the preferred technique for investigating molecular underpinnings of adipogenesis. In chapter 4, I perform systematic characterization of transcript enrichment and detection biases in snRNA-seq as compared to scRNA-seq for mapping human adipocyte lineages and illustrate the applicability of snRNA-seq in recovering similar biology as scRNA-seq within the adipose tissue. Finally, In chapter 5, I utilize large-scale snRNA-seq to create a temporal transcriptional landscape of white and brown fat development in humans.

Chapter 2

Quantitative imaging of lipid droplets in single cells

2.1 Introduction

In Chapter 1, we discussed how lipid droplet (LD) morphology, its composition, and spatial distribution are key features that play critical role in defining adipocyte identity and function. Besides its role in energy storage, LDs also regulate lipid metabolism for processes such as construction of cell membranes [41, 42], key signaling pathways [41, 43], and protein degradation [44] & function [45]. Typically, LDs store a wide distribution of lipid molecules with structural variations in their hydrophobic and hydrophilic regions, and alterations in lipid metabolism directly changes this composition [99, 100]. The field of lipidomics aims to study such changes in lipid metabolism in response to physiological (adipogenesis), pathological (obesity, lipodytrophy), and environmental (nutritional intake) conditions by characterizing the compositional distribution of all cellular lipids residing in LDs. Established techniques for lipidomic analysis include gas or liquid chromatography-mass spectrometry (GC/LC-MS)[101–103] and shotgun mass spectrometry [104, 105]. GC/LC-MS and shotgun techniques allow for targeted and untargeted detection of lipid molecules, respectively, when implemented on biological extracts from a population of cells (Fig. 2.1A). Recent advancements in sample preparation and ionization techniques have further enabled researchers to profile the lipidome at the single-cell level based on microarray for MS (MAMS), single-cell matrix assisted laser desorption/ionization-MS (MALDI-MS), and subcellular content aspiration-based MS techniques [106–109]. Imaging mass spectrometry (IMS) is an imaging method that allows for visualization and quantification of spatial distribution of lipids in intact biological systems [110–114]. Implementation of IMS techniques requires extensive sample preparation [115] with spatial resolution ranging from submicron to hundreds of microns depending on the ion source [116]. Typically, sensitivity of MS-based techniques lies in the picomolar range with detection specificity of hundreds of lipid species simultaneously [117]. Such high sensitivity and specificity of MS-based techniques comes at the cost of

destructive measurements.

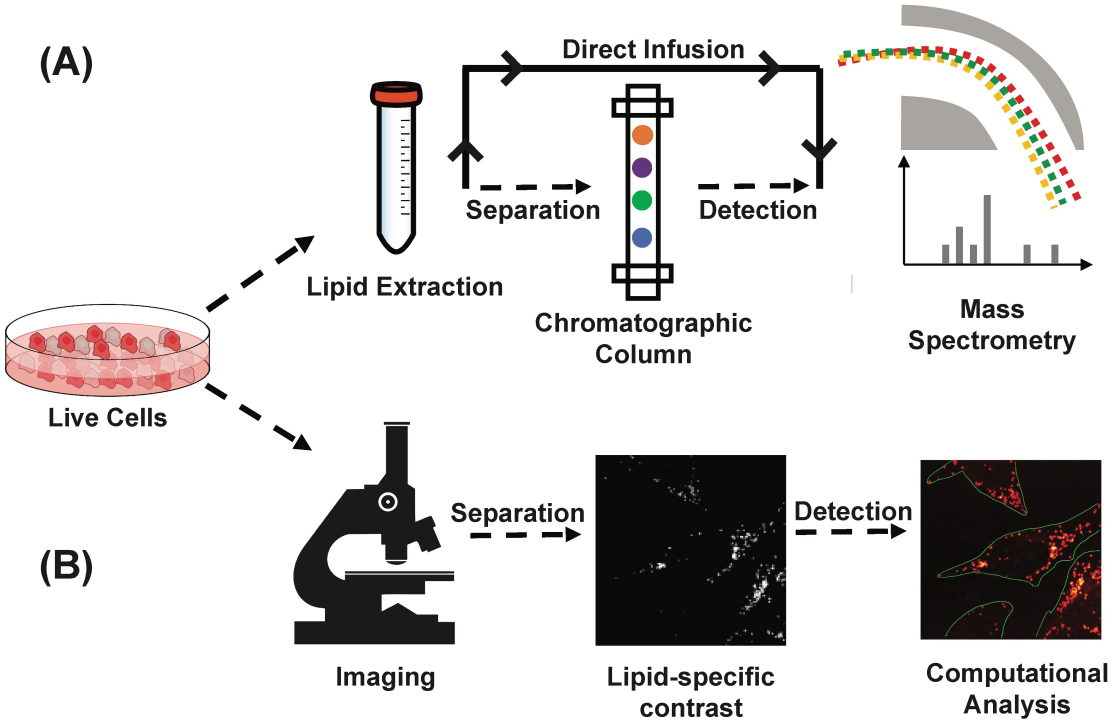

Figure 2.1: **Pipeline of mass spectrometry (MS) and microscopic quantitative imaging for lipidomic analysis** (A) In MS-based techniques, lipid is extracted from bulk cells. Extracted lipid can be separated using a gas/liquid chromatographic column before mass spectrometric detection, or directly infused in mass spectrometer for untargeted detection. (B) In quantitative imaging-based techniques, multiple live cells in the field of view are first imaged non-destructively to generate a lipid-specific contrast. The image is then computationally analyzed to segment cells and quantify properties of subcellular lipid droplets at the single-cell level.

Quantitative microscopic imaging techniques are complementary to MS-based technology and allow for non-destructive spatial characterization of LDs in live cells but with less lipid specificity. The non-destructive nature of optical microscopy allows researchers to perform time-resolved imaging to investigate dynamic cellular behavior. Furthermore, live-cell imaging can be coupled with subsequent molecular measurements such as sequencing or mass-spectrometry. Also, when combined with image processing algorithms, microscopy enables researchers to gather subcellular information such as LD morphology, or LD composition. This circumvents the need for physical isolation of single cells, thereby increasing the speed of data acquisition (Fig. 2.1B).

Amongst quantitative microscopic imaging techniques, fluorescence imaging allows for quantification down to a single molecule level. Fluorescence imaging with lipid-soluble dyes, lipid-binding probes, or fluorophore-conjugated lipids, has been used to study the compo-

sition and morphology of LDs [118, 119]. In some cases, the process of labeling can alter the distribution of cellular lipids. For example, Yen et al. showed that staining based on both Nile red and BODIPY does not correlate with fat stores for the model organism C. elegans [120]. Complementary to fluorescence imaging are label-free optical techniques such as phase contrast [121], differential interference contrast [122], quantitative phase-imaging [123], and third harmonic generation microscopy [124] that have been used to visualize LDs. In order to extend the capabilities of label-free imaging techniques for lipid profiling and quantification, magnetic resonance imaging (MRI) and coherent Raman scattering (CRS) techniques have been implemented to provide a lipid-specific contrast. MRI is an imaging technique based on nuclear magnetic resonance that has been implemented for quantification of total fat content and lipid accumulation [125–127]. The high penetration depth achieved from near-IR imaging allows researchers to implement MRI techniques *in vivo*. For *in vitro* and *in vivo* label-free mapping of LD composition, CRS imaging techniques are used. CRS techniques include coherent anti-Stokes Raman scattering (CARS) imaging and stimulated Raman scattering (SRS) imaging, both of which have been widely used to quantify LDs at the single-cell level with high spatial and temporal resolution. In this review, we will highlight applications of CRS techniques for quantifying LDs. We will also discuss object recognition algorithms for identification of LDs and cellular boundaries in an image. Such segmentation analysis is necessary for microscopy to be used for quantitative single-cell analysis. We will conclude by discussing the implications of non-destructive CRS techniques towards promises of multi-omic analysis at the single-cell level.

2.2 Coherent Raman scattering (CRS) microscopy

CRS microscopy provides a label-free approach for profiling the chemical composition of biological specimens by probing the characteristic vibrational modes of molecular bonds. Because of the strong vibrational modes associated with CH_2, CRS is particularly powerful for imaging intracellular lipids. For selective imaging of lipids, the asymmetric-stretching vibrational mode of the carbon–hydrogen bond is probed at 2845 cm^{-1} (Fig. 2.2A). CRS is induced by simultaneously illuminating the specimen with two photons at frequencies ω_{p} (pump) and ω_{s} (Stokes). When the difference in frequency between the two photons equals a vibrational frequency that is characteristic of the target molecule

$$\Omega = \omega_p - \omega_s,$$

the Raman scattering cross-section is resonantly enhanced giving rise to a strong CRS signal. Coherent anti-Stokes Raman scattering (CARS) and stimulated Raman scattering (SRS) are two imaging modalities that operate on this principle. In CARS, a signal is detected at the anti-Stokes frequency, ω_{AS}, given by

$$\omega_{AS} = 2 \times \omega_p - \omega_s$$

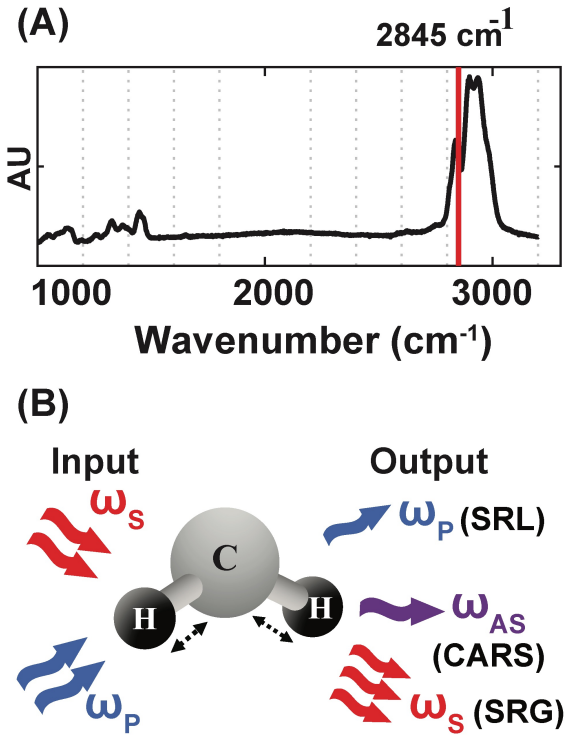

Figure 2.2: **Vibrational imaging of lipids using coherent Raman scattering.**(A) Spontaneous Raman spectra of oleic acid. The red solid line indicates asymmetric stretching vibrational mode of the carbon–hydrogen bond at 2845 cm^{-1}.(B) Schematic of excitation and detection for coherent Raman scattering. For both coherent anti-stokes Raman scattering (CARS) and stimulated Raman scattering (SRS) imaging, a characteristic vibrational mode of the CH_2 bond in lipids is excited with two incoming photons at the pump (ω_p) and stokes (ω_s) frequency. Stimulated raman loss (SRL) is detected as a loss in the pump intensity and stimulated Raman gain (SRG) is detected as a gain in the stokes intensity. CARS is detected at the anti-stokes frequency, ω_{AS}

CARS relies on homodyne detection, as ω_{AS} can be separated from both the incoming frequencies ω_p and ω_s using a dichroic mirror or optical filters [128]. In SRS, one of the two incoming photons, ω_p or ω_s, is amplitude modulated and the signal is detected as a loss or gain in the intensity of the pump or Stokes photon respectively (Fig. 2.2B). Therefore, SRS techniques utilize heterodyne detection schemes and require a lock-in amplifier to amplify the modulated stimulated Raman loss or gain [129]. As CARS and SRS are nonlinear optical processes, signal is only generated at the focal plane of the objective, enabling intrinsic three-dimensional sectioning by scanning in the x, y, and z axes. Long-term live cell imaging is also possible as CRS contrast is not limited by photobleaching. CARS and SRS are diffraction limited techniques and therefore offer quantification at a sub-cellular level with resolution as low as hundreds of nanometers. The CARS signal is quadratic with respect to

the concentration of resonant chemical bonds and the SRS signal is linear. SRS also has a higher signal to noise ratio (SNR) as compared to CARS because there is no non-resonant background. However, heterodyne detection in SRS requires additional instrumentation (lock-in amplifier) which is bypassed in CARS by using appropriate filters for homodyne detection.

Multiphoton excitation techniques like CRS imaging employ ultrashort pulsed lasers to obtain high concentrations of laser power inside the sample, which is necessary for efficient excitation of the targeted vibrational mode. A possible consequence of this elevated laser irradiance is photodamage to cells and tissues. Schönle and Hell developed a model for investigating the effects of optical absorption (in near-IR, by water in biological specimens) on focal heating during multiphoton excitation microscopy [130]. Their results showed an increase in focal temperature by not more than 3 K for an average laser power of 100 mW at the focal plane, suggesting that heating through linear absorption does not play a destructive role. However, the required peak laser power, to maintain an average laser power of 100 mW, may lead to nonlinear photodamage. Other studies have shown that maintaining laser power below 10 mW at the focal plane is considered to be a safe range for sample integrity [131, 132]. Some applications of CRS imaging may require higher laser power for fast and efficient excitation of the resonant mode [133, 134]. For such purposes, optimizing the average and peak laser power should be the first step towards maintaining a strong signal while minimizing photodamage to the sample [135]. Work has been done by several research groups to identify and define criterias for characterization of photodamage induced by nonlinear imaging [136–139].

In this section, we discuss investigations using CRS techniques for quantifying LDs. In section 2.3, we will then discuss object recognition algorithms applicable for cell and LD boundary determination. Section 2.4 will focus on biological investigations using CRS techniques coupled with segmentation algorithms for quantitative single-cell and single-lipid droplet analysis.

CARS and SRS

As CARS signal is quadratic with molecular concentration of the resonant bond, quantification using CARS requires processing of signal intensity. For example, Chen et al. derived a formula to calibrate CARS intensity to accurately report the number of lipid molecules in the scattering volume [140]. In this study, they developed an automated image analysis algorithm for quantification of lipid content in single cells. Rinia et al. adopted another strategy where they implemented spectral-analysis tools in conjunction with multiplex CARS for retrieval of spontaneous Raman-like spectra which is linear with the number of vibrating molecules [141]. In this study, they analyzed the retrieved spontaneous Raman-like spectra to map the acyl chain unsaturation and acyl chain order within individual LDs in adipocytes, which were incubated with exogenous free fatty acids (FFA) of varying compositions (Fig. 2.3). They found heterogeneity in lipid composition and packing in individual LDs and demonstrated that this heterogeneity was dependent on the FFA composition of incubation

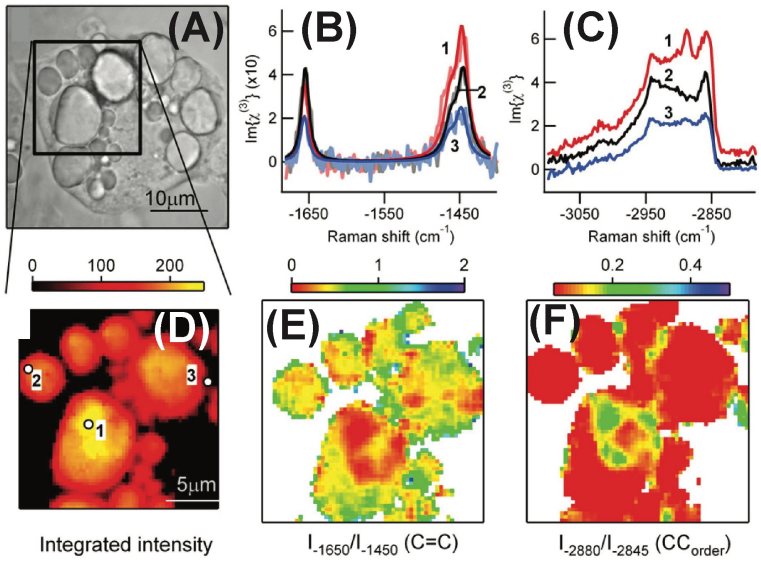

Figure 2.3: **Multiplex coherent anti-stokes Raman scattering (CARS) imaging of 3T3-L1-derived adipocyte to map the composition and packing of individual lipid droplets. Cells were incubated in a 1:3 mix of unsaturated:saturated fatty acid** (A) Brightfield image of an adipocyte. Spontaneous Raman-like spectra in the (B) CC-stretch and (C) CH-stretch regions for locations indicated (in D). Retrieved spectra was then analyzed for mapping the (D) lipid concentration, (E) acyl chain unsaturation and (F) acyl chain order on the same adipocyte. Reprinted from ref. [141], Copyright (2021), with permission from Elsevier.

mixture. In contrast to CARS, SRS signal is linear with the number of vibrating molecules, thereby making quantification more straightforward. Freudiger et al. demonstrated SRS as a contrast mechanism for imaging biological specimens [129]. They monitored the uptake and metabolism of unsaturated FFA by imaging at 3015 cm^{-1} wavenumber specific to the =C–H bond in unsaturated fatty acids. Wang et al. used SRS microscopy combined with RNA interference screening to determine lipid storage regulatory genes in C. elegans [142]. Lipid storage capacity was quantified based on mean SRS intensity. Using this technique, they were able to screen for 272 genes and found 8 new regulatory genes for fat storage. Besides quantifying LDs, CRS techniques have been critical towards visualizing LD growth and formation thereby revealing new lipid functions in cellular environment [143, 144]. Nan et al. demonstrated vibrational imaging of LDs using CARS and monitored LD formation during differentiation of 3T3-L1 fibroblast cells into adipocytes [145]. They found that after adding adipogenic differentiation media, there was an initial clearance of LDs at the early stage of differentiation followed by formation of large LDs (Fig. 2.4). Le and Cheng combined CARS microscopy with fluorescence imaging and flow cytometry to investigate heterogeneity in rates of LD formation in differentiating 3T3-L1 cells [146]. They found that phenotypic variability among differentiating 3T3-L1 cells was dependent on the kinetics of

an insulin signaling cascade.

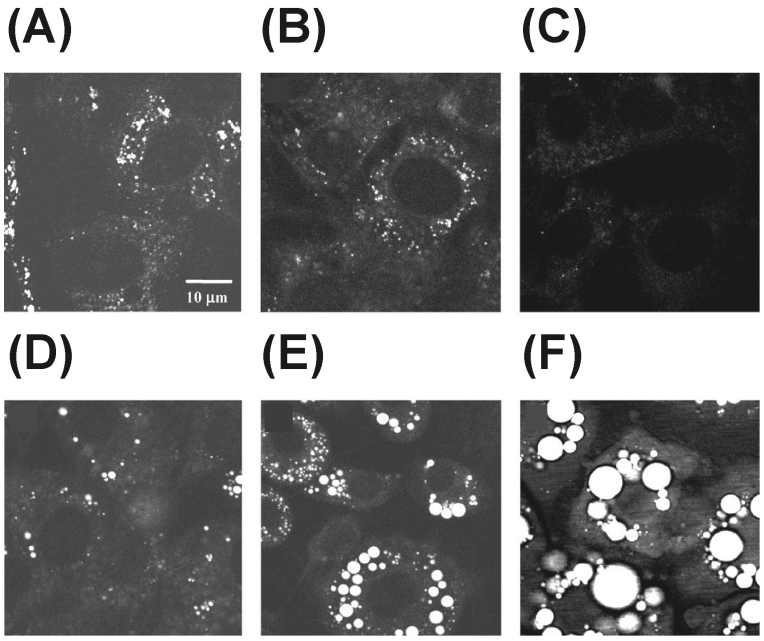

Figure 2.4: **Monitoring lipid droplet formation during differentiation of 3T3-L1 cells using CARS at 2845 cm^{-1}**. Images were taken at different times after adding differentiation induction media: (A) 0 h, (B) 24 h, (C) 48 h, (D) 60 h, (E) 96 h, and (F) 192 h. Republished with permission of American Soc for Biochemistry & Molecular Biology, from vibrational imaging of lipid droplets in live fibroblast cells with coherent anti-stokes Raman scattering microscopy. Reprinted from [145] under the terms of the Creative Commons CC-BY license, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Vibrational Raman tags

Imaging at a single frequency is insufficient for monitoring the uptake of saturated fatty acids because all vibrational markers of saturated fatty acids are shared by unsaturated fatty acids. However, no endogenous molecular species, including lipids, vibrate in the range from 1800 cm^{-1} to 2800 cm^{-1}, known as the "Raman-silent region" in cells. Raman tags are biorthogonal vibrational labels that consist of chemical bonds having a unique Raman shift in the cell's silent region. Fatty acids have been conjugated with Raman tags for tracking their uptake dynamics. Stable isotope substitution using ^{2}H [147, 148] or conjugation with alkyne tags [149, 150] are the two major strategies employed with CRS techniques. Wei et al. demonstrated metabolic incorporation of saturated FFA into triglycerides and its storage in LDs using alkyne tagging together with SRS [151]. Li and Cheng demonstrated direct visualization and quantification of glucose metabolism in single cells using SRS microscopy

coupled with isotope labeling (glucose-d7) [152]. They demonstrated up-regulation of *de novo* lipogenesis in pancreatic and prostate cancer cell lines as compared to healthy cell lines. They also showed that compared to pancreatic cancer cells, prostate cancer cells have lower level of *de novo* lipogenesis but higher level of dietary lipid uptake. On the other hand, Hu et al. monitored glucose uptake activity in live cells using a glucose analogue labeled with an alkyne tag 9(3-O-propargyl-d-glucose, 3-OPG) [153]. In their study, they found that glioblastoma cells have a higher level of *de novo* lipogenesis as compared to cervical cancer cells. These studies demonstrated that cancer cells with differing metabolic activities can be distinguished using Raman tagging strategies. It will be interesting to see if the reported results can be validated for prostate and pancreatic cells using alkyne tagging and for cervical and glioblastoma cells using isotope labeling.

Hyperspectral SRS

Single-channel imaging of deuterated or alkyne-tagged lipids has been demonstrated as a useful tool for tracking uptake dynamics of a targeted lipid molecule. For unbiased profiling of the distribution of cellular lipids in response to changes in cellular metabolic states, hyperspectral SRS (hSRS) imaging is implemented. hSRS imaging enables researchers to separately quantify lipid molecules with overlapping Raman spectra by utilizing subtle differences in the spectral intensity across a range of wavenumbers [154, 155]. hSRS techniques are often used in conjunction with spectral-analysis tools to retrieve the Raman spectra of different molecules from the convoluted SRS spectra. The retrieved spectra can be used to reconstruct the compositional distribution images for each lipid species (Fig. 2.5) [156, 157]. Li et al. employed hSRS imaging to quantitatively analyze the composition of intracellular lipids inside single ovarian cancer and non-cancer stem cells and reported higher levels of unsaturated lipids in cancer cells based on the ratio of intensities at 3002 cm^{-1} and 2900 cm^{-1} wavenumber [158]. Alfonso-García et al. used hSRS coupled with unsupervised vertex component spectral analysis to study the metabolism and storage of deuterated cholesterol (D38-cholesterol) [159]. They utilized the spectral differences in the CH fingerprint region between D38-cholesterol and natural cholesterol to map the distribution of esterified and unesterified cholesterol in LDs. They found that subpopulations of LDs exist each with a predominant storage of esterified or free cholesterol. They also found that steroidogenic Y1 cells store triacylglycerol (TAG) and cholesteryl esters (CE) in different LDs. It is known that steroidogenic cells and macrophages primarily accumulate CE in LDs and liver cells primarily accumulate TAG in LDs [160, 161]. This study observed accumulation of TAG in steroidogenic cells but didn't perform any investigation in macrophages or liver cells [159]. In contrast, Fu et al. detected only CE containing LDs in macrophages and only TAG containing LDs in hepatocytes [162]. In this study, spectral differences between TAG and CE were utilized to quantitatively profile the two classes of neutral lipids. Based on these observations, it will be interesting to see whether lipid sorting occurs in macrophages, liver cells and other cell types using Alfonso-García's methodology. Fu et al. also characterized lipid compositional changes associated with metabolic disorders and further extended hSRS

coupled with isotope labeling to simultaneously trace saturated and unsaturated fatty acids [162].

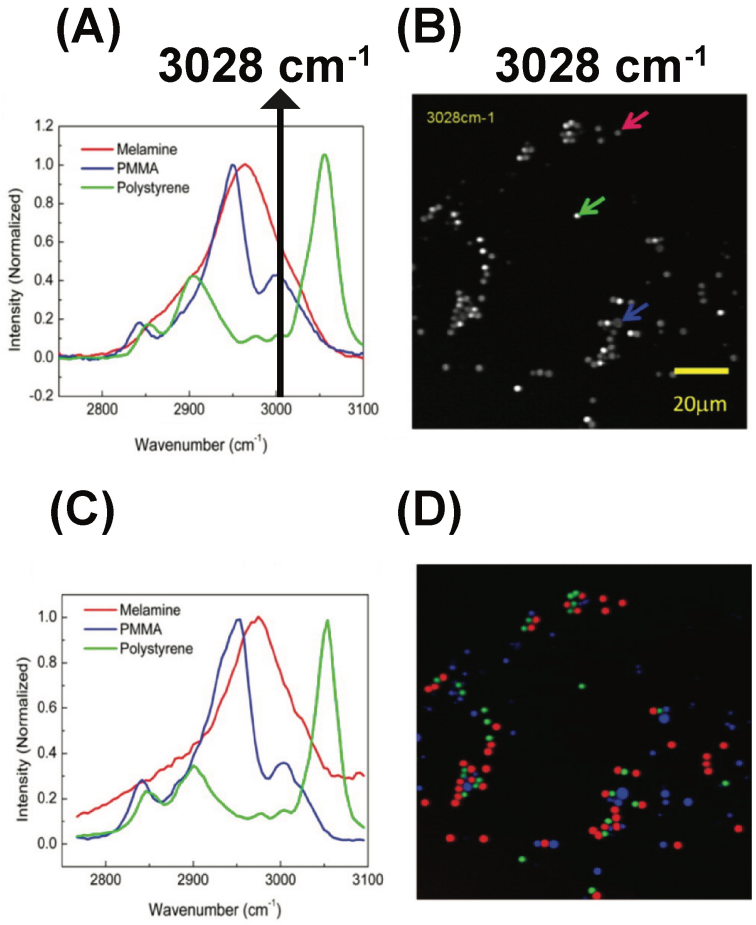

Figure 2.5: **Hyperspectral stimulated Raman scattering (hSRS) imaging for mapping three types of polymer beads with overlapping but distinct Raman spectra** (A) spontaneous Raman spectra of the three polymer beads. The black solid line indicates overlapping Raman spectra at 3028 cm^{-1} (B) stimulated Raman scattering (SRS) imaging of the three polymer beads at 3028 cm^{-1} with different color arrows pointing out corresponding beads (C) SRS spectra for the three polymer beads pointed out by the arrows (in B). (D) Color-code distribution of the three polymer beads generated using hSRS imaging coupled with spectral decomposition. PMMA: poly (methyl methacrylate). Reprinted with permission from ref. [156] Copyright (2021) American Chemical Society.

2.3 Object recognition algorithms

While optical microscopy has the spatial resolution necessary to be an inherently single-cell measurement, interpretation of micrographs in the single-cell paradigm is not always straightforward. Historically, microscopy has been used in low volume, manual, and generally qualitative, descriptions of biological samples. Such an approach, in addition to being susceptible to interpretation bias, is now increasingly impractical as image data has become larger and more complex. Furthermore, with the push in the life sciences towards generating results with greater statistical power, there is more demand for quantitative analyses of images, which all but necessitates computation. Image analysis algorithms have been under development since the pre-digital age, and the past two decades have seen many improvements in their application to biological datasets.

One of the most basic, and arguably most important, questions that can be asked about an image is where are the boundaries between objects? When quantifying metabolic composition of cells, it is important to have an objective methodology for defining objects. In tissue this amounts to cell boundaries, in individual cells, the subcellular structures and organelles such as LDs. Traditional techniques for answering this question often start with contrast enhancement and gradient-based edge detection methods. The simplest approach is thresholding, with automatic threshold determination by algorithms such as Otsu's method[163], or balanced histogram thresholding [164]. Thresholding tends to separate objects and is also often employed to aid in background correction. Convolution with operators such as the Sobel [165], Canny [166], or other gradient operators can provide information on sharp line boundaries. For more general shape extraction, the Hough transform has been a popular choice in a wide variety of fields. First patented in 1962 for line identification [167], it was then generalized to arbitrary shapes [168]. It is well-suited to identifying regularly shaped features which can vary in dimension across an image.

These gradient or edge detection techniques are then frequently combined with a watershed based algorithm [169], which imagines filling basins from minima in the images and draws boundaries where the watersheds meet. Implementations of these techniques can be found in all major programming languages, and are also included in many widely available image analysis software suites, like Fiji [170]. They have therefore been applied, in a number of combinations and variations, for analysis of LD size and number distribution [171, 172].

More recently, the field of computer vision has shifted focus to machine learning approaches for everything from automatic feature extraction to image classification. This has been driven in large part by the success of convolutional neural networks (CNN), and their rapid development in the past decade. First introduced over 20 years ago [173, 174], initial adoption was slow, but the list of current variations and applications is now constantly growing. CNNs work similarly to conventional, or 'fully-connected', neural networks but reduce the number of parameters that need to be learned by using convolutions rather than transformation matrices that relate every point in the image to every point in the output. This is in some ways analogous to some traditional methods listed above, but instead of pre-selecting, e.g. a gradient filter, the filter is learned by the network, and there are many filtering steps.

While generally more computationally intensive, fully-connected neural networks have also found use in image analysis.

The major drawback for using CNNs or deep learning architectures generally is the need for training data. This has slowed adoption in the field of lipidomics, although CNNs have been successfully applied to numerous types of microscopy data. Medical imaging has been a recent adopter, with hundreds of successful demonstrations in the last three years [175]. Importantly, these demonstrations span a wide-variety of disciplines but utilize similar network architectures. Many are straightforward modifications of well-known networks, and often rely on already trained networks as starting points, suggesting a similar strategy may be effective for lipidomics. Single-cell segmentation, cell cycle progression and disease state identification, have been recently demonstrated using CNNs on fluorescent images [176]. Chen et al. also recently showed algal cell classification based on lipid content, using time-stretch quantitative phase imaging and deep neural networks [177].

A final consideration, is that many of the imaging techniques used for lipid characterization contain additional information beyond the purely morphological. Most of the analysis algorithms discussed thus far have focused on segmentation and object identification. This makes them generalizable to all types of images, but also makes them blind to the additional information that can be encoded in some microscopy datasets. In some cases, it is therefore advantageous to utilize more specialized algorithms for analysis, hyperspectral coherent Raman imaging being a prime example. Fu and Xie demonstrated the ability to segment subcellular structures, including lipid droplets, from a hSRS dataset using a spectral phasor method adapted from the fluorescence lifetime imaging field [178]. Di Napoli et al. were also able to monitor uptake of different lipid components using hyperspectral CARS [179], using an unsupervised retrieval algorithm [180].

2.4 Quantitative CRS for single-cell and single-LD analysis

High signal to noise ratio (SNR) associated with concentrated CH_2 bonds in lipids allows researchers to monitor the dynamics of LDs in a straightforward fashion using CRS techniques coupled with LD recognition and trajectory tracking packages. Jüngst et al. demonstrated tracking of LDs using fast, long-term three-dimensional CARS imaging at 2850 cm^{-1} in order to investigate the dynamics of LD fusion in living adipocytes undergoing differentiation [181]. They used the Imaris software package for detection and tracking of LDs. In Imaris, thresholding is performed for automated segmentation of LDs. Morphological characterization of identified LDs is then performed including radius and volume rendering. Detected LDs are then tracked by selecting for appropriate three-dimensional tracking algorithm. Based on the lipid transfer rates obtained, researchers suggested a model in which lipid transfer is driven by the pressure difference between participating LDs through a putative fusion pore, whose size depends on the size of the donor LD.

Zhang et al. used SRS microscopy to study the dynamics of LDs using three-dimensional SRS imaging at 2850 cm^{-1} [182]. They implemented a feature point tracking algorithm, as developed for the Particle Tracker software [183], for monitoring LD movements. In this software, feature points are localized by finding local intensity maxima in the filtered image. The retrieved positions are then refined to reduce the standard deviation of the position measurement, which takes into consideration a user-provided threshold. Once point location matrices have been defined for each frame in the time-resolved image, a cost function is minimized to find a set of associations for tracking each point. Using this software, researchers demonstrated that the dynamics of LDs, quantified using maximum displacement and speed as the parameters, can be used to differentiate changes in lipid metabolism in living cells. They studied changes in lipid metabolism upon glucose starvation and refeeding and showed that their methodology could predict increase in lipolysis upon starvation as expected.

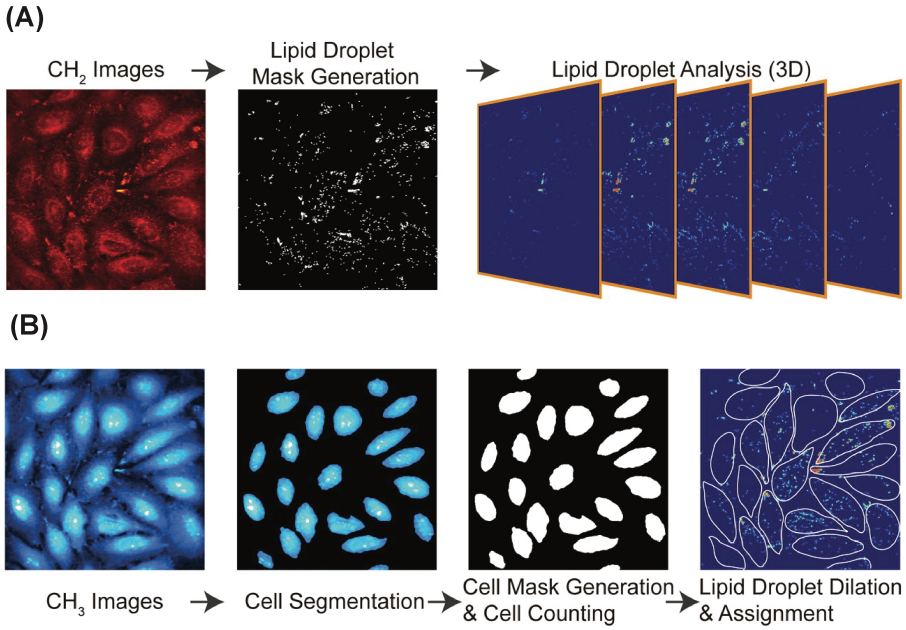

Figure 2.6: **Stimulated Raman Scattering (SRS) image processing pipeline for determining cellular boundaries and characterizing lipid droplets (LDs) in single cells.** (A) Three-dimensional lipid-specific images were acquired at 2850 cm^{-1}. The signal was processed to generate a lipid droplet mask. The lipid droplet mask was analyzed for three-dimensional morphological characterization. (B) Three-dimensional protein-specific images were acquired at 2950 cm^{-1} for cell boundary segmentation and cell mask generation. The position of each LD was then recorded and assigned to an individual cell. Reprinted with permission from ref. [184] Copyright (2021) American Chemical Society.

Medyukhina et al. developed an image processing approach for detection of nuclear and cellular boundaries from co-registered two-photon excited fluorescence (TPEF) and CARS images respectively [185]. For nuclei boundary determination, they first used the local gray-scale minimum from denoised TPEF images for localization of nuclei centers. The gradient maxima from each nucleus location was used to detect the nuclear boundary. Once nuclei locations and boundaries were validated, they subsequently used TPEF images to delineate the cellular boundaries in the denoised CARS images. They assumed that the cellular boundary corresponds to the first local gradient minimum behind the nuclear boundary. Finally, they demonstrated the implementation of this approach for automated segmentation of cells and nuclei in brain tumor samples.

In order to reveal single-cell heterogeneity, data has to be acquired from multiple single cells for statistically significant conclusions. Cao et al. characterized the mechanisms of LD growth and formation upon lipid accumulation, as induced by exogenous FFA, at the single-cell level using SRS microscopy [184]. LD growth and formation was monitored by tracking the number, average size, and average SRS intensity of LDs in a single cell under various concentrations of FFA. To increase throughput and therefore statistical power, all experiments were performed on a microfluidic platform capable of delivering controlled concentration of FFA to uniquely addressable nanoliter cell culture colonies. Images were obtained at 2850 cm^{-1} to identify LDs (Fig. 2.6A). A second set of images were taken at the protein-rich CH3 stretching vibration at 2950 cm^{-1} to extract boundaries of single cells. Thresholding was performed to generate a LD and cell mask. The position and morphology of each LD was then recorded and assigned to an individual cell (Fig. 2.6B). In this investigation, researchers found that lipid accumulation in nonadipocyte cells is mainly reflected in the increase of LD number, as opposed to an increase in their size or lipid concentration.

2.5 From lipidomic to multiomic analysis

Highly-multiplexed barcoding strategies and automated fluid handling has now made it possible to profile the transcriptome from thousands of single cells in one experiment. However, in order to understand the correlation between gene expression and metabolic states at the single-cell level, multiple measurements must be made on the same single cell. Because CRS imaging is non-destructive, cells can be sequenced directly downstream of lipidomic analysis, thereby making implementation of multi-omic approaches possible. In this section, we will discuss the applicability of utilizing the developed microfluidic and microscopic platforms for combined single-cell genomic and lipidomic analysis.

Microfluidic platforms

Microfluidic technology has proven critical for increasing the throughput of NGS techniques permitting profiling of genome-wide features from a large number of single cells. Implementation of single-cell sequencing requires single-cell isolation. In microfluidic platforms, this is

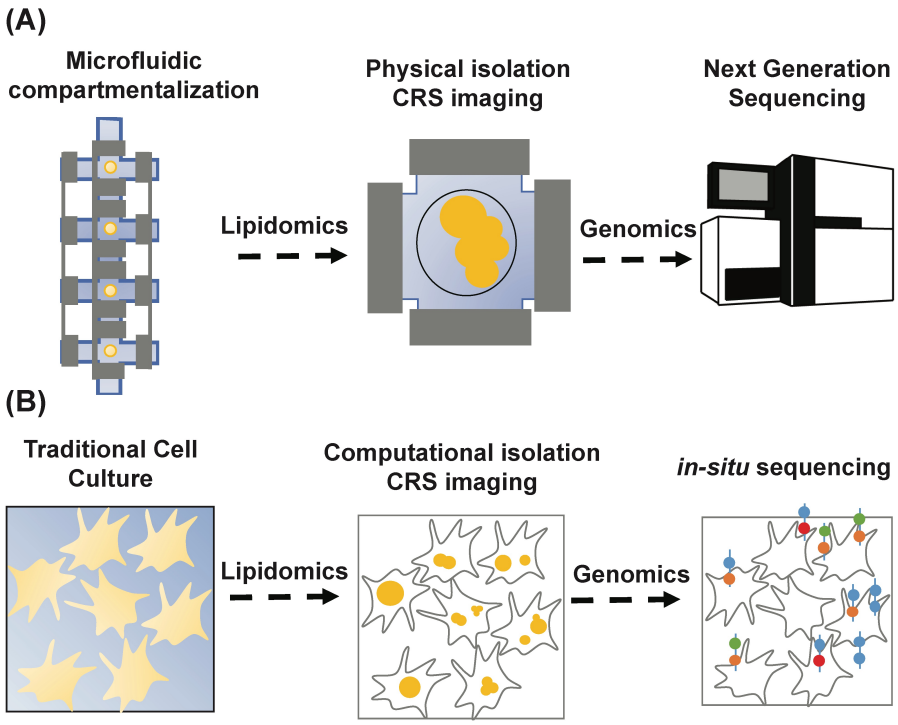

Figure 2.7: **Combining lipidomic and genomic analysis at the single-cell level.** (A) Lipidomic and genomic analysis using microfluidic single-cell isolation. A single cell is physically isolated in a small chamber using valve-based compartmentalization. While the cell is trapped, images are acquired in a non-destructive fashion using coherent Raman scattering (CRS) imaging for lipidomic analysis. The cell is then pushed downstream for library preparation and finally sequenced using next generation sequencing (NGS) techniques. (B) Lipidomic and genomic analysis using microscopy and computational cell-segmentation. Multiple live cells are imaged on a coverglass using CRS. Individual cells are then computationally isolated using object recognition algorithms and images are analyzed for lipidomic analysis at the single-cell level. The transcriptome of the same cells is then profiled using in situ sequencing techniques.

typically achieved by valve-based compartmentalization [17, 18], droplet encapsulation [32, 33], and microwell separation [30, 186]. After single-cell isolation, downstream library preparation reactions are implemented. Another advantage of microfluidic devices is the optical transparency of the polymer used for chip fabrication, polydimethylsiloxane (PDMS), which enables researchers to visualize sequencing protocols in real-time using a microscope. Because of the optical transparent nature of microfluidic devices and the necessity to physically isolate single cells, lipidomic and genomic analysis can be performed on the same single cell by acquiring images upstream of library preparation reactions (Fig. 2.7A).

Streets et al. developed a microfluidic platform for whole-transcriptome profiling of single

cells [17]. In this device, cells were isolated in nanoliter-scale trapping chambers using a valve-based strategy. CRS imaging can be performed while cells are trapped thereby allowing researchers to perform combined lipidomic and genomic analysis on the same cell. Lane et al. integrated epifluorescence microscopy with scRNA-seq on a commercial microfluidic platform, Fluidigm C1 [187]. They used this approach to measure both the dynamics of activation for a specific transcription factor and the global transcriptional response in the same individual cell. Instead of fluorescence microscopy, label-free CRS imaging can be implemented on this platform for combined lipidomic and genomic analysis on the same cell. Gierahn et al [188]. and Bose et al. [189] developed platforms for massively parallel scRNA-seq based on gravitational settling of single cells in subnanoliter and picoliter-scale microwells respectively. As cells are stationary while isolated in microwells, this solid-phase capture can be utilized for high-resolution CRS imaging upstream of library preparation reactions. Zhang et al. developed a flow cytometer based on Raman scattering for fast, high-throughput single-cell analysis [190]. They developed a multiplex stimulated Raman scattering flow cytometry (SRS-FC) technique for measuring chemical contents of single cells. This technique can be extended for quantifying lipid content in single cells. These cells can then be isolated using droplet encapsulation platforms [32, 33] for scRNA-seq. Thus coupling SRS-FC with droplet encapsulation-based microfluidic platforms will allow researchers to perform combined lipidomic and genomic analysis on the same cell. Such coupled datasets will transform the way we understand single-cell biology by enabling researchers to study the correlation between single-cell phenotype and gene expression profile.

High-speed CRS imaging

Microfluidic devices have enabled researchers to perform single-cell analysis in a high-throughput fashion. In the previous paragraph, we discussed the applicability of microfluidic platforms for retrieving lipidomic (CRS imaging) as well as transcriptomic (scRNA-seq) information from the same single cell, thereby advancing towards multi-omic approaches. However, implementation of such coupled experiments on hundreds to thousands of single cells will require application of high-speed CRS imaging techniques for fast single-cell lipidome profiling. As discussed previously, hyperspectral imaging techniques are essential for profiling the distribution of multiple cellular lipids simultaneously. Thus, it becomes critical to employ hyperspectral CRS techniques capable of rapid spectral acquisition at microsecond scale. Recent developments in CARS and SRS instrumentation have been influential in accelerating the spectral acquisition rate. For example, Liao et al. demonstrated parallel acquisition of SRS signal over 180 cm^{-1} bandwidth ($\sim$ 20 spectral data points) with 42 µs pixel dwell time using spectrally focused laser pulses and a homebuilt microsecond optical delay-line tuner [191]. He et al. integrated a galvanometer mirror-based rapid-scanning optical delay line with spectrally focused laser pulses to acquire a spectrum with 20 data points in 40 µs [192]. Liao et al. built an array of tuned amplifiers for lock-in free parallel acquisition of SRS signal over 180 cm^{-1} bandwidth ($\sim$ 20 spectral data points) with 32 µs pixel dwell time using multiplexed SRS [193, 194]. Alshaykh et al. integrated a rapid acousto-optic

delay line with spectrally focused laser pulses to achieve parallel acquisition of SRS signal over 180 cm^{-1} bandwidth ($\sim$ 20 spectral data points) with 12.8 µs pixel dwell time [195]. Hashimoto et al. coupled a rapid-scanning retro-reflective optical path length scanner with a Fourier-transform CARS (FT-CARS) system to accomplish spectral acquisition rate of 20,000 spectra per second over 1300 cm^{-1} bandwidth ($\sim$ 130 spectral data points) [196]. Tamamitsu et al. updated this system to incorporate a more rapidly scanning optical delay line thereby achieving spectral acquisition rate of 50,000 spectra per s ($\sim$ 500 spectral data points) [197]. Recently, Coluccelli et al. demonstrated parallel detection of CARS signal with Raman shifts of $\sim$ 3000 cm^{-1} using FT-CARS. The system was based on a single high-power Yb-fiber laser source coupled to a FT interferometer with pixel dwell time of 160 µs ($\sim$ 675 spectral data points) [198]. He et al. achieved simultaneous two-color SRS imaging by engineering the profile of Stokes beams and utilizing the output of a dual-phase lock-in amplifier, thereby reaching the maximum speed as in a single-color SRS [199]. Thus, such studies focused on development of rapid CRS imaging techniques demonstrate the promise of coupling high-content spectral imaging with high-throughput single-cell analysis.

Microscopic platforms

An alternative to physical isolation for single-cell genomic analysis is to employ techniques that turn the genomic information into optical information *in situ*. Fluorescence *in situ* hybridization (FISH) is a technique that uses fluorescent probes that bind specifically to complementary nucleic acid sequences. Thus, researchers can obtain spatial information about the distribution and subcellular localization of specific DNA or RNA molecules. In situ sequencing leverages FISH to extract sequence information from tens to hundreds of targeted transcripts for large scale gene expression profiling in single cells [200–202]. Such methods preserve the microenvironment of the biological sample allowing single molecule RNA sequencing and localization without removing cells from their original context. These emerging technologies are enabling a new-wave of spatial transcriptomic studies, which link single-cell gene expression to cellular niche in a tissue or organ. Since FISH techniques are fundamentally based on imaging, quantitative CRS techniques for lipidomic analysis can be combined with in situ sequencing for multi-omic single-cell analysis. Fig. 2.7B illustrates how single-cell transcriptomics might be combined with CRS-based single-cell lipidomics.

2.6 Conclusion

Coherent Raman scattering (CRS) techniques have become an essential tool for profiling LDs in single-cells by enabling researchers to quantify intracellular lipids in a non-destructive and time-resolved fashion. As the development of CRS instrumentation progresses towards higher specificity, sensitivity, and faster hyperspectral imaging, and next generation sequencing techniques advance towards higher throughput single-cell genomic analysis with lesser bias, coupling these techniques will lead to a more acute understanding of the regulation

of metabolic pathways. Particularly in the context of adipose tissue, adipocytes display a wide range of functions and phenotypes, from energy storage in large unilocular LDs (white adipocytes) to thermogenic lipolysis of small LDs (brown adipocytes). Adult humans were thought to only have white adipose tissue with brown adipose tissue being essentially absent after infancy [203, 204]. In the early 2000s, observations in the field of nuclear medicine started challenging this notion [205, 206]. Multiple studies performing positron emission tomography (PET) with [18F]-fluorodeoxyglucose (FDG) for staging of cancer observed increased uptake of glucose in tumor-unrelated areas [205, 206]. These areas were found in the neck and shoulder region and presented itself with features of adipose tissue. It was hypothesized that this FDG uptake could represent activated brown adipose tissue in adult humans and this was finally demonstrated by three independent studies in 2009 [207–209]. Now, the existence of brown adipose tissue in adult humans is a well-accepted fact in the research community. Rodents also have a third kind of adipocyte called beige adipocyte, which has a different developmental origin from brown adipocytes [48]. This fact naturally raises the question of whether humans also possess beige adipocytes. Interestingly, recent investigations of human brown adipocytes have reported the mixed presence of presumed beige adipocytes [51, 210]. These claims have been reported based on the upregulation of beige adipocyte markers as identified in rodents. Consequently, it is clear that we are only just beginning to understand and appreciate the vast cellular diversity of human adipose tissue. These data raise some critical questions about the composition of human adipose tissue that might only be addressed with single-cell measurements. Technology that couples CRS for lipid profiling and RNA-sequencing for gene expression analysis in single cells could greatly advance our understanding of adipocyte heterogeneity. We anticipate that imaging and sequencing single cells will be the next wave of multi-omic single-cell analysis.

2.7 Conflicts of interest

There are no conflicts to declare

2.8 Acknowledgements

This publication was supported by the National Institute of General Medical Sciences of the National Institutes of Health under award number R35GM124916. AMS is a Chan Zuckerberg Investigator. Credits to Gabriel Dorlhiac for writing section 2.3 of this chapter. This work is now published as Gupta et al. 2020:
Anushka Gupta, Gabriel F Dorlhiac, and Aaron M Streets. "Quantitative imaging of lipid droplets in single cells". *Analyst*, 144.3 (2019), pp. 753–765.

Chapter 3

μCB-seq: microfluidic cell barcoding and sequencing for high-resolution imaging and sequencing of single cells

3.1 Introduction

In Chapter 2, we ended with how the scientific community is just beginning to understand and appreciate the vast cellular diversity of human adipose tissue, and technologies that couple imaging and sequencing measurements for lipid and gene expression profiling respectively, in single cells, could greatly advance our understanding of adipocyte heterogeneity. Broadly speaking, in almost all cell-types, identity is not entirely described by the transcriptome alone. Rather, phenotypic features such as morphology, protein localization, and metabolic composition provide critical information about the identity, function, or state of cells but are not directly encoded in the genome, and therefore cannot be measured by sequencing. Thus, optical microscopy remains an indispensable tool for characterizing phenotypic and functional features of single cells in a wide variety of biological systems [211–213]. Combining microscopy with scRNA-seq in such multicellular systems can provide valuable insights into the relationship between gene expression and cellular phenotype.

Performing optical imaging and sequencing measurements on the same single cell is technically challenging because it requires precise cell manipulation and tracking. A cell's volume is ~ 7 orders of magnitude smaller than that of a typical well in a microwell plate, which makes it difficult to locate and image a single cell using a high magnification objective in tube- or plate-based scRNA-seq protocols. A previous study demonstrated imaging and downstream gene expression analysis using RT-qPCR for adherent cells which can be confined to the bottom plane of a well, though the process of adherence and imaging takes multiple hours [214]. A more recent study used a commercial dissection microscope to capture images of single yeast cells at recorded coordinates, which were then selected by an automated micro-manipulator and dispensed in a tube-based array for gene expression

analysis [215]. These examples demonstrate the challenge of imaging and sequencing single cells with traditional "benchtop" techniques. Microfluidic technology is well-suited to address such technical challenges, as it provides low Reynolds number, laminar flow, and programmable fluidic control at the microscale. Specifically, multi-layer microfluidic devices with integrated valves allow for the trapping of single cells in nanoliter volumes which allows for rapid imaging and sorting for downstream genomic analysis. For example, Lane et al. used the Fluidigm C1 for microfluidic scRNA-seq with optical microscopy to combine fluorescent measurements of transcription factor dynamics with gene expression profiling in single cells [187]. In this study, the link between a cell's image and its transcriptome was preserved by carrying out individual library preparation for each cell, making library preparation the rate limiting step. Furthermore, imaging was limited to low-magnification with a long working distance objective. When imaging is not required, higher-throughput methods such as microwell- and droplet-based techniques allow for multiplexed processing of many cells at once, thus drastically reducing library preparation time [30, 32–34, 188, 216]. These methods use microfabricated devices to isolate cells in nanoliter volumes, in which cellular barcodes are incorporated into cDNA during RT to allow for pooling of many cells into a single sequencing library. However, these techniques are currently not compatible with imaging because cellular barcodes are assigned randomly, making it impossible to know which transcriptome belongs to which cell image. Yuan et al. recently demonstrated a promising solution to this challenge, in which the random barcode sequences were optically decoded using fluorescence microscopy [217]. Spectrally-encoded beads [218] or printed droplet microfluidics [219] may provide yet other solutions for imaging and sequencing single cells. Zhang et al. used a microfluidic droplet generator to acquire fluorescence intensity measurements of encapsulated cells before dispensing them in nanowells preloaded with "coordinate-oligos" for sequencing [220]. However, these studies have not demonstrated high-resolution imaging to reveal subcellular structure. Thus, further developments are needed to realize the benefits of combined high-resolution imaging and high-sensitivity RNA-seq on single cells.

In this chapter, we present microfluidic cell barcoding and sequencing (µCB-seq), a microfluidic platform that enables paired imaging and sequencing measurements of single cells. Our platform uses integrated microfluidic valves to precisely manipulate single cells for isolation, imaging, and multistep library preparation on-chip. In µCB-seq, independently addressable microfluidic reaction chambers are preloaded with known barcoded primers, which are used to capture genomic material from single cells. This approach provides the ability to couple genomic information with phenotypic information that requires high-resolution imaging or even time-resolved imaging to investigate dynamic cellular behavior. Here, we demonstrate the capabilities of µCB-seq by performing scRNA-seq using the molecular crowding single-cell RNA barcoding and sequencing (mcSCRB-seq) protocol [13]. We find that µCB-seq improves upon the high sensitivity of mcSCRB-seq by utilizing the benefits of microscale volume library preparation reactions [17]. We then combine multiplexed scRNA-seq with live-cell fluorescence imaging on-chip to demonstrate µCB-seq as a scalable platform for extracting high-resolution phenotypic data and high-sensitivity genomic data from single cells.

3.2 Results

Microfluidic device design and µCB-seq workflow

µCB-seq is implemented on a PDMS microfluidic device with integrated elastomeric valves fabricated by multilayer soft-lithography [221]. The device has two functional layers, an upper control layer, and a lower flow layer (Fig. 3.1A). The control valves are pneumatically actuated by a solenoid valve array that is operated with the KATARA controller and a programmable computer interface [222]. The device design was inspired by a previous

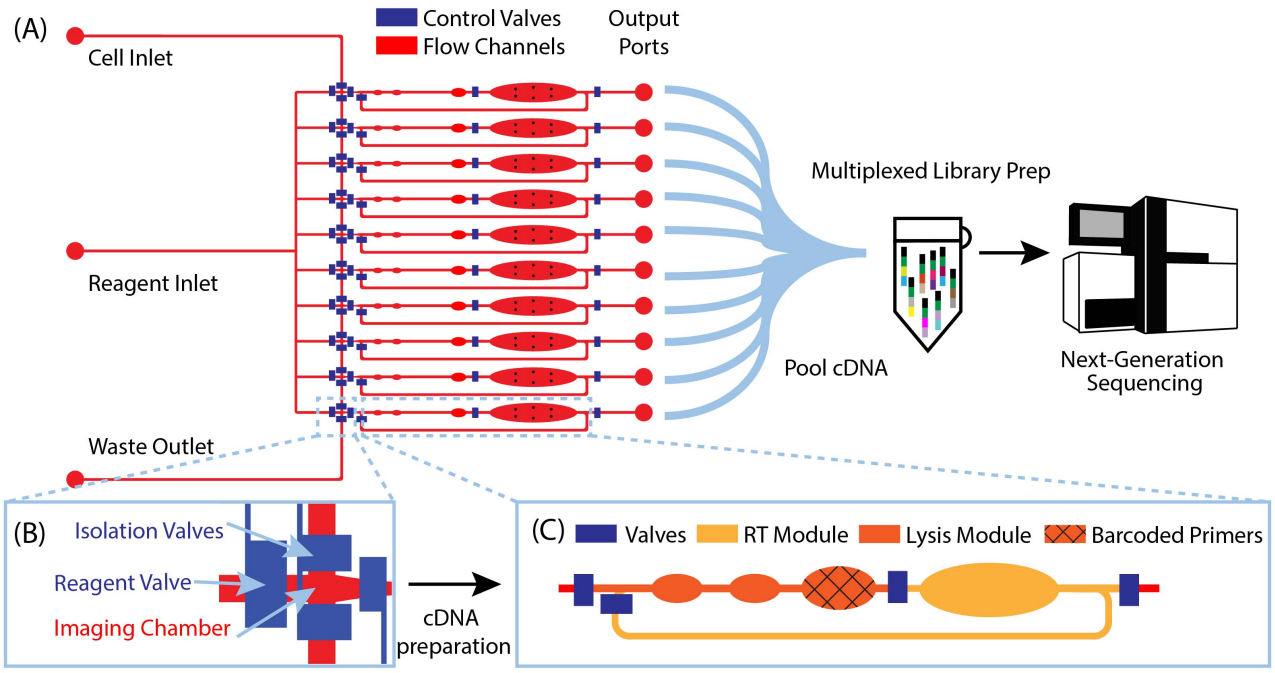

Figure 3.1: **µCB-seq device design and workflow** (A) schematic of the microfluidic device with control valves in blue and flow layer in red. Cells are loaded into the cell inlet and reagent is introduced through the reagent inlet. The device processes 10 cells in 10 individual reaction lanes, each ending in an output port. Reverse-transcribed cDNA is recovered from output ports for all cells, pooled in a single tube for off-chip library preparation using the mcSCRB-seq protocol, and sequenced using next-generation sequencing platforms. (B) Detailed diagram of the imaging module showing the imaging chamber. The two isolation valves can be actuated to actively capture a cell of interest in the imaging chamber. (C) Detailed diagram of one reaction lane showing the lysis and RT modules separated by valves. The textured reaction chamber in the lysis module is preloaded with barcoded RT primers.

scRNA-seq platform [17], and in this demonstration, can process 10 cells simultaneously in parallel reaction lanes. Each reaction lane has a modular design to allow for imaging, cell

lysis, and implementation of a wide range of multistep library preparation protocols. The imaging module consists of an imaging chamber flanked by two isolation valves (Fig. 3.1B), and the lysis and reverse transcription (RT) modules consist of isolated reaction chambers separated by valves (Fig. 3.1C, Fig. A.1). During chip operation, a suspension of single cells is loaded into the cell inlet and directed towards the imaging module using pressure-driven flow. Once a cell reaches an imaging chamber, it is actively trapped, imaged, and then sorted into its respective reaction lane or sent to waste, allowing for the enrichment of cell subpopulations or the selection of rare cells. After imaging, the selected cell is ejected from the imaging chamber into the lysis module of its reaction lane by a flow of lysis buffer from the reagent inlet. After all 10 lysis modules are filled with lysis buffer, processing proceeds in parallel for all 10 cells.

RT primers with known barcode sequences are preloaded in the lysis module of each reaction lane (Fig. 3.1C, device fabrication). Each reaction lane is indexed by two pieces of information: a known barcode sequence and its lane index on the device. As a result, all sequencing reads with a unique cell barcode sequence can be linked to cell images with the corresponding lane index. Barcode sequences used in this study are a subset of 8-nt long Hamming-correctable barcodes [223] designed for 50% GC content and minimal sequence redundancy (Table A.1). The unique molecular identifier (UMI) sequence in the RT primers is 10-nt long.

Positioned above the reaction chambers in the lysis module are mixing paddles (Fig. A.1), which are used to accelerate mixing as demonstrated previously [17]. After dead-end filling of the lysis module, barcoded RT primers are resuspended in cell lysate by active mixing, after which the entire chip is placed on a temperature-controlled platform to hybridize suspended RT primers to cellular mRNA transcripts. The reagent input line is then flushed and filled with RT buffer, which is injected into all reaction lanes to dead-end fill the RT module. The RT buffer contains 7.5% PEG 8000, which has been demonstrated to increase RT efficiency through molecular crowding [224, 225]. Reverse transcription is carried out for 1.5 hours at 42 °C, during which the mixing paddles are actuated in a peristaltic manner to circulate the relatively viscous RT mix throughout the mixing channel of each reaction lane (Fig. A.1).

The total reaction volume of each lane is 227 nL, which is 1–2 orders of magnitude smaller than typical plate-based protocols [13]. After RT, all lanes are independently flushed with 1.7 µL of nuclease-free water to recover cDNA, and pooled into a single tube using gel-loading pipette tips for a total volume of 17 µL. Additional off-chip steps including exonuclease digestion and cDNA amplification followed by purification and Nextera library preparation are performed in a single tube using the conventional mcSCRB-seq protocol (Material and methods). cDNA libraries representing whole single-cell transcriptomes are then sequenced on a next-generation sequencing platform. µCB-seq's ability to multiplex off-chip library preparation reactions significantly reduces the cost of Nextera reagents, which dominates library preparation reagent cost for commercial integrated microfluidic platforms. Consequently, a 96-cell implementation of µCB-seq stands to reduce reagent cost by almost 2 orders of magnitude as compared to non-multiplexed protocols. We performed a line-by-line library preparation cost analysis for µCB-seq, including the cost of consumables and

reagents, in Supplemental Table 1. Comparing this analysis to a cost estimate for commercial platforms, we found a $\sim$ 50-fold reduction in total library preparation cost-per-cell [226].

Microfluidic device fabrication with addressable barcode spotting

Multilayer chip fabrication is necessary to create microfluidic devices with integrated valves and pumps that can be actuated for precise fluidic manipulation of cells, buffer exchange, and continuous-flow mixing of reagents [227]. These capabilities enable the implementation of multistep reactions for library preparation on such devices [18, 228, 229]. However, as the number of cells is increased, "world-to-chip" interfacing becomes more complex and off-chip library preparation steps are increased proportionally [230]. For example, commercial devices which can process 50–100 single cells require researchers to prepare an equivalent number of individual sequencing libraries, which increases cost and processing time [231]. A sophisticated fluidic circuit architecture and combinatorial barcoding have been implemented to increase throughput of these devices and process up to 800 cells with only 20 individual libraries off-chip [232]. µCB-seq offers an improved fabrication method that obviates the need for complex routing of barcoded reagents, and could be incorporated in existing devices to process hundreds of cells with only two inlet and two outlet ports and a single low-cost off-chip library preparation.

In order to increase multiplexing throughput while minimizing the complexity of device operation, µCB-seq utilizes a fabrication method that combines multilayer soft lithography and DNA array printing to preload the lysis module of each lane with known barcoded RT primers. This approach is similar to previous microfluidic devices for high-throughput screening of protein-DNA interactions [233]. To verify that RT primers can be successfully resuspended from PDMS after baking, 2 µL droplets of 2 ng μL^{-1} primer were manually spotted on PDMS slabs, allowed to dry, baked at 80 °C for 2 h, and incubated at room temperature for 24 h. Primers were manually resuspended in 2 µL of nuclease-free water and analyzed for fragment length. The RT primers showed no noticeable degradation during the final baking at 80 °C and can be resuspended with high efficiency (Fig. A.2).

The µCB-seq device was designed in the push-down configuration with three layers: a thick upper control layer, a thin middle flow layer, and a thin lower dummy layer. We used on-ratio PDMS–PDMS bonding to avoid PDMS waste and provide a stable seal by partial crosslinking of a 10:1 base:crosslinker mixture with each new layer of the microfluidic device [234]. The control and flow molds were first patterned using standard photolithography techniques (Fig. 3.2A, Material and methods). The 10:1 PDMS mixture was then separately cast onto the two molds and baked. The partially crosslinked control layer was peeled from the mold and placed atop the thin flow layer for alignment, after which the two-layer assembly was baked to achieve undercured PDMS–PDMS bonding (Fig. 3.2B). The two-layer assembly was trimmed and inverted, exposing the open-faced flow layer of the device. 0.2 µL of 1.5 µM barcoded RT primers were then spotted into the lysis module of each reaction lane and allowed to dry (Fig. 3.2C). By spotting the primers directly into the lysis modules, we avoid subsequent alignment steps. The two-layer chip (still undercured) with dried primers was

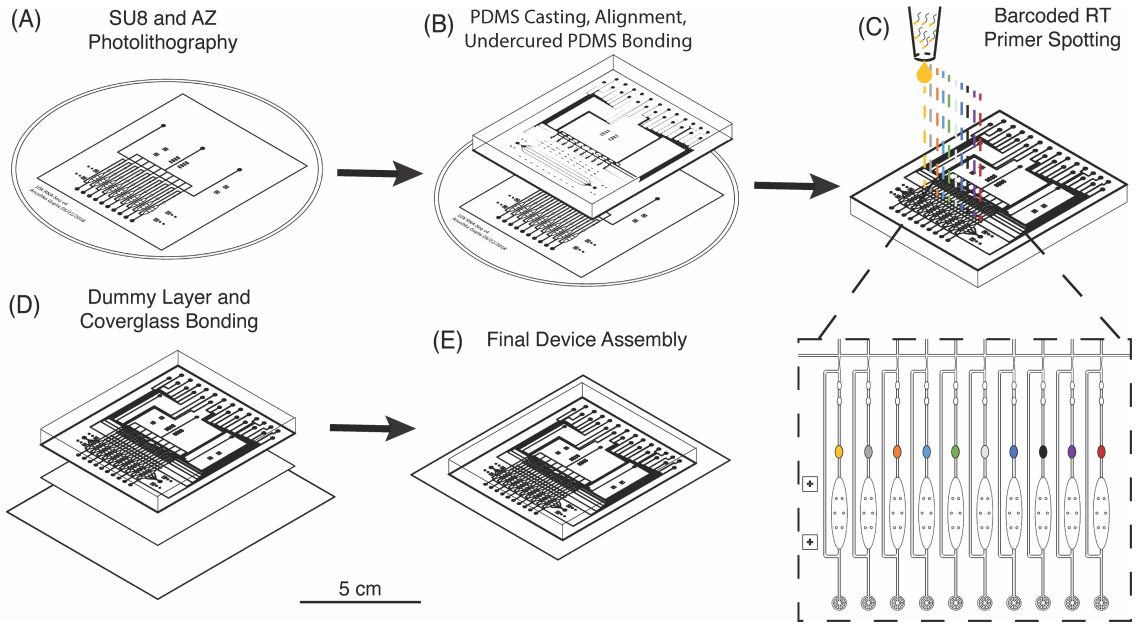

Figure 3.2: **Fabrication of µCB-seq devices with barcoded RT primer spotting.** (A) Photolithographic patterning of control and flow molds on Si wafers. (B) Diagram of PDMS casting and undercured PDMS bonding between the control and flow layers. (C) Detailed diagram of barcoded RT primer spotting. Unique primers are delivered to each lysis module and dried before the device is closed (D) by bonding to a PDMS dummy layer. (E) PDMS devices are then plasma bonded to a coverglass for final assembly. The scale bar refers to the panels (A) to (E).

placed atop an undercured dummy layer and bonded with heat to complete crosslinking between the layers (Fig. 3.2D). The PDMS–PDMS bond between the spotted flow layer and the bottom dummy layer in the µCB-seq device is achieved without the use of oxygen plasma, thereby preserving primer integrity. After complete curing, the three-layered µCB-seq device was cut from the dummy wafer and oxygen plasma was used to bond this final device assembly onto a #1.5 glass coverslip. The result of this fabrication protocol was a valve-based multilayer microfluidic device, preloaded with intact barcoded RT primers at addressable locations (Fig. 3.2E, Material and methods).

µCB-seq yields high-quality scRNA-seq libraries

µCB-seq was designed to be compatible with most barcoded single-cell library preparation protocols. In this demonstration of µCB-seq, single-cell cDNA libraries were prepared by implementing the highly sensitive mcSCRB-seq protocol within the microfluidic device. mcSCRB-seq is a multiplexed 3' counting method using cell barcodes and UMIs to acquire an absolute transcript count from each cell [13]. We first evaluated the effectiveness of µCB-seq by generating cDNA libraries from 20 replicates of 10 pg total RNA isolated from HEK293T

cells. Total RNA extracted from HEK293T cells was injected into the cell inlet and the 10 sets of isolation valves were simultaneously actuated to trap 10 pg RNA in each imaging chamber (Note A.3). The contents of each imaging chamber were then pushed into their respective reaction lanes for cDNA processing (Material and methods). The cDNA libraries were then collected from the chip, pooled, and prepared for high-throughput sequencing. The libraries were sequenced with read 1 (R1) encoding the 8-nt long known barcode sequence and 10-nt long UMI and read 2 encoding the cDNA fragment. After sequencing, all raw fastq files were analyzed using the zUMIs pipeline (Material and methods) [235]. In zUMIs, reads with all R1 bases having quality score >20 were mapped to the human reference genome (GRCh38) using STAR [236]. Gene annotations were obtained from Ensembl (GRCh38.93) and filtered to remove biotypes such as pseudogenes [237]. Quantification of aligned reads was done using the Subread package to generate expression profiles for each library [238]. Throughout this study, genes detected were defined as those for which at least one UMI was detected. In total, all 20 libraries of purified RNA were sequenced to an average depth of 65,000 reads (Table A.2).

We first characterized the mapping statistics for each of the 20 total RNA libraries, which allowed us to evaluate the percentage of useful reads for downstream analysis. Across all the replicates, a median of 53% of reads mapped to exons, 11% to introns, 16% to intergenic regions, and 17% to no region in the human genome (Fig. 3.3A). These statistics are comparable to other 3'-barcoding-based sequencing protocols with a range of 29–57% exonic reads, 2–15% intronic reads and 6–23% unmapped reads [239]. Detection of reads from unspliced transcripts makes µCB-seq data compatible with single-cell analyses utilizing splicing events such as RNA velocity [240]. Here, reads mapping to the exonic regions of the genome were quantified to generate a UMI count expression matrix. These 10 pg total RNA sequencing libraries generated with µCB-seq detected a median of 3008 unique genes with only 30,000 reads per sample (Fig. 3.3B). Transcript abundance was strongly correlated between µCB-seq libraries, with a median pairwise Pearson coefficient of 0.84 (n = 190 pairs) across reaction lanes and devices (Fig. 3.3C).

Next, we compared transcript abundance in these pseudo-single-cell libraries with typical gene expression in HEK293T cells as measured by bulk RNA-seq of HEK total RNA (1 µg, Material and methods). For comparison, we pooled the reads from all 20 µCB-seq libraries of 10 pg total RNA for a total of 1.3 million reads (Table A.2) and compared the genes detected against those present in 1.3 million bulk sample reads (TPM > 0). With the same total number of reads, ∼70% of genes that were present in bulk RNA-seq library of 1 µg total RNA were also detected in pooled µCB-seq libraries consisting of 200 pg RNA in total (Fig. 3.3D). There were over 700 genes that were detected in µCB-seq but not in bulk RNA-seq. These are likely a combination of low-abundance transcripts and transcripts that are not primed or reverse-transcribed in bulk due to molecular differences in the protocols. Transcript abundance in an average 10 pg total RNA library (averaged counts per million over all 20 replicates) correlated well with the bulk measurement (Pearson correlation = 0.65, p-value < 0.05, Fig. 3.3E). This demonstrates that µCB-seq can recapitulate expected gene expression profiles with low quantities of mRNA.

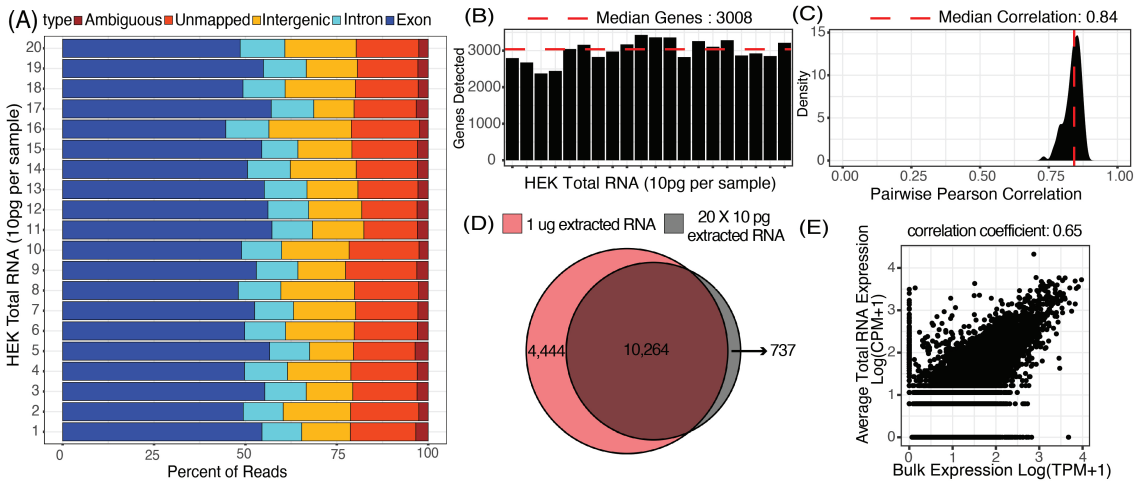

Figure 3.3: **20 libraries of 10 pg total RNA extracted from HEK293T cells were sequenced using µCB-seq.** (A) Distribution of percent exonic, intronic, intergenic, ambiguous and unmapped reads in each of the 20 libraries sequenced to an average depth of 65,000 reads per sample. (B) Number of genes detected (UMI count >0) in each of the 20 libraries subsampled to a depth of 30,000 reads per sample. (C) Distribution of correlation in gene expression profile for all possible pairs of the 20 libraries (n = 190 pairs) subsampled to a depth of 30,000 reads per sample. Pearson correlation coefficients were calculated for genes detected in at least one of the 20 libraries. (D) Genes detected in a pool of the 20 libraries for a total sequencing depth of $\sim$ 1.3 million reads (grey circle) compared with the genes detected in a bulk library (TPM > 0) prepared using 1 µg total RNA and sequenced to the same depth (red circle). (E) Scatter plot shows correlation in gene expression profile between an average 10 pg library of total RNA and the bulk library prepared using 1 µg total RNA. Pearson correlation coefficient was calculated using genes detected in either bulk sample or one of the 20 total RNA libraries.

µCB-seq offers improved gene detection sensitivity

The sensitivity of a scRNA-seq protocol can be understood as the efficiency of mRNA capture and conversion into sequenceable cDNA molecules. More practically, the number of genes detected from a single cell is commonly used as a proxy for sensitivity. Gene detection sensitivity can be reduced by many sources of inefficiency, including adsorption of molecules to reaction chamber walls, inefficient reverse transcription, and transcript loss during bead cleanup steps. When molecules are lost after PCR, the information content of the library is not reduced significantly, since each transcript has many duplicates in the pool that contain the same information. Transcript loss before PCR, however, reduces the overall library complexity and severely reduces the sensitivity of the protocol. Multiplexed plate-based scRNA-seq protocols often rely on lossy bead-based cleanup to pool and concentrate single-cell cDNA libraries after RT but before PCR, a process which necessarily loses unique cDNA molecules during bead binding and elution [13, 241, 242]. This loss of molecules before PCR

reduces the sensitivity and gene detection capability of multiplexed scRNA-seq protocols compared to their theoretical maximum. Here, we show that microfluidic library preparation allows us to improve performance of a highly sensitive protocol by eliminating post-RT bead-based pooling altogether, because cDNA only occupies nanoliter-scale volumes on-chip. We evaluated the sensitivity of scRNA-seq on the µCB-seq platform by sequencing the transcriptomes of single HEK cells and comparing the genes detected to single HEK cell libraries generated by mcSCRB-seq in a standard 0.3 mL 96-well plate (also described as in-tube, Material and methods). We prepared scRNA-seq libraries from 18 single cells on µCB-seq devices and 16 single cells using mcSCRB-seq in-tube. All libraries were sequenced to an average depth of 500,000 total reads per cell (Table A.3) and downsampled to evaluate gene detection as a function of sequencing depth. The zUMIs pipeline was used to generate the count matrix for all sequencing depths, which included only exonic reads. µCB-seq consistently detected more genes and UMIs (Fig. A.3), with significantly higher genes for depths $\geq$ 40,000 reads per cell (p-value < 0.01, two-group Mann–Whitney U-test, Fig. 3.4A). Moreover, µCB-seq libraries had a median of 21% intronic reads as compared to 15% in mcSCRB-seq (Fig. A.7) which were not counted during transcript quantification, making Fig. 3.4A a conservative estimate of the sensitivity improvements offered by the microfluidic protocol (Fig. A.4).

We further evaluated the sensitivity of µCB-seq and mcSCRB-seq in-tube by comparing gene detection efficiency as a function of transcript abundance across all expression levels. Detection efficiency was calculated as the fraction of genes detected in bulk that were also detected in a single cell for a given abundance bin. Bulk library was prepared using 1 µg total RNA extracted from HEK293T cells and sequenced to a depth of 63 million reads (Material and methods). We downsampled all µCB-seq and mcSCRB-seq libraries to 200,000 reads per cell with 16 cells in each protocol. µCB-seq detected more genes than mcSCRB-seq across all expression levels, with a substantial increase in our ability to detect low- and medium-abundance transcripts (Fig. 3.4B and 3.4C).

Next, we assessed measurement precision in the µCB-seq protocol as compared to mcSCRB-seq in-tube. Variation in gene count measurements between single-cell cDNA library preparations is caused by technical variation such as pipetting, human handling errors, and sampling statistics, as well as true biological variation between cells. With microfluidics, it is possible to minimize the technical noise by automating and parallelizing library preparation reactions in lithographically defined volumes [17, 243]. As the noise associated with technical artifacts decreases, we gain statistical power to parse out real biological variation. To quantify this, we calculated the coefficient of variation (CV) for common genes detected across bulk RNA-seq, µCB-seq, and mcSCRB-seq libraries as a function of bulk expression levels. We observed slightly lower variation in µCB-seq compared to mcSCRB-seq across the entire range of bulk expression except for very highly abundant genes (TPM $\geq$ 560, Fig. A.8). These results indicate that µCB-seq offers improved gene detection sensitivity with comparable measurement precision by eliminating lossy post-RT bead-based cleanup and carrying out library preparation in lithographically defined nanoliter-scale volumes. Furthermore, µCB-seq demonstrates similar or improved performance as compared to commercial microfluidic

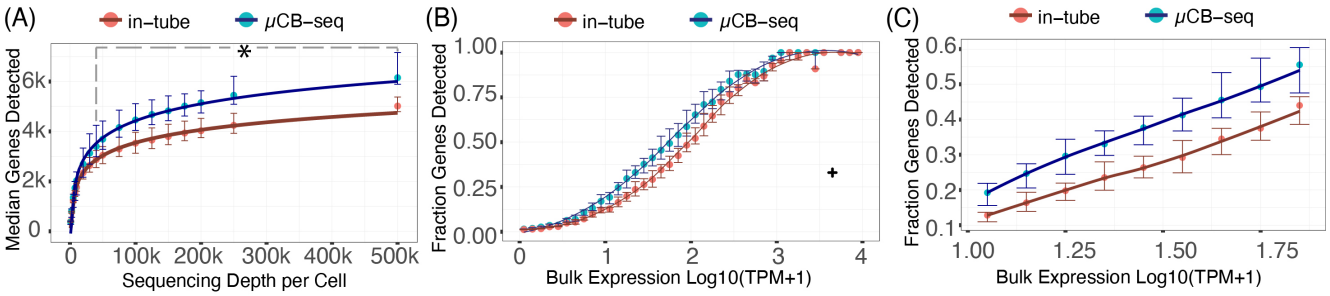

Figure 3.4: **µCB-seq is more sensitive than in-tube mcSCRB-seq protocol.** (A) Median genes detected for downsampled read depth across single HEK cells sequenced using µCB-seq and mcSCRB-seq. µCB-seq detected significantly higher genes for read depth $\geq$ 40,000 as tested by two-group Mann–Whitney U-test (p-value < 0.01). Error bars indicate the interquartile range. (B) The ratio of genes detected (UMI count >0) in the single-cell libraries subsampled to an average depth of 200,000 reads to the genes detected in the bulk library (TPM > 0) binned by expression level (bin width $= 0.1$). Bulk library was prepared using 1 µg total RNA and sequenced to a depth of 63 million reads. Error bars indicate interquartile range (n = 16 cells for each protocol). For a single bin (marked by +), only one out of three genes were detected in all single cells across both protocols and was considered an outlier. A Loess regression was used as a guide to the eye for this plot. (C) A magnified plot of panel (B) comparing the fraction of genes detected in the two protocols with low- and medium-abundance in bulk measurement ($9 <$ bulk TPM < 79).

platforms when modified to implement UMI based scRNA-seq protocol (Note A.3) [244].

µCB-seq links high-resolution optical images with the transcriptome of the same single cell

µCB-seq enables the collection of both imaging and sequencing data from single cells by associating known barcodes with microfluidic lane indices. As a proof-of-concept demonstration of µCB-seq, we captured high-resolution confocal images and sequenced the transcriptomes of single cells from a population of two differentially labeled cell types. We stained HEK293T cells and human adipocyte precursor cells (preadipocytes) [76] with CellBrite Green and Red cytoplasmic membrane dyes respectively (Material and methods). The cells were then suspended and processed in three µCB-seq devices. One device processed a mix of both HEK cells (n = 4 cells) and preadipocytes (n = 3 cells). The other two devices processed just HEK cells (n = 7 cells), or just preadipocytes (n = 6 cells) separately. Fluorescence confocal imaging was performed while cells were isolated in the imaging chambers using 488 nm and 633 nm lasers. The µCB-seq device is mounted on a #1.5 coverslip (170 µm thickness) and has a 50 µm thick dummy layer, for a total of 220 µm distance between the focal plane and the objective. We used a high-magnification (63x, 0.7 NA) air objective to enable high-resolution imaging (Material and methods). After imaging, the cells were ejected into their respective

reaction lanes for library preparation on-chip followed by pooled PCR. These 20 libraries were sequenced to saturation in order to characterize the sensitivity of µCB-seq (Material and methods). After sequencing, we demultiplexed reads based on their cell barcodes, which allowed us to assign each cDNA read to the lane index and thus to the image of the cell from which the molecule originated. In this analysis, both intronic and exonic reads were used for generating a count matrix to utilize the introns detected by µCB-seq. Fig. 3.5A displays

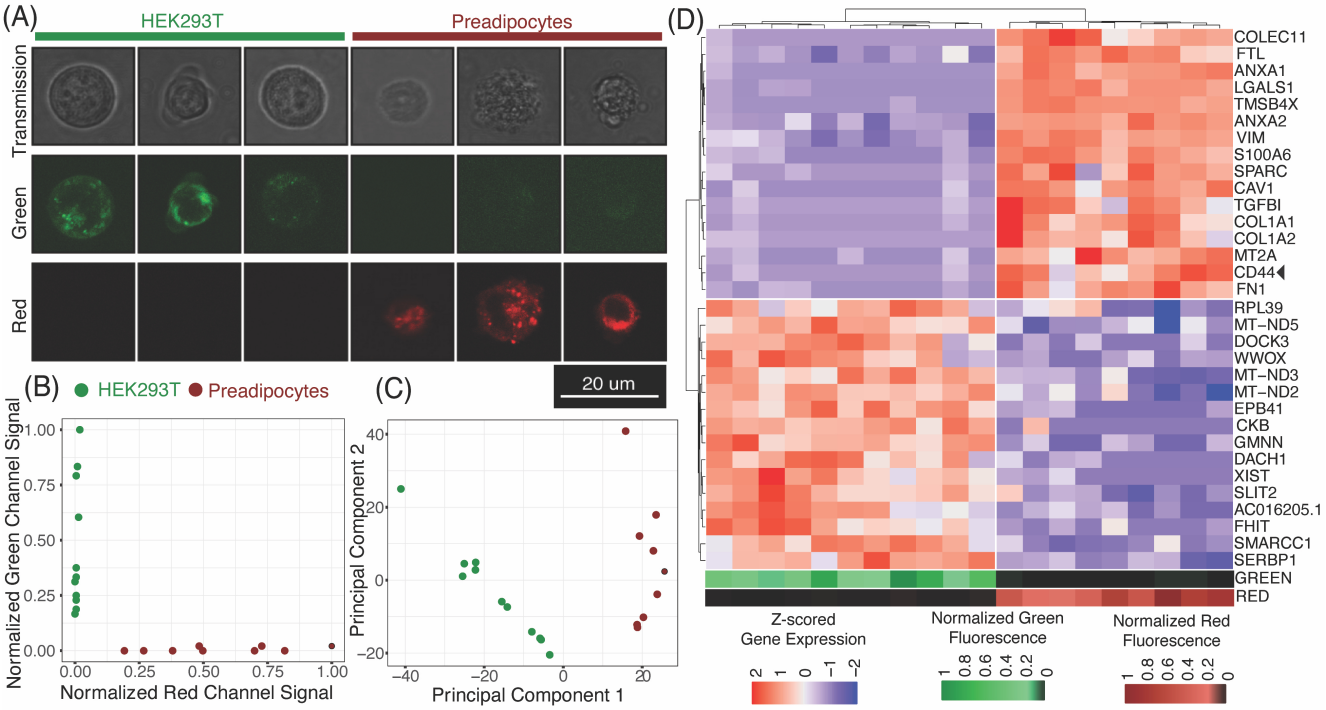

Figure 3.5: **Linked imaging and sequencing using µCB-seq** (A) montage of representative images of HEK cells and preadipocytes acquired using scanning transmission and scanning confocal microscopy in the green and red channels. HEK cells and preadipocytes were stained with CellBrite Green and Red cytoplasmic membrane dye respectively. (B) Normalized fluorescence signal in the green and red channel confocal images of both HEK cells and preadipocytes. Analysis of images for cell-mask generation and quantification of fluorescent intensities is explained in the Material and methods section. (C) Accurate identification of HEK cells and preadipocytes as two cell populations using unsupervised hierarchical clustering in the principal component space. Top 2000 most variable features were used as an input for determining the first two principal components. (D) Unsupervised hierarchical clustering using scaled expression values of top-16 upregulated genes in HEK cells and preadipocytes. Heat map shows z-scored expression values for the 32 genes. On the bottom are heat map visualizations of normalized fluorescence intensities plotted in panel (B). The heat maps for the green and red channels are ordered to accurately reflect a one-on-one correspondence between imaging and sequencing data points.

representative transmission and scanning-confocal images of HEK cells and preadipocytes in

both green and red channels confirming differential labeling of the two cell types (Fig. A.6). We estimated the spatial resolution of acquired confocal fluorescent images to be 959 nm on average, by performing decorrelation analysis (Material and methods, Fig. A.10). With this resolution subcellular features can be reliably resolved (Fig. 3.5A, A.6). With this magnification, transmission images revealed a distinct texture for preadipocytes as compared to HEK cells (Fig. A.6). We quantified the textural features in individual transmission images by calculating the correlation and variance of grayscale intensities54 and observed that these two features partially separated preadipocytes and HEK cells (Material and methods, Fig. A.11). These results demonstrate that µCB-seq allows for high-resolution imaging, which provides the potential to draw connections between subcellular features and gene expression.

Using distinct fluorescent stains on HEK cells and preadipocytes allowed us to determine the cell type of each captured cell prior to sequencing-based analysis. As expected, quantification of the fluorescence signal in the green and red channels completely separated the two cell-types along those two axes (Fig. 3.5B, Material and methods). Groups of HEK cells and preadipocytes identified using image analysis also presented as two distinct cell populations upon unsupervised clustering in the principal component space (Fig. 3.5C, Material and methods). No technical artifacts associated with the three different devices were observed in the reduced space (Fig. A.12). In this case, µCB-seq optical imaging serves as a ground truth for naïve clustering of transcriptomic data from the same cells.

We further analyzed the sequencing dataset to understand the transcriptomic variations in this heterogeneous group of 20 cells. Differential gene expression analysis revealed 103 genes with logFC > 0.5 and adjusted p-value < 0.05 (Material and methods). Interestingly, preadipocytes had an enriched expression of CD44, a mesenchymal stem cell surface marker which has been suggested to be expressed in adipogenic cells [245, 246]. We also performed unsupervised hierarchical clustering on the expression levels of the top 16 upregulated genes in each of the two cell types. All twenty cells were sorted into two distinct groups that accurately reflected their known cell type (Fig. 3.5D). These data demonstrate that µCB-seq can successfully pair high-sensitivity gene expression profiles with high-resolution fluorescence images from single cells.

3.3 Conclusion

Microfluidic technologies have been at the core of the recent exponential increase in the throughput of scRNA-seq techniques, paving the way for undertakings such as the Human Cell Atlas project [35]. However, because scRNA-seq can only record information encoded as a sequence of nucleotides, orthogonal measurements enabled by quantitative live-cell imaging, such as fluorescence staining, subcellular lipid quantification [184], or organelle-level pH measurements [247], will play an important role in the generation of a comprehensive human cell atlas. In this report, we present µCB-seq, a scalable microfluidic platform which allows us to acquire high-resolution images and generate RNA-sequencing libraries from the same single cells. µCB-seq links optical and genomic measurements with known barcodes, which

are pre-delivered to addressable locations on-chip and recovered with high efficiency during device operation, even after fabrication at 80 °C. By preloading barcoded primers to reaction chambers, the µCB-seq fabrication process obviates the need for complex fluidic routing of multiple barcoded reagents. By combining the final reagent outlets to pool all single-cell libraries on-chip, the µCB-seq device can easily be scaled up to process hundreds of cells with only two inlet and two outlet ports. The device architecture needed to scale µCB-seq to this throughput has been readily demonstrated in both academic3 [18, 228, 229] and commercial [231, 232] microfluidic platforms. This increased throughput can be achieved by using a microfluidic multiplexing strategy which requires only a minimal increase in the peripheral operating equipment [248, 249]. Additionally, high-precision, low-volume array spotters can be used to automate barcode preloading, enabling throughput at the level of existing commercial devices with a far simpler microfluidic circuit. Due to its ability to pool all cells and perform a single off-chip library preparation step, implementation of the µCB-seq barcoding strategy in commercial platforms could significantly reduce the cost per cell of sequencing library preparation (Supplemental Table 1). Ultimately, the throughput of linked imaging and sequencing measurements by µCB-seq will be limited by imaging time. Automated stage-scanning can be implemented in µCB-seq to reduce imaging time, as cells are immobilized in a linear array of nanoliter-scale imaging chambers. µCB-seq devices have a modular microfluidic circuit design allowing for the implementation of other multistep scRNA-seq library preparation protocols on-chip. µCB-seq's ability to correlate optical measurements with gene expression on the single-cell level has the potential to provide insight into the relationship between genome regulation and cellular phenotypes. While this scRNA-seq demonstration uses a single barcoding step, we believe our µCB-seq barcoding approach may prove useful for many-step reactions in which aqueous samples can be automatically directed to multiple preloaded chambers for combinatorial spatial barcoding [250], targeted gene expression [30], or CRISPR-based gene editing [251].

By using a microfluidic approach in µCB-seq for library preparation, we have eliminated post-RT bead-based cleanup, minimized operational errors, and achieved nanoliter-scale, reproducible reaction volumes. Our microfluidic approach offers improvements in sensitivity, as demonstrated by an increased gene detection efficiency. Using µCB-seq, we were also able to effectively reconstruct a large portion of the bulk transcriptome by sequencing 200 pg total RNA to a total depth of ∼1.3 million reads. The integration of on-chip valves in the device allowed us to actively select cells of interest, making the µCB-seq platform applicable for studies that focus on rare cell populations [252]. On-chip isolation valves prevent cellular motion due to fluid flow, thereby allowing the acquisition of even prolonged spectroscopic measurements [253] on our device. Compatibility of µCB-seq with a standard inverted microscope configuration enables the implementation of any single-objective imaging technique with working distance of 220 µm, such as coherent Raman scattering microscopy [184] or super-resolution microscopy [254]. For example, µCB-seq could be paired with super-resolution microscopy to investigate phase separation of super-enhancers and its effect on gene expression across the whole transcriptome of individual cells [255]. Another implementation could pair µCB-seq with microfluidic DamID [256] to investigate the bidi-

rectional interplay between gene expression and chromatin organization in the same single cell. We believe the µCB-seq platform will be a powerful tool for investigations aiming to understand the association between a phenotype and the transcriptome, thereby gaining a high-resolution fingerprint for a particular cell population identified using other higher-throughput scRNA-seq protocols.

3.4 Materials and Methods

HEK293T cell culture and single-cell suspension preparation

HEK293T cells were obtained from the UCSF cell repository, and cultured in DMEM medium (Gibco, 10566-016) supplemented with 10% vol/vol FBS and containing 1% vol/vol penicillin–streptomycin (Gibco). The cell culture was maintained at 37 °C in a humidified incubator containing 5% vol/vol CO_2. Confluent cells were passaged using TrypLE (Gibco, 12563011) with a 1:25 split in a new T25 flask (Falcon, 353109). For generating HEK293T single-cell suspensions for µCB-seq vs. mcSCRB-seq comparisons (Fig. 3.4), cells were first grown to 100% confluence. The cells were then resuspended in 1 mL TrypLE and 5 mL of growth media and centrifuged at 1200 rpm for 4 min. After centrifugation, the supernatant was removed and the cell pellet was washed with 1 mL of PBS (Corning, 21-040-CV). The cells were centrifuged again and this process was repeated for a total of three PBS washes to remove cell debris. Finally, the concentration of the cell suspension was adjusted in ice-cold PBS to 700 cells per µL using a hemocytometer (Hausser Scientific). After this, the cell suspension was always stored on ice throughout the course of device operation. In most experiments, around 50 µL of the single-cell suspension was aspirated into a gel-loading pipette tip and placed into the device, although the full volume was rarely completely used, and it is possible to decrease this volume in situations where the sample is limited.

Preadipocyte cell culture

Human preadipocytes were provided by our collaborators in the Tseng lab at Joslin Diabetes Center at Harvard. The cells were isolated from the deep neck region of a deidentified individual using the protocol in Xue et al. and immortalized to allow for cell culture and expansion [76]. For culturing, preadipocytes were grown in DMEM medium (Corning, 10-017-CV) supplemented with 10% vol/vol FBS and containing 1% vol/vol penicillin–streptomycin (Gibco). The cell culture was maintained at 37 °C in a humidified incubator containing 5% vol/vol CO_2. 80% confluent cells were passaged using 0.25% trypsin with 0.1% EDTA (Gibco; 25200-056) for a 1:3 split in a new 100 mm cell culture dish (Corning).

HEK293T and preadipocyte membrane staining protocol

HEK293T cells and preadipocytes were stained with CellBrite Green (30021) and Red (30023) cytoplasmic membrane labeling kits respectively using manufacturer's protocol. Briefly, cells were suspended at a density of 1,000,000 cells per mL in their respective normal growth medium. 5 µL or 10 µL of the cell labeling solution was then added per 1 mL of cell suspension for HEKs and preadipocytes respectively. Cells were then incubated for 20 minutes (HEKs) or 40–60 minutes (preadipocytes) in a humidified incubator containing 5% vol/vol CO_2. Cells were then pelleted by centrifugation at 1200 rpm for 4 min. After centrifugation, the supernatant was removed, and cells were washed in warm (37 °C) medium. Cells were centrifuged again, and the process was repeated for a total of 3 growth medium washes for HEKs and 1–3 growth medium washes for preadipocytes. Cells were then centrifuged a final time at 1200 rpm for 4 minutes and resuspended in ice-cold PBS (Corning, 21-040-CV) for a final concentration of 700 cells per µL adjusted using a hemocytometer (Hausser Scientific). The cells were then stored on ice throughout the µCB-seq device operation.

Bulk RNA-sequencing and data analysis: Credit Annie Maslan

Total RNA was extracted from HEK293T cells using the RNeasy Mini Kit from Qiagen (74104) with the QIAshredder (79654) for homogenization. RNA library preparation was performed with 1µg of total RNA input quantified by Qubit fluorometer using the NEBNext poly(A) mRNA magnetic isolation module (E7335S) followed by NEBNext Ultra II RNA Library Prep Kit for Illumina (E7770S). Paired-end 2 × 150 bp sequencing for the bulk library was performed on the Illumina Novaseq platform for a coverage of approximately 63 million read pairs. For analyzing the dataset, adapters were first trimmed using trimmomatic [257](v0.36; ILLUMINACLIP:adapters-PE.fa:2:30:10 LEADING:3 TRAILING:3 SLIDINGWINDOW:4:15 MINLEN:36, where adapters-PE.fa is:
>PrefixPE/1 TACACTCTTTCCCTACACGACGCTCTTCCGATCT
>PrefixPE/2 GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCT).

After trimming, reads were then aligned to the GRCh38 index generated using STAR. We provided the GTF file that is recommended for the 10X CellRanger pipeline as an input in STAR while generating the index. Paired-reads aligning to the exonic regions were then quantified using the featurecounts command in the Subread package. Chimeric reads and primary hits of multi-mapping reads were also counted towards gene expression levels. The same GTF file as in STAR was used as the input for transcript quantification. The fragment-counts matrix so obtained was converted to transcripts per kilobase million mapped reads (TPM) using the lengths for each gene as calculated by the featurecounts command in the Subread package. For analysis in Fig. 3.3, reads were subsampled to a depth of 1.3 million reads using the Seqtk package (v1.3) [258]. These subsampled reads were then analysed in the exact same fashion as described above.

Confocal imaging of HEK293Ts and Preadipocytes

Fluorescence confocal imaging of cells was performed in the imaging chamber of the µCB-seq device using an inverted scanning confocal microscope (Leica, Germany), and with a 63x 0.7 NA long-working-distance air objective. As outlined before, HEKs were stained using CellBrite Green dye and preadipocytes were stained using CellBrite Red dye. Each cell was excited by two continuous-wave lasers, a 488 nm Ar/Kr laser and a 633 nm He/Ne laser, for concurrent imaging in the green and red channels respectively. Bandpass filters captured backscattered light from 490–590 nm at the photomultiplier tube in the green channel (Green-PMT), and from 660–732 nm at the photomultiplier tube in the red channel (Red-PMT), with the pinhole set to 1 Airy unit. A third PMT simultaneously captured a scanning transmission image using the unfiltered forward-scattered light. The imaging resolution was Rayleigh-limited, with a scanning zoom of 2.2x to achieve a Nyquist sampling rate of 207 nm per pixel (as calculated for the Ar/Kr laser with a shorter wavelength). Each image was 8-bit, grayscale and 512 × 512 pixels in size. Since individual HEK cells and preadipocytes internalized varying amounts of membrane stain, the PMT gain which utilized the entire range of bit-depth (0–255) differed from one cell to another. Therefore, stained HEK and preadipocyte cell suspensions were first imaged on a #1.5 coverslip for adjusting the range of Green-PMT gain (range: 524.6) and Red-PMT gain (range: 512–582). We measured a maximum gain of 524.6 in the green channel and 582 in the red channel to observe cellular features, and therefore set the background PMT gain to an even higher value of 600, to validate that lack of features in background images was not because of low PMT gain. In all our images, the focal plane was positioned at the cross-section with maximum fluorescence intensity. The final images were Kalman-integrated over 6 frames to remove noise. Images in Fig. 3.5A have been adjusted to highlight cellular features. However, no adjustment was done for quantitative image processing.

Spatial resolution quantification of confocal fluorescent images

To quantify the spatial resolution of confocal fluorescent images, we implemented decorrelation analysis [259] using the image-decorrelation-analysis plugin [260] on ImageJ (v2.0.0). For analysis on ImageJ, unsaturated confocal images (with maximum pixel intensity ¡255) were first cropped to frame the cell in the region of interest. The resolution was then computed with the cropped images as input to the image-decorrelation-analysis plugin, using these settings: radius-min = 0, radius-max = 1, Nr = 50, and Ng = 10. The median resolution across 18 images was 959 nm (Fig. A.10).

Texture analysis of brightfield images

To quantify the correlation and variance of grayscale intensities in the brightfield images, we used the Measure-Texture module of CellProfiler (v3.1.9) [261]. In this module, correlation and variance are image parameters that were calculated as defined by Haralick et al [262].

For analysis, the brightfield images were first cropped to frame the cell in the region of interest using ImageJ (v2.0.0). Correlation and variance were then computed with cropped images as the input to the Measure-Texture module, and scale was set to 2 pixels.

Image processing for fluorescence signal quantification

To quantify the fluorescence signal intensity in individual HEKs and preadipocytes labeled using the CellBrite Green and Red dye respectively, we wrote a custom image analysis script in Python (v3.7.1) using the skimage package (v0.20.2) and multi-dimensional image processing (ndimage) package from the SciPy (v1.2.1) ecosystem. As explained in the confocal imaging section above, each cell had two fluorescence images, one green-channel confocal image, and one red-channel confocal image. Depending on the cell-type, one of the channels exhibited cellular signal (green for HEK and red for preadipocytes) and the second channel conversely was a control image. For images of individual HEK cells and preadipocytes, all green-channel and red-channel images respectively were analyzed to generate a cell mask (as detailed below). The pixels constituting the cell mask were designated as foreground pixels and the remaining pixels were designated as background pixels. The fluorescence signal to noise ratio (SNR) was then quantified as the ratio of mean foreground pixel intensity over mean background pixel intensity. The same pixel annotation (for foreground and background pixels) was also used in the control images to quantify SNR in the second channel. In essence, we quantified the SNR in both green and red channels for each cell and these values were normalized to linearly scale between 0 and 1 for Fig. 3.5B and 3.5D. For cell mask generation, grayscale images were first Gaussian filtered to remove noise using the ndimage.gaussianfilter command with sigma set as 1. The filtered images were converted into binary images using Otsu thresholding from the skimage package. Pixels with value 1 in the binarized images were annotated as foreground and pixels with value 0 were annotated as background (Fig. A.6).

Principal component analysis, clustering and differential gene expression analysis

Single HEK cells and preadipocytes were sequenced on the MiniSeq platform to an average depth of 346,000 reads per cell (Table A.4). For consistency, reads per cell were downsampled to 125,000 reads across both cell types. For membrane-stained HEK cells and preadipocytes, principal component analysis (PCA), clustering, and differential gene expression analysis were performed using the Seurat package (v3.1.1) [263] in the R programming language (v3.5.2). First, the umi-count matrix generated using zUMIs at a read depth of 125,000 reads per cell was read using the readRDS command. The count matrix was then used to create a Seurat object with no filtering for either cells or genes. The umi-count matrix was log-normalized with a scaling factor of 10,000 using the NormalizeData command. The top 2000 most variable genes in the full dataset were identified using the variance-stabilizing

transformation (vst) method implemented by the FindVariableFeatures command. The normalized count matrix was then scaled and centered to generate the Z-scored matrix using the ScaleData command. The first and second principal components were then calculated based on the Z-scored expression values of the 2000 variable genes using the RunPCA command and the reduced space visualization was plotted using the ggplot2 package (v3.1.0) in R.

For clustering using Seurat, first, a K-nearest neighbor graph (KNN) was constructed using the cell embeddings in the PCA space (K = 5). The generated KNN graph was then used to construct a shared nearest neighbor (SNN) graph by calculating the Jaccard index between every cell and its nearest neighbors using the FindNeighbors command. Using the SNN graph, the clusters were then identified using the FindClusters command with the resolution parameter set to 0.1. At this resolution, HEKs and preadipocytes separated into two clusters as visualized in the PCA space (Fig. 3.5C). After clustering, differentially expressed genes (logFC $>$ 0.5 and adjusted p-value $<$ 0.05) between the two clusters were identified by fitting a negative binomial generalized linear model (negbinom test) on the raw umi-count matrix as implemented in the FindAllMarkers command. Z-scored expression values of the top 16 upregulated genes for each cell-type were then color mapped in a Heatmap plot using the ComplexHeatmap package [264]. ComplexHeatmap was also used to perform unsupervised hierarchical clustering of single cells and genes using the Euclidean distance metric and complete linkage classification method. Imaging heatmaps, with normalized green- and red-channel fluorescence signal as the data points, were also plotted using the ComplexHeatmap package.

Control and flow mold fabrication

Two molds, a control mold and a flow mold, were patterned on silicon wafers (University Wafers, S4P01SP) with photolithography. Patterns for the control and flow molds were designed in AutoCAD (Autodesk) and printed onto 25,400 dpi photomasks (CAD/Art Services, Inc., Bandon, Oregon). The silicon wafers were first thoroughly cleaned using acetone, isopropyl alcohol, and water. The wafers were then baked at 150 °C for 10 min to dehydrate the surface. For the control mold, a 5 µm dummy layer of SU8-2005 (MicroChem) was first spin-coated at 3000 rpm for 30 s. The resist-coated mold was then baked at 65 °C for 1 min and 95 °C for 2 min and exposed to UV radiation with no mask for 10 s. After exposure, the mold was again baked at 65 °C for 1 min and at 95 °C for 3 min and allowed to cool to room temperature. After dummy layer deposition, a dollop of SU8-2025 negative photoresist (MicroChem) was poured onto the control mold directly and then spun at 3000 rpm for 30 s, yielding a 25 µm layer. Then, the wafer was baked on a hotplate at 65 °C for 1 min and then at 95 °C for 5 min. The resist-coated wafer was exposed to a 150 mJ cm^2 dose of UV radiation through a negative mask (clear features and opaque background) imprinted with the control circuit using a photolithography aligner. After exposure, the wafer was again baked at 65 °C for 1 min and 95 °C for 5 min. The wafer was then submerged in SU-8 developer and gently agitated until the unexposed photoresist was removed, leaving the positive control features. Then, the wafer was carefully washed with isopropyl alcohol and blow-

dried. The mold was baked at 150 °C for at least 20 min before further use. The flow mold was fabricated using two photoresists to achieve multiple feature heights. The flow channels were fabricated using the positive photoresist AZ 40XT-11D (Integrated Micro Materials, Argyle, TX) and the taller reaction chambers were fabricated using the negative SU8-2025 photoresist. The flow mold was first spin-coated with a 5 µm dummy layer of SU8-2005 and processed the same as described for the control mold above. After dummy layer deposition, a dollop of AZ 40XT-11D positive photoresist was poured onto the flow wafer directly and then spun at 3000 rpm for 30 s, yielding a 20 µm layer. After baking at 65 °C for 1 min and 125 °C for 6 min, the photoresist was then exposed to a 420 mJ cm^2 dose of UV light through a high-resolution positive mask containing the flow circuit design and developed in AZ400K developer. We then baked the mold again at 65 °C for 1 min and at 105 °C for 100 s to reflow the positive photoresist and create rounded channels. Negative photoresist (SU8-2025) was then used for building the reaction chambers using the same protocol as described for the control mold above.

PDMS device fabrication

Each layer of the multilayer µCB-seq device was bonded together by on-ratio (10:1) bonding of RTV-615 (GE Advanced Materials). The control and flow molds were exposed to chlorotrimethylsilane (Sigma-Aldrich) vapor for 30 minutes before soft lithography to facilitate PDMS releasing from the mold. After mixing and degassing of PDMS, 50 g of PDMS was cast onto each control mold and baked at 80 °C for 15 min to partially cure the PDMS slabs. Control ports were punched and flow molds were spin-coated with a PDMS layer at a speed of 2000 rpm for 60 s. Flow layers were partially cured at 80 °C for 5 min, after which control slabs were aligned and placed atop flow PDMS. PDMS assemblies were cured at 80 °C for a further 10 min, after which devices were peeled off of the Si wafer. Flow ports were punched, and assemblies were placed upside-down in preparation for primer spotting. In a clean hood, 0.2 µL of 1.5 µM barcoded RT primer was manually spotted in lysis chambers using a P2 pipette, with each lane receiving a unique, known barcode sequence (Table A.1). Primers were allowed to dry while a PDMS dummy layer was spin-coated and partially cured on a blank, silanized Si wafer. Control + flow-layer PDMS assemblies were then placed onto the PDMS dummy layer for a 1.5 h hard bake at 80 °C. Final devices were bonded to 1.5 glass coverslips by O_2 plasma (PETS Inc.) and placed at 4 °C for storage.

Microfluidic device operation

Microfluidic devices were attached to an Arduino-based pneumatic controller (KATARA) in preparation for running on-chip library preparation. Prior to single-cell experiments, the cell trapping line was flushed with nuclease-free water (nfH2O) and incubated with 0.2% (wt/wt) Pluronic F-127 (Invitrogen, P6867) for 1 h, leaving downstream chambers containing barcoded primers empty. A single cell suspension was prepared and drawn into the cell trapping line by peristaltic pumping action of the integrated microfluidic valves.

Triton buffer was first prepared by combining 0.2 µL RNase inhibitor (40 U μL^{-1}, Takara 2313A) and 3.8 µL 0.2% (v/v) Triton X-100 (Sigma, X-100). Lysis buffer was then prepared by mixing 1 µL 1:100 5x Phusion HF buffer (NEB, B0518S) 2.5 µL Triton buffer, 0.7 µL nfH2O, and 0.8 µL 1% (v/v) Tween 20 (Sigma, P7949) in a 0.2 mL PCR tube. Lysis buffer was aspirated into a gel-loading pipette tip, which was inserted into the reagent inlet and pressurized. The reagent tree was dead-end filled with lysis buffer, and the device was transferred to a confocal microscope (Leica) for cell trapping and imaging.

Cells were drawn along the cell input line by the peristaltic pump and manually trapped in the imaging chamber for imaging, which was carried out by the protocol described in confocal imaging. After imaging, the lane's reagent valves were opened, allowing lysis buffer to push the trapped cell into the lysis module containing dried, uniquely barcoded RT primers. After the dead-end filling of the lysis module, primers were resuspended by pumping action of the microfluidic paddle above the lysis chamber. The microfluidic device was transferred to a thermal block for cell lysis at 72 °C for 1 min, after which the block was cooled to 4 °C. During cooling, the reagent inlet was flushed with 20 µL nuclease-free water and dried with air. Reverse transcription mix was then prepared in a 0.2 mL tube by mixing 0.8 µL 25 mM each dNTP mix (Thermo Fisher, R0181), 4 µL 5x Maxima H- buffer (Thermo Fisher EP0751), 0.4 µL 100 µM E5V6 TSO (Table A.5), 5 µL 30% PEG 8000 (Sigma Aldrich, 89510-250G-F), 6.4 µL nfH2O, 0.2 µL 1% Tween 20, and 0.2 µL 200 U μL^{-1} Maxima H- Reverse Transcriptase (Thermo Fisher EP0751). Reverse transcription mix was injected into the reagent inlet to dead-end fill the reagent tree. The isolation valves were then closed and reagent valves were opened to allow the RT mix to dead-end fill all lanes. Reverse-transcription was carried out for 90 min at 42 °C, with the peristaltic pump operating at 1 Hz to accelerate diffusive mixing of cell lysate, reverse transcription mix, and barcoded primers. Following reverse transcription, the chip was cooled to 4 °C and the reagent inlet was washed and dead-end filled with nuclease-free water. Barcoded cDNA was eluted in a volume of 1.7 µL per lane into gel loading pipette tips and pooled in a single PCR tube for downstream single-pot reactions.

Exonuclease digestion was carried out on the 17 µL of pooled library by adding 2 µL exonuclease buffer (10x) and 1 µL 20 U μL^{-1} ExoI (Thermo Fisher, EN0581), with no concentration steps required, followed by incubation at 37 °C for 20 min, 80 °C for 10 min, and cooling to 4 °C. Following exonuclease digestion, the following reagents were added to the library tube for PCR: 1.5 µL 1.25 U μL^{-1} Terra direct polymerase (Clontech, 639270), 37.5 µL 2x Terra direct buffer, 1.5 µL 10 µM SINGV6 primer (Table A.5), and 14.5 µL nfH2O. PCR was carried out with the following protocol: 3 min at 98 °C followed by 17 cycles of (15 s at 98 °C, 30 s at 65 °C, 4 min at 68 °C), followed by 10 min at 72 °C and a 4 °C hold. Post-PCR libraries were size-selected with AmPure XP beads (Beckman Coulter, A63880) using a 0.6:1 beads:library volume ratio. Final libraries were run through the Nextera XT tagmentation protocol (Illumina), with the PNEXTPT5 custom primer (Table A.5) substituted for the P5 index primer as in mcSCRB-seq. Indexed libraries were pooled and sequenced on an Illumina MiniSeq platform.

mcSCRB-seq in-tube library preparation

For mcSCRB-seq in-tube experiments, 96-well plates were first prepared with 10 barcoded primers and lysis buffer according to the mcSCRB-seq protocol, with the only difference being the use of µCB-seq RT primers instead of standard mcSCRB-seq ones. For single HEK cell experiments, the CellenONE X1 instrument was used to individually deliver a single HEK cell into each well. Following cell delivery, the mcSCRB-seq protocol was followed directly, but with a 1:1 ratio of AmPure XP beads to pool all cDNA after RT as opposed to the manual bead formulation from standard mcSCRB-seq. After library preparation, HEK single-cell mcSCRB-seq libraries were sequenced on the NovaSeq platform to an average depth of 500,000 reads per cell.

HEK single-cell and HEK total RNA sequencing data processing

HEK single-cell and total RNA libraries were sequenced on the MiniSeq platform to an average depth of 500[thin space (1/6-em)]000 and 65[thin space (1/6-em)]000 reads per sample respectively (Table A.2 and A.3). Filtering, demultiplexing, alignment, and UMI/gene counting were carried out on the zUMIs pipeline for all samples, using the GRCh38 index for STAR alignment. We provided the GTF file that is recommended for the 10X CellRanger pipeline for standardization of gene counts. Reads with any barcode or UMI bases under the quality threshold of 20 were filtered out, and known barcode sequences were supplied in an external text file. UMIs within 1 hamming distance were collapsed to ensure that molecules were not double-counted due to PCR or sequencing errors. For this analysis, cell barcodes were not collapsed based on their hamming codes. For the Total RNA µCB-seq dataset (TC012), the quality of the 3rd base of read 1 was poor due to the fact that all barcodes in the sequencing run had an Adenine at that position. Therefore, fastq files for this dataset were edited to remove the third base, and truncated barcode sequences were provided to zUMIs to match. This modification did not affect the information content or quality of the processed library.

For comparison, all HEK total RNA libraries were subsampled to 30,000 reads (Fig. 3.3B and 3.3C). For benchmarking against bulk RNA-seq library, all the reads across all samples were pooled together resulting in a total of approximately 1.3 million reads for the analyses (Fig. 3.3D and 3.3E; A.2).

3.5 Data Access

Yaml files for zUMIs analysis of HEK Total RNA, single HEK cells and single HEK and preadipocyte datasets are provided in the streetslab GitHub repository. Downstream data tidying and analysis was carried out in a Jupyter notebook with an R kernel, which can also be found in the repository. The CAD file with µCB-seq device design can be downloaded from the same GitHub repository.

3.6 Declaration of Interests

There are no conflicts to declare.

3.7 Acknowledgements

The authors would like to thank Prof. Yu-Hua Tseng for providing adipocyte precursors. This publication was supported by the National Institute of General Medical Sciences of the National Institutes of Health under award number R35GM124916. AG is supported by the UC Berkeley Lloyd Fellowship in Bioengineering. AS is a Chan Zuckerberg Investigator. Tyler N. Chen contributed equally to this work. Credits to Annie Maslan for Bulk RNA-seq library preprataion. This work is now published as Chen*, Gupta* et al. 2020:
Chen*, Gupta* et al., "µCB-seq: microfluidic cell barcoding and sequencing for high-resolution imaging and sequencing of single cells". *Lab Chip*, 20 (2020), pp. 3899-3913.

Chapter 4

Characterization of transcript enrichment and detection bias in single-nuclei RNA-seq for mapping of distinct human adipocyte lineages

4.1 Introduction

In the last two chapters, I focused on technological developments that would enable investigations into adipocyte sub-type discovery. Another key aspect of adipose tissue biology in maintaining system-wide energy balance is its expansion via adipogenesis. As outlined in Chapter 1, the regulation of whole-body energy homeostasis is primarily maintained by two functionally different types of fat: white adipose tissue (WAT), the primary site of lipid storage, and brown adipose tissue (BAT), which specializes in thermogenic energy expenditure. An imbalance in the expansion of WAT and BAT is implicated in the emergence of varied metabolic syndromes such as lipodystrophy or obesity and associated comorbidities like cardiovascular diseases and type 2 diabetes. Therefore, understanding the molecular pathways of adipose tissue expansion (adipogenesis) in humans is necessary for understanding the tissue's contribution in the pathology of such metabolic diseases.

As discussed previously, scRNA-seq has proven to be a powerful tool for transcriptomic profiling of complex tissues in an unbiased manner [25, 265, 266]. This technological revolution has been facilitated by the development of droplet-microfluidics- and micro-well-based workflows for scRNA-seq that make it possible to analyze hundreds to thousands of single cells in one experiment [25]. Indeed, multiple recent studies using droplet-microfluidic scRNA-seq approaches are investigating the heterogeneity of primary preadipocytes in mice [90, 267]. However, applicability of such high-throughput microfluidic approaches is limited in mature adipocytes, since adipocytes can easily rupture in microchips and during droplet formation due to their fragile nature. Such fragility further makes it challenging to generate a

single-cell suspension of intact adipocytes at the first place, when starting from *primary* adipose tissue samples. Existing protocols for adipose tissue digestion and single-cell suspension preparation often result in complete or partial adipocyte lysis and therefore are not compatible with scRNA-seq library preparation. Consequently, transcriptomic analysis of primary adipocytes has relied on bulk RNA-sequencing of clonal cell populations [76, 268–271] or scRNA-sequencing of adipocytes harvested by precise pipetting [54], making generation of individual clones or isolates the rate limiting step. More recently, microfluidic scRNA-seq was used to identify transcriptomic heterogeneity within murine brown adipocytes [53], with library preparation limited to adipocytes relatively smaller in size as bigger adipocytes can easily rupture during cellular encapsulation. Such size-fractionated application of scRNA-seq, however, results in loss of transcriptional patterns uniquely associated with bigger adipocytes [272]. To address the challenge of working with fragile tissues, recent studies have turned to single-nucleus RNA-sequencing (snRNA-seq) as an alternative approach for transcriptomic profiling of cellular heterogeneity. Indeed, investigations have already started reporting the existence of multiple adipocyte subtypes in humans using snRNA-seq [94, 273]. However, a single nucleus contains 10-100-fold less mRNA than whole-cells, raising the question whether the composition of mRNA transcripts in the nucleus is sufficient to enable identification of the same cell populations as whole-cells. Previous comparisons of single-cell and single-nucleus approaches suggest that in certain tissues, sampling the nuclear transcriptome is sufficient to characterize cellular composition [93, 274–277]. However, collectively these studies also demonstrate that the relationship between nuclear and cytoplasmic mRNA is tissue-specific [277, 278]. Therefore, there is a need to understand the transcriptomic similarities and differences between single-cell and single-nucleus profiles in the context of the human adipose tissue, for which there is growing need to rely on snRNA-seq.

In this study, we explored the ability of snRNA-seq to recapitulate the transcriptional profiles observed by scRNA-seq in the human adipose tissue white and brown lineages. We focused our study on a well-controlled *in vitro* system of human white and brown adipogenesis [76, 279] (4.1A). In this *in vitro* model, paired white and brown primary preadipocytes were isolated from a defined anatomical location (the neck depot) of a single individual. This system allowed us to measure cell-to-cell transcriptomic variations within and between lineages, while controlling for inter-individual variabilities that are typically associated with transcriptomic profiling of primary human adipose tissue, such as body mass index, genotype, and gender. Preadipocytes from both lineages were isolated while preserving their intrinsic cellular heterogeneity and were then immortalized to allow for long-term *in vitro* cell-culture. Previously reported data demonstrated that the preadipocyte populations could be differentiated into mature adipocytes with gene expression profiles that correspond to the adipogenic and thermogenic function of primary tissue from human neck BAT and WAT [76]. Moreover, the *in vitro* cell-culture system allows for isolation of intact nuclei as well as intact single cells across well-defined stages of adipogenesis including mature, lipid-laden white and brown adipocytes. Using this system, we first mapped the cellular heterogeneity at the preadipocyte stage. Both white and brown preadipocytes were processed using a commercial high-throughput single-cell sequencing platform (10x Genomics). We then ex-

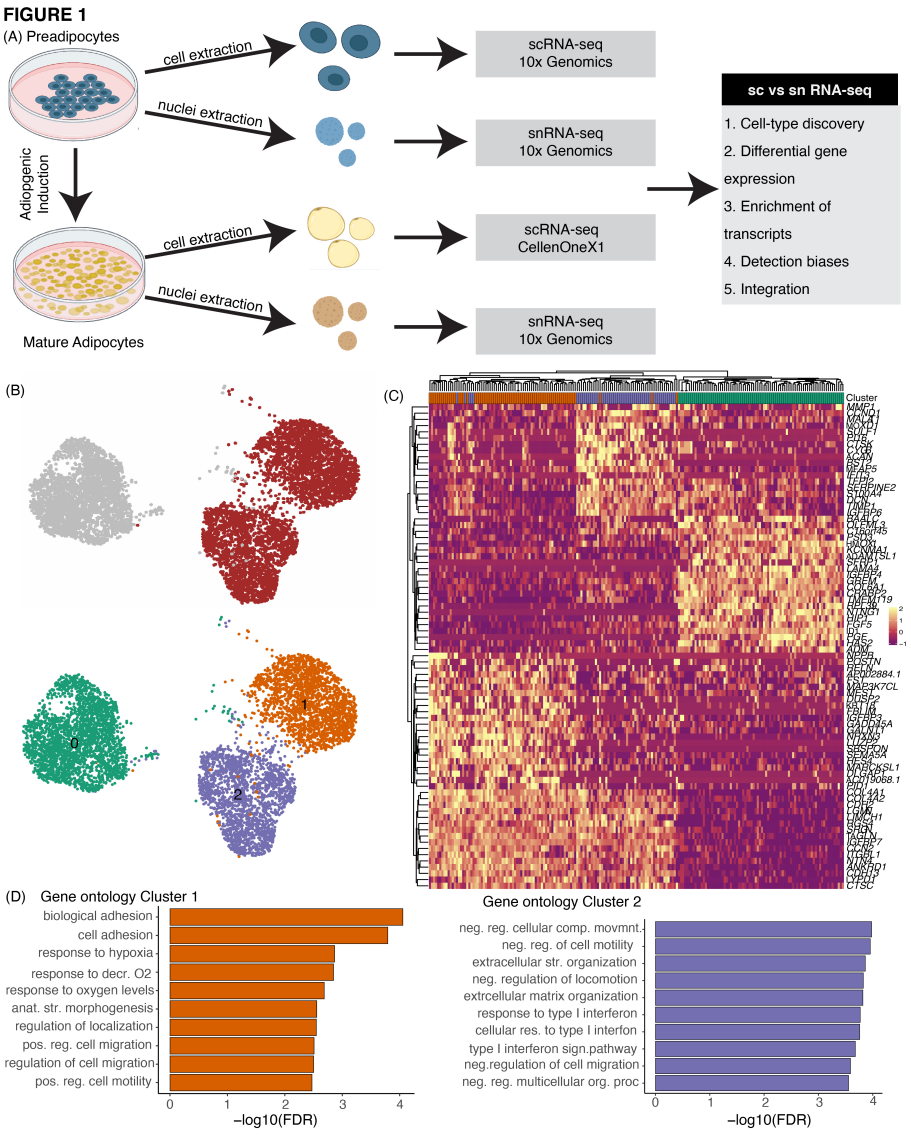

Figure 4.1: **scRNA-seq reveals transcriptional and compositional landscape of white and brown preadipocytes** (A) A schematic representation of scRNA-seq vs snRNA-seq characterizations performed in our study. (B) UMAP visualization of white and brown preadipocytes annotated either manually to reflect the sample of origin (top panel) or based on unsupervised clustering (bottom panel). (C) Heat map of top 20 differentially expressed genes between white and brown preadipocytes, and cluster-1 and cluster-2 preadipocytes based on log fold-change values. Topmost row reflects cluster assignment as in panel (B). (D) Top 10 gene ontology terms enriched in brown preadipocyte cluster 1 (left panel) and cluster 2 (right panel). decr. = decreased; anat. str. = anatomical structure; pos. reg. = positive regulation; neg. reg = negative regulation; str. = structural.

tracted nuclei from these populations and performed snRNA-seq using the same isolation and sequencing protocol. We sequenced snRNA-seq libraries to saturation and compared their transcriptomic profiles with those obtained from scRNA-seq across different cellular subtypes. We next developed a single-adipocyte whole-cell isolation protocol and mapped cellular heterogeneity in mature white adipocytes using the molecular single-cell RNA barcoding and sequencing (mcSCRB-seq) protocol [13]. The transcriptomic profiles obtained were compared with molecular profiles of single nuclei isolated from the same population of adipocytes. Our analyses characterized the accuracy with which snRNA-seq can identify cell types present at the precursor and mature stages of adipogenesis. We identified both technical and biological artifacts that can introduce gene detection biases in snRNA-seq, and we systematically evaluate the limitations of these biases in the context of human adipogenesis. Finally, we propose a normalization strategy for the removal of systematic technical biases between scRNA-seq and snRNA-seq and demonstrate recovery of shared biology by integrating the two datasets using scVI, a variational autoencoder based framework for analysis of scRNA-seq data [280].

4.2 Results

scRNA-seq reveals transcriptional landscape of white and brown preadipocytes

Unsupervised clustering of white and brown preadipocyte scRNA-seq library grouped the cells into three clusters, referred to as populations 0, 1 and 2 (Fig. 4.1B). White preadipocytes organized into a single homogeneous cell population, cluster 0, whereas brown preadipocytes revealed two cell populations, cluster 1 and cluster 2 (Fig. 4.1B). As expected, clusters of white and brown preadipocytes were highly concordant with molecular features of respective primary preadipocytes [281] (Fig. B.2C and B.2S1D). All populations were devoid of endothelial (CD31) and hematopoietic marker genes (CD45, Fig. B.2E and B.2F) and reflected a preadipocyte state on the basis of their high expression for common mesenchymal stem cell markers ITGB1 (CD29), THY1 (CD90), CD44, and ENG [282, 283] (Fig. B.2G to B.2J). All populations also had positive expression for adipogenesis regulators CEBPB & PPARG [284], and ZEB1 [285], further verifying an adipogenic fate for these cells (Fig. B.3A to B.3C).

Differential gene expression (DGE) analysis confirmed that white preadipocytes showed enrichment of genes that are reported to be involved in establishing white preadipocytes' identity (Supplemental Table 2B) such as TCF21 [286], PAX3 [287], and PDGFRA [288]. The most upregulated gene in white preadipocytes was ID1 (Fig. 4.1C), which is known to maintain progenitor state in preadipocytes by positively regulating the progression of cell cycle for sustained growth and proliferation [289, 290]. Consequently, enriched expression of ID1 in white preadipocytes suggested ongoing signaling for maintenance of cellular proliferation. In brown preadipocytes, the top upregulated genes included ANKRD1 and CCN2 (Fig.

4.1C), which are well-characterized YAP target genes [291]. YAP/TAZ are mechanosensitive transcriptional co-activators that regulate proliferation and differentiation at precursor state [292–294], while also maintaining thermogenic activity at mature adipocyte state in brown lineage [295]. Therefore, our results suggest that brown preadipocytes may have ongoing YAP/TAZ activity for maintenance of brown-lineage progenitor state. DGE analysis also revealed upregulation of smooth-muscle lineage marker genes in brown preadipocytes, such as TAGLN (Fig. 4.1C), ACTA2, MYL9, and CNN1 (Supplemental Table 2B). These findings are consistent with a recent study that demonstrated abundant expression of smooth muscle lineage–selective genes in clonal human brown preadipocytes [268], suggesting that brown preadipocytes derived from human neck depot may share this lineage.

Interestingly, we identified two distinct cell populations within brown preadipocytes (cluster 1 and cluster 2, Fig. 4.1B). Gene ontology (GO) analysis identified cellular adhesion, and regulation of cellular motility as the most enriched terms in cluster 1 (Fig. 4.1D), suggesting the prevalence of stem-cell-like migratory behavior in these cells. Transforming growth factor superfamily genes (BMP4 and TGFB2) were also enriched in cluster 1 (Supplemental Table 2C), which play an important role in regulating adipocyte commitment in mesenchymal stem cells [296, 297]. Investigating differential activity of transcription factors (TFs) in cluster 1, transcription factor enrichment analysis (TFEA) identified FOX (FOXC2 and FOXL1) and FOSL1 transcription factors (TFs) with high activity (Supplemental Table 2D). FOXC2 participates in the early regulation of preadipocyte differentiation [72, 298] while FOSL1 proteins have been implicated as regulators of cell differentiation, and transformation [299, 300]. Therefore, our results indicate that cluster 1 cells may exhibit migratory behavior with ongoing signaling similar to adipogenic fate commitment in mesenchymal stem cells, a behavior we refer to here as stem-cell-like. Enrichment of multiple regulators of adipose tissue development was also detected in cluster 1, such as SEMA5A [301], NPPB [302], MEST [303], and FST [304], further suggesting the existence of adipogenic commitment activity in this cell population.

Cluster 2 cells were marked by the expression of S100A4 gene, also known as the fibroblast specific protein 1 (FSP1, Fig. 4.1C, Fig. B.4D and B.4F), which is considered a reliable marker of fibroblasts [305]. GO analysis showed enrichment of immune response, extracellular structure and matrix organization, and negative regulation of cell migration terms in this cell population (Fig. 4.1D). Multiple genes encoding for extracellular matrix (ECM) components such as MFAP5, ECM1, COL6A2, and ACAN were also enriched in cluster 2 (Supplemental Table 2C). Recent investigations have reported the presence of Fsp1+ fibroblasts in the adipogenic niche, with potential role in maintaining adipose homeostasis [294, 306, 307]. The markers identified for fibroblasts in these investigations FBN1, IGFBP6, MFAP5, S100A4, and PI16 were some of the most enriched markers of cluster 2 cells (Fig. B.4A to B.4F). Therefore, these results indicate that cluster 2 cells are fibroblast-like, with negative regulation of cellular migration and an ongoing activity for ECM organization. Existence of two phenotypically distinct brown preadipocytes was further corroborated by performing single-molecule fluorescent in situ hybridization (smFISH) imaging of cluster-2 enriched gene MMP1 (Note B.2, Fig. B.1).

Recently, snRNA-seq of primary human WAT harvested from the same anatomical location as our model system (neck) identified a single white preadipocyte population [273] while multiple scRNA-seq studies identified 2-3 adipocyte progenitor populations in adult murine abdominal WAT [308]. These differences in WAT preadipocyte composition could arise from species-specific and/or depot-specific variation. Interestingly, in contrast to our findings, snRNA-seq of primary human neck BAT also revealed a homogenous brown preadipocyte population [273], possibly because of differences during tissue biopsy collection, or poor cell-capture efficiencies during single-nuclei isolation.

snRNA-seq identifies the same preadipocyte populations as scRNA-seq and detects biologically relevant differential expression

To evaluate the efficacy of snRNA-seq for recovering transcriptional heterogeneity, we sequenced the nuclear transcriptome of single preadipocytes from the white and brown lineages. Unsupervised clustering of the two lineages grouped nuclei into four clusters, referred to as populations 0, 1, 2 and 3 (Fig. B.5A and B.5B). Cluster 3 nuclei, however, had enriched expression for stress response genes and mitochondrial genes, along with high background RNA contamination (Fig. B.5D), and hence were removed from downstream analyses. In the remaining clusters, brown nuclei were primarily grouped into clusters 1 and 2 whereas white nuclei grouped into a single cluster 0 (Fig. 4.2A). Similarity between clusters identified in snRNA-seq and scRNA-seq was assessed using the concept of transcriptional signatures [309, 310], defined as genes differentially expressed in either white vs brown preadipocytes, or cluster 1 vs cluster 2, in the scRNA-seq dataset (Supplemental Table 2B and 2C). As expected, the transcriptional signature scores, calculated using Vision [311], were enriched in the corresponding preadipocyte-type/clusters in the snRNA-seq dataset (Fig. 4.2B), thereby demonstrating a high concordance between transcriptional features uncovered by the two techniques.

As was observed with scRNA-seq, white nuclei were enriched for genes TCF21, PAX3 and PDGFRA (Supplemental Table 3A), and brown nuclei were enriched for YAP/TAZ target genes ANKRD1 and CCN2 (Supplemental Table 3A), and smooth muscle lineage marker genes TAGLN, MYL9, CNN1, and MYH11 (Supplemental Table 3A). Gene ID1, however, was not differentially enriched in white nuclei, because of a lack of differential enrichment in the nuclear compartment between white and brown preadipocyte (Note B.2 and Fig. B.6). In scRNA-seq dataset, we had classified certain DE genes as markers for white and brown preadipocytes based on their highly enriched and specific expression (Note B.2). All such white- and brown-preadipocyte specific marker genes were also enriched in white and brown nuclei respectively (Supplemental Table 3A). Of the 50 genes with maximum enrichment (ordered by logFC) in white and brown preadipocytes in scRNA-seq dataset, over 94% were also differentially expressed in white and brown nuclei respectively (Fig. 4.2C). This analysis demonstrates that snRNA-seq has sufficient sensitivity to recover same molecular differences as scRNA-seq between white and brown preadipocytes.

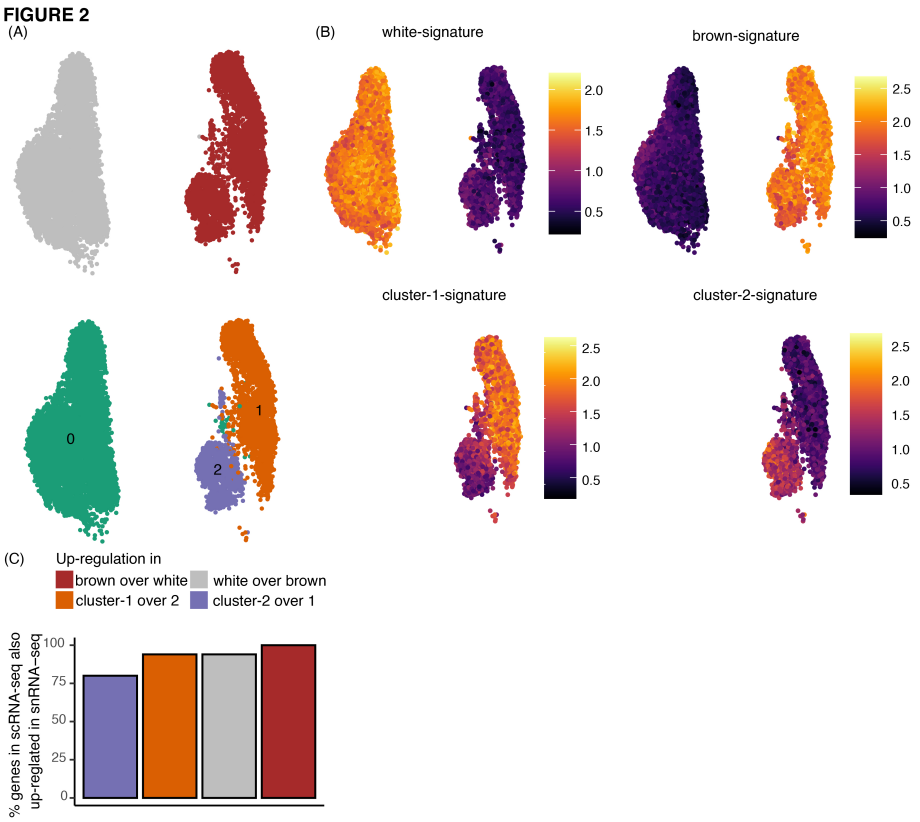

Figure 4.2: **snRNA-sequencing identifies the same preadipocyte populations as scRNA-seq and detects biologically relevant differential expression** (A) UMAP visualization of white and brown preadipocytes annotated either manually to reflect the sample of origin (top panel) or based on unsupervised clustering (bottom panel). 6556 white and 3891 brown nuclei were detected. Of these nuclei, 6578 were in cluster 0, 2716 in cluster 1, and 1153 in cluster 2. (B) Heatmap of transcriptional signature scores for white preadipocyte (top left panel), brown preadipocyte (top right panel), brown preadipocyte cluster 1 (bottom left panel), and brown preadipocyte cluster 2 (bottom right panel) as plotted on the UMAP visualization of snRNA-seq data (C) Bar plot of percent top-50 genes differentially enriched (DE) in scRNA-seq dataset that are also DE in snRNA-seq dataset. Top-50 genes were evaluated based on log fold-change values using scRNA-seq dataset.

GO analysis identified enrichment of cellular adhesion, and regulation of cellular localization terms in brown cluster 1 nuclei, corresponding with the findings in scRNA-seq dataset (Fig. B.5E). Transforming growth factor superfamily genes BMP4 and TGFB2 were also enriched in cluster 1, along with regulators of adipose tissue development SEMA5A, MEST, and FST (Supplemental Table 3B). All 6 cluster-1-specific marker genes (Note B.2) identified were also enriched in cluster 1 nuclei (Supplemental Table 3B). Of the 50 genes with maximum enrichment (ordered by logFC) in cluster 1 cells in scRNA-seq dataset, 94% were also differentially expressed in the nuclear dataset (Fig. 4.2C). In cluster 2 brown nuclei,

enrichment of S100A4 was observed (Supplemental Table 3B), as well as regulation of extracellular matrix organization terms based on GO analysis (Fig. B.5F). Genes encoding for extracellular matrix components COL6A2, MFAP5, ACAN, and ECM1 were all upregulated in cluster 2 (Supplemental Table 3B). Of the 50 genes with maximum enrichment (logFC) in cluster 2 brown preadipocytes (scRNA-seq dataset), 80% were also differentially expressed in the nuclear dataset (Fig. 4.2C). All cluster-2-specific marker genes (Note B.2) identified were also enriched in cluster 2 nuclei (Supplemental Table 3B). Overall, our snRNA-seq analyses indicated the emergence of stem-cell-like behavior in cluster 1 and fibroblast-like behavior in cluster 2, in agreement with the whole-cell dataset. Finally, preadipocyte-/cluster-specific transcriptional signatures now defined using snRNA-seq dataset revealed enrichment in corresponding preadipocyte-type/clusters in the scRNA-seq dataset, thereby validating that markers derived from snRNA-seq can be used to identify the same populations in whole-cell analysis (Fig. B.5H).

snRNA-seq achieves informational saturation at similar sequencing depth as scRNA-seq

Typical 3' scRNA-seq protocols target a sequencing depth of 20,000 to 50,000 reads per cell for identifying diverse cell-types in a moderately heterogeneous sample [312, 313]. However, it is unclear whether similar sequencing depth in enough to recover relevant, discerning information from snRNA-seq datasets. Here, using scRNA-seq and snRNA-seq datasets from matched cell-types, we investigated the trends for multiple sequencing metrics as well as cluster separation as a function of sequencing depth across the two techniques. It is critical to mention that snRNA-seq libraries in our datasets were sequenced using the more sensitive 10x-v3 chemistry as compared to the less sensitive 10x-v2 chemistry for scRNA-seq datasets, which may influence this comparison, as the v3 technology can detect more genes with fewer reads.

When comparing the two techniques in white preadipocytes, mean number of UMIs detected at a given sequencing depth were higher in scRNA-seq as compared to snRNA-seq (Fig. 4.3A). This makes sense since nuclear mRNA is only a subset of whole-cell mRNA and hence is inherently less complex than scRNA-seq. However, our analysis also revealed a higher number of mean genes detected in snRNA-seq as compared to scRNA-seq (Fig. 4.3B), which could be an artifact of increased 10x-v3 sensitivity. Next, we compared recovery of relevant biological information as a function of sequencing depth for the two techniques. Recovery of relevant biological information was quantified by the separation resolution between the two brown preadipocyte clusters as quantified by the Silhouette coefficient. Our analysis revealed saturation of Silhouette coefficient in both scRNA-seq and snRNA-seq, although with better clustering resolution in scRNA-seq at all sequencing depths (Fig. 4.3C). This is interesting since snRNA-seq detected more genes as compared to scRNA-seq at any given sequencing depth, thereby suggesting that increased gene detection sensitivity in snRNA-seq was not relevant to separating the 2 brown preadipocyte clusters. Our analysis also demon-

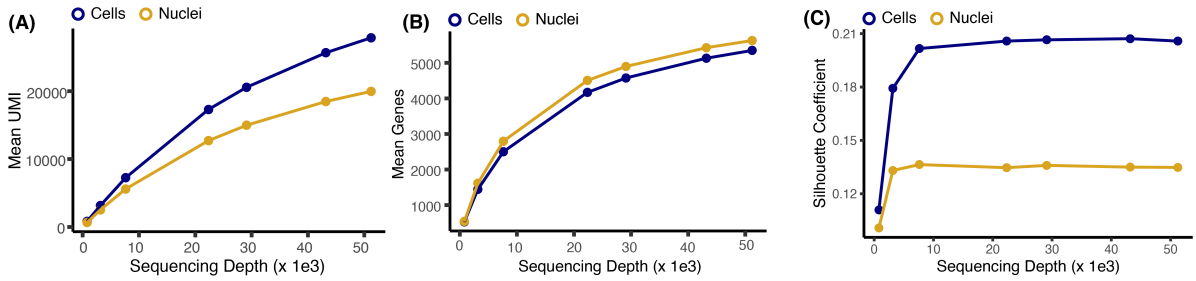

Figure 4.3: **snRNA-seq achieves informational saturation at similar sequencing depth as scRNA-seq** (A) Mean UMIs detected in cells and nuclei isolated from white preadipocytes as a function of sequencing depth (B) Mean genes detected in cells and nuclei isolated from white preadipocytes as a function of sequencing depth (C) Cluster separation resolution quantification between brown cluster 2 vs cluster 1 in scRNA-seq and snRNA-seq dataset. Both datasets were subsampled to have the same number of cells/nuclei and mean transcriptome mapped reads.

strates that Silhouette coefficient at saturation in snRNA-seq is smaller than Silhouette coefficient in scRNA-seq dataset at typical sequencing depth of 20,000 to 50,000 reads, suggesting that increasing sequencing depth for snRNA-seq libraries with the aim of achieving similar clustering resolution as scRNA-seq may not be worthwhile. Although, snRNA-seq had lower cluster separation than scRNA-seq, both techniques achieved saturation of the Silhouette coefficient at ∼10,000 reads, suggesting that any increase in information after 10,000 reads is marginal for both scRNA-seq and snRNA-seq datasets. Overall, our analysis demonstrates that in our dataset, ∼10,000 reads are enough to achieve informational saturation (as quantified by cluster separation) in both scRNA-seq and snRNA-seq datasets, although with a smaller cluster separation in the latter technique.

Gene length-associated detection bias in single-nuclei RNA-sequencing

Typical scRNA-seq data analysis pipelines often filter intronic reads for downstream count matrix generation. More recently, however, evidence has suggested that intronic reads originate from nascent transcripts [314–316], and hence are informative about expression levels in single-cell data. Furthermore, the additional read counts improve gene detection sensitivity and can improve cell-cluster resolution [93, 275]. Multiple recent studies have suggested internal hybridization of polyT RT-primer to intronic polyA stretches in nascent transcripts as the primary mechanism for the capture and detection of intronic reads [240, 317, 318]. Consequently, intronic reads are more readily detected in genes with more intronic polyA stretches, which are more likely to be longer in length (Fig. B.7A). This bias is increased in nuclear libraries where up to 40% of all the reads map to intronic regions as compared to only 9% in scRNA-seq (Fig. 4.4A). Consequently, recent studies have reported enrichment of longer genes [275, 277] and poor detection of shorter genes [278] in nuclei.

To examine the enrichment of long genes in nuclei, we first performed DGE analysis between cells and nuclei in white preadipocytes. Using both intronic and exonic reads, our analysis identified 493 genes enriched in cells and 568 genes enriched in nuclei (logFC >1and FDR <0.05). Notably, nuclear-enriched genes were significantly longer than genes enriched in whole-cells (two- group Mann–Whitney U-test, p-value < 0.01, Fig. 4.4B). Next, we performed DGE analysis between white and brown nuclei, with and without intronic reads, for genes that are enriched in white preadipocytes in the scRNA-seq dataset. Notably, we identified certain long genes such as KCNMA1 (99th percentile), AK5 (96th percentile), and BAALC (85th percentile) that become differentially expressed (in white nuclei over brown nuclei; DE) only upon inclusion of intronic reads (non-DE with only exonic reads), likely because of their preferential detection (Fig. 4.4C and Fig. 4.4D, highlighted in red). Conversely, we also identified certain short genes such as CRABP2 (32nd percentile), TM4SF1 (40th percentile), and EVA1B (17th percentile), that remain non-differentially expressed, even with inclusion of intronic reads in the snRNA-seq dataset (also non-DE with only exonic reads; Fig. 4.4C and Fig. 4.4D, highlighted in blue). We also performed DGE analysis between white cells and white nuclei using only exonic reads (logFC >0.25 and FDR <0.05). Notably, the logFC differential enrichment for nuclear-enriched genes was poorly correlated with counting exons or exons and introns (Fig. 4.4E, Pearson $R = 0.50$, p-value < 0.01). logFC values for some of the longest genes were artificially inflated, possibly because of their preferential detection upon inclusion of intronic reads (Fig. 4.4E Right panel). Conversely, logFC values for some of the shortest genes were artificially deflated because of their poor detection (Fig. 4.4E Right panel). Consequently, the ratio of the logFC values with counting exons or exons and introns, was strongly correlated with gene length (Fig. 4.4E, Pearson $R = 0.69$, p-value < 0.01). Overall, our results demonstrate technical artifacts induced by gene-length associated detection bias in snRNA-seq, upon inclusion of intronic reads. We therefore developed a normalization strategy to address this technical, length-associated detection bias (Note B.2 and Fig. B.7). After normalization, the logFC differential enrichment of nuclear-enriched genes was highly correlated with counting exons or exons and introns (Fig. 4.4F, Pearson $R = 0.94$, p-value < 0.01). Moreover, the ratio of the logFC values with counting exons or exons and introns, after normalization, was poorly correlated with gene-length (Fig. 4.4F, Pearson $R = 0.15$, p-value < 0.01). Nuclear and cellular transcriptomes were also better correlated after removal of technical biases using our normalization strategy (Fig. 4.4G and 4.4H).

DGE analysis, between white cells and white nuclei, with normalized read counts identified 382 enriched genes in cells and 249 enriched genes in nuclei (logFC $>$ 1and FDR <0.05), with nuclear-enriched genes still significantly longer than whole-cells (two-group Mann–Whitney U-test, p-value < 0.01, Fig. B.8A). However, the genes enriched in nuclei were on average 14-fold longer than genes enriched in cells (as compared to 32-fold difference before normalization), which is comparable to the difference observed when using only exonic reads (11-fold difference, Fig. B.8B), suggesting that after accounting for technical bias, there also exists biological enrichment of longer genes in nuclei. Overall, our observations demonstrate that length-normalization removes artificial detection biases thereby improving

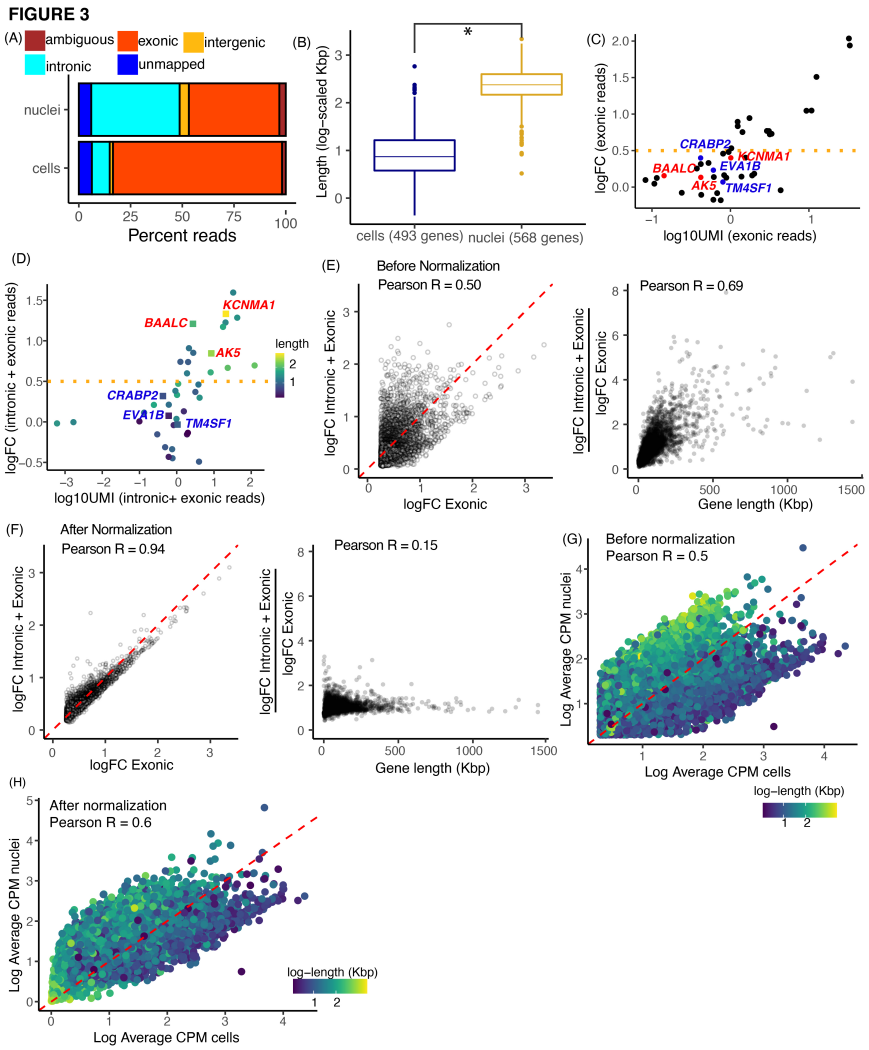

Figure 4.4: **Gene length associated detection bias in the nuclear transcriptome** (A) Distribution of reads in scRNA-seq and snRNA-seq (B) Distribution of gene length for genes enriched in cells (in blue) and nuclei (in yellow) including both intronic and exonic reads (C) Log-fold-change vs log-UMI counts in white nuclei, where each dot represents a white-preadipocyte-enriched gene in scRNA-seq dataset. Horizontal dotted line indicates logFC cutoff value of 0.5 (D) Log-fold-change vs log-UMI counts in white nuclei when using both intronic and exonic reads. Each dot is the same as in panel (C) (E) Left panel: Log-fold-change for nuclear-enriched genes when using only exonic reads, or both intronic and exonic reads before normalization. Red dotted line indicates $y = x$ axis. Right panel: Ratio of y−axis−value over x−axis−value for genes in left panel, plotted as a function of their length. (F) Same plot as in (E) but after normalization. (G) and (H) Average expression of genes in white cells and white nuclei when using both intronic and exonic reads, without normalization (G), and with normalization (H). Red dotted line has slope = 1

UMI count estimation accuracy, while also preserving improved gene detection sensitivity afforded by inclusion of intronic reads.

To further understand differential transcript enrichment between whole-cell and nuclear transcriptomes, we next focused on genes enriched in whole-cells after normalization. GO analysis identified protein translation associated terms as most enriched in whole cells (Fig. B.8D). Genes contributing to the enrichment of translational terms primarily included mRNAs encoding for ribosomal proteins. This enrichment of ribosomal-protein mRNAs in whole-cells is consistent with their very low cytoplasmic decay rates and selective nuclear export machinery [319, 320]. Yet, poor detection of ribosomal proteins in the nuclear transcriptome did not affect the ability to resolve cellular populations in snRNA-seq, as evident by the score of the transcriptional signature consisting of top 100 genes enriched in cells based on logFC values ($\sim$ 53/100 ribosomal protein genes; Fig. B.8C).

Nuclear transcriptome is enriched for long non-coding RNAs that regulate adipogenesis and drive cell-type differences

Long non-coding RNAs (lncRNAs) function in regulating diverse biological processes, including regulation of transcription, proliferation, pluripotency, and cellular differentiation [321–323]. Because of their regulatory function, lncRNAs predominantly remain localized in the nucleus [324, 325]. snRNA-seq intrinsically enriches for nuclear localized transcripts, and previous studies have reported enrichment of lncRNAs in snRNA-seq libraries over scRNA-seq [326, 327]. We hypothesized that nuclear enrichment of lncRNAs could be advantageous for characterizing adipose tissue because multiple lncRNAs have also been implicated in regulating adipogenesis [328–332]. We tested this hypothesis in our *in vitro* system by profiling adipogenic regulatory lncRNAs in our whole-cell and nuclear libraries derived from white preadipocytes, after normalization. We identified significant enrichment of lncRNAs NEAT1 [330], MEG3 [333], MIR31HG [334], and PVT1 [220] in white nuclei, which are previously reported regulators of adipogenesis (Fig. 4.5A). All four lncRNAs were also enriched in brown nuclei as compared to brown whole-cells (Fig. B.9A to B.9D). Generally, snRNA-seq consistently detected a greater number of lncRNAs at all read depths than scRNA-seq (Fig. 4.5B, p-value < 0.01, two-group Mann–Whitney U-test). Of the 111 differentially expressed lncRNAs between white nuclei and white cells, $\sim$86% (96/111 genes) were upregulated in nuclei, thereby validating a higher prevalence of this class of genes in the nuclear compartment. 7 out of 15 lncRNAs that were enriched in white cells were snoRNA host genes (SNHGs), that have been shown to have various functions in cytoplasm such as repressing mRNA translation, miRNA sponging, and protein ubiquitination [335]. Overall, our results suggest a higher likelihood to deconstruct the functional roles of adipogenic regulatory lncRNAs (and other lncRNAs in general) using snRNA-seq.

Next, we evaluated the sensitivity of snRNA-seq for detection of lncRNAs driving molecular heterogeneity between brown preadipocyte cluster 1 and 2, two cell-types most closely related to each other. At $\sim$ 50,000 reads per cell/nuclei, DGE analysis identified over 40

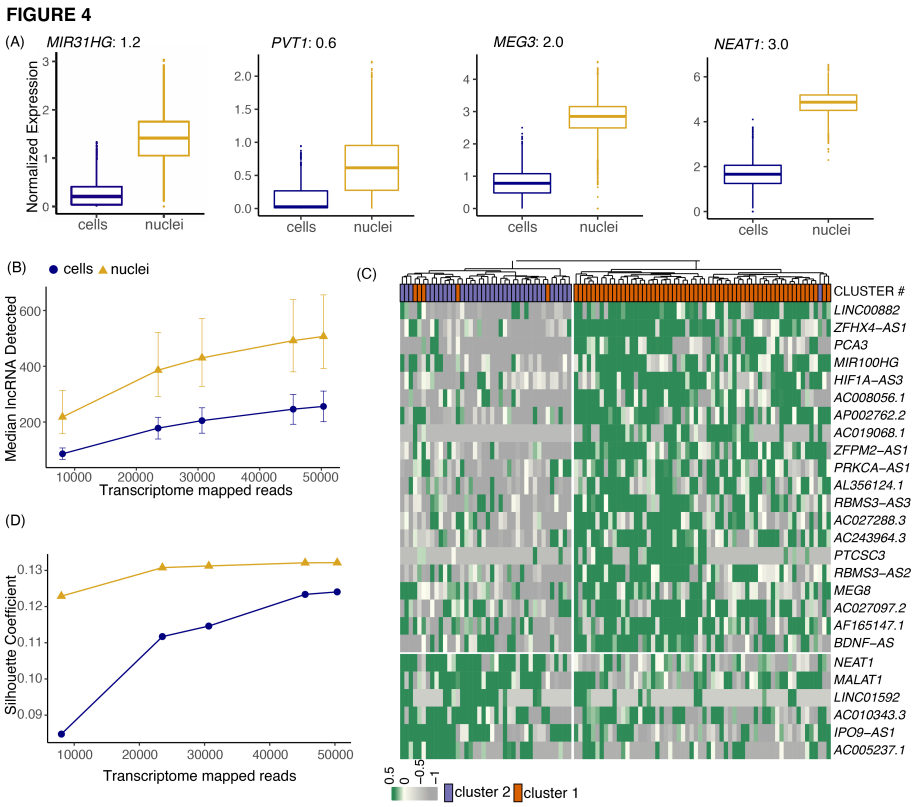

Figure 4.5: **Nuclear transcriptome is enriched for lncRNAs that regulate adipogenesis and drive cell-type differences** (A) Boxplots of lncRNAs reported as regulators of adipogenesis. Black text indicates logFC value for white nuclei vs. white cell DE test in preadipocytes with FDR < 0.05 after normalization (B) Median lncRNAs detected as a function of read depth across single cells and nuclei (both white and brown lineages). Error bars indicate the interquartile range (C) Hierarchical clustering using scaled expression values of top-20 upregulated lncRNAs in brown cluster 1 and cluster 2 in snRNA-seq dataset. 100 random barcodes were chosen for this analysis. Topmost row reflects original cluster assignment for the selected barcodes (D) Cluster separation resolution quantification between brown cluster 2 vs cluster 1 in scRNA-seq and snRNA-seq dataset. Only lncRNAs were considered for PCA manifold generation. Both datasets were subsampled to have the same number of cells/nuclei and same number of mean transcriptome mapped reads.

lncRNAs distinctively regulated between cluster 1 and 2 in the snRNA-seq dataset as compared to only 15 lncRNAs in scRNA-seq dataset. Unsupervised hierarchical clustering in the snRNA-seq dataset based on the expression of top 20 upregulated lncRNAs in cluster 1 and 2 each revealed sorting of nuclei into two distinct groups that predominantly reflected their original cluster assignment (Fig. 4.5C). Moreover, Silhouette coefficient analysis (a method for evaluating clustering performance) revealed better cluster separation performance for snRNA-seq as compared to scRNA-seq between cluster 1 and 2 for all downsampled read

depths (Fig. 4.5D). Silhouette coefficients were calculated based on Euclidean distance between cells/nuclei in the principal component space generated using only lncRNAs (see Methods). To validate that the observed performance features were not metric dependent, we quantified two more indices, the Calinski-Harabasz Index, and the Davies-Bouldin Index to compute inter-cluster separation and found similar trends (Fig. B.9F and B.9G). A similar analysis performed by normalizing for the same number of mean unique molecules (UMI) per sample revealed a similar trend for the three separation indices (Fig. B.9H to B.9J). Together, our results suggest that snRNA-seq is superior for learning heterogeneity governed by lncRNAs as compared to scRNA-seq.

snRNA-seq detects relevant transcriptional regulation during adipogenesis in white preadipocytes

After identifying transcriptomic similarities and differences between scRNA-seq and snRNA-seq in preadipocyte state, we next focused on evaluating molecular correspondence between the two techniques in mature adipocytes. We leveraged our *in vitro* model of white adipogenesis that enabled us to prepare a single-cell suspension of mature adipocytes without the need of implementing harsh tissue dissociation protocols (see Methods). Following single-cell suspension preparation, one of the most common ways to sort single cells is using flow cytometry. Recently, FACS gating strategies have been tailored to isolate mature adipocytes [336, 337], although only a small percentage of adipocytes are able to survive the shear stress associated with flow sorting [337]. Therefore, to enable gentle sorting of single adipocytes for downstream scRNA-seq, we developed a new protocol using the cellenONE X1 single-cell isolation platform. This automated liquid-handling robot uses gentle piezo-acoustic technology for dispensing cells encapsulated in a picoliter-volume droplet, ensuring minimal cellular perturbation and background RNA contamination. To harvest adipocytes *in vitro*, human white preadipocytes were cultured and differentiated using a chemical adipogenic induction cocktail for 20 days [338]. Coherent anti-stokes Raman imaging established successful differentiation of white preadipocytes, with distinctly visible signal from round lipid droplets [62] (Fig. B.10A). After creating a single-cell suspension of white adipocytes, 200 cells were spotted using the cellenONEX1 machine, onto 96-well plates preloaded with lysis buffer and barcoded polyT primer. Library preparation was then performed using the mcSCRB-seq chemistry [13]. Transcriptomic profiles of these cells were then compared with a snRNA-seq library of $\sim$ 12,000 nuclei isolated from 20-days differentiated white adipocytes.

Independent unsupervised clustering revealed organization of both cells and nuclei into primarily two clusters, referred to as cluster 0 and 1 (Fig. 4.6A). snRNA-seq identified an additional cluster 2, which exhibited characteristics of mitotic preadipocytes with ongoing cell cycle progression, suggesting that these cells could be preadipocytes that never underwent growth arrest (Note B.2; Fig. B.11A to B.11H). Cluster 0 in both datasets was marked by the expression of mesenchymal marker THY1 (Fig. 4.6A), suggesting that these cells/nuclei were differentiating preadipocytes. Cluster 1, on the other hand, had high expression of adi-

pogenic gene ADIPOQ, indicating that cells/nuclei in this cluster were mature adipocytes (Fig. 4.6A). DGE analysis further identified enrichment of other adipogenic marker genes (along with ADIPOQ) in cluster 1 (Fig. 4.6B and 4.6C, highlighted in red), confirming a transition from differentiating preadipocytes to mature adipocytes from cluster 0 to cluster 1 in both datasets. GO analysis identified enrichment of extracellular matrix organization terms in cluster 0 and lipid metabolism in cluster 1, independently in both scRNA-seq and snRNA-seq datasets (Fig. B.10B to B.10E). Moreover, ∼80% genes (106/133) upregulated in cluster 1 in the scRNA-seq dataset, were also differentially expressed in the snRNA-seq dataset. Notably, the remaining 20% genes (27/133) that were not differentially expressed in the snRNA-seq dataset primarily included genes associated with the mitochondrial respiratory chain process (Fig. B.10F), suggesting that adipocytes' enhanced mitochondrial activity may not be captured in the snRNA-seq dataset. Correspondingly, snRNA-seq dataset lacked manifestation of mitochondrial biological processes such as oxidative phosphorylation, and electron transport chain in cluster 1 upon GO (Fig. B.10B vs B.10D). This observation was also supported by the fact that these 27 genes had a median length of ∼11 Kbp, the same order of magnitude as length of genes with poor detection in nuclei over whole cells (Fig. 4.4B). As expected, scores of clusters 0 and 1 transcriptional signatures in the scRNA-seq dataset were observed to be highly conserved and enriched in corresponding cluster types (Fig. 4.6D and 4.6E), further validating the conservation of information in the nuclear transcriptome. Overall, our results reveal a comparable molecular landscape in white adipocytes between scRNA-seq and snRNA-seq datasets.

We next looked to investigate any differential adipogenic capacity between the two brown preadipocyte clusters identified in our study (cluster 1 and 2, Fig. 4.1B) by performing scRNA-seq on mature brown adipocytes using the cellenONEX1 for gentle isolation of intact mature adipocytes. These adipocytes were derived by differentiating brown preadipocytes for a period of 20 days. After sorting ∼200 cells, library preparation was performed using the mcSCRB-seq protocol [13]. Transcriptomic profiling revealed detection of mature brown adipocytes, along with recovery of multiple cells that were not terminally differentiated, but rather distributed along a continuum of differentiation states (Fig. B.4G and B.4H). Consequently, our analysis revealed a range of adipogenic gene expression (ADIPOQ) in our dataset (Fig. 4.6F), which was mutually exclusive from the expression of cluster-1-enriched gene SEMA5A (Fig. 4.6F). On the other hand, we identified multiple cells with shared expression of cluster-2-enriched gene S100A4 and ADIPOQ (Fig. 4.6G). These results supported the observation of two brown preadipocyte populations and indicate that cluster-2 cells are more likely to differentiate into mature brown adipocytes. Additionally, we also compared transcriptomic similarities between mature brown adipocytes (Fig. B.4H, highlighted in red) and cluster-1/cluster-2 cells using transcriptional signatures defined for respective clusters using the day-0 scRNA-seq dataset (Supplemental Table 2C). Mature adipocytes had a significantly higher score for cluster 2 cells as compared to cluster 1 (Fig. B.4I), thereby providing additional evidence that the former cell-type is more likely to be adipogenic.

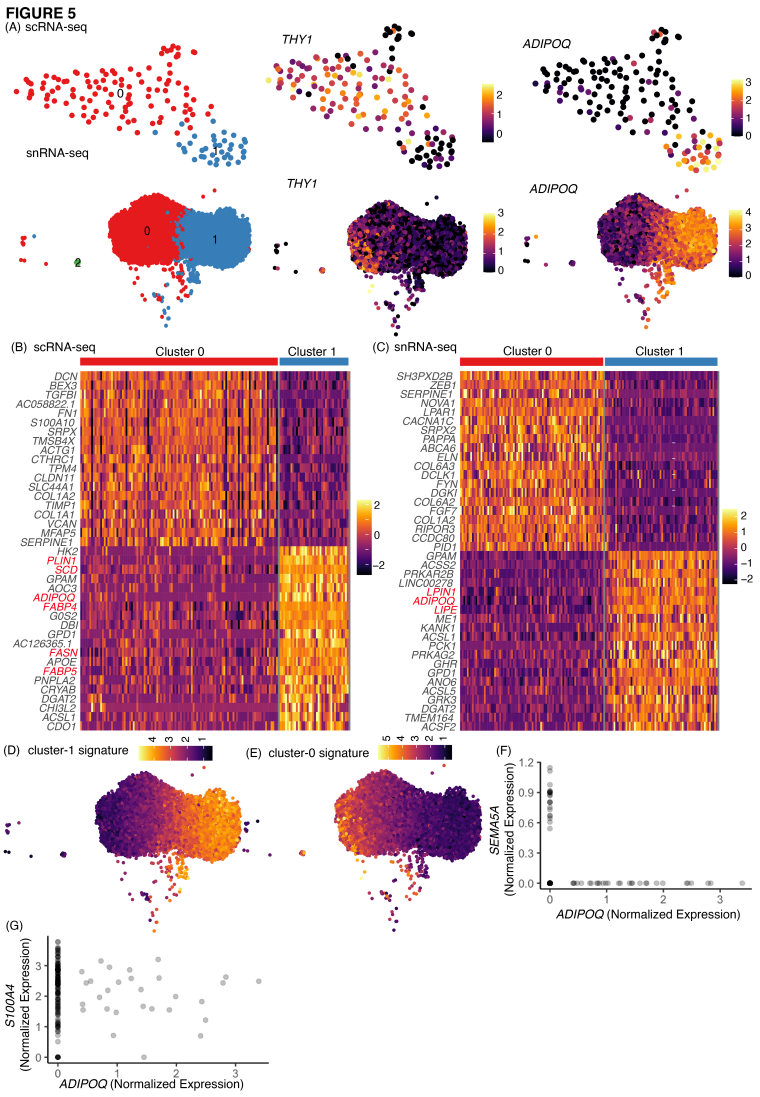

Figure 4.6: **snRNA-seq detects important transcriptional regulation during adipogenesis in white preadipocytes** (A) UMAP of scRNA-seq and snRNA-seq white adipocyte datasets after unsupervised clustering (leftmost panels). Expression profile for mesenchymal marker THY1 and mature-adipocyte marker ADIPOQ in both scRNA-seq and snRNA-seq datasets (middle and rightmost panels) (B) and (C) Heat map of z-scored expression of top 20 differentially expressed genes between cluster 0 and cluster 1 in scRNA-seq (B) and snRNA-seq (C) white adipocyte dataset. Highlighted in red are markers of adipogenesis. (D) and (E) Heatmap of transcriptional signature scores for cluster 1 (D) and cluster 0 (E) as plotted on the UMAP visualization of snRNA-seq white adipocyte data (F) Normalized expression of genes ADIPOQ and SEMA5A and (G) ADIPOQ and S100A4 in differentiating brown preadipocytes (day-20) scRNA-seq dataset. Also see Fig. B.4G to B.4I.

Integration of snRNA-seq and scRNA-seq datasets

A comprehensive cell atlas of the adipose tissue will require joint analyses of datasets generated using both scRNA-seq and snRNA-seq. However, technical biases and differential transcript enrichment in snRNA-seq leads to significant batch effects between snRNA-seq and scRNA-seq experiments, thereby reducing clusterability of cells from these two protocols [339]. Multiple bioinformatic tools are now available to remove covariates that lead to technical batch effects and facilitate integration of scRNA-seq datasets generated across different days, laboratories, individuals, or technologies [340]. We used single-cell variational inference (scVI), a deep generative modeling-based tool [280, 341], to explore the possibility of integrating snRNA-seq and scRNA-seq datasets for joint analysis. Four datasets of white preadipocytes were integrated in total: day-0 scRNA-seq snRNA seq (cluster 0 in Fig. 4.1B and Fig. 4.2A), and day-20 scRNA-seq & snRNA-seq (top and bottom left panels in Fig. 4.6A).

Without batch correction, all four datasets arranged into distinct individual clusters, with no shared population identified at the same time point across different techniques, or same technique but across different time-points (Fig. 4.7A). A dendrogram, based on the Euclidean distance in dimensionally reduce space, grouped clusters first by sequencing chemistry (mcSCRB-seq vs 10x), followed by technique type (snRNA-seq vs scRNA-seq), and finally by time point (day-0 vs day-20, Fig. 4.7A). After integration, matching adipocyte populations from day-20 and preadipocyte populations from both day-0 and day-20 in nuclear and whole-cell datasets were primarily nearest neighbors in a dendrogram based on the Euclidean distance in dimensionally reduced space (Fig. 4.7B). UMAP visualization further revealed proximal placements of similar cell populations (Fig. 4.7B and 4.7C). Of note, we observed that preadipocytes from both day-0-snRNA-seq and day-0-scRNA-seq datasets localized into two distinct groups, which was driven by differences in proliferation state with one cluster composed of mitotic cells and another composed of growth arrested cells (Note B.2; Fig. B.11A to B.11H). Unsupervised clustering of the integrated dataset revealed adipocytes, day-20 preadipocytes, and the two groups of day-0 preadipocytes as distinct cell-types, illustrating scVI's abilities to remove batch effects while retaining biological variation (Fig. 4.7C, right-most panel). Indeed, top marker genes for each cluster recovered previously reported expression trends such as enrichment of ECM components in day-0 preadipocytes [342], enrichment of insulin-binding proteins in day-20 preadipocytes [343], and enrichment of adipogenic genes in mature adipocytes (Fig. B.11J). Although, multiple markers for each cluster had conserved expression across scRNA-seq and snRNA-seq, some markers were exclusively enriched in either one of the datasets (Fig. B.11J), thereby highlighting the importance of performing joint analysis. Finally, integration of the same 4 datasets using Seurat [344] revealed minimal overlap of single-cell and single-nuclei datasets for both day-0 and day-20 (Fig. B.11I). Recently, benchmarking of distinct integration methodologies indeed revealed effective performance by scVI on complex integration tasks, with Seurat v3 performing well on simpler tasks with distinct biological signals [345]. Overall, our results demonstrate scVI's integration abilities by identifying functionally similar preadipocyte and

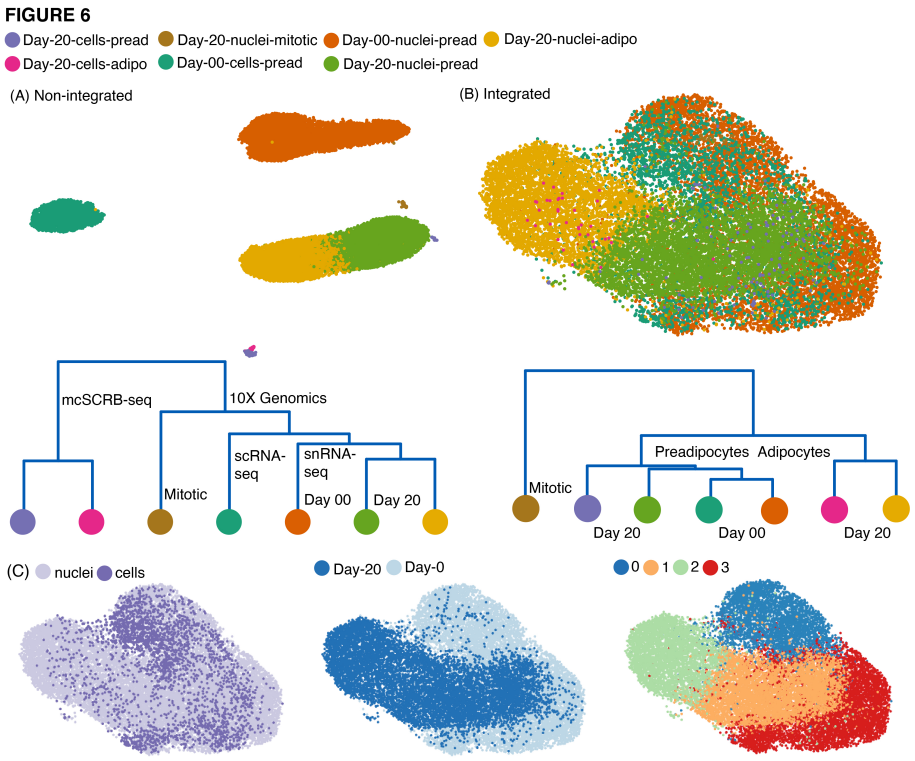

Figure 4.7: **Integration of snRNA-seq and scRNA-seq datasets** (A) UMAP visualization of non-integrated scRNA-seq and snRNA-seq datasets for both white preadipocyte (day-0) and mature adipocyte (day-20), for a total of 4 batches (top panel, total 18717 barcodes). Cluster dendrogram for non-integrated datasets based on the eigenvalue-weighted Euclidean distance matrix constructed in latent-dimension space inferred using scVI (bottom panel) (B) UMAP visualization and cluster dendrogram of scRNA-seq and snRNA-seq datasets as in panel A after integration using scVI-tools (total 18717 barcodes). See also Note S4 and Fig. S10. (C) UMAP visualization of scVI integrated dataset with barcodes annotated by sequencing technique (left panel), harvestation day (middle panel), and joint unsupervised clustering (right panel).

adipocyte populations shared across single-cell and single-nuclei RNA-sequencing techniques.

4.3 Conclusion

In this investigation, we evaluated the ability of snRNA-seq to recapitulate the molecular and compositional landscape of distinct lineages in human adipose tissue. We avoided confounding variability associated with inter-depot and inter-subject transcriptional variation by performing a direct comparison of snRNA-seq and scRNA-seq on a pair of immortalized white and brown human preadipocytes isolated from the neck region of the same individual. We found that snRNA-seq was able to recover the same cell-types as scRNA-seq at both

preadipocyte and mature adipocyte states. Furthermore, we provided evidence for recovering similar expression profiles of biologically relevant genes, and attributing similar functional annotations to cell-types by nuclear transcriptome profiling as compared to whole-cells. At the preadipocyte stage, brown preadipocytes were a heterogeneous mix of two distinct cell populations, cluster 1 and cluster 2. However, cell-type enrichment followed by differentiation and metabolic assays will need to be further performed to identify their individual functions in maintaining adipose tissue homeostasis. To date, different scRNA-seq studies of mouse stromal vascular fraction have identified multiple subpopulations of adipose progenitor cells (APCs) expressing distinct markers [346–349]. Integrated analysis of these datasets primarily identified two common populations of APCs in mice referred to as Asc1 and Asc2 [267, 308]. Similar to cluster 2 cells, Asc2 exhibited pro-inflammatory and pro-fibrotic phenotype and positive expression of genes PI16 and MFAP5. Functional investigations into the two cell types revealed Asc2 cells inhibiting the differentiation of Asc1 cells in vitro [267]. This agrees with the pro-adipogenic nature of cluster 2 cells identified in our study. Therefore, it is plausible that cluster 1 and cluster 2 cells identified in our study may be functioning in a manner similar to Asc1 and Asc2 to maintain adipocyte turnover.

snRNA-seq is the preferred technique to study samples whose compositional landscape may be biased by the differential efficiency of cell-type recovery when using scRNA-seq. Adipose tissue is one such sample where isolation of intact, single adipocytes is complicated by their fragile nature. Indeed, most adipose scRNA-seq studies to date derive the transcriptomes of the cell types within stromal vascular fraction (SVF) only, with minimal detection efficiencies for mature adipocytes either in animal models [89, 350, 351] or humans [307, 348, 352, 353]. Here, we developed a new single-adipocyte isolation protocol using piezo-acoustic-based gentle dispensing technology for improved recovery with downstream scRNA-seq. Using this strategy, $\sim$26% barcodes recovered were annotated as adipocytes. However, this adipocyte capture efficiency was still limited as compared to snRNA-seq where $\sim$ 48% barcodes were identified as adipocytes. Conversely, at the preadipocyte stage, where cell-type recovery is efficient, scRNA-seq recovered equal proportions of the two brown preadipocyte clusters. However, analysis of snRNA-seq data revealed $\sim$1.5-fold enrichment of cluster 1 over cluster 2, suggesting a bias in compositional sampling in snRNA-seq. Therefore, such cell-level sampling biases must be considered when evaluating the composition of complex tissues with snRNA-seq.

Understanding the advantages and drawbacks of using snRNA-seq, a nuclear transcriptome is inherently enriched for nascent transcripts, thereby predominantly reflecting changes in gene expression as a result of differences in transcription rates alone [354]. In contrast, a cellular transcriptome is fundamentally enriched for mature transcripts, thereby capturing gene expression changes driven by both transcriptional and post-transcriptional regulatory processes such as mRNA processing and degradation. Higher relative proportion of nascent to mature transcripts in the nucleus also results in a large fraction of intronic reads in snRNA-seq, which when considered for count matrix generation, gives rise to detection bias against short genes with few intronic polyA stretches. Consequently, for a biological system compatible with both techniques, scRNA-seq may be better for identifying cellular subpop-

ulations. scRNA-seq will also be better for assessing gene expression changes as a result of post-transcriptional regulation. However, nuclear transcriptome is preferentially enriched for lncRNAs, indicating that functional investigations of these genes will be enhanced by sequencing nuclei. Moreover, some studies of specific nuclear functions may be enhanced by directly accessing nuclei for example, changes in gene expression profile as a result of targeted transcriptional activation mediated by epigenetic modifications. Therefore, it is important to evaluate each approach depending on the task at hand. However, for tissues such as the adipose tissue, snRNA-seq may be the only option. In our investigation, lncRNAs regulating adipogenesis were enriched in the nuclear transcriptome. lncRNAs driving differences between cluster 1 and 2 in brown preadipocytes were also better detected in the snRNA-seq dataset. However, we also identified poor detection of shorter genes in nuclei, some of which were key to driving heterogeneity between distinct cell-types.

Including intronic reads for UMI quantification presents researchers with both advantages and drawbacks. polyA stretches are found randomly dispersed along the length of the genome, and introns become the predominant site for the localization of such stretches because of their extensive length (21-fold longer than exons) [355]. These polyA stretches present additional priming sites (besides the 3' polyA tail) for the polyT RT primer, thereby enabling more efficient transcript capture. Conversely, most intronic reads are therefore derived from genes with multiple polyA stretches (long genes), thereby introducing technical detection bias. This bias gets further magnified in snRNA-seq libraries that are inherently enriched for nascent transcripts (and hence intronic reads), and filtering such reads would mean reduced gene detection sensitivity, shallower sequencing depth and under-utilized sequencing cost. Here, we provided a normalization strategy for UMI counts derived from intronic reads that can remove gene-length associated technical biases. Implementation of this normalization strategy removes technical artifacts while retaining true biological features, thereby improving integration and enabling joint analysis of scRNA-seq and snRNA-seq datasets. In such joint analysis, our normalization strategy would also improve the accuracy of differential expression testing between any technique-specific clusters identified.

Finally, we demonstrated applicability of scVI for integration of scRNA-seq and snRNA-seq datasets. This is critical for the generation of a comprehensive adipose tissue atlas since investigations into the stromal vascular fraction heterogeneity have been performed using scRNA-seq whereas snRNA-seq is favorable for investigations into the existence of adipocyte subtypes. Therefore, any efforts to identify shared subpopulations across such datasets, and the lineages therein would demand data integration. However, our findings here are based on an *in-vitro* adipogenic model system, with a less heterogeneous cellular composition than primary tissue. Therefore, integration of scRNA-seq and snRNA-seq datasets with this in vitro adipogenic model system is likely more robust than in a primary sample. Indeed, multiple previous reports have demonstrated strong batch effects between scRNA-seq and snRNA-seq datasets derived from the same primary tissue, resulting in suboptimal integration performance [345, 356, 357]. As a demonstration of this phenomenon, we integrated the human heart cell atlas dataset using scVI-tools, the algorithm demonstrated to be most effective in integrating scRNA-seq and snRNA-seq datasets, and observed sub-grouping of

samples derived from the two techniques within clusters of major cell-types identified (Fig. B.13). Such batch effects are likely rooted in technical differences across the two techniques, such as gene-length associated detection biases, and high background mRNA levels in nuclear libraries (Fig. B.12) [358, 359]. For Integrative algorithms like scvi-tools, which utilize deep generative models for batch effect correction, users could run posterior predictive checks to quantitatively compare integration performance for two scRNA-seq datasets vs scRNA-seq snRNA-seq datasets. Lower predictive power for integration of scRNA-seq and snRNA-seq datasets (as compared to two scRNA-seq datasets) would indicate a need for engineering more accurate machine learning models specifically developed for the task of integrating single-cell and single-nuclei RNA-seq datasets.

Overall, snRNA-seq provides an effective method for characterizing cellular heterogeneity and functionally relevant gene expression profiles within human preadipocytes and adipocytes. We expect that snRNA-seq will be actively adopted by the adipose community for high-throughput transcriptomic profiling of the tissue and aid in increasing its representation in initiatives such as the Human Cell Atlas. Ultimately, joint analysis of datasets acquired using multiple sequencing techniques will aid in the creation of a comprehensive human adipose tissue atlas, thereby enabling us to dissect its critical role in health and disease.

4.4 Materials and Methods

Preadipocyte culture and adipogenic differentiation

Detailed protocol for maintenance, cryopreservation, and differentiation of white and brown preadipocytes are outlined in a different study (Shamsi and Tseng 2017). Briefly, for culturing preadipocytes, cells were grown in DMEM medium (Corning, 10-017-CV) supplemented with 10% vol/vol FBS and containing 1% vol/vol Penicillin-Streptomycin (Gibco). Cell culture was maintained at 37°C in a humidified incubator containing 5% vol/vol CO2. 80% confluent cells were passaged using 0.25% trypsin with 0.1% EDTA (Gibco, 25200-056) for a 1:3 split in a new 100 mm cell culture dish (Corning).

Prior to adipogenic differentiation, white and brown preadipocytes were allowed to grow up to 100% confluence in a 100 mm cell culture dish (Corning). After 48 hours at 100% confluence, growth media was replaced with adipogenic induction media every 48 hours for the next 20 days. Induction media was prepared by adding 1 mL FBS, 500 µl Penicillin-Streptomycin, 15 µl human Insulin (0.5 µM, Sigma-Aldrich, I2643-50MG), 10 µl T3 (2 nM, Sigma-Aldrich,T6397-100MG), 50 µl Biotin (33 µM, Sigma-Aldrich, B4639-100MG), 100 µl Pantothenate (17 µM, Sigma-Aldrich, P5155-100G), 1 µl Dexamethasone (0.1 µM, Sigma-Aldrich, D2915-100MG), 500 µl IBMX (500 µM, Sigma-Aldrich, I7018-100mg), and 12.5 µl Indomethacin (30 µM, Sigma-Aldrich, I7378-5G) to 48.5 mL DMEM medium and sterile filter.

Harvesting preadipocyte and mature adipocyte for scRNA-seq

At preadipocyte stage, cells were harvested from 100 mm plates, labeled with hashtag antibodies (Supplemental Table 1A, Note B.2), and finally suspended in PBS with 0.04%BSA at ∼1000 cells/µL concentration for downstream sequencing. At mature adipocyte stage, cells were first washed with PBS (Corning, 21-040-CV) and incubated with a monolayer of .25%trypsin with 0.1%EDTA (Gibco; 25200-056; monolayer obtained by adding and removing 1 mL of trypsin) for 2-3 minutes in a tissue culture incubator. When adipocytes started to become round and detached from the plate, trypsin was neutralized by adding 1 mL of FBS. Clumps of adipocytes were dislodged using a wide bore 1 mL pipette tip and filtered using a 70 µm cell strainer. Concentration of adipocyte suspension was adjusted to ∼ 200 cells/µL using FBS for downstream spotting using the CellenOne X1 machine.

Nuclei isolation from preadipocytes and mature adipocytes for snRNA-seq

Nuclei were isolated from white and brown preadipocytes using an NP-40 based lysis buffer: To 14.7 mL nuclease-free water (Qiagen), 150 µL of Tris-Hydrochloride (Sigma, T2194), 30 µL of Sodium Chloride (5M; Sigma, 59222C), 45 µL of Magnesium Chloride (1M; Sigma, M1028), and 75 uL of NP-40 (Sigma, 74385) was added. Two 100 mm dishes were used for nuclei isolation from each preadipocyte type. 500 µL of NP-40 based lysis buffer was added to each 100 mm dish and a cell scraper was employed to release adherent cells from the plates. Cells were then incubated with the lysis buffer for 5 minutes on ice in a pre-chilled 15 mL falcon tube. Cells were washed with ice-cold PBS supplemented with .2 U/µL RNase Inhibitor (Protector RNase Inhibitor; henceforth called wash buffer) 4 times by centrifuging at 500 rcf for 5 minutes at 4°C. Wash buffer was aspirated after the final round of centrifugation and nuclei were resuspended in the ice-cold wash buffer and filtered using a 40 um cell strainer. Final concentration was adjusted to ∼ 1000 nuclei/µL using a hemocytometer for downstream sequencing. Nuclei were also stained using 0.08% trypan blue dye to assess nuclear membrane integrity under brightfield imaging. For nuclear isolation at the mature adipocyte stage, the same protocol was implemented as mentioned above with the modification of using 1 mL lysis buffer for each 100 mm dish.

Single-cell and single-nuclei sequencing

For mcSCRB-seq experiment with white adipocytes (day 20), 96-well plates were first preloaded with rows of 10 uniquely barcoded primers and lysis buffer according to the mcSCRB-seq protocol, with the only difference being the use of µCB-seq RT primers [63] instead of standard mcSCRB-seq ones. The sequence of barcodes used were: TCACAGCA, GTAGCACT, ATAGCGTC, CTAGCTGA, CTACGACA, GTACGCAT, ACATGCGT, GCATGTAC, ATACGTGC, and GCAGTATC. CellenONE X1 instrument was used to individually deliver a single adipocyte into each well for a total of 200 cells. Following cell delivery, the mcSCRB-

Table 4.1: Sequencing metrics for individual libraries used in our study. All sequencing was performed on the Illumina NovaSeq platform.

Technique	Sample	Chemistry	Depth	No. cells
scRNA-seq	White, Brown Pread.	v2 + CITE-seq	78,000	8,000
scRNA-seq	White Adipocytes	mcSCRB-seq	100,000	200
snRNA-seq	White Pread.	v3	174,738	8,000
snRNA-seq	Brown Pread.	v3	216,700	7,000
snRNA-seq	White Adipocytes	v3	123,700	12,000

seq protocol was followed directly, but with the following two modifications:

- A 1:1 ratio of AmPure XP beads was used to pool all cDNA after RT as opposed to the manual bead formulation from standard mcSCRB-seq
- NEBNext i5 indexed primers (NEB, E7600 and E7645) were used as opposed to the non-indexed P5NEXTPT5 primer during library PCR and indexing step to generate dual indexed libraries for multiplexing

scRNA-seq and snRNA-seq data analysis

scRNA-seq white brown preadipocytes dataset was processed using cellranger-3.0.2 with default parameters, and the human GRCh38-3.0.0 genome (November 19, 2018) as input. A custom pre-mRNA GTF file was created using the GRCh38-3.0.0 FASTA file as input to include intronic reads in UMI counts. Sample demultiplexing, doublet removal, and empty droplet removal was performed using the Seurat [263] function HTODemux (Note S1). Cell barcodes were further filtered to have more than 200 genes. Post demultiplexing and filtering, scVI [280, 341] was used to infer a 20-dimensional latent space based on the expression of the top 2000 most variable genes. This latent space was then used in Seurat to generate the UMAP visualization using the RunUMAP command. Downstream clustering (resolution = 0.4) and differential expression analysis (logFC $>$ 0.5) was performed using Seurat's SCTransform pipeline [360]. Clusters with $>$ 5% mean mitochondrial content were removed from downstream analyses. In the identified high-quality clusters, cells had minimal cell-cycle effects as calculated using Seurat (Fig. B.3G). Gene ontology analysis was performed at geneontology.org [361–363] and results were further confirmed using the goana package in R with genome wide human annotation derived from org.Hs.eg.db Bioconductor package. Transcription factor enrichment analysis was performed using the ChEA3 tool [364]. GRCh38-ref20202A (2020) reference was used for analysis involving lncRNAs, keeping everything else the same. Independent sub-clustering of cluster 0 and cluster 1 identified differences in cellular states based on cell-cycle only, suggesting the absence of

any cellular subtypes (Fig. B.3D and B.3E). However, sub-clustering of cluster 2 revealed a PI16+ adipocyte progenitor population (Fig. B.3F) [348, 365], which was also identified in snRNA-seq dataset (Fig. B.5G). In this manuscript, we focused on only the major cell-types identified within human white and brown preadipocytes (Fig. 4.1B). For sub-clustering, resolution was set to 0.3, the smallest value at which distinct clusters were first identified within clusters 0, 1, and 2.

snRNA-seq white and brown preadipocyte dataset was also processed using cellranger-3.0.2. For white preadipocyte, barcodes with < 200 genes were removed from downstream analyses. CellBender [359] was used to remove empty droplets. For downstream analyses, only barcodes called as cells by both cellranger and CellBender were used and barcodes with UMI count > 49000 were filtered out as possible doublets. For brown preadipocyte, barcodes with < 200 genes were removed and scVI was used to infer a 20-dimensional latent space. First round of clustering was performed in Seurat with the resolution set to 0.06. We identified 3 clusters, with cluster 1 having most of the barcodes called as empty by CellBender. Therefore, cluster 1 was removed from downstream analysis as well other barcodes that were called as "cell-containing" by cellranger but not by CellBender. Cluster 2 was marked with high mitochondrial content ($> 20\%$) and hence was also removed from downstream analyses. After filtering out low-quality barcodes and clusters, Scrublet [366] was used to remove any potential doublets. After individual QC of white and brown preadipocyte libraries, the two datasets were integrated together using scVI with no batch effect correction. The output from scVI analysis was a 20-dimensional latent space representation with cell embeddings for both white and brown nuclei. This latent space was then used in Seurat to generate the UMAP visualization using the RunUMAP command. Downstream clustering (resolution $= 0.24$) and differential expression analysis (logFC > 0.5) was performed using Seurat's SCTransform pipeline (see Fig. B.5). Cells in each cluster had no significant cell-cycle effects (Fig. B.5C). For gene ontology, and differential expression analyses, the same tools as mentioned in the above paragraph were used. GRCh38-ref20202A (2020) reference was used for analysis involving lncRNAs, keeping everything else the same.

mcSCRB-seq white and brown adipocyte dataset was processed using zUMIs [235] using the GRCh38 index for STAR alignment. We provided the 10X CellRanger recommended GRCh38-3.0.0 GTF file as input for standardization of gene counts. Reads with any barcode or UMI bases under the quality threshold of 20 were filtered out and known barcode sequences were supplied in an external text file. UMIs within 1 hamming distance were collapsed to ensure that molecules were not double-counted due to PCR or sequencing errors. Only exonic reads were counted towards UMI quantification. The umi-count matrix generated using zUMIs was read using the readRDS command in Seurat. The CellenOne X1 machine acquires an image of every cell spotted and the presence of a single cell was further validated by analyzing these images to remove possibly empty or doublet barcodes. The Seurat object was analyzed using a standard Seurat pipeline with resolution set to 0.6 for clustering of white adipocytes and 1.1 for brown adipocytes. snRNA-seq white adipocyte dataset was processed using cellranger-3.1.0. Barcodes with < 200 genes were removed from downstream analyses and scVI was used to infer a 20-dimensional latent space. For clustering using Seurat,

the resolution parameter was set to 0.45. We identified 7 clusters, with cluster 3 having most of the barcodes called as empty by CellBender. Therefore, cluster 3 was removed from downstream analysis as well other barcodes that were called as "cell-containing" by cellranger but not by CellBender. Cluster 5 was marked with high mitochondrial content and hence was also removed from downstream analyses. Cluster 2 had the greatest number of doublets identified by the doubletDetection [367] tools and was filtered out, as well as cluster 4 which was enriched for ribosomal proteins suggesting cellular debris contamination.

Transcriptional signature analysis using primary white and brown preadipocytes

Primary white and brown preadipocytes were isolated from the neck region of 6 individuals and subjected to microarray gene expression profiling [281]. Data was accessed using GEO Accession GSE54280 and analyzed using GEO2R. Differentially expressed genes were identified using a white vs brown test. List of genes enriched in white or brown primary preadipocytes were defined as signatures for respective cell-types and used as input in Vision to assign score to in vitro preadipocytes analyzed in our study (Fig. 4.1).

RNA smFISH and Spot Counting Analysis

To perform RNA FISH, we followed the protocol in Raj et al. 2008 [368] with minor modifications. We pre-washed cells with wash buffer containing 10% formamide and 2X saline-sodium citrate (SSC). We then performed hybridization by adding 1 µL of probe (6.25 µM) to 50 µL of hybridization buffer consisting of 10% formamide, 2X SSC, and 10% dextran sulfate (w/v). The final probe concentration for overnight hybridization was 125 nM. We hybridized the samples overnight in a humidified chamber at 37°C. Following hybridization, we washed the samples twice with wash buffer for 30 minutes at 37°C. We then washed the samples 2X SSC, anti-fade buffer. Imaging was done in anti-fade buffer supplemented with catalase and glucose oxidase.

For quantification of number of RNA spots per cell, Find Foci tool [369] was used in Fiji. For analysis, imaged were first cropped to only have one cell per field of view. Then, the Find Foci plugin was used using the GUI, with Max Size = 100, Peak parameter = 0.2, Max peaks = 1000, and Minimum size = 5 (Advanced settings). For image binarization, a manually selected value was used for thresholding, with visibly best performance in selecting RNA spots as foreground over background. With total RNA spots calculated for each cell, gaussian mixture model fitting was performed using the mclust package in R [370], and negative binomial mixture model fitting was performed using the fitNB command in SIBERG package [371] in R.

Identifying number of lncRNAs detected as a function of sequencing depth

For identifying the number of lncRNAs detected as a function of sequencing depth, the fastq files for scRNA-seq preadipocyte dataset only were subsampled using seqtk v1.3 with the random seed = 100. For each subsample depth, fastq files were processed using cellranger-3.1.0 with GRCh38-ref2020A pre-mRNA as the reference. snRNA-seq data for white and brown nuclei (as processed with cellranger at full depth) were then aggregated with the output of scRNA-seq preadipocyte data at varying sequencing depth using the cellranger aggr command to achieve same number of average transcriptome mapped reads. Number of lncRNAs detected were then calculated as a function of sequencing depth, with lncRNA assumed as detected in a given cell/nuclei if UMI count > 0.

Silhouette coefficient analysis

Both scRNA-seq and snRNA-seq datasets for brown preadipocytes were subsampled as described above. snRNA-seq dataset was further randomly subset to have the same number of total barcodes as scRNA-seq. At each sequencing depth, top 20 principal components were calculated using Seurat's standard pipeline. Three resolution coefficients based on the Silhouette index, Calinski Harabasz index, and Davies Bouldin index were then calculated based on Euclidean distance between cells in the PCA space using the clusterCrit package in R. For analyzing cluster separation resolution between brown cluster 1 and 2 as a function of UMI count, exactly the same analysis was performed except that downsampling was performed to have the same number of UMI rather than reads between scRNA-seq and snRNA-seq dataset using the downsampleMatrix command in the DropletUtils package in R [372].

Integration of snRNA-seq and scRNA-seq data using scVI

For integrating scRNA-seq white preadipocyte (day-0) white-adipocyte (day-20) and snRNA-seq white preadipocyte (day-0) white-adipocyte (day-20) datasets (a total of 4 datasets), we first created a single anndata object with UMI count-matrices from each dataset as input. Each of the four UMI matrices were generated by processing the originals fastq files (no downsampling of reads), and subset to only have high-quality barcodes as outlined in Methods above. During concatenation, each of the four datasets was assigned a "batch" key. The concatenated anndata object was then used as input to scvi-tools for integration using the commands outlined in the tutorial here. The output of following these steps was a 10-dimensional latent space with batch-corrected embedding for cells from each of the four datasets. UMAP visualization was then generated using the RunUMAP command in Seurat with the 10-dimensional latent space as input. The dendrogram was generated using the BuildClusterTree command in Seurat, which constructs a phylogenetic tree relating the

'average' cell from each identity class. Tree is estimated based on the eigenvalue-weighted euclidean distance matrix constructed in latent-dimension space.

Unsupervised clustering of integrated dataset was performed using Seurat at a resolution of 0.3. Marker genes were then identified using the FindAllMarkers command. To investigate if the identified markers were conserved in their differential expression in scRNA-seq or snRNA-seq datasets, the integrated object (post-clustering) was first split based on sequencing techniques using the SplitObject command. Then, the identified marker genes were tested for differential expression using the FindAllMarkers command, with a logFC threshold of 0.25.

Integration of snRNA-seq and scRNA-seq data using Seurat

For integration with Seurat, scRNA-seq white preadipocyte (day-0) white-adipocyte (day-20) and snRNA-seq white preadipocyte (day-0) & white-adipocyte (day-20) datasets were defined as individual batches (a total of 4 batches). Integration was performed following the commands outline in this tutorial.

4.5 Data Access

Data related to this study is available upon request to the corresponding author. Analysis scripts are available upon request to the first author and corresponding author.

4.6 Declaration of Interests

There are no conflicts to declare.

4.7 Acknowledgements

This publication was supported by the National Institute of General Medical Sciences of the National Institutes of Health under award number R35GM124916. This publication was also supported by Grant No. CZF2019-002454 from the Chan Zuckerberg Foundation, and Grant No. R01DK102898 and K01DK125608 from the National Institutes of Health. Aaron Streets is a Chan-Zuckerberg Biohub Investigator and a Pew Scholar in the Biomedical Sciences, supported by the Pew Charitable Trusts. AG is supported by the UC Berkeley Lloyd Fellowship in Bioengineering. We would also like to thank Joshua Cantlon from Scienion for lending help with the CellenOne instrument, and the Gartner lab at UCSF for providing us with Multi-seq probes. This work is has been submitted to *biorxiv* and is currently under revision at *Genome Research* as Gupta et al. 2021:
Gupta et al., "Characterization of transcript enrichment and detection bias

in single-nuclei RNA-seq for mapping of distinct human adipocyte lineages". *bioRxiv*, 2021.03.24.435852

Chapter 5

Mapping the temporal transcriptional landscape of human white and brown adipogenesis using single-nuclei RNA-seq

5.1 Introduction

Adipogenesis is a highly orchestrated process, where networks of transcription factors (TFs) induce differentiation of adipose precursor cells (called preadipocytes) into mature adipocytes. This differentiation process is central to maintaining systemic energy balance in response to varying nutritional needs, with mature white adipocytes participating in energy storage and mature brown adipocytes participating in thermogenic energy expenditure. Notably, excess fat manifests itself in the development of syndromes such as Obesity, whereas deficiency of fat is associated with disorders such as lipodystrophy, thereby implicating a pathogenic role of imbalanced white and brown adipose tissue (WAT and BAT) expansion in such metabolic disorders. Consequently, a firm understanding of the molecular underpinnings of healthy adipogenic expansion is key in elucidating the pathophysiology and potential treatment modalities of such pathological cases.

Although studies in rodent models of adipogenesis have offered significant insights (see Chapter 1), recent studies have started focusing on adipogenic regulation using human model systems due to existing metabolic, functional, and physiological differences between the two species (see Chapter 1). For example, studies have compared transcriptomic profiles of human-derived adipose stem cells (ASCs) at multiple stages of adipogenic differentiation using techniques such as microarray analysis [373–375], bulk RNA-seq [74], siRNA screens [376], and RT-qPCR [376]. This has resulted in the identification of novel adipogenic TFs such as FGF11, and SOX4 [374]. However, such techniques provide a bulk gene expression measurement which are population-level ensemble measurements that do not take into ac-

count the inherent heterogeneity of asynchronously differentiating biological systems. With this approach, cellular heterogeneity cannot be resolved since variably expressed genes will be averaged or – if exclusively expressed in rare cells – completely missed. Moreover, experimental feasibility permits transcriptomic sampling at only coarse time-intervals during the differentiation period, thereby providing an incomplete and inaccurate view of gene expression dynamics. Recently, single-cell RNA-sequencing (scRNA-seq) has proven to be a powerful tool for unbiased transcriptomic profiling of complex tissues at an unprecedented resolution [25, 265, 266]. Furthermore, scRNA-seq has enabled high-resolution investigations into the transcriptional dynamics of differentiation in multiple biological systems, by capturing molecular differences in individual cells distributed along a continuum of maturation state (*pseudo-time* as a proxy; see Chapter 1). Indeed, within primary adipose tissue, recent investigations utilizing this technique have reported the molecular dynamics of adipocyte development in mice [88, 89]. However, inability to sample cells at all stages of differentiation, *in vivo*, has hindered the possibility of creating a high-resolution transcriptional map of adipogenesis using primary samples. Therefore, in this study, we mapped the transcriptional landscape of human white and brown adipogenesis using a unique, well-controlled, *in vitro* model system [76, 279], which enables isolation of differentiating preadipocytes at well-defined, multiple stages of development. In this *in vitro* system, paired white and brown primary preadipocytes were isolated from a defined anatomical location (the neck depot) of a single individual. This system, therefore, allowed us to measure transcriptional dynamics within and between white and brown lineages, while controlling for inter-individual variability that are typically associated with transcriptomic profiling of primary human adipose tissue, such as body mass index, genotype, and gender [98]. Preadipocytes from both lineages were isolated while preserving their intrinsic cellular heterogeneity and were then immortalized to allow for long-term *in vitro* cell-culture. Previously reported data demonstrated highly concordant molecular features between primary and immortalized preadipocytes, with *in vitro* differentiated adipocytes recovering gene expression profiles and functions of primary human neck BAT and WAT [76].

Using this *in vitro* human model system, we chose single-nuclei RNA-seq (snRNA-seq) as the preferred modality for performing a large-scale, time-course experiment on differentiating white and brown preadipocytes. Using snRNA-seq is important for equitable cell-recovery, since adipocytes suffer capture losses during single-cell extraction and isolation [98]. Furthermore, our previous work demonstrated the applicability of snRNA-seq in recovering similar cellular diversity and molecular differences as scRNA-seq in the adipose tissue [98]. Therefore, in this study, we isolated intact nuclei from differentiating white and brown preadipocytes at 5 stages of adipogenesis and performed droplet-microfluidics-based high throughput snRNA-seq. We then defined custom, white-/brown-lineage-specific adipogenic gene signatures that enabled high-resolution ordering of individual nuclei in increasing order of maturity. Using this nuclear ordering, our analyses revealed temporal regulation of distinct gene modules in both white and brown adipogenesis, each module highlighting the dynamics of biologically relevant functional processes. We investigated potential roles of temporally regulated genes in Obesity and further identified novel adipogenic as well thermogenic

transcription factors in humans. We also demonstrated the applicability of our adipogenic signature in assessing differentiation maturity of preadipocytes identified in publicly available scRNA-seq datasets. Overall, our study, for the first time, provides a comprehensive molecular understanding of both white and brown adipogenesis in humans. We believe this dataset to be an important resource and a reference to map the future *in vivo* adipogenic studies onto, both in healthy as well as metabolically diseased state.

5.2 Results

Large-scale snRNA-seq reveals a continuum of gene-expression during human white and brown adipogenesis

Intact nuclei were harvested from differentiating white and brown preadipocytes (see Methods) at 5 equally-spaced time-points during the 20-day adipogenic induction period (Fig. 5.1A). Nuclei harvested on day-0 were isolated from preadipocytes prior to adipogenic dif-

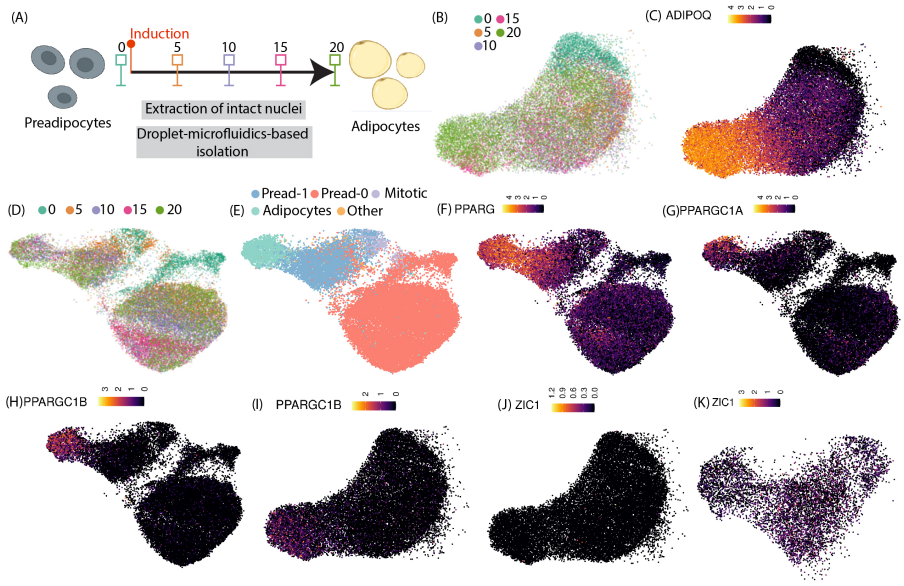

Figure 5.1: **snRNA-seq of differentiating white and brown preadipocytes** (A) Schematic of experimental outline for white and brown preadipocytes (B) UMAP visualization of white adipogenesis dataset integrated using scVI. Nuclei are colored by the day of harvestation (C) Normalized Expression of gene ADIPOQ (D) UMAP visualization of brown adipogenesis dataset integrated using scVI. Nuclei are colored by the day of harvestation. (E) Same plot as (D) but nuclei are colored by clusters identified using unsupervised clustering. Also see Fig. C.2. (H) to (K) Normalized expression of marker genes in white or brown adipogenesis dataset.

ferentiation. Isolated nuclei were subjected to droplet-based snRNA-seq, followed by QC

analyses (see Methods). In total, we recovered 25,339 high-quality white nuclei and 27,568 high-quality brown-nuclei, with 2000-6000 genes detected per nuclei.

Independent unsupervised clustering of differentiating white preadipocytes revealed a homogenous cluster of nuclei on day-0 and day-5, with capture and detection of differentiating adipocytes after day-10 (Fig. C.2A). All 5 white adipogenic libraries were integrated using scVI-tools, revealing ordering of cells with increasing maturity, starting from early precursor state (day-0) to mature adipocyte state (day-20; Fig. 5.1B and 5.1C). Adipogenic transcriptional signature analysis, where each cell is assigned a score based on expression of previously identified adipogenic genes, revealed a monotonically increasing trend with day of harvestation, thereby confirming a higher fraction of mature adipocytes on later days (Fig. C.2C). Notably, the spread of adipogenic signature score also increased with the day of harvestation, highlighting the asynchronous behavior of adipogenic differentiation in our in-vitro model system.

For differentiating brown preadipocytes, we had previously demonstrated underlying heterogeneity within day-0 (preadipocyte-0 and preadipocyte-1, Fig. C.2B). Independent analysis of brown nuclei harvested at later time-points (day-5 to day-20) further revealed similar preadipocyte heterogeneity, along with capture and detection of mature adipocytes (Fig. C.2B). Notably, integrative analysis of all 5 time-points revealed two contrary transcriptional responses to the induction media, with preadipocyte-1 differentiating into mature adipocytes, and preadipocyte-0 exhibiting a non-adipogenic response (Fig. 5.1D and 5.1E). Furthermore, differential expression analysis revealed strong up-regulation of adipogenic master regulator PPARG, and thermogenic transcription factors PGC1A and PGC1B in the adipogenic trajectory only (Fig. 5.1F to 5.1H). Notably, pathway analysis using genes up regulated as part of the non-adipogenic response revealed enrichment of FOXO1's transcriptional activity (Fig. C.2D and C.2E, see Methods), a known repressor of PPARG [377, 378], suggesting a possible role of FOXO1 in inhibiting an adipogenic response in preadipocyte-0 cell-type.

Focusing on the adipogenic response in differentiating brown preadipocytes, mean score for the adipogenic transcriptional signature was mostly increasing with the day of harvestation, except for day-15 which had the highest score because of the highest differentiation efficiency. Joint unsupervised clustering of all adipogenic cells revealed four clusters, referred to as populations 0, 1, 2, and 3 (Fig. C.2F), and indeed, when considering only mature adipocytes (cluster 3), day-20 nuclei had the highest adipogenic score, validating recovery of more mature adipocytes (albeit at a lower rate) as compared to day-15 (Fig. C.2G and C.2H). As expected, contrary to white adipocytes, brown adipocytes had an up-regulated expression of the thermogenic marker gene PGC1B (Fig. 1F, 1H and 1I) and brown-adipocyte-specific marker gene ZIC1 (Fig. 5.1J and 5.1K). Moreover, differential expression followed by transcription factor enrichment analysis between mature white and brown adipocytes (day-20) identified highest activity for thermogenic TFs FOXS1 (Heglind et al. 2005) and FOXC2 (Cederberg et al. 2001) in brown adipocytes. Therefore, molecular profiling of brown adipocytes validates the emergence of a functionally active thermogenic response in this cell-type. Overall, our experimental and sampling strategy allows us to capture a spectrum of cell-states undergoing differentiation into human white or brown fat.

Pseudotemporal ordering of differentiating preadipocytes identifies dynamics of key biological processes during adipogenesis

Acquisition of many single-nuclei transcriptomes with high temporal resolution created the possibility of reconstructing adipogenic developmental trajectory by ordering individual nuclei along a pseudo-time. To achieve this, we defined custom white-/brown-adipogenesis-specific gene signatures (Note S1), whose score was used as a proxy for pseudo-time (Note C.2). Notably, our custom-defined gene signatures only consisted of genes monotonically increasing in expression from immature preadipocytes to mature adipocytes (Note C.2), thereby providing a high dynamic range as well as pseudo-temporal resolution to identify minor transcriptomic differences along cellular differentiation states.

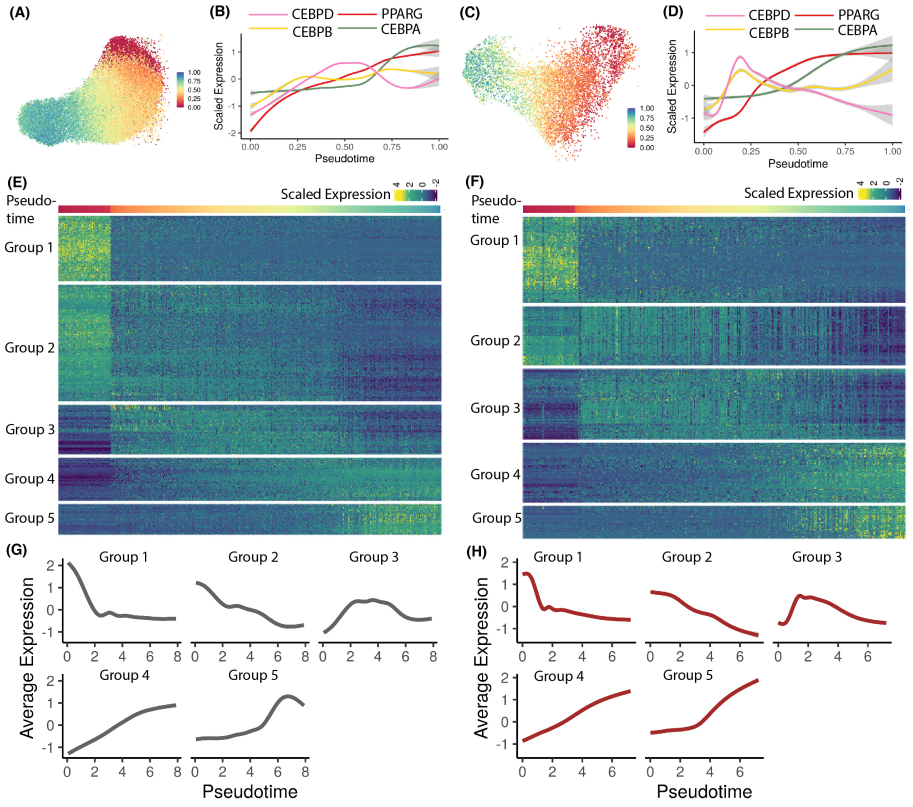

Figure 5.2: **pseudo-temporal ordering of differentiating white and brown preadipocytes** (A) and (C) ordering of differentiating white and brown preadipocytes. (B) and (D) Expression dynamics of adipogenic TFs with pseudotime in white dataset (B) and brown dataset (D). (E) and (F) Expression dynamics of temporally regulated genes in white (D) and brown (F) dataset. Genes are rows and nuclei are column, ordered by increasing pseud-time. (G) and (H) Smoothed expression dynamics of genes in each module in white (G) and brown (H) dataset.

Using our white-/brown-adipogenesis-specific gene signature, differentiating preadipocytes

were ordered in increasing maturation state (Fig. 5.2A and 5.2C). As expected, when presented in pseudotime, expression dynamics of key adipogenic TFs CEBPB, CEBPD, PPARG and CEBPA accurately reflected known biology (Fig. 5.2B and 5.2D), with early induction of CEBPB, followed by sustained expression in response to insulin (see Methods) [379], early induction and transient expression of CEBPD, stable increase in expression of PPARG, and late induction of CEBPA, thereby validating our cell-ordering strategy for both white and brown fat development. Next, dynamically regulated genes were identified by grouping nuclei into distinct pseudo-temporal bins (Fig. C.3A and C.3B, see Methods) and performing differential expression testing for each bin against the first and last pseudo-temporal bins (logFC > 1 and FDR < 0.05). In total, we identified 596 and 454 temporally expressed genes during white and brown adipogenesis respectively, which grouped primarily into five clusters using unsupervised clustering (Fig. 5.2E to 5.2H): immediately down-regulated (Group 1), gradually down-regulated (Group 2), transiently up-regulated (Group 3), gradually up-regulated (Group 4), and lately up-regulated (Group 5).

Focusing on gene clusters identified in white adipogenesis, genes undergoing immediate down-regulation (Group 1) primarily included cell adhesion molecules (CAMs) such as ITGB8, ITGA11, and ITGBL1, as well as growth factors such as VEGFA, VEGFC, FGF2, and FGF5 (Fig. C.3C). Findings from previous studies agree with our observations, with reported downregulation of such integrin-associated genes during adipogenesis [380–382], and known anti-adipogenic traits of above-mentioned growth factors [383–385]. Notably, extensive remodeling of the extracellular matrix (ECM) is critical for adipogenesis [386, 387], and CAMs serve as contact points between cells and ECM. Therefore, the disruption of cellular-ECM contacts, achieved via down-regulation of CAMS, becomes critical for ECM remodeling [388]. Interestingly, Group-1 genes also included ECM components itself such as COL1A1 and FN1, which agrees with previous reports of progressive degradation for collagen type 1 and fibronectin [342, 388] during adipogenesis.

Group 2 genes primarily included ECM structural components (Fig. C.3D) such as collagen types -1, -3, and -5, undergoing gradual down-regulation. Notably, multiple previous reports have demonstrated adipogenic remodeling of the ECM, with such fibrillar ECM components degrading to pave way for basement-membrane-type ECM components such as collagen-4 [389, 390]. Indeed, collagen-4 was observed to be gradually increasing during white fat development in our dataset (Fig. C.3E). Our findings also suggest that cell-ECM contact disruption is followed by ECM remodeling during early adipogenesis. Moreover, Group 2 also included down-regulated cytoskeletal components such as ACTB, TUBB, and VIM (Fig. C.3D), which agrees with our understanding that transition from preadipocytes to mature adipocytes involves significant reorganization of the cytoskeleton, involving down regulation of proteins such as actin and tubulin [391, 392].

Group 3 genes (transient up-regulation) mostly consisted of protease inhibitors such as TIMP3 and SERPINF1 (Fig. C.3F). Protease inhibitors serve as ECM constructive enzymes, antagonizing the ECM degradation activity of metalloproteases such as MMPs, ADAMs, and ADAMTSs [393]. Notably, during white adipogenesis, all such metalloproteases were down-regulated in our dataset (Fig. C.3G). Therefore, our results indicate an initial ECM

degeneration activity by metalloproteases, followed by a shift toward ECM regeneration via activity of protease inhibitors. Overall, our analysis of dynamically down-regulated genes (Groups -1, -2, and -3) highlights the importance of an interplay between cell adhesion contact disruption, ECM turnover and cytoskeletal remodeling in the progression of human white adipogenesis.

As expected, Group-4 primarily consisted of canonical adipogenic genes such as PLIN1, FABP4, CD36 and adipogenic transcriptional regulators such as PPARG, MLXIPL [394], and ZBED3 [395] undergoing gradual up-regulation. Group-5, on the other hand, included lipogenic genes such as FASN, ACSL1, GPAM and lipogenic transcription factor NR1H3 [396], suggesting a delayed onset of lipid biosynthesis response as compared to an adipogenic response. This agrees with pathway analysis which revealed enrichment of adipogenic regulation terms in Group-4 (Fig. C.3H) and fatty acid biosynthesis terms in Group-5 (Fig. C.3I). Therefore, our results indicate upregulation of adipogenic and lipogenic response during white fat development, with a possible delay between the two.

Next, focusing on gene modules identified in brown adipogenesis, Group 1 included proliferation marker genes such as TOP2A, CCND1, and MKI67, that undergo immediate down-regulation as differentiating preadipocytes exit from a proliferative state to growth arrested state, a transition required for the progression of adipogenesis [397]. A lack of such regulation for cell-cycle markers during white adipogenesis is likely due to an already growth-arrested state of day-0 nuclei isolated from white preadipocytes, as compared to a more proliferative state of day-0 nuclei isolated from brown preadipocytes (Fig. C.3J). Notably, like white adipogenesis, an immediate down-regulation for ITGA11, as well as multiple CAMs was also observed during brown adipogenesis.

Focusing on Group 2, like white adipogenesis, gradually down-regulated genes primarily included cytoskeletal components such as ACTB, TUBB, and VIM (Fig. C.3K). However, unlike white adipogenesis, fibrillar collagen components such as collagen types -1, -3, and -5 were clustered in Group-3 undergoing initial up-regulation (Fig. C.3L). This initial increase in expression was likely to provide a fibrillary-type ECM to early proliferating brown preadipocytes, until such cells reached a growth arrested state [389]. Indeed, such fibrillar collagen components were enriched in day-0 white preadipocytes as compared to day-0 brown preadipocytes, with the former cells having a more growth arrested state (Fig. C.3M and C.3J). Finally, like white adipogenesis, collagen types -1, -3, and -5 eventually undergo down-regulation, with a consistent increase in expression of basement-membrane-type collagen-4 (Fig. C.3N). Moreover, like white adipogenesis, protease inhibitors were also enriched in Group-3 (Fig. C.3L), suggesting a similar interplay of ECM degradation and construction during early stages of brown adipogenesis.

Group 4 genes were enriched for transcriptional regulators of adipogenic and lipogenic response such as PPARG, FABP4, SREBF1 [398, 399], and NR1H3 [396], whose expression stably increases during the course of brown fat development. Pathway analysis also identified enrichment of fatty acid biosynthesis, and lipid metabolism associated terms in Group 4 genes (Fig. C.3O). However, this observation was different from white adipogenesis, where an adipogenic response was followed by a more delayed lipogenic response. Instead, brown

adipogenesis was marked by a delayed onset of a thermogenic response (Group 5), as observed by a late induction of genes such as PGC1A, PGC1B, and PRKAG2. Furthermore, pathway analysis identified an enrichment of AMPK-associated lipolytic pathways [400, 401], as well as mitochondrial biogenesis pathways, further confirming the emergence of a thermogenic response (Fig. C.3P). Overall, our results suggest a consistent adipogenic & lipogenic response during brown fat development, followed by a delayed thermogenic response.

Next, we focused on genes exclusively regulated during brown adipogenesis, to better understand the molecular underpinnings that regulate the development of energy-spending fat. Notably, examination of temporally expressed genes in white and brown adipogenesis revealed exclusive regulation of autophagic pathways in the latter (Fig. C.3Q), with autophagic genes such as heat-shock proteins (HSPs) undergoing down-regulation as brown fat development progresses. Notably, autophagic pathways are reported to regulate early brown fat development [402], with subsequent down-regulation to suppress mitochondrial clearance via activity of HSPs [403, 404], thereby improving energy metabolism [405, 406]. Therefore, our results highlight an exclusive role of autophagy in regulating the development and thermogenic response of brown fat in humans.

High-resolution map of transcription factor dynamics identifies novel regulators of adipogenic and thermogenic response in humans

Given the key role of transcription factors in the formation and maintenance of different cell-types during development, we next focused on characterizing the dynamics of temporally regulated TFs during white fat development. In total, we identified 49 TFs with dynamic gene expression profiles during differentiation of white preadipocytes (Fig. 5.3A). As expected, 32/49 TFs identified in humans were observed to have similar expression dynamics as previously reported in murine models, thereby suggesting a high concordance of molecular features between the two species (Fig. 5.3A). This included Group 1 anti-adipogenic TFs such as GLI2 (hedgehog signaling mediator) [407, 408], RBPJ (Notch signaling mediator) [409, 410], and AHRR [411], Group 2 anti-adipogenic TFs such as TCF4 TCF12 (mediator of Wnt/B-catenin) [412], and SMAD3 (mediator of TGFB pathway) [413], Group 4 pro-adipogenic TFs such as PPARG, MLXIPL [394], and ZBED3 [395], and Group 5 pro-lipogenic TF NR1H3 [396].

Notably, we also identified 17 TFs with no previous reports of their adipogenic regulatory behavior, either in mice or in humans (Fig. 5.3A). This included TFs such as KLF12, ZEB2, CREB3L2, and MEF2A, whose homologous partners KLF8 [414], ZEB1 [285], CREB5 [69], and MEF2D [415] are known regulators of adipogenesis in rodents. Of the 17 TFs, 6 were also regulated during brown adipogenesis (Fig. 5.3C), thereby suggesting a common regulatory approach for these TFs in both white and brown fat development. Notably, based on the transcription factor binding site analysis, of these 6 TFs commonly regulated in both white and brown adipogenesis, only RFX8 was also commonly enriched in both lineages

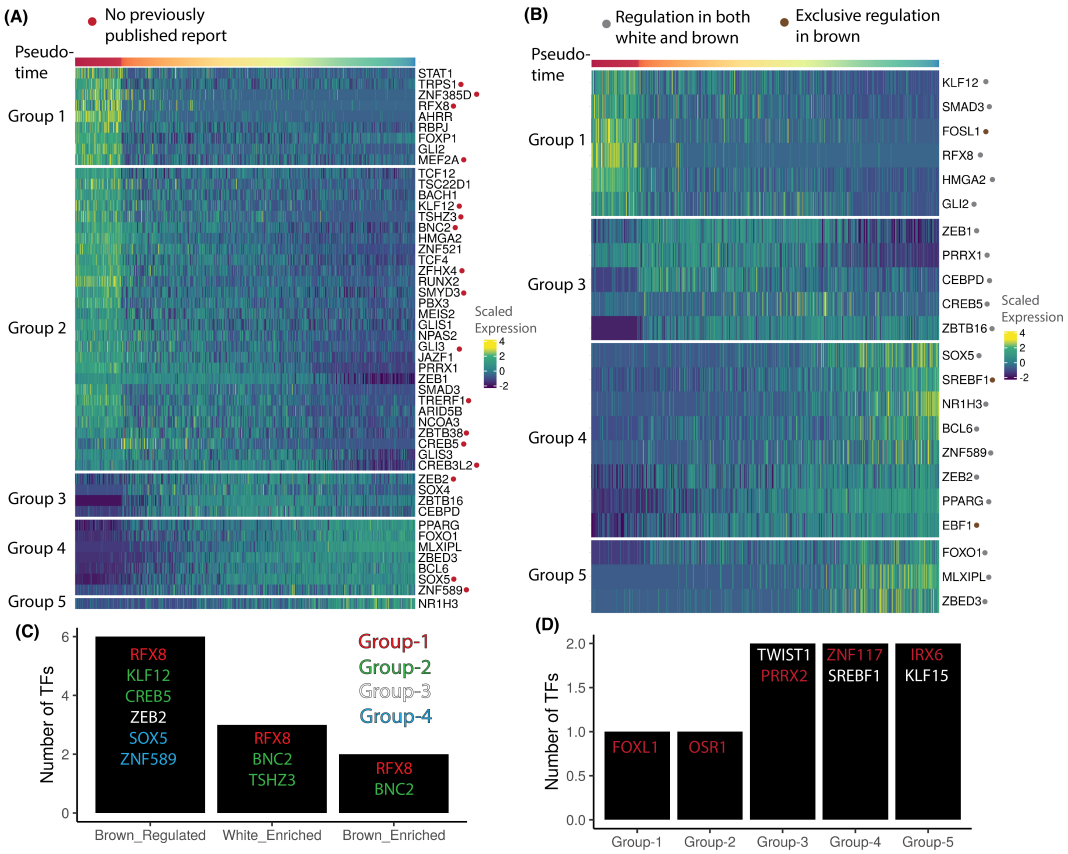

Figure 5.3: **Temporally regulated TFs during white and brown adipogenesis** (A) and (B) TFs dynamically regulated during white (A) and brown (B) adipogenesis, grouped based on their module annotation. (C) Characterization of novel TFs identified in white adipogenesis for involvement in brown adipogenesis, as well as TF enrichment analysis. (D) Distribution of TFs enriched in brown adipogenesis (identified using TFEA) by their module annotation. Highlighted in red are TFs with no prior literature for their involvement in regulating a thermogenic response.

(Fig. 5.3C), thereby providing a potentially novel lineage-agnostic target to investigate fat development in humans.

Besides identifying adipogenic TFs, another major goal in the adipose community is to identify novel thermogenic TFs that can augment an energy-spending response in mature adipocytes. Typically, thermogenic TFs are identified based on differential enrichment of genes in BAT over WAT. While such a strategy is applicable in rodents, where BAT form a homogenous interscapular depot, it is inapplicable in humans where BAT is found interspersed within WAT. Moreover, such a strategy lends weight to TFs that are highly expressed in BAT, with no information on temporal regulation of these TFs during brown adipogenesis. Here, using our time-resolved dataset, we identified TFs SREBF1, EBF1, and FOSL1 that

were exclusively regulated during brown adipogenesis (Fig. 5.3B). Notably, multiple previous studies have demonstrated a pro-thermogenic role for TFs SREBF1 [398, 399] and EBF1 [416], but no previous study has reported a thermogenic role for FOSL1, thereby suggesting a potentially novel approach to investigate thermogenesis using this gene. Next, we identified 8 TFs with exclusive activity in brown adipogenesis using binding-site enrichment analysis (Fig. 5.3D). This included thermogenic regulatory TFs such as SREBF1 [398, 399], KLF15 [417], and TWIST1 [418, 419], along with TFs FOXL1, ZNF117, IRX6, OSR1, and PRRX2, whose involvement in the context of thermogenic response has never been previously reported (Fig. 5.3D). Notably, ZNF117 and IRX6 were enriched in up-regulated genes (Group-4 and Group-5), thereby suggesting a potentially pro-thermogenic role for these genes. This is in agreement with a pro-thermogenic role for TFs SREBF1 and KLF15, which are also enriched in up-regulated genes (Group-4 and Group-5; Fig. 5.3D).

Temporally regulated white and brown adipogenic genes are implicated in Obesity based on GWAS and Bulk RNA-seq studies

A fundamental motivation behind generating the transcriptional landscape of human adipogenesis is to better our understanding of metabolic disease pathology, with the broader goal of potentially defining novel molecular targets for therapeutic intervention. Besides gene expression profiling, genome-wide association studies have also been critical in linking genetic variants to metabolic disease risk, thereby vastly improving our understanding of obesity genetics. Here, we took an integrative approach to further analyze our adipogenic-molecular findings in light of recent GWAS studies and asked the question: are genes dynamically regulated in human adipogenesis also linked to metabolic traits associated with increased obesity risk?

Using publicly available GWAS datasets, we identified SNPs located within a gene that are associated with metabolic traits such as the BMI, waist circumference, and hip circumference. In total, we worked with datasets from over 15 studies identifying over 1000 SNPs localized within the genic regions of 984 distinct genes. 77/984 genes were observed to be temporally regulated during differentiation of white preadipocytes in our dataset. A majority of these genes belonged to Group-2 (Fig. 5.4A), which was associated with ECM and cytoskeletal remodeling during adipogenesis. This is in line with dysfunctional ECM remodeling being a hallmark of Obesity, where excessive lipid accumulation in adipocytes provokes an excess of deposition of ECM components such as collagens, elastin, and fibronectin in the adipose tissue [420, 421]. Of the 77 genes, 15 genes were transcription factors that were both temporally regulated during white adipogenesis and associated with metabolic disease risk traits (Fig. 5.4B). Focusing on brown adipogenesis, exclusively regulated TF EBF1 was also associated with metabolic disease risk traits, thereby highlighting the importance of both white and brown fat in maintaining a healthy metabolic function.

Besides GWAS, RNA-seq and microarray-based investigations have also been critical in identifying potentially therapeutic molecular targets based on differential gene regulation

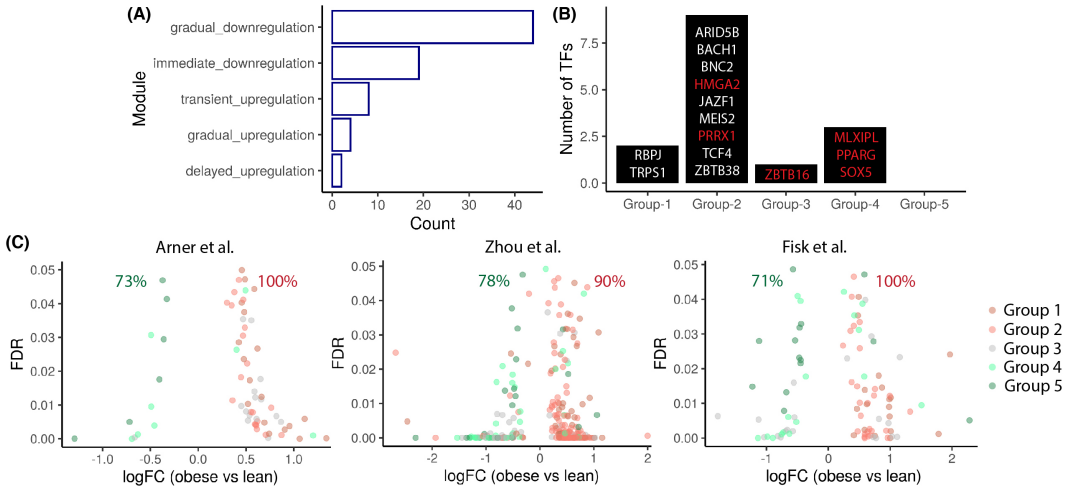

Figure 5.4: **Implication of white adipogenesis temporally regulated genes in Obesity using GWAS and RNA-seq** (A) Distribution of genes temporally regulated during white adipogenesis, that are also associated with high metabolic disease risk using GWAS (B) Distribution of TFs temporally regulated during white adipogenesis, that are also associated with high metabolic disease risk using GWAS (C) Volcano plot of genes DE in lean vs obese human phenotype across 3 different studies. The percentage indicates percent of Group-1 and Group-2 DE genes that are enriched in obese samples (in red) or percent of Group-4 and Group-5 DE genes that are enriched in lean samples (in green).

in lean vs obese humans. Typically, such studies utilize tissue specimens isolated from the subcutaneous abdominal fat depot in humans as a representative WAT sample. Therefore, utilizing our transcriptome profiling dataset in white adipogenesis, we investigated differential expression of temporally regulated genes in the context of human lean vs obese metabolic phenotype. We worked with 3 publicly available datasets in total, 2 of which profiled gene expression using bulk RNA-sequencing[422, 423] and the third using microarray [424]. Notably, of all the temporally downregulated genes that were differentially expressed, a large majority were enriched in obese samples (Fig. 5.4C). This is expected since downregulated genes primarily include CAMs and fibrillar ECM components, which exhibit increased accumulation and production during adipose tissue fibrosis in Obesity. Surprisingly, of all the temporally upregulated genes that were differentially expressed, which primarily included adipogenic and lipogenic markers, a large majority were enriched in lean samples in all 3 studies (Fig. 5.4C). Notably, previous studies report similar trends [425, 426], with possible downregulation of genes characteristic of adipocyte differentiation in obese samples due to hypertrophic adipose tissue expansion [427, 428], as well as adipocyte de-differentiation [429].

Adipogenic transcriptional signature enables assessment of cell maturation state amongst distinct metabolic phenotypes and cell-types

In this study, we defined custom white and brown adipogenic gene signatures, whose score (pseudo-time) was used for ordering differentiating preadipocytes along a continuum of maturation states. Since genes were selected for their monotonically increasing expression, our custom-defined signatures provided a high pseudo-temporal resolution for ordering cells. Hence, we utilized our high-resolution adipogenic signatures to quantitatively investigate differences in cell maturation state in primary adipose tissue isolated from mice and humans.

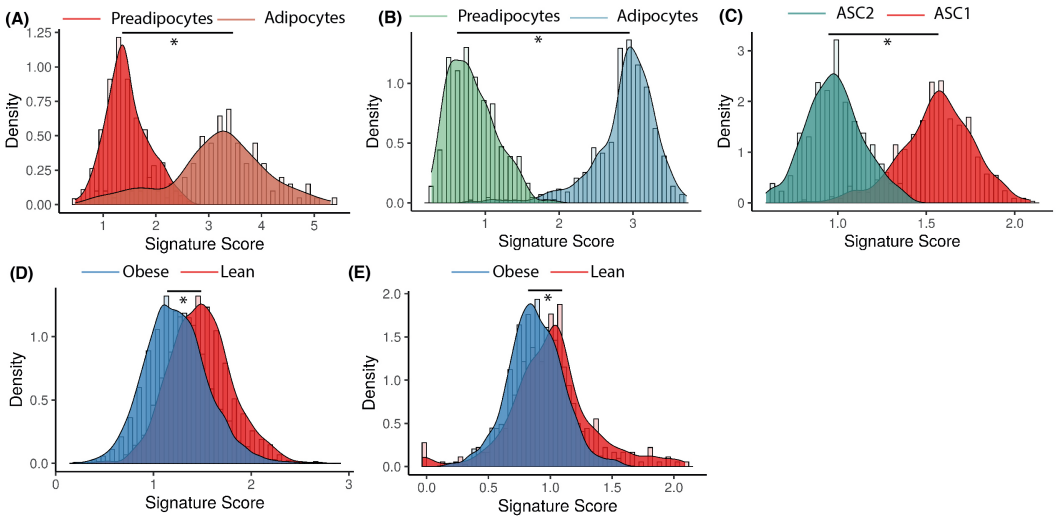

Figure 5.5: **Application of lineage-specific gene signatures to publicly available scRNA-seq datasets** (A) and (B) Distribution of signature score between preadipocytes and adipocytes in WAT (A) and BAT (B). (C) Distribution of signature score between ASC1 and ASC2 cell-types identified in Hepler et al. [347] (D) and (E) Distribution of signature score in Preadipocytes (D) and APCs (E) derived from lean and obese patients in Hildreth et al. [430]

Applicability of adipogenic gene signatures in capturing differences in cellular maturation state was first validated using a snRNA-seq dataset of primary white and brown adipose tissue [273] isolated from the same anatomical location as our *in vitro* model system (neck region, see Methods, Fig. C.4A and C.4B). As expected, our analysis revealed a significantly higher signature score for mature white and brown adipocytes, as compared to respective preadipocytes (Fig. 5.5A and 5.5B). Recently, multiple scRNA-seq studies identified ASC1 and ASC2 cells as two types of white adipocyte precursors (APCs) in mice [347–349] (Fig. C.4C to C.4E), with *in vivo* studies revealing a transition of ASC2 into ASC1 prior to becoming adipocytes, thereby suggesting a less committed progenitor state for ASC2 cells

[308, 348]. Indeed, our analysis also confirmed a similar trend, with a significantly higher signature score for ASC1 over ASC2 from three different studies (Fig. 5.5C, C.4F and C.4G). Therefore, our results demonstrate the applicability of our signatures in revealing major (preadipocytes vs adipocytes) as well as minor (ASc1 vs ASC2) maturation differences within adipose tissue.

After validating our signatures, next, we looked to investigate differences in cell maturation state between lean vs obese metabolic phenotypes in humans [430]. As mentioned previously, obese adipose samples were marked with a downregulation of adipogenic genes over lean samples, suggesting that Obesity may put brakes on healthy adipogenic differentiation. Indeed, a recent scRNA-seq study of lean and obese patients reported a higher fraction of preadipocytes, and a lower fraction of APCs in the SVF (WAT) of lean patients, suggesting an accelerated development from APCs to preadipocytes (Fig. C.1A to C.1D) in the healthy metabolic phenotype [430]. Therefore, we hypothesized that Preadipocytes and APCs themselves are less mature in obese patients because of a decelerated adipogenic development in this phenotype. We tested this hypothesis by calculating white adipogenic signature score in lean and obese patients for each cell-type. Indeed, our analysis revealed a significantly higher score for both Preadipocytes and APCs in lean samples (Fig. 5.5D and 5.5E), thereby validating our hypothesis.

5.3 Conclusion

In our work, we show the power of single nuclei transcriptome analysis to decipher the transcriptional dynamics of human white and brown fat development using an in vitro model system. We ordered white and brown cells by their progression through differentiation based on score for custom-defined adipogenic gene signatures. Our analysis illustrates the continual nature of adipogenesis, where cells progressively transit through five transcriptional modules that result in the generation of mature white and brown adipocytes. Through trajectory comparison, we also identify novel TFs potentially involved in regulation of an adipogenic or a thermogenic response.

Interestingly, integrative analysis of differentiating brown preadipocytes identified a precursor population with non-adipogenic response (Preadipocyte-0). Previously, non-adipogenic cell-types called Aregs were identified, which inhibit adipogenesis of other APCs in a paracrine fashion. However, Aregs are CD142+ which, in our dataset, was not expressed in the Preadipocyte-0 population. Recently, multiple scRNA-seq studies reported the existence of primarily two APC populations in mice referred to as Asc1 and Asc2. Similar to adipogenic brown preadipocytes (Preadipocyte-1), Asc2 exhibited pro-inflammatory and pro-fibrotic phenotype and positive expression of genes PI16 and MFAP5. Interestingly, functional investigations into Asc2 and Asc1 cells revealed Asc2 inhibiting the differentiation of Asc1 cells in vitro. Therefore, it is plausible that Preadipocyte-1 cells identified in our study may be functioning in a manner similar to Asc2, possibly to maintain adipocyte turnover. However, further functional investigations are necessary to validate the existence and functionality of

the two cell types in primary brown adipose tissue.

Within both white and brown adipogenesis, the underlying transcriptional program was revealed to be highly coordinated. Notably, ECM remodelling was one of the earliest exhibited response upon induction of differentiation. Although down-regulation of fibrillar ECM and up-regulation of basement membrane ECM has been demonstrated during 3T3-L1 adipogenesis, the dynamics of ECM reorganization and its regulation during human adipogenesis are not well understood. Our findings provide a high-resolution view into the expression dynamics of specific ECM components during healthy adipose tissue expansion, as well as potential ECM remodeling regulated by an interplay of metalloproteases and its inhibitors. Therefore, our dataset here could be used as a reference to better understand Obesity-induced ECM remodeling, with the goal of identifying therapeutic targets to prevent the inflammation, and in the end fibrosis, in diseased adipose tissue.

One unique aspect of our model system is the isolation of white and brown preadipocytes from a single individual, and a single anatomical location. Such a paired isolation strategy accounts for confounding inter-subject and inter-depot variability, and enables investigation not just within, but also across white and brown lineages. Using our model system system, we identified RFX8 and SOX5 as new TFs with potential adipogenic regulatory activity in both white and brown fat development. Moreover, based on exclusive regulation/enrichment in brown adipogenesis, we also identified novel TFs ZNF117 and IRX6 with potential involvement in thermogenic regulation. Of course, rigorous functional investigations would be required to confirm the regulatory roles of these TFs in healthy as well as diseased conditions. Ultimately, identification of novel TFs helps discern specific pharmacological targets for stimulating metabolically healthy, as well as thermogenic fat development.

In order to assign a pseudo-time to differentiating human preadipocytes, we defined custom transcriptional signatures specifically associated with white or brown fat development. Scores for these transcriptional signatures could be utilized as a quantitative metric to investigate differences in cellular maturity across varying metabolic conditions, and for different anatomical locations. Such an analysis could help us better understand the differential roles of fat depots towards Obesity development and progression. Moreover, our gene signature provides a targeted list of temporally, and biologically relevant genes, which could be specifically profiled using techniques such as spatial transcriptomics, or in situ hybridization/sequencing, to better understand the spatial context of adipose tissue development in primary samples. Finally, our gene signature could also be used to understand adipocyte dedifferentiation, and investigate molecular differences between dedifferentiated preadipocytes, and differentiating preadipocytes, at similar stages of maturity.

Our findings here are based on key experimental aspects that must be considered critically. In this work, adipogenic transcriptional dynamics were investigated using an immortalized, in vitro system of human white and brown preadipocytes. Although this model system accurately recovers primary human WAT and BAT functional behavior, an important question that still remains is to what extent the same transcriptional mechanisms are also operational in vivo. Consequently, there is a need to comprehensively investigate the molecular circuitry of fat development in primary adipose tissue samples, and for such under-

takings, our adipogenic transcriptional landscape could serve as a reference. Interestingly, cellular profiling of adult primary adipose tissue rarely identifies cell-types across all stages of adipogenic differentiation, and hence, pediatric adipose tissue biopsies become more apt for such investigations. A second detail that is relevant for our study is that the current model system was isolated from a single individual, and from a single anatomical location. And as such, the transcriptional landscape generated here could be made even more comprehensive by isolating similar model systems across multiple individuals and depot locations. The third aspect is technical, and pertains to varying background mRNA levels associated with individual single-nuclei libraries. Single-nuclei extraction involves breaking apart the cellular matrix to isolate nuclei, which releases high amounts of debris and cytoplasmic mRNA. During droplet-based single nuclei isolation, this debris gets encapsulated in the droplet along with the nuclei, leading to background mRNA contamination. This varying mRNA contamination makes it challenging to identify nuclei that are at similar stages of differentiation but distributed across different harvestation days (different single-nuclei libraries). scRNA-seq dataset integration algorithms do mitigate this challenge partially, but there is a need for better snRNA-seq integration strategies, with algorithms to model background mRNA distribution.

Despite these considerations, our study takes the first steps towards understanding the nature of adipogenic differentiation at a high temporal and cellular resolution in humans. These findings will therefore serve as a resource for multiple efforts into investigating adipose tissue biology in health, as well as disease, ultimately enabling newer therapeutics for improved clinical tackling of Obesity.

5.4 Materials and Methods

Preadipocyte culture and adipogenic differentiation

Detailed protocol for maintenance, cryopreservation, and differentiation of white and brown preadipocytes are outlined in a different study (Shamsi and Tseng 2017). Briefly, for culturing preadipocytes, cells were grown in DMEM medium (Corning, 10-017-CV) supplemented with 10% vol/vol FBS and containing 1% vol/vol Penicillin-Streptomycin (Gibco). Cell culture was maintained at 37°C in a humidified incubator containing 5% vol/vol CO2. 80% confluent cells were passaged using 0.25% trypsin with 0.1% EDTA (Gibco, 25200-056) for a 1:3 split in a new 100 mm cell culture dish (Corning).

Prior to adipogenic differentiation, white and brown preadipocytes were allowed to grow up to 100% confluence in a 100 mm cell culture dish (Corning). After 48 hours at 100% confluence, growth media was replaced with adipogenic induction media every 48 hours for the next 20 days. Induction media was prepared by adding 1 mL FBS, 500 µl Penicillin-Streptomycin, 15 µl human Insulin (0.5 µM, Sigma-Aldrich, I2643-50MG), 10 µl T3 (2 nM, Sigma-Aldrich,T6397-100MG), 50 µl Biotin (33 µM, Sigma-Aldrich, B4639-100MG), 100 µl Pantothenate (17 µM, Sigma-Aldrich, P5155-100G), 1 µl Dexamethasone (0.1 µM, Sigma-

Aldrich, D2915-100MG), 500 µl IBMX (500 µM, Sigma-Aldrich, I7018-100mg), and 12.5 µl Indomethacin (30 µM, Sigma-Aldrich, I7378-5G) to 48.5 mL DMEM medium and sterile filter.

Nuclei isolation from differentiating preadipocytes

Nuclei were isolated from differentiating white and brown preadipocytes using an NP-40 based lysis buffer: To 14.7 mL nuclease-free water (Qiagen), 150 µL of Tris-Hydrochloride (Sigma, T2194), 30 µL of Sodium Chloride (5M; Sigma, 59222C), 45 µL of Magnesium Chloride (1M; Sigma, M1028), and 75 uL of NP-40 (Sigma, 74385) was added. Two 100 mm dishes were used for nuclei isolation from each preadipocyte type. 500 µL of NP-40 based lysis buffer was added to each 100 mm dish and a cell scraper was employed to release adherent cells from the plates. On day 10, 15, and 20, where cells had visible lipid droplet accumulation, dounce homogenizer was used on scraped out cells to separate out the lipids. Cells were then incubated with the lysis buffer for 5 minutes on ice in a pre-chilled 15 mL falcon tube. Cells were washed with ice-cold PBS supplemented with .2 U/µL RNase Inhibitor (Protector RNase Inhibitor; henceforth called wash buffer) 4 times by centrifuging at 500 rcf for 5 minutes at 4°C. Wash buffer was aspirated after the final round of centrifugation and nuclei were resuspended in the ice-cold wash buffer and filtered using a 40 um cell strainer. Final concentration was adjusted to $\sim$ 1000 nuclei/µL using a hemocytometer for downstream sequencing. Nuclei were also stained using 0.08% trypan blue dye to assess nuclear membrane integrity under brightfield imaging. For nuclear isolation on day 10, 15, and 20, the same protocol was implemented as mentioned above with the modification of using 1 mL lysis buffer for each 100 mm dish.

After preparing nuclei suspension, isolation was performed on the 10x Chromium platform and libraries prepared as per the manufacturer's protocol using v3 sequencing chemistry. All final libraries were sequenced on the Illumina NovaSeq platform to ˜ 100,000 reads per nuclei.

Sequencing Data Analysis

In total, we had 5 libraries each for the white and brown adipogenesis dataset. For each library, empty droplets were removed using CellBender [359], and doublets were removed using Scrublet [366] or DoubletDetection [367]. Using Seurat, low-quality clusters such as clusters with high MT content, clusters with cellular debris (as marked by the enrichment of translation terms in GO analysis), clusters enriched for empty/doublet barcodes were removed from downstream analysis. Integration of all 5 time points for white and brown dataset was performed using scVI-tools [341]. Post-integration, Seurat was used for unsupervised clustering, and differential gene expression analysis.

Pseudo-temporal Ordering

Slingshot Analysis

For white and brown adipogenesis dataset, integrated Seurat object was clustered at resolution = 0.4. Slingshot was then used to infer the trajectory using the cluster with highest contribution from day 0 as the starting cluster. For identifying temporally regulated genes, cells were clustered into 6 equally spaced pseudo-temporal bins. DGE was then performed for each bin against the 1st and last bin, and all genes with logFC > 1 and FDR < 0.05 were considered as temporally regulated. For identifying monotonically increasing genes when cells are ordered using Slingshot, genes were clustered using the ComplexHeatmap package, with k-means clustering algorithm, and the number of clusters set to 5.

Vision analysis

For both white and brown adipogenesis dataset, Vision was used to assign a score to each cell in the integrated Seurat object using the "HallmarkAdipogenesis" MSigDB signature. This score was used as a proxy for pseudotime. For identifying temporally regulated genes in white adipogenesis dataset, cells were distributed into bins defined using the command *cutpoints=c(-Inf,seq(0.15,0.6,0.15),Inf)*. For brown adipogenesis dataset, cells were distributed into bins defined using the command *cutpoints=c(-Inf,0.2,0.3,seq(0.4,0.6,0.2),Inf)*. Temporally regulated genes were identified using the same strategy as defined above. For identifying monotonically increasing genes when cells are ordered using Vision, genes were clustered using the ComplexHeatmap package, with k-means clustering algorithm, and the number of clusters set to 5.

Identifying lineage-specific gene signatures

Once monotonically increasing genes were identified using both Vision and Slingshot, the intersection of the two was taken to define lineage-specific gene signatures. These signatures were used as input to Vision to assign a score to differentiating white and brown preadipocytes, and used as a proxy for pseudotime.

Gene Module Clustering

For identifying temporally regulated genes in white dataset, when cells are ordered using lineage-specific gene signatures, cells were distributed into bins defined using the command *cutpoints=c(-Inf,0.1,seq(0.3,0.8,0.1),Inf)*. DGE was then performed for each bin against the 1st and last bin, and all genes with logFC > 1 and FDR < 0.05 were considered as temporally regulated. Clustering for genes was performed using Seurat with resolution set to 0.5. For brown adipogenesis dataset, same steps were used with cutpoints defined using the command *cutpoints=c(-Inf,0.15,0.225,seq(0.3,0.7,0.2),Inf)*.

GWAS Analysis

The GWAS dataset was downloaded from the GWAS catalog (gwascatalogv1.0-associationse100r2021-04-20.tsv) and subset to metabolic traits defined in Locke et al. [431] and Shungin et al. [432]. The catalog was further subset to SNPs that were mapped to a single gene.

Bulk RNA-seq and Microarray Analysis

RNA-seq datasets were downloaded from the GEO Accession Viewer using Accession GSE25401 GSE162653. Microarray data was downloaded from the journal's website (oby22950-sup-0007-TableS1.xlsx). Differential expression analysis for RNA-seq datasets was performed using the DESeq package in R.

Signature Scoring Analysis

For assessing the maturation of cells in publicly available scRNA-seq datasets, Vision was used to assign score to cells using our lineage-specific signatures. Datasets were downloaded from locations as mentioned in the original manuscript.

5.5 Data Access

Data related to this study is available upon request to the corresponding author. Analysis scripts are available upon request to the first author and corresponding author.

5.6 Declaration of Interests

There are no conflicts to declare.

5.7 Acknowledgements

This publication was supported by the National Institute of General Medical Sciences of the National Institutes of Health under award number R35GM124916. This publication was also supported by Grant No. CZF2019-002454 from the Chan Zuckerberg Foundation, and Grant No. R01DK102898 and K01DK125608 from the National Institutes of Health. Aaron Streets is a Chan-Zuckerberg Biohub Investigator and a Pew Scholar in the Biomedical Sciences, supported by the Pew Charitable Trusts. This work is currently unpublished. We would also like to that Prof. Yu-Hua Tseng, Prof. Farnaz Shamsi, Dr. Mary-Elizabeth Patti, and Dr. Arionas Efthymiou for their continuous feedback.

Chapter 6

Concluding Remarks

In my dissertation, I have focused on building novel platforms and approaches to investigate the heterogeneity and developmental lineages within the human white and brown adipose tissue. These investigations have been made possible by recent advancements in the field NGS, microfluidics, and single-cell RNA-sequencing, all together enabling high-throughput, unbiased, transcriptomic measurements in individual cells. Thus far, I have worked on comprehensively characterizing adipocyte identity using multiomic measurements, characterizing advantages and biases in single-nuclei RNA-seq as compared to single-cell RNA-seq in the context of adipose tissue, and generating a high-resolution transcriptional landscape of human white and brown adipogenesis. In the next paragraphs, I briefly discuss promising directions for my previously discussed work.

In Chapter 3, I present µCB-seq, a microfluidic platform that combines high-resolution imaging and sequencing of single cells. Such measurements are critical for comprehensively characterizing the cellular identity, by pairing phenotypic and genotypic measurements. However, with the current throughput of the platform limited to 10 cells, statistically relevant measurements remain challenging. I anticipate future directions to focus on increasing µCB-seq's throughput which can be achieved by microfluidic multiplexing strategies [248, 249], accompanied with automated barcode dispensing using platforms such as the cellenOneX1. Ultimately, throughput of µCB-seq is limited by imaging time. Automated stage-scanning can be implemented in µCB-seq to reduce imaging time, as cells are immobilized in a linear array of nanoliter-scale imaging chambers. Another strategy to reduce imaging time would be to obviating stage scanning completely, by trapping cells in a single chamber, followed by automated sorting into indivudal reaction lanes. More sophisticated AI-assisted strategies could also be used for fully automated cell-trapping and sorting. With an increased throughput in the next-generation of µCB-seq devices, I anticipate the application of this platform for investigating adipocyte heterogeneity using a variety of *in vitro* model systems. In order to extend paired imaging and sequencing measurements to primary samples, I anticipate future studies to focus on *in situ* hybridization/sequencing techniques, which bypass the need for tissue digestion, thereby preserving the integrity of resident adipocytes in the tissue, as discussed in Chapter 2.

In Chapter 4, I presented my work on characterizing transcript enrichment and detection biases in snRNA-seq as compared to scRNA-seq. One critical finding of this study was the presence of gene-length associated detection biases when including intronic reads for UMI quantification, an effect which gets exacerbated in snRNA-seq measurements because of high fraction of intronic mapped reads. Our analyses also revealed the role of background mRNA contamination in abating biological heterogeneity, when preparing snRNA-seq libraries. Both these technical artifacts contribute towards systemic differences between scRNA-seq and snRNA-seq datasets, resulting in relatively poor performance by existing integrative algorithms, particularly in highly heterogeneous primary tissue samples. I anticipate future work to focus on developing exclusive integrative algorithms for scRNA-seq and snRNA-seq datasets derived from same tissues, taking into account the above-mentioned technical artifacts for batch correction. Although my work in this chapter focused on the adipose tissue as the model system, my findings from this study are generalizable to other tissues. In future, I anticipate the utilization of the framework provided in this study for anticipating the appropriate sequencing technique depending on the biological question at hand.

In Chapter 5, I, for the first time, present a high-resolution temporal transcriptional landscape of human white and brown adipogenesis using snRNA-seq. In this study, I relied on using an *in vitro* system of human adipogenesis derived from a single individual, and from a single anatomical location (neck region). Naturally, future steps in this investigation could include extending this study to other individuals, as well as other anatomical locations, thereby enabling the generation of a robust, comprehensive adipogenic landscape across humans & depots. I also anticipate the utilization of our dataset to serve as a reference for *in vivo* investigation into the primary adipose tissue samples. Such samples could be derived across metabolic phenotype such as varying BMI, blood glucose level, diseased condition, etc. Future work would also benefit from connecting transcriptomic profiles derived in our study with the spatial locations of cells *in situ*. Such spatial measurements could also be targeted towards lineage-specific gene signatures defined in this study, thereby reducing the experimental as well as financial logistics associated with spatial transcriptomic measurements. Beyond transcriptomic measurements, measurement of chromatin accessibility, transcription factors, and surface proteins could additionally inform the regulatory networks governing commitment to human white and brown adipogenic lineages.

Overall, I hope that the technologies and data resources generated in my dissertation inspire future investigations into the role of adipose tissue in maintaining a healthy metabolic phenotype, so that we, as a civilization can get closer to providing novel therapeutic interventions for metabolic disorders such as Obesity.

Bibliography

[1] R. Hooke, *Micrographia*, BoD–Books on Demand, **1665**.

[2] H. Baker, *Philosophical Transactions of the Royal Society of London* **1739**, *41*, 503–519.

[3] A. J. Wollman, R. Nudd, E. G. Hedlund, M. C. Leake, *Open biology* **2015**, *5*, 150019.

[4] B. Xia, I. Yanai, *Development* **2019**, *146*, dev169854.

[5] A. Tolkovsky, C. Richards, *Neuroscience* **1987**, *22*, 1093–1102.

[6] L. B. Cohen, B. M. Salzberg, *Reviews of Physiology Biochemistry and Pharmacology Volume 83* **1978**, 35–88.

[7] G. Childs, **2014**.

[8] E. T. Floyd, M. D. JAMES, E. B. THOMPSON, *DNA* **1983**, *2*, 309–327.

[9] M. Bustin, D. Goldblatt, R. Sperling, *Cell* **1976**, *7*, 297–304.

[10] E. A. O'donnell, D. N. Ernst, R. Hingorani, *Immune network* **2013**, *13*, 43–54.

[11] D. C. Koboldt, K. M. Steinberg, D. E. Larson, R. K. Wilson, E. R. Mardis, *Cell* **2013**, *155*, 27–38.

[12] F. Tang, C. Barbacioru, Y. Wang, E. Nordman, C. Lee, N. Xu, X. Wang, J. Bodeau, B. B. Tuch, A. Siddiqui, et al., *Nature methods* **2009**, *6*, 377–382.

[13] J. W. Bagnoli, C. Ziegenhain, A. Janjic, L. E. Wange, B. Vieth, S. Parekh, J. Geuder, I. Hellmann, W. Enard, *Nature communications* **2018**, *9*, 1–8.

[14] S. Picelli, Å. K. Björklund, O. R. Faridani, S. Sagasser, G. Winberg, R. Sandberg, *Nature methods* **2013**, *10*, 1096–1098.

[15] T. Luo, L. Fan, R. Zhu, D. Sun, *Micromachines* **2019**, *10*, 104.

[16] A. M. Streets, Y. Huang, *Biomicrofluidics* **2013**, *7*, 011302.

[17] A. Streets, X. Zhang, C. Cao, Y. Pang, X. Wu, L. Xiong, L. Yang, Y. Fu, L. Zhao, F. Tang, Y. Huang, *Proc. Natl. Acad. Sci. U. S. A*, *2014*, 7048–7053.

[18] Y. Marcy, C. Ouverney, E. Bik, T. Losekann, N. Ivanova, H. Martin, E. Szeto, D. Platt, P. Hugenholtz, D. Relman, R. S. Quake, *Proc. Natl. Acad. Sci. U. S. A* **2007**, *104*, 11889–11894.

[19] S. J. Pamp, E. D. Harrington, S. R. Quake, D. A. Relman, P. C. Blainey, *Genome research* **2012**, *22*, 1107–1119.

[20] H. C. Fan, J. Wang, A. Potanina, S. R. Quake, *Nature biotechnology* **2011**, *29*, 51–57.

[21] J. Wang, H. C. Fan, B. Behr, S. R. Quake, *Cell* **2012**, *150*, 402–412.

[22] V. Lecault, A. K. White, A. Singhal, C. L. Hansen, *Current opinion in chemical biology* **2012**, *16*, 381–390.

[23] A. R. Wu, N. F. Neff, T. Kalisky, P. Dalerba, B. Treutlein, M. E. Rothenberg, F. M. Mburu, G. L. Mantalas, S. Sim, M. F. Clarke, S. R. Quake, *Nature Methods* **2014**, *11*, 41–46.

[24] A. A. Pollen, T. J. Nowakowski, J. Shuga, X. Wang, A. A. Leyrat, J. H. Lui, N. Li, L. Szpankowski, B. Fowler, P. Chen, N. Ramalingam, G. Sun, M. Thu, M. Norris, R. Lebofsky, D. Toppani, D. W. Kemp, M. Wong, B. Clerkson, B. N. Jones, S. Wu, L. Knutsson, B. Alvarado, J. Wang, L. S. Weaver, A. P. May, R. C. Jones, M. A. Unger, A. R. Kriegstein, J. A. West, *Nature Biotechnology* **2014**, *32*, 1053–1058.

[25] K. D. Birnbaum, *Annual review of genetics* **2018**, *52*, 203–221.

[26] A. R. Wu, J. Wang, A. M. Streets, Y. Huang, *Annual Review of Analytical Chemistry* **2017**, *10*, 439–462.

[27] A. A. Manzoor, L. Romita, D. K. Hwang, *The Canadian Journal of Chemical Engineering* **2021**, *99*, 61–96.

[28] T. Thorsen, R. W. Roberts, F. H. Arnold, S. R. Quake, *Physical review letters* **2001**, *86*, 4163.

[29] P. Umbanhowar, V. Prasad, D. A. Weitz, *Langmuir* **2000**, *16*, 347–351.

[30] H. Fan, G. Fu, S. Fodor, *Science* **2015**, *347*, 1258367.

[31] X. Han, R. Wang, Y. Zhou, L. Fei, H. Sun, S. Lai, A. Saadatpour, Z. Zhou, H. Chen, F. Ye, et al., *Cell* **2018**, *172*, 1091–1107.

[32] E. Macosko, A. Basu, R. Satija, J. Nemesh, K. Shekhar, M. Goldman, I. Tirosh, A. Bialas, N. Kamitaki, E. Martersteck, J. Trombetta, D. Weitz, J. Sanes, A. Shalek, A. Regev, S. McCarroll, *Cell* **2015**, *161*, 1202–1214.

[33] A. Klein, L. Mazutis, I. Akartuna, N. Tallapragada, A. Veres, V. Li, L. Peshkin, D. Weitz, M. Kirschner, *Cell* **2015**, *161*, 1187–1201.

[34] G. X. Zheng, J. M. Terry, P. Belgrader, P. Ryvkin, Z. W. Bent, R. Wilson, S. B. Ziraldo, T. D. Wheeler, G. P. McDermott, J. Zhu, M. T. Gregory, J. Shuga, L. Montesclaros, J. G. Underwood, D. A. Masquelier, S. Y. Nishimura, M. Schnall-Levin, P. W. Wyatt, C. M. Hindson, R. Bharadwaj, A. Wong, K. D. Ness, L. W. Beppu, H. J. Deeg, C. McFarland, K. R. Loeb, W. J. Valente, N. G. Ericson, E. A. Stevens, J. P. Radich, T. S. Mikkelsen, B. J. Hindson, J. H. Bielas, *Nature Communications* **2017**, *8*, DOI 10.1038/ncomms14049.

[35] A. Regev, S. A. Teichmann, E. S. Lander, I. Amit, C. Benoist, E. Birney, B. Bodenmiller, P. Campbell, P. Carninci, M. Clatworthy, et al., *elife* **2017**, *6*, e27041.

[36] S. Darmanis, S. A. Sloan, Y. Zhang, M. Enge, C. Caneda, L. M. Shuer, M. G. H. Gephart, B. A. Barres, S. R. Quake, *Proceedings of the National Academy of Sciences* **2015**, *112*, 7285–7290.

[37] A. Zeisel, A. B. Muñoz-Manchado, S. Codeluppi, P. Lönnerberg, G. La Manno, A. Juréus, S. Marques, H. Munguba, L. He, C. Betsholtz, et al., *Science* **2015**, *347*, 1138–1142.

[38] E. M. Kernfeld, R. M. Genga, K. Neherin, M. E. Magaletta, P. Xu, R. Maehr, *Immunity* **2018**, *48*, 1258–1270.

[39] M. J. Muraro, G. Dharmadhikari, D. Grün, N. Groen, T. Dielen, E. Jansen, L. Van Gurp, M. A. Engelse, F. Carlotti, E. J. De Koning, et al., *Cell systems* **2016**, *3*, 385–394.

[40] T. M. Consortium et al., *Nature* **2018**, *562*, 367–372.

[41] A. Fernandis, M. Wenk, Curr, *Opin. Lipidol* **2007**, *18*, 121–128.

[42] H. Sunshine, M. Iruela-Arispe, Curr, *Opin. Lipidol*, *2017*, 408–413.

[43] A. Higdon, A. Diers, J. Oh, A. Landar, M. V. Darley-Usmar, *Biochem. J* **2012**, *442*, 453–464.

[44] K. Bersuker, J. Olzmann, *Biochim. Biophys. Acta Mol. Cell Biol. Lipids*, *2017*, 1166–1177.

[45] A.-E. Saliba, I. Vonkova, A. Gavin, *Nat. Rev. Mol. Cell Biol* **2015**, *16*, 753–761.

[46] K. Ikeda, T. Yamada, *Frontiers in Endocrinology* **2020**, *11*.

[47] J. Tordjman in *Physiology and Physiopathology of Adipose Tissue*, Springer, **2013**, pp. 67–75.

[48] P. Seale, B. Bjork, W. Yang, S. Kajimura, S. Chin, S. Kuang, A. Scime, S. Devarakonda, H. M. Conroe, H. Erdjument-Bromage, et al., *Nature* **2008**, *454*, 961–967.

[49] P. Trayhurn, *Frontiers in physiology* **2018**, *9*, 1672.

[50] F. Lizcano, *International journal of molecular sciences* **2019**, *20*, 5058.

[51] J. Wu, P. Boström, L. M. Sparks, L. Ye, J. H. Choi, A.-H. Giang, M. Khandekar, K. A. Virtanen, P. Nuutila, G. Schaart, et al., *Cell* **2012**, *150*, 366–376.

[52] A. K. Ramirez, S. N. Dankel, B. Rastegarpanah, W. Cai, R. Xue, M. Crovella, Y.-H. Tseng, C. R. Kahn, S. Kasif, *Nature communications* **2020**, *11*, 1–9.

[53] A. Song, W. Dai, M. J. Jang, L. Medrano, Z. Li, H. Zhao, M. Shao, J. Tan, A. Li, T. Ning, et al., *The Journal of clinical investigation* **2020**, *130*, 247–257.

[54] J. M. Spaethling, M. Sanchez-Alavez, J. Lee, F. C. Xia, H. Dueck, W. Wang, S. A. Fisher, J.-Y. Sul, P. Seale, J. Kim, et al., *The FASEB Journal* **2016**, *30*, 81–92.

[55] H. Green, O. Kehinde, *Cell* **1975**, *5*, 19–27.

[56] V. Rizzatti, F. Boschi, M. Pedrotti, E. Zoico, A. Sbarbati, M. Zamboni, *European journal of histochemistry: EJH* **2013**, *57*.

[57] J. G. Granneman, P. Li, Y. Lu, J. Tilak, *American Journal of Physiology-Endocrinology and Metabolism* **2004**, *287*, E574–E582.

[58] M. Nagayama, T. Uchida, K. Gohara, *Journal of lipid research* **2007**, *48*, 9–18.

[59] L. B. Salans, J. W. Dougherty, et al., *The Journal of clinical investigation* **1971**, *50*, 1399–1410.

[60] O. Varlamov, R. Somwar, A. Cornea, P. Kievit, K. L. Grove, C. T. Roberts Jr, *American Journal of Physiology-Endocrinology and Metabolism* **2010**, *299*, E486–E496.

[61] H. A. Rinia, K. N. Burger, M. Bonn, M. Müller, *Biophysical journal* **2008**, *95*, 4908–4914.

[62] A. Gupta, G. F. Dorlhiac, A. M. Streets, *Analyst* **2019**, *144*, 753–765.

[63] T. N. Chen, A. Gupta, M. D. Zalavadia, A. Streets, *Lab on a Chip* **2020**, *20*, 3899–3913.

[64] M. I. Lefterova, Y. Zhang, D. J. Steger, M. Schupp, J. Schug, A. Cristancho, D. Feng, D. Zhuo, C. J. Stoeckert, X. S. Liu, et al., *Genes & development* **2008**, *22*, 2941–2952.

[65] E. D. Rosen, C. J. Walkey, P. Puigserver, B. M. Spiegelman, *Genes & development* **2000**, *14*, 1293–1307.

[66] M. Harms, P. Seale, *Nature medicine* **2013**, *19*, 1252–1263.

[67] Z. Wu, S. Wang, *Developmental biology* **2013**, *373*, 235–243.

[68] V. A. Payne, W.-S. Au, C. E. Lowe, S. M. Rahman, J. E. Friedman, S. O'Rahilly, J. J. Rochford, *Biochemical Journal* **2010**, *425*, 215–224.

[69] J. E. Reusch, L. A. Colton, D. J. Klemm, *Molecular and cellular biology* **2000**, *20*, 1008–1020.

[70] M. Shao, J. Ishibashi, C. M. Kusminski, Q. A. Wang, C. Hepler, L. Vishvanath, K. A. MacPherson, S. B. Spurgin, K. Sun, W. L. Holland, et al., *Cell metabolism* **2016**, *23*, 1167–1184.

[71] Q. Tong, G. Dalgin, H. Xu, C.-N. Ting, J. M. Leiden, G. S. Hotamisligil, *Science* **2000**, *290*, 134–138.

[72] I. Gerin, G. T. Bommer, M. E. Lidell, A. Cederberg, S. Enerback, O. A. MacDougald, *Journal of Biological Chemistry* **2009**, *284*, 10755–10763.

[73] A. Loft, I. Forss, S. Mandrup, *Trends in Endocrinology & Metabolism* **2017**, *28*, 104–120.

[74] A. Ehrlund, N. Mejhert, C. Björk, R. Andersson, A. Kulyté, G. Åström, M. Itoh, H. Kawaji, T. Lassmann, C. O. Daub, et al., *Diabetes* **2017**, *66*, 218–230.

[75] M. J. Betz, S. Enerbäck, *Diabetes* **2015**, *64*, 2352–2360.

[76] R. Xue, M. D. Lynes, J. M. Dreyfuss, F. Shamsi, T. J. Schulz, H. Zhang, T. L. Huang, K. L. Townsend, Y. Li, H. Takahashi, et al., *Nature medicine* **2015**, *21*, 760.

[77] D. E. Chusyd, D. Wang, D. M. Huffman, T. R. Nagy, *Frontiers in nutrition* **2016**, *3*, 10.

[78] S. J. Fitzgerald, A. V. Janorkar, A. Barnes, R. O. Maranon, *Journal of biomedical science* **2018**, *25*, 1–12.

[79] X. Liu, C. Cervantes, F. Liu, *Protein & cell* **2017**, *8*, 446–454.

[80] H. F. Bahmad, R. Daouk, J. Azar, J. Sapudom, J. Teo, W. Abou-Kheir, M. Al-Sayegh, *Cells* **2020**, *9*, 2326.

[81] S. Tritschler, M. Büttner, D. S. Fischer, M. Lange, V. Bergen, H. Lickert, F. J. Theis, *Development* **2019**, *146*, dev170506.

[82] W. Saelens, R. Cannoodt, H. Todorov, Y. Saeys, *Nature biotechnology* **2019**, *37*, 547–554.

[83] L. E. Byrnes, D. M. Wong, M. Subramaniam, N. P. Meyer, C. L. Gilchrist, S. M. Knox, A. D. Tward, J. Y. Chun, J. B. Sneddon, *Nature communications* **2018**, *9*, 1–17.

[84] J. A. Farrell, Y. Wang, S. J. Riesenfeld, K. Shekhar, A. Regev, A. F. Schier, *Science* **2018**, *360*.

[85] A. C. Habermann, A. J. Gutierrez, L. T. Bui, S. L. Yahn, N. I. Winters, C. L. Calvi, L. Peter, M.-I. Chung, C. J. Taylor, C. Jetter, et al., *Science advances* **2020**, *6*, eaba1972.

[86] Z. Steier, L. L. McIntyre, L. K. Lutes, T.-S. Huang, E. A. Robey, N. Yosef, A. Streets, *bioRxiv* **2021**.

[87] D. Meistermann, A. Bruneau, S. Loubersac, A. Reignier, J. Firmin, V. François-Campion, S. Kilens, Y. Lelièvre, J. Lammers, M. Feyeux, et al., *Cell Stem Cell* **2021**.

[88] A. K. Sárvári, E. L. Van Hauwaert, L. K. Markussen, E. Gammelmark, A.-B. Marcher, M. F. Ebbesen, R. Nielsen, J. R. Brewer, J. G. S. Madsen, S. Mandrup, *Cell Metabolism* **2021**, *33*, 437–453.

[89] R. B. Burl, V. D. Ramseyer, E. A. Rondini, R. Pique-Regi, Y.-H. Lee, J. G. Granneman, *Cell metabolism* **2018**, *28*, 300–309.

[90] A. Deutsch, D. Feng, J. E. Pessin, K. Shinoda, *International Journal of Molecular Sciences* **2020**, *21*, 4773.

[91] R. Gao, C. Kim, E. Sei, T. Foukakis, N. Crosetto, L.-K. Chan, M. Srinivasan, H. Zhang, F. Meric-Bernstam, N. Navin, *Nature communications* **2017**, *8*, 1–12.

[92] B. B. Lake, R. Ai, G. E. Kaeser, N. S. Salathia, Y. C. Yung, R. Liu, A. Wildberg, D. Gao, H.-L. Fung, S. Chen, et al., *Science* **2016**, *352*, 1586–1590.

[93] H. Wu, Y. Kirita, E. L. Donnelly, B. D. Humphreys, *Journal of the American Society of Nephrology* **2019**, *30*, 23–32.

[94] P. Rajbhandari, D. Arneson, S. K. Hart, I. S. Ahn, G. Diamante, L. C. Santos, N. Zaghari, A.-C. Feng, B. J. Thomas, L. Vergnes, et al., *Elife* **2019**, *8*, e49501.

[95] B. Lacar, S. B. Linker, B. N. Jaeger, S. R. Krishnaswami, J. J. Barron, M. J. Kelder, S. L. Parylak, A. C. Paquola, P. Venepally, M. Novotny, et al., *Nature communications* **2016**, *7*, 1–13.

[96] N. Habib, Y. Li, M. Heidenreich, L. Swiech, I. Avraham-Davidi, J. J. Trombetta, C. Hession, F. Zhang, A. Regev, *Science* **2016**, *353*, 925–928.

[97] Q. Liang, R. Dharmat, L. Owen, A. Shakoor, Y. Li, S. Kim, A. Vitale, I. Kim, D. Morgan, S. Liang, et al., *Nature communications* **2019**, *10*, 1–12.

[98] A. Gupta, F. Shamsi, N. Altemose, G. F. Dorlhiac, A. M. Cypess, A. P. White, M. E. Patti, Y.-H. Tseng, A. M. Streets, *bioRxiv* **2021**.

[99] M. Visram, M. Radulovic, S. Steiner, N. Malanovic, T. O. Eichmann, H. Wolinski, G. N. Rechberger, O. Tehlivets, *Journal of Biological Chemistry* **2018**, *293*, 5544–5555.

[100] F. Chiappini, A. Coilly, H. Kadar, P. Gual, A. Tran, C. Desterke, D. Samuel, J.-C. Duclos-Vallée, D. Touboul, J. Bertrand-Michel, et al., *Scientific reports* **2017**, *7*, 1–17.

[101] T. Cajka, O. Fiehn, *Chem*, *2014*, 192–206.

[102] H. Köfeler, A. Fauland, G. Rechberger, M. Trötzmüller, *Metabolites* **2012**, *2*, 19–38.

[103] L. Li, J. Han, Z. Wang, J. Liu, J. Wei, S. Xiong, Z. Zhao, *Int. J. Mol. Sci*, *2014*, 10492–10507.

[104] X. Han, R. Gross, *Mas Spectrom. Rev* **2005**, *24*, 367–412.

[105] M. Wang, C. Wang, R. Han, X. Han, *Prog. Lipid Res* **2016**, *61*, 83–108.

[106] L. Zhang, A. Vertes, *Angew. Chem. Int. Ed*, *2018*, 4466–4477.

[107] R. Zenobi, *Science* **2013**, *342*, 1243259.

[108] N. Tsuyama, H. Mizuno, T. Masujima, *Biol. Pharm. Bull* **2012**, *35*, 1425–1431.

[109] S. Rubakhin, E. Lanni, J. Sweedler, *Curr. Opin. Biotechnol* **2013**, *24*, 95–104.

[110] E. Lanni, S. Rubakhin, J. Sweedler, J., *Proteomics* **2012**, *75*, 5036–5051.

[111] K. Boggio, E. Obasuyi, K. Sugino, S. Nelson, N. Agar, N. J. Agar, *Expert Rev. Proteomics* **2011**, *8*, 591–604.

[112] N. Goto-Inoue, T. Hayasaka, N. Zaima, M. Setou, *Biochim. Biophys. Acta Mol. Cell Biol. Lipids* **2011**, *1811*, 961–969.

[113] D. Gode, D. Volmer, *Analyst* **2013**, *138*, 1289–1315.

[114] D. Touboul, A. Brunelle, O. Laprévote, *Biochimie.* **2011**, *93*, 113–119.

[115] A. Römpp, B. Spengler, *Histochem. Cell Biol* **2013**, *139*, 759–783.

[116] V. H. J. Pól, M. Strohalm, M. Volný, *Histochem. Cell Biol* **2010**, *134*, 423–443.

[117] S. Blanksby, T. Mitchell, *Annu. Rev. Anal. Chem* **2010**, *3*, 433–465.

[118] M. Maekawa, G. Fairn, J., *Cell Sci*, *2014*, 4812.

[119] S. Daemen, M. van Zandvoort, S. Parekh, M. Hesselink, *Mol. Metab* **2016**, *5*, 153–163.

[120] K. Yen, T. Le, A. Bansal, S. Narasimhan, J. Cheng, H. A, *Tissenbaum PLoS One* **2010**, *5*, 1–10.

[121] J. Sims, B. Rohr, E. Miller, K. Lee, T. Eng., P. C, *Tissue Eng. Part C* **2015**, *21*, 605–613.

[122] C. McPhee, G. Zoriniants, W. Langbein, P. Borri, *J. Biophys* **2013**, *105*, 1414–1420.

[123] K. Kim, S. Lee, J. Yoon, J. Heo, C. Choi, Y. Park, *Sci. Rep* **2016**, *6*, 36815.

[124] D. Débarre, W. Supatto, A.-M. Pena, A. Fabre, T. Tordjmann, L. Combettes, M. Schanne-Klein, E. Beaurepaire, *Nat. Methods* **2006**, *3*, 47–53.

[125] T. Bley, O. Wieben, C. François, J. Brittain, S. Reeder, *J. Magn. Reson. Imaging* **2010**, *31*, 4–18.

[126] J.-H. Hwang, C. Choi, *Exp. Mol. Med* **2015**, *47*, 139.

[127] K. Jurowski, K. Kochan, J. Walczak, M. Barańska, W. Piekoszewski, B. Buszewski, *Critical Reviews in Analytical Chemistry* **2017**, *47*, 418–437.

[128] J. Cheng, Y. Jia, G. Zheng, X. Xie, *Biophys. J* **2002**, *83*, 502–509.

[129] C. Freudiger, W. Min, B. Saar, S. Lu, G. Holtom, C. He, J. Tsai, J. Kang, X. Xie, *Science* **2008**, *322*, 1857–1861.

[130] A. Schönle, S. Hell, *Opt. Lett* **1998**, *23*, 325–327.

[131] R. Galli, O. Uckermann, E. Andresen, K. Geiger, E. Koch, G. Schackert, G. Steiner, M. Kirsch, *PLoS One*, *2014*, 110295.

[132] X. Nan, E. Potma, X. Xie, J, *Biophys* **2006**, *91*, 728–735.

[133] L. Gao, H. Zhou, M. Thrall, F. Li, Y. Yang, Z. Wang, P. Luo, K. Wong, G. Palapattu, W. S. C, *Biomed. Opt. Express* **2011**, *2*, 915.

[134] F.-K. Lu, S. Basu, V. Igras, M. Hoang, M. Ji, D. Fu, G. Holtom, V. Neel, C. Freudiger, D. Fisher, X. Xie, *Proc. Natl. Acad. Sci. U. S. A* **2015**, *112*, 11629, 11624.

[135] D. Zhang, M. Slipchenko, J. Cheng, *J. Phys. Chem. Lett* **2011**, *2*, 1248–1253.

[136] Y. Fu, H. Wang, R. Shi, J.-X. Cheng, *Opt. Express* **2006**, *14*, 3942.

[137] A. Hopt, E. Neher, *J. Biophys* **2001**, *80*, 2029–2036.

[138] K. Konig, U. Simon, K. Halbhuber, *Cell Mol. Biol* **1996**, *42*, 1181–1194.

[139] K. König, P. So, W. Mantulin, E. Gratton, *Opt. Lett* **1997**, *22*, 135.

[140] W. Chen, C. Chien, C. Wang, H. Wang, Y. Wang, S. Ding, T. Lee, T. C, Chang, *Anal. Bioanal. Chem* **2013**, *405*, 8549–8559.

[141] H. Rinia, K. Burger, M. Bonn, M. Müller, *Biophys. J* **2008**, 95–4908.

[142] M. Wang, W. Min, C. Freudiger, G. Ruvkun, X. S, Xie, *Nat. Methods* **2011**, *8*, 135–138.

[143] T. Le, S. Yue, J.-X. Cheng, *J. Lipid Res* **2010**, *51*, 3091–3102.

[144] Y. Yu, P. Ramachandran, M. Wang, *Biochim. Biophys. Acta Mol. Cell Biol. Lipids*, *2014*, 1120–1129.

[145] X. Nan, J.-X. Cheng, X. Xie, *J. Lipid Res* **2003**, *44*, 2202–2208.

[146] T. Le, J. X, Cheng, *PLoS One* **2009**, *4*, 5189.

[147] H.-J. van Manen, Y. Kraan, D. Roos, C. Otto, *Proc. Natl. Acad. Sci. U. S. A* **2005**, *102*, 10159–10164.

[148] X. Xie, J. Yu, W. Yang, *Science* **2006**, *312*, 228–230.

[149] H. Yamakoshi, K. Dodo, M. Okada, J. Ando, A. Palonpon, K. Fujita, S. Kawata, M. Sodeoka, *J. Am. Chem. Soc* **2011**, *133*, 6102–6105.

[150] H. Yamakoshi, K. Dodo, A. Palonpon, J. Ando, K. Fujita, S. Kawata, M. Sodeoka, *J. Am. Chem. Soc* **2012**, *134*, 20681–20689.

[151] L. Wei, F. Hu, Y. Shen, Z. Chen, Y. Yu, C.-C. Lin, M. C, Wang, W. Min, *Nat. Methods*, *2014*, 410–412.

[152] J. Li, J.-X. Cheng, *Sci. Rep* **2015**, *4*, 6807.

[153] F. Hu, Z. Chen, L. Zhang, Y. Shen, L. Wei, W. Min, *Angew. Chem. Int. Ed* **2015**, *54*, 9821–9825.

[154] E. Andresen, P. Berto, H. Rigneault, *Opt. Lett* **2011**, *36*, 2387.

[155] H. Beier, G. Noojin, B. Rockwell, *Opt. Express* **2011**, *19*, 18885.

[156] D. Fu, G. Holtom, C. Freudiger, X. Zhang, X. Xie, *J. Phys. Chem. B* **2013**, *117*, 4634–4640.

[157] D. Zhang, P. Wang, M. Slipchenko, D. Ben-Amotz, A. Weiner, J.-X. Cheng, *Anal. Chem* **2013**, *85*, 106.

[158] J. Li, S. Condello, J. Thomes-Pepin, X. Ma, Y. Xia, T. Hurley, D. Matei, X. J. Cheng, *Cell Stem Cell* **2017**, *20*, 303–314.

[159] A. Alfonso-García, S. Pfisterer, H. Riezman, E. Ikonen, E. Potma, *J. Biomed. Opt* **2015**, *21*, 061003.

[160] F. Kraemer, Mol, *Cell. Endocrinol* **2007**, 265–266.

[161] V. Khor, R. Ahrends, Y. Lin, W. Shen, C. Adams, A. Roseman, Y. Cortez, M. Teruel, S. Azhar, B. F. Kraemer, *PLoS One*, *2014*, 105047.

[162] D. Fu, Y. Yu, A. Folick, E. Currie, R. Farese, T. Tsai, X. Xie, M. Wang, *J. Am. Chem. Soc*, *2014*, 8820–8828.

[163] N. Otsu, Man, *I.E.E.E.Trans Syst Cybern*, *1979*, 62–66.

[164] A. dos Anjos, H. Shahbazkia in *Proc. of ICBED, BIOSIGNALS*, *Vol. 2*, **2008**, pp. 70–76.

[165] R. Szeliski in *Computer Vision: Algorithms and Applications (Texts in Computer Science)*, Springer, London, **2011**.

[166] J. Canny, *Trans I.E.E.E. Pattern Anal. Mach. Intell*, *1986*, 679–698.

[167] P. Hough, *US Pat* **1962**, *3069654*.

[168] D. Ballard, *Pattern Recognit.* **1981**, *13*, 111–122.

[169] S. Beucher, C. Lantuejoul, *Real-time Edge Motion Detect* **1979**, 12–21.

[170] J. Schindelin, I. Arganda-Carreras, E. Frise, V. Kaynig, M. Longair, T. Pietzsch, S. Preibisch, C. Rueden, S. Saalfeld, B. Schmid, J. Tinevez, D. White, V. Hartenstein, K. Eliceiri, P. Tomancak, A. Cardona, *Nat. Methods* **2012**, *9*, 676–682.

[171] S. Dejgaard, J. Presley, *J. Histochem. Cytochem*, *2014*, 889–901.

[172] H. Varinli, M. Osmond-McLeod, P. Molloy, P. Vallotton, *J. Lipid Res* **2015**, *56*, 2206–2216.

[173] K. Fukushima, *Biol Cybern* **1980**, *36*, 193–202.

[174] Y. LeCun, L. Bottou, Y. Bengio, P. Haffner, *Proc IEEE* **1998**, *86*, 2278–2323.

[175] G. Litjens, T. Kooi, B. Bejnordi, A. Setio, F. Ciompi, M. Ghafoorian, J. van der Laak, B. van Ginneken, C. Sánchez, *Med. Image Anal*, *2017*, 60–88.

[176] P. Eulenberg, N. Köhler, T. Blasi, A. Filby, A. Carpenter, P. Rees, F. Theis, A. F. Wolf, *Nat. Commun*, *2017*, 463.

[177] C. Chen, A. Mahjoubfar, L. Tai, I. Blaby, A. Huang, K. Niazi, B. Jalali, *Sci. Rep* **2016**, *6*, 21471.

[178] D. Fu, X. Xie, *Anal. Chem*, *2014*, 4119.

[179] C. Napoli, I. Pope, F. Masia, W. Langbein, P. Watson, P. Borri, *Anal. Chem* **2016**, *88*, 3677–3685.

[180] F. Masia, A. Glen, P. Stephens, P. Borri, W. Langbein, *Anal. Chem* **2013**, *85*, 10820–10828.

[181] C. Jüngst, M. Klein, A. Zumbusch, *J. Lipid Res* **2013**, *54*, 3419–3429.

[182] C. Zhang, J. Li, L. Lan, J. Cheng, *Anal. Chem* **2017**, *89*, 4502–4507.

[183] I. Sbalzarini, P. Koumoutsakos, *J. Struct. Biol* **2005**, *151*, 182–195.

[184] C. Cao, D. Zhou, T. Chen, A. Streets, Y. Huang, *Anal. Chem* **2016**, *88*, 4931–4939.

[185] A. Medyukhina, T. Meyer, M. Schmitt, B. Romeike, B. Dietzek, J. Popp, *J. Biophotonics* **2012**, *5*, 878–888.

[186] J. Yuan, P. Sims, *Sci. Rep* **2016**, *6*, 33883.

[187] K. Lane, D. Valen, M. DeFelice, D. Macklin, T. Kudo, A. Jaimovich, A. Carr, T. Meyer, D. Pe'er, S. Boutet, W. M. Covert, *Cell Syst*, *2017*, 458–469.

[188] T. Gierahn, M. Wadsworth, T. Hughes, B. Bryson, A. Butler, R. Satija, S. Fortune, J. Love, K. A. Shalek, *Nat. Methods*, *2017*, 395–398.

[189] S. Bose, Z. Wan, A. Carr, A. Rizvi, G. Vieira, D. Pe'er, A. P. Sims, *Genome Biol* **2015**, *16*, 120.

[190] C. Zhang, K.-C. Huang, B. Rajwa, J. Li, S. Yang, H. Lin, C. Liao, G. Eakins, S. Kuang, V. Patsekin, J. Robinson, J.-X. Cheng, **2017**, *4*, 103.

[191] C.-S. Liao, K.-C. Huang, W. Hong, A. Chen, C. Karanja, P. Wang, G. Eakins, J. Cheng, *Optica* **2016**, *3*, 1377.

[192] R. He, Y. Xu, L. Zhang, S. Ma, X. Wang, D. Ye, M. Ji, *Optica* **2017**, *4*, 44.

[193] C.-S. Liao, M. Slipchenko, P. Wang, J. Li, S.-Y. Lee, R. Oglesbee, J. Cheng, *Light: Sci. Appl* **2015**, *4*, 265.

[194] M. Slipchenko, R. Oglesbee, D. Zhang, W. Wu, J. Cheng, *J. Biophotonics* **2012**, *5*, 801–807.

[195] M. Alshaykh, C.-S. Liao, O. Sandoval, G. Gitzinger, N. Forget, D. Leaird, J.-X. Cheng, M. A. Weiner, *Opt. Lett*, *2017*, 1548.

[196] K. Hashimoto, M. Takahashi, T. Ideguchi, K. Goda, *Sci Rep* **2016**, *6*, 21036.

[197] M. Tamamitsu, Y. Sakaki, T. Nakamura, G. Podagatlapalli, T. Ideguchi, K. Goda, *Vib. Spectrosc*, *2017*, 163–169.

[198] N. Coluccelli, E. Vicentini, A. Gambetta, C. Howle, K. Mcewan, P. Laporta, G. Galzerano, *Opt. Express* **2018**, *26*, 18855.

[199] R. He, Z. Liu, Y. Xu, W. Huang, H. Ma, M. Ji, *Opt Lett*, *2017*, 659.

[200] K. Chen, A. Boettiger, J. Moffitt, S. Wang, X. Zhuang, *Science* **2015**, *348*, 6090.

[201] E. Lubeck, A. Coskun, T. Zhiyentayev, M. Ahmad, L. Cai, *Nat. Methods*, *2014*, 360–361.

[202] J. Lee, E. Daugharthy, J. Scheiman, R. Kalhor, T. Ferrante, R. Terry, B. Turczyk, J. Yang, H. Lee, J. Aach, K. Zhang, M. G. Church, *Nat. Protoc* **2015**, *10*, 442–458.

[203] S. Cunningham, P. Leslie, D. Hopwood, P. Illingworth, R. Jung, D. Nicholls, N. Peden, J. Rafael, E. Rial, *Clin. Sci*, *1985*, 343–348.

[204] M. Lean, *Proc. Nutr. Soc* **1989**, *48*, 243–256.

[205] T. Hany, E. Gharehpapagh, E. Kamel, A. Buck, J. Himms-Hagen, G. von Schulthess, E. J, *Nucl. Med. Mol. Imaging* **2002**, *29*, 1393–1398.

[206] J. Nedergaard, T. Bengtsson, B. Cannon, *Am. J. Physiol.: Endocrinol. Metab* **2007**, *293*, 444–452.

[207] K. Virtanen, M. Lidell, J. Orava, M. Heglind, R. Westergren, T. Niemi, M. Taittonen, J. Laine, N.-J. Savisto, S. Enerbäck, P. Nuutila, *N.Engl J. Med* **2009**, *360*, 1518–1525.

[208] W. van Marken Lichtenbelt, J. Vanhommerig, N. Smulders, J. Drossaerts, G. Kemerink, N. Bouvy, P. Schrauwen, G. Teule, *N.Engl J. Med* **2009**, *360*, 1500–1508.

[209] A. Cypess, S. Lehman, G. Williams, I. Tal, D. Rodman, A. Goldfine, F. Kuo, E. Palmer, Y. Tseng, A. Doria, G. Kolodny, C. Kahn, *Obstet. Gynecol. Surv* **2009**, *64*, 519–520.

[210] A. Cypess, A. White, C. Vernochet, T. Schulz, R. Xue, C. Sass, T. Huang, C. Roberts-Toler, L. Weiner, C. Sze, A. Chacko, L. Deschamps, L. Herder, N. Truchan, A. Glasgow, A. Holman, A. Gavrila, P.-O. Hasselgren, M. Mori, M. Molla, Y. Tseng, *Nat. Med* **2013**, *19*, 635–639.

[211] P. Lang, K. Yeow, A. Nichols, A. Scheer, *Nature Reviews Drug Discovery* **2006**, *5*, 343–356.

[212] F. S. Wouters, P. J. Verveer, P. I. Bastiaens, *Trends in cell biology* **2001**, *11*, 203–211.

[213] M. Boutros, F. Heigwer, C. Laufer, *Cell* **2015**, *163*, 1314–1325.

[214] J. R. Yaron, C. P. Ziegler, T. H. Tran, H. L. Glenn, D. R. Meldrum, *Biological Procedures Online* **2014**, *16*, DOI 10.1186/1480-9222-16-9.

[215] M. Saint, F. Bertaux, W. Tang, X. M. Sun, L. Game, A. Köferle, J. Bähler, V. Shahrezaei, S. Marguerat, *Nature Microbiology* **2019**, *4*, 480–491.

[216] D. A. Jaitin, E. Kenigsberg, H. Keren-Shaul, N. Elefant, F. Paul, I. Zaretsky, A. Mildner, N. Cohen, S. Jung, A. Tanay, I. Amit, *Science* **2014**, *343*, 776–779.

[217] J. Yuan, J. Sheng, P. A. Sims, *Genome Biology* **2018**, *19*, 227.

[218] H. Q. Nguyen, B. C. Baxter, K. Brower, C. A. Diaz-Botia, J. L. DeRisi, P. M. Fordyce, K. S. Thorn, *Advanced Optical Materials* **2017**, *5*, DOI 10.1002/adom.201600548.

[219] R. H. Cole, S. Y. Tang, C. A. Siltanen, P. Shahi, J. Q. Zhang, S. Poust, Z. J. Gartner, A. R. Abate, *Proceedings of the National Academy of Sciences of the United States of America* **2017**, *114*, 8728–8733.

[220] J. Q. Zhang, C. A. Siltanen, L. Liu, K.-C. Chang, Z. J. Gartner, A. R. Abate, *Genome Biology* **2020**, *21*.

[221] M. A. Unger, H.-P. Chou, T. Thorsen, A. Scherer, S. R. Quake, *Science* **2000**, *288*, 113–116.

[222] J. A. White, A. M. Streets, *HardwareX* **2017**, *3*, 135–145.

[223] L. V. Bystrykh, *PLoS ONE* **2012**, *7*, (Ed.: J.-A. L. Stanton), e36852.

[224] I. M. Kuznetsova, K. K. Turoverov, V. N. Uversky, **2014**, *15*, 23090–23140.

[225] G. B. Ralston, Effects of "Crowding" in Protein Solutions, tech. rep.

[226] C. Ziegenhain, B. Vieth, S. Parekh, B. Reinius, A. Guillaumet-Adkins, M. Smets, H. Leonhardt, H. Heyn, I. Hellmann, W. Enard, *Molecular Cell* **2017**, *65*, 631–643.e4.

[227] J. Melin, S. R. Quake, *Annual Review of Biophysics and Biomolecular Structure* **2007**, *36*, 213–231.

[228] S. Kim, J. De Jonghe, A. B. Kulesa, D. Feldman, T. Vatanen, R. P. Bhattacharyya, B. Berdy, J. Gomez, J. Nolan, S. Epstein, P. C. Blainey, *Nature Communications* **2017**, *8*, 1–10.

[229] H. C. Fan, J. Wang, A. Potanina, S. R. Quake, *Nature Biotechnology* **2011**, *29*, 51–59.

[230] J. Liu, C. Hansen, S. R. Quake, *Analytical Chemistry* **2003**, *75*, 4718–4723.

[231] N. Ramalingam, B. Fowler, L. Szpankowski, A. A. Leyrat, K. Hukari, M. T. Maung, W. Yorza, M. Norris, C. Cesar, J. Shuga, M. L. Gonzales, C. D. Sanada, X. Wang, R. Yeung, W. Hwang, J. Axsom, N. S. G. K. Devaraju, N. D. Angeles, C. Greene, M. F. Zhou, E. S. Ong, C. C. Poh, M. Lam, H. Choi, Z. Htoo, L. Lee, C. S. Chin, Z. W. Shen, C. T. Lu, I. Holcomb, A. Ooi, C. Stolarczyk, T. Shuga, K. J. Livak, M. Unger, J. A. West, *Frontiers in Bioengineering and Biotechnology* **2016**, *4*, 70.

[232] B. Magella, M. Adam, A. S. Potter, M. Venkatasubramanian, K. Chetal, S. B. Hay, N. Salomonis, S. S. Potter, *Developmental Biology* **2018**, *434*, 36–47.

[233] S. J. Maerkl, S. R. Quake, *Science* **2007**, *315*, 233–237.

[234] A. Lai, N. Altemose, J. A. White, A. M. Streets, *Journal of Micromechanics and Microengineering* **2019**, *29*, DOI 10.1088/1361-6439/ab341e.

[235] S. Parekh, C. Ziegenhain, B. Vieth, W. Enard, I. Hellmann, **2018**, *7*, DOI 10.1093/gigascience/giy059.

[236] A. Dobin, C. A. Davis, F. Schlesinger, J. Drenkow, C. Zaleski, S. Jha, P. Batut, M. Chaisson, T. R. Gingeras, *Bioinformatics* **2013**, *29*, 15–21.

[237] Creating a Reference Package with cellranger mkref -Software -Single Cell Gene Expression -Official 10x Genomics Support.

[238] Y. Liao, G. K. Smyth, W. Shi, *Nucleic Acids Research* **2013**, *41*, DOI 10.1093/nar/gkt214.

[239] J. Ding, X. Adiconis, S. K. Simmons, M. S. Kowalczyk, C. C. Hession, N. D. Marjanovic, T. K. Hughes, M. H. Wadsworth, T. Burks, L. T. Nguyen, J. Y. H. Kwon, B. Barak, W. Ge, A. J. Kedaigle, S. Carroll, S. Li, N. Hacohen, O. Rozenblatt-Rosen, A. K. Shalek, A.-C. Villani, A. Regev, J. Z. Levin, *bioRxiv* **2019**, 632216.

[240] G. La Manno, R. Soldatov, A. Zeisel, E. Braun, H. Hochgerner, V. Petukhov, K. Lidschreiber, M. E. Kastriti, P. Lönnerberg, A. Furlan, J. Fan, L. E. Borm, Z. Liu, D. van Bruggen, J. Guo, X. He, R. Barker, E. Sundström, G. Castelo-Branco, P. Cramer, I. Adameyko, S. Linnarsson, P. V. Kharchenko, **2018**, *560*, 494–498.

[241] T. Hashimshony, F. Wagner, N. Sher, I. Yanai, *Cell Reports* **2012**, *2*, 666–673.

[242] H. Keren-Shaul, E. Kenigsberg, D. A. Jaitin, E. David, F. Paul, A. Tanay, I. Amit, *Nature Protocols* **2019**, *14*, 1841–1862.

[243] P. Brennecke, S. Anders, J. K. Kim, A. A. Kołodziejczyk, X. Zhang, V. Proserpio, B. Baying, V. Benes, S. A. Teichmann, J. C. Marioni, M. G. Heisler, *Nature Methods* **2013**, *10*, 1093–1098.

[244] M. J. Arguel, K. Lebrigand, A. Paquet, S. R. García, L. E. Zaragosi, P. Barbry, R. Waldmann, *Nucleic Acids Research* **2017**, *45*, DOI `10.1093/nar/gkw1242`.

[245] Y. H. Lee, R. I. Thacker, B. E. Hall, R. Kong, J. G. Granneman, *Cell Cycle* **2014**, *13*, 184–190.

[246] Y. H. Lee, S. N. Kim, H. J. Kwon, K. R. Maddipati, J. G. Granneman, *American Journal of Physiology - Regulatory Integrative and Comparative Physiology* **2016**, *310*, R55–R65.

[247] H. Hou, Y. Zhao, C. Li, M. Wang, X. Xu, Y. Jin, *Scientific Reports* **2017**, *7*, DOI `10.1038/s41598-017-01956-1`.

[248] T. Thorsen, S. J. Maerkl, S. R. Quake, *Science* **2002**, *298*, 580–584.

[249] W. H. Grover, R. H. Ivester, E. C. Jensen, R. A. Mathies, *Lab on a Chip* **2006**, *6*, 623–631.

[250] A. B. Rosenberg, C. M. Roco, R. A. Muscat, A. Kuchina, P. Sample, Z. Yao, L. T. Graybuck, D. J. Peeler, S. Mukherjee, W. Chen, S. H. Pun, D. L. Sellers, B. Tasic, G. Seelig, *Science* **2018**, *360*, 176–182.

[251] H. Sinha, A. B. Quach, P. Q. Vo, S. C. Shih, *Lab on a Chip* **2018**, *18*, 2300–2312.

[252] Y. Chen, P. Li, P. H. Huang, Y. Xie, J. D. Mai, L. Wang, N. T. Nguyen, T. J. Huang, **2014**, *14*, 626–645.

[253] K. J. Kobayashi-Kirschvink, H. Nakaoka, A. Oda, K. i. F. Kamei, K. Nosho, H. Fukushima, Y. Kanesaki, S. Yajima, H. Masaki, K. Ohta, Y. Wakamoto, *Cell Systems* **2018**, *7*, 104–117.e4.

[254] Y. Zhou, S. Basu, K. J. Wohlfahrt, S. F. Lee, D. Klenerman, E. D. Laue, A. A. Seshia, *Sensors and Actuators B: Chemical* **2016**, *232*, 680–691.

[255] W. K. Cho, J. H. Spille, M. Hecht, C. Lee, C. Li, V. Grube, I. I. Cisse, *Science* **2018**, *361*, 412–415.

[256] N. Altemose, A. Maslan, A. Lai, J. A. White, A. M. Streets, *bioRxiv* **2019**, 706903.

[257] A. M. Bolger, M. Lohse, B. Usadel, *Bioinformatics* **2014**, DOI 10.1093/bioinformatics/btu170.

[258] lh3/seqtk: Toolkit for processing sequences in FASTA/Q formats.

[259] A. Descloux, K. S. Grußmayer, A. Radenovic, *Nature Methods* **2019**, *16*, 918–924.

[260] GitHub - Ades91/ImDecorr.

[261] C. McQuin, A. Goodman, V. Chernyshev, L. Kamentsky, B. A. Cimini, K. W. Karhohs, M. Doan, L. Ding, S. M. Rafelski, D. Thirstrup, W. Wiegraebe, S. Singh, T. Becker, J. C. Caicedo, A. E. Carpenter, *PLoS Biology* **2018**, *16*, e2005970.

[262] R. M. Haralick, I. Dinstein, K. Shanmugam, *IEEE Transactions on Systems Man and Cybernetics* **1973**, *SMC-3*, 610–621.

[263] A. Butler, P. Hoffman, P. Smibert, E. Papalexi, R. Satija, *Nature Biotechnology* **2018**, *36*, 411–420.

[264] Z. Gu, R. Eils, M. Schlesner, *Bioinformatics* **2016**, *32*, 2847–2849.

[265] C. Trapnell, **2015**, DOI 10.1101/gr.190595.115.

[266] X. Chen, S. A. Teichmann, K. B. Meyer, *Annual Review of Biomedical Data Science* **2018**, *1*, 29–51.

[267] E. A. Rondini, J. G. Granneman, *The Biochemical journal* **2020**, *477*, 583–600.

[268] K. Shinoda, I. H. Luijten, Y. Hasegawa, H. Hong, S. B. Sonne, M. Kim, R. Xue, M. Chondronikola, A. M. Cypess, Y. H. Tseng, J. Nedergaard, L. S. Sidossis, S. Kajimura, *Nature Medicine* **2015**, *21*, 389–394.

[269] S. Y. Min, J. Kady, M. Nam, R. Rojas-Rodriguez, A. Berkenwald, J. H. Kim, H. L. Noh, J. K. Kim, M. P. Cooper, T. Fitzgibbons, M. A. Brehm, S. Corvera, *Nature Medicine* **2016**, *22*, 312–318.

[270] K. Y. Lee, Q. Luong, R. Sharma, J. M. Dreyfuss, S. Ussar, C. R. Kahn, *The EMBO Journal* **2019**, *38*, DOI 10.15252/embj.201899291.

[271] R. Gao, C. Kim, E. Sei, T. Foukakis, N. Crosetto, L. K. Chan, M. Srinivasan, H. Zhang, F. Meric-Bernstam, N. Navin, *Nature Communications* **2017**, *8*, DOI 10.1038/s41467-017-00244-w.

[272] M. Blüher, L. Wilson-Fritch, J. Leszyk, P. G. Laustsen, S. Corvera, C. R. Kahn, *Journal of Biological Chemistry* **2004**, *279*, 31902–31909.

[273] W. Sun, H. Dong, M. Balaz, M. Slyper, E. Drokhlyansky, G. Colleluori, A. Giordano, Z. Kovanicova, P. Stefanicka, L. Ding, G. Rudofsky, J. Ukropec, S. Cinti, A. Regev, C. Wolfrum, **2020**, DOI 10.1101/2020.01.20.890327.

[274] A. Selewa, R. Dohn, H. Eckart, S. Lozano, B. Xie, E. Gauchat, R. Elorbany, K. Rhodes, J. Burnett, Y. Gilad, S. Pott, A. Basu, *Scientific Reports* **2020**, *10*, 1–13.

[275] *PLoS ONE* **2018**, *13*, (Ed.: E. Soriano), e0209648.

[276] N. Habib, I. Avraham-Davidi, A. Basu, T. Burks, K. Shekhar, M. Hofree, S. R. Choudhury, F. Aguet, E. Gelfand, K. Ardlie, D. A. Weitz, O. Rozenblatt-Rosen, F. Zhang, A. Regev, *Nature Methods* **2017**, *14*, 955–958.

[277] B. B. Lake, S. Codeluppi, Y. C. Yung, D. Gao, J. Chun, P. V. Kharchenko, S. Linnarsson, K. Zhang, *Scientific Reports* **2017**, *7*, 6031.

[278] N. Thrupp, C. S. Frigerio, L. Wolfs, R. Mancuso, N. G. Skene, N. Fattorelli, S. Poovathingal, Y. Fourne, P. M. Matthews, T. Theys, B. De Strooper, M. Fiers, **2020**, DOI 10.1016/j.celrep.2020.108189.

[279] *Scientific Reports* **2017**, *7*, DOI 10.1038/s41598-017-00291-9.

[280] R. Lopez, J. Regier, M. B. Cole, M. I. Jordan, N. Yosef, *Nature Methods* **2018**, *15*, 1053–1058.

[281] D. Tews, V. Schwar, M. Scheithauer, T. Weber, T. Fromme, M. Klingenspor, T. F. Barth, P. Möller, K. Holzmann, K. M. Debatin, P. Fischer-Posovszky, M. Wabitsch, *Molecular and Cellular Endocrinology* **2014**, *395*, 41–50.

[282] W. P. Cawthorn, E. L. Scheller, O. A. MacDougald, *Journal of Lipid Research* **2012**, *53*, 227–246.

[283] A. Mildmay-White, W. Khan, *Current Stem Cell Research Therapy* **2017**, *12*, DOI 10.2174/1574888x11666160429122133.

[284] A. L. Ghaben, P. E. Scherer, *Nature Reviews Molecular Cell Biology* **2019**, *20*, 242–258.

[285] C. Gubelmann, P. C. Schwalie, S. K. Raghav, E. Röder, T. Delessa, E. Kiehlmann, S. M. Waszak, A. Corsinotti, G. Udin, W. Holcombe, G. Rudofsky, D. Trono, C. Wolfrum, B. Deplancke, *eLife* **2014**, *3*, 1–30.

[286] J. M. de Jong, O. Larsson, B. Cannon, J. Nedergaard, *American Journal of Physiology - Endocrinology and Metabolism* **2015**, *308*, E1085–E1105.

[287] J. Sanchez-Gurmaches, C. M. Hung, D. A. Guertin, *Trends in Cell Biology* **2016**, *26*, 313–326.

[288] R. Berry, M. S. Rodeheffer, *Nature Cell Biology* **2013**, *15*, 302–308.

[289] M. Patil, B. K. Sharma, A. Satyanarayana, *Frontiers in Bioscience - Landmark* **2014**, *19*, 1386–1397.

[290] A. Satyanarayana, K. D. Klarmann, O. Gavrilova, J. R. Keller, *The FASEB Journal* **2012**, *26*, 309–323.

[291] F. X. Yu, B. Zhao, N. Panupinthu, J. L. Jewell, I. Lian, L. H. Wang, J. Zhao, H. Yuan, K. Tumaneng, H. Li, X. D. Fu, G. B. Mills, K. L. Guan, *Cell* **2012**, DOI 10.1016/j.cell.2012.06.037.

[292] S. Dupont, L. Morsut, M. Aragona, E. Enzo, S. Giulitti, M. Cordenonsi, F. Zanconato, J. Le Digabel, M. Forcato, S. Bicciato, N. Elvassore, S. Piccolo, *Nature* **2011**, *474*, 179–184.

[293] C. G. Hansen, T. Moroishi, K. L. Guan, **2015**, DOI 10.1016/j.tcb.2015.05.002.

[294] R. Zhang, Y. Gao, X. Zhao, M. Gao, Y. Wu, Y. Han, Y. Qiao, Z. Luo, L. Yang, J. Chen, G. Ge, *PLoS Biology* **2018**, DOI 10.1371/journal.pbio.2001493.

[295] K. M. Tharp, M. S. Kang, G. A. Timblin, J. Dempersmier, G. E. Dempsey, P. J. H. Zushin, J. Benavides, C. Choi, C. X. Li, A. K. Jha, S. Kajimura, K. E. Healy, H. S. Sul, K. Saijo, S. Kumar, A. Stahl, *Cell Metabolism* **2018**, *27*, 602–615.e4.

[296] S. Modica, C. Wolfrum, *Adipocyte* **2017**, *6*, 141–146.

[297] S. N. Li, J. F. Wu, **2020**, *11*, 41.

[298] M. E. Lidell, M. J. Betz, O. D. Leinhard, M. Heglind, L. Elander, M. Slawik, T. Mussack, D. Nilsson, T. Romu, P. Nuutila, K. A. Virtanen, F. Beuschlein, A. Persson, M. Borga, S. Enerbäck, *Nature Medicine* **2013**, *19*, 631–634.

[299] J. Luther, K. Ubieta, N. Hannemann, M. Jimenez, M. Garcia, C. Zech, G. Schett, E. F. Wagner, A. Bozec, *Cell Death and Differentiation* **2014**, *21*, 655–664.

[300] J. Luther, F. Driessler, M. Megges, A. Hess, B. Herbort, V. Mandic, M. M. Zaiss, A. Reichardt, C. Zech, J. P. Tuckermann, C. F. Calkhoven, E. F. Wagner, G. Schett, J. P. David, *Journal of Cell Science* **2011**, *124*, 1465–1476.

[301] A. Giordano, R. Coppari, M. Castellucci, S. Cinti, *Journal of Neurocytology* **2001**, DOI 10.1023/A:1011916822633.

[302] F. Villarroya, A. Vidal-Puig, **2013**, *17*, 638–643.

[303] M. Karbiener, C. Glantschnig, D. F. Pisani, J. Laurencikiene, I. Dahlman, S. Herzig, E. Z. Amri, M. Scheideler, *International Journal of Obesity* **2015**, *39*, 1733–1741.

[304] M. Braga, S. T. Reddy, L. Vergnes, S. Pervin, V. Grijalva, D. Stout, J. David, X. Li, V. Tomasian, C. B. Reid, K. C. Norris, S. U. Devaskar, K. Reue, R. Singh, *Journal of Lipid Research* **2014**, *55*, 375–384.

[305] F. Strutz, H. Okada, C. W. Lo, T. Danoff, R. L. Carone, J. E. Tomaszewski, E. G. Neilson, *Journal of Cell Biology* **1995**, *130*, 393–405.

[306] S. Hou, Y. Jiao, Q. Yuan, J. Zhai, T. Tian, K. Sun, Z. Chen, Z. Wu, J. Zhang, *Laboratory Investigation* **2018**, *98*, 1025–1038.

[307] J. Vijay, M. F. Gauthier, R. L. Biswell, D. A. Louiselle, J. J. Johnston, W. A. Cheung, B. Belden, A. Pramatarova, L. Biertho, M. Gibson, M. M. Simon, H. Djambazian, A. Staffa, G. Bourque, A. Laitinen, J. Nystedt, M. C. Vohl, J. D. Fraser, T. Pastinen, A. Tchernof, E. Grundberg, *Nature Metabolism* **2020**, *2*, 97–109.

[308] R. Ferrero, P. Rainer, B. Deplancke, **2020**, *30*, 937–950.

[309] J. T. Gaublomme, B. Li, C. McCabe, A. Knecht, Y. Yang, E. Drokhlyansky, N. Van Wittenberghe, J. Waldman, D. Dionne, L. Nguyen, P. L. De Jager, B. Yeung, X. Zhao, N. Habib, O. Rozenblatt-Rosen, A. Regev, *Nature Communications* **2019**, *10*, 1–8.

[310] D. DeTomaso, N. Yosef, *BMC Bioinformatics* **2016**, *17*, 315.

[311] D. DeTomaso, M. G. Jones, M. Subramaniam, T. Ashuach, C. J. Ye, N. Yosef, *Nature Communications* **2019**, *10*, DOI 10.1038/s41467-019-12235-0.

[312] A. Haque, J. Engel, S. A. Teichmann, T. Lönnberg, *Genome medicine* **2017**, *9*, 1–12.

[313] V. Svensson, E. da Veiga Beltrame, L. Pachter, *bioRxiv* **2019**, 762773.

[314] A. Ameur, A. Zaghlool, J. Halvardson, A. Wetterbom, U. Gyllensten, L. Cavelier, L. Feuk, *Nature Structural and Molecular Biology* **2011**, *18*, 1435–1440.

[315] J. M. Gray, D. A. Harmin, S. A. Boswell, N. Cloonan, T. E. Mullen, J. J. Ling, N. Miller, S. Kuersten, Y.-C. Ma, S. A. McCarroll, S. M. Grimmond, M. Springer, *PLoS ONE* **2014**, *9*, (Ed.: Z. Zuo), e89673.

[316] G. J. Hendriks, D. Gaidatzis, F. Aeschimann, H. Großhans, *Molecular Cell* **2014**, *53*, 380–392.

[317] R. Patrick, D. T. Humphreys, V. Janbandhu, A. Oshlack, J. W. Ho, R. P. Harvey, K. K. Lo, *Genome Biology* **2020**, *21*, 167.

[318] E. D. Shulman, R. Elkon, *Nucleic acids research* **2019**, *47*, 10027–10039.

[319] V. O. Wickramasinghe, R. Andrews, P. Ellis, C. Langford, J. B. Gurdon, M. Stewart, A. R. Venkitaraman, R. A. Laskey, *Nucleic Acids Research* **2014**, *42*, 5059–5071.

[320] T. Chen, B. van Steensel, *PLoS Genetics* **2017**, *13*, e1006929.

[321] S. Quinodoz, M. Guttman, **2014**, *24*, 651–663.

[322] V. V. Sherstyuk, S. P. Medvedev, S. M. Zakian, **2018**, *14*, 58–70.

[323] M. J. Delá, B. T. Jackson, N. Erard, S. R. V. Knott, G. J. Hannon, T. Kovacevic, S. Vangelisti, E. M. Maravilla, S. A. Wild, E. M. Stork, **2019**, DOI 10.1016/j.celrep.2019.03.080.

[324] M. N. Cabili, M. C. Dunagin, P. D. McClanahan, A. Biaesch, O. Padovan-Merhar, A. Regev, J. L. Rinn, A. Raj, *Genome Biology* **2015**, *16*, DOI 10.1186/s13059-015-0586-4.

[325] X. Wen, L. Gao, X. Guo, X. Li, X. Huang, Y. Wang, H. Xu, R. He, C. Jia, F. Liang, *Database* **2018**, *2018*, 85.

[326] R. V. Grindberg, J. L. Yee-Greenbaum, M. J. McConnell, M. Novotny, A. L. O'Shaughnessy, G. M. Lambert, M. J. Araúzo-Bravo, J. Lee, M. Fishman, G. E. Robbins, X. Lin, P. Venepally, J. H. Badger, D. W. Galbraith, F. H. Gage, R. S. Lasken, *Proceedings of the National Academy of Sciences of the United States of America* **2013**, *110*, 19802–19807.

[327] W. Zeng, S. Jiang, X. Kong, N. El-Ali, A. R. Ball, C. I. Ma, N. Hashimoto, K. Yokomori, A. Mortazavi, *Nucleic Acids Research* **2016**, *44*, DOI 10.1093/nar/gkw739.

[328] L. Sun, L. A. Goff, C. Trapnell, R. Alexander, K. A. Lo, E. Hacisuleyman, M. Sauvageau, B. Tazon-Vega, D. R. Kelley, D. G. Hendrickson, B. Yuan, M. Kellis, H. F. Lodish, J. L. Rinn, *Proceedings of the National Academy of Sciences of the United States of America* **2013**, *110*, 3387–3392.

[329] C. Ding, Y. C. Lim, S. Y. Chia, A. C. E. Walet, S. Xu, K. A. Lo, Y. Zhao, D. Zhu, Z. Shan, Q. Chen, M. K. S. Leow, D. Xu, L. Sun, *Nature Communications* **2018**, *9*, 1–14.

[330] S. Wei, M. Du, Z. Jiang, G. J. Hausman, L. Zhang, M. V. Dodson, *Cellular and Molecular Life Sciences* **2016**, *73*, 2079–2087.

[331] L. Sun, J. D. Lin, *Diabetes* **2019**, *68*, 887–896.

[332] Q. Zhou, Q. Wan, Y. Jiang, J. Liu, L. Qiang, L. Sun, *Cell Reports* **2020**, *31*, DOI 10.1016/j.celrep.2020.107694.

[333] Z. Li, C. Jin, S. Chen, Y. Zheng, Y. Huang, L. Jia, W. Ge, Y. Zhou, *Molecular and Cellular Biochemistry* **2017**, *433*, 51–60.

[334] Y. Huang, C. Jin, Y. Zheng, X. Li, S. Zhang, Y. Zhang, L. Jia, W. Li, *Scientific Reports* **2017**, *7*, 1–13.

[335] A. A. Zimta, A. B. Tigu, C. Braicu, C. Stefan, C. Ionescu, I. Berindan-Neagoe, **2020**, *10*, 389.

[336] C. E. Hagberg, Q. Li, M. Kutschke, D. Bhowmick, E. Kiss, I. G. Shabalina, M. J. Harms, O. Shilkova, V. Kozina, J. Nedergaard, J. Boucher, A. Thorell, K. L. Spalding, *Cell Reports* **2018**, *24*, 2746–2756.e5.

[337] S. M. Majka, H. L. Miller, K. M. Helm, A. S. Acosta, C. R. Childs, R. Kong, D. J. Klemm in *Methods in Enzymology*, *Vol. 537*, Academic Press Inc., **2014**, pp. 281–296.

[338] F. Shamsi, Y. H. Tseng in *Methods in Molecular Biology*, *Vol. 1566*, Humana Press Inc., **2017**, pp. 77–85.

[339] E. Mereu, A. Lafzi, C. Moutinho, C. Ziegenhain, D. J. McCarthy, A. Álvarez-Varela, E. Batlle, Sagar, D. Grün, J. K. Lau, S. C. Boutet, C. Sanada, A. Ooi, R. C. Jones, K. Kaihara, C. Brampton, Y. Talaga, Y. Sasagawa, K. Tanaka, T. Hayashi, C. Braeuning, C. Fischer, S. Sauer, T. Trefzer, C. Conrad, X. Adiconis, L. T. Nguyen, A. Regev, J. Z. Levin, S. Parekh, A. Janjic, L. E. Wange, J. W. Bagnoli, W. Enard, M. Gut, R. Sandberg, I. Nikaido, I. Gut, O. Stegle, H. Heyn, *Nature Biotechnology* **2020**, *38*, 747–755.

[340] L. Zappia, B. Phipson, A. Oshlack, *PLOS Computational Biology* **2018**, *14*, (Ed.: D. Schneidman), e1006245.

[341] A. Gayoso, R. Lopez, G. Xing, P. Boyeau, K. Wu, M. Jayasuriya, E. Melhman, M. Langevin, Y. Liu, J. Samaran, G. Misrachi, A. Nazaret, O. Clivio, C. Xu, T. Ashuach, M. Lotfollahi, V. Svensson, E. Da Veiga Beltrame, C. Talavera-López, L. Pachter, F. J. Theis, A. Streets, M. I. Jordan, J. Regier, N. Yosef, *bioRxiv* **2021**, 2021.04.28.441833.

[342] F. R. Weiner, A. Shah, P. J. Smith, C. S. Rubin, M. A. Zern, *Biochemistry* **1989**, *28*, 4094–4099.

[343] C. E. Gleason, Y. Ning, T. P. Cominski, R. Gupta, K. H. Kaestner, J. E. Pintar, M. J. Birnbaum, *Molecular Endocrinology* **2010**, *24*, 178–192.

[344] T. Stuart, A. Butler, P. Hoffman, C. Hafemeister, E. Papalexi, W. M. Mauck, Y. Hao, M. Stoeckius, P. Smibert, R. Satija, *Cell* **2019**, *177*, 1888–1902.e21.

[345] M. D. Luecken, M. Büttner, K. Chaichoompu, A. Danese, M. Interlandi, M. F. Mueller, D. C. Strobl, L. Zappia, M. Dugas, M. Colomé-Tatché, F. J. Theis, *bioRxiv* **2020**, *1*, 2020.05.22.111161.

[346] R. B. Burl, V. D. Ramseyer, E. A. Rondini, R. Pique-Regi, Y. H. Lee, J. G. Granneman, *Cell Metabolism* **2018**, *28*, 300–309.e4.

[347] C. Hepler, B. Shan, Q. Zhang, G. H. Henry, M. Shao, L. Vishvanath, A. L. Ghaben, A. B. Mobley, D. Strand, G. C. Hon, R. K. Gupta, *eLife* **2018**, *7*, DOI `10.7554/eLife.39636`.

[348] D. Merrick, A. Sakers, Z. Irgebay, C. Okada, C. Calvert, M. P. Morley, I. Percec, P. Seale, *Science* **2019**, *364*, eaav2501.

[349] P. C. Schwalie, H. Dong, M. Zachara, J. Russeil, D. Alpern, N. Akchiche, C. Caprara, W. Sun, K. U. Schlaudraff, G. Soldati, C. Wolfrum, B. Deplancke, *Nature* **2018**, *559*, 103–108.

[350] D. S. Cho, B. Lee, J. D. Doles, *Life Science Alliance* **2019**, *2*, DOI `10.26508/lsa.201900561`.

[351] W. Gu, W. N. Nowak, Y. Xie, A. Le Bras, Y. Hu, J. Deng, S. Issa Bhaloo, Y. Lu, H. Yuan, E. Fidanis, A. Saxena, T. Kanno, A. J. Mason, J. Dulak, J. Cai, Q. Xu, *Arteriosclerosis Thrombosis and Vascular Biology* **2019**, *39*, 2049–2066.

[352] J. R. Acosta, S. Joost, K. Karlsson, A. Ehrlund, X. Li, M. Aouadi, M. Kasper, P. Arner, M. Rydén, J. Laurencikiene, *Stem Cell Research and Therapy* **2017**, *8*, DOI 10.1186/s13287-017-0701-4.

[353] D. A. Jaitin, L. Adlung, C. A. Thaiss, A. Weiner, B. Li, H. Descamps, P. Lundgren, C. Bleriot, Z. Liu, A. Deczkowska, H. Keren-Shaul, E. David, N. Zmora, S. M. Eldar, N. Lubezky, O. Shibolet, D. A. Hill, M. A. Lazar, M. Colonna, F. Ginhoux, H. Shapiro, E. Elinav, I. Amit, *Cell* **2019**, *178*, 686–698.e14.

[354] D. Gaidatzis, L. Burger, M. Florescu, M. B. Stadler, *Nature Biotechnology* **2015**, *33*, 722–729.

[355] A. Piovesan, M. Caracausi, F. Antonaros, M. C. Pelleri, L. Vitale, *Database* **2016**, DOI 10.1093/database/baw153.

[356] T. S. Andrews, J. Atif, J. C. Liu, C. T. Perciani, X.-Z. Ma, C. Thoeni, M. Slyper, G. Eraslan, A. Segerstolpe, J. Manuel, S. Chung, E. Winter, I. Cirlan, N. Khuu, S. Fischer, O. Rozenblatt-Rosen, A. Regev, I. D. McGilvray, G. D. Bader, S. A. MacParland, *bioRxiv* **2021**, 2021.03.27.436882.

[357] M. Slyper, C. B. Porter, O. Ashenberg, J. Waldman, E. Drokhlyansky, I. Wakiro, C. Smillie, G. Smith-Rosario, J. Wu, D. Dionne, S. Vigneau, J. Jané-Valbuena, T. L. Tickle, S. Napolitano, M. J. Su, A. G. Patel, A. Karlstrom, S. Gritsch, M. Nomura, A. Waghray, S. H. Gohil, A. M. Tsankov, L. Jerby-Arnon, O. Cohen, J. Klughammer, Y. Rosen, J. Gould, L. Nguyen, M. Hofree, P. J. Tramontozzi, B. Li, C. J. Wu, B. Izar, R. Haq, F. S. Hodi, C. H. Yoon, A. N. Hata, S. J. Baker, M. L. Suvà, R. Bueno, E. H. Stover, M. R. Clay, M. A. Dyer, N. B. Collins, U. A. Matulonis, N. Wagle, B. E. Johnson, A. Rotem, O. Rozenblatt-Rosen, A. Regev, *Nature Medicine* **2020**, *26*, 792–802.

[358] M. Alvarez, E. Rahmani, B. Jew, K. M. Garske, Z. Miao, J. N. Benhammou, C. J. Ye, J. R. Pisegna, K. H. Pietiläinen, E. Halperin, P. Pajukanta, *Scientific Reports* **2020**, *10*, 1–16.

[359] S. J. Fleming, J. C. Marioni, M. Babadi, *bioRxiv* **2019**, 791699.

[360] C. Hafemeister, R. Satija, *Genome Biology* **2019**, *20*, 1–15.

[361] M. Ashburner, C. A. Ball, J. A. Blake, D. Botstein, H. Butler, J. M. Cherry, A. P. Davis, K. Dolinski, S. S. Dwight, J. T. Eppig, M. A. Harris, D. P. Hill, L. Issel-Tarver, A. Kasarskis, S. Lewis, J. C. Matese, J. E. Richardson, M. Ringwald, G. M. Rubin, G. Sherlock, **2000**, *25*, 25–29.

[362] S. Carbon, E. Douglass, N. Dunn, B. Good, N. L. Harris, S. E. Lewis, C. J. Mungall, S. Basu, R. L. Chisholm, R. J. Dodson, E. Hartline, P. Fey, P. D. Thomas, L. P. Albou, D. Ebert, M. J. Kesling, H. Mi, A. Muruganujan, X. Huang, S. Poudel, T. Mushayahama, J. C. Hu, S. A. LaBonte, D. A. Siegele, G. Antonazzo, H. Attrill, N. H. Brown, S. Fexova, P. Garapati, T. E. Jones, S. J. Marygold, G. H. Millburn, A. J.

Rey, V. Trovisco, G. Dos Santos, D. B. Emmert, K. Falls, P. Zhou, J. L. Goodman, V. B. Strelets, J. Thurmond, M. Courtot, D. S. Osumi, H. Parkinson, P. Roncaglia, M. L. Acencio, M. Kuiper, A. Lreid, C. Logie, R. C. Lovering, R. P. Huntley, P. Denny, N. H. Campbell, B. Kramarz, V. Acquaah, S. H. Ahmad, H. Chen, J. H. Rawson, M. C. Chibucos, M. Giglio, S. Nadendla, R. Tauber, M. J. Duesbury, N. T. Del, B. H. Meldal, L. Perfetto, P. Porras, S. Orchard, A. Shrivastava, Z. Xie, H. Y. Chang, R. D. Finn, A. L. Mitchell, N. D. Rawlings, L. Richardson, A. Sangrador-Vegas, J. A. Blake, K. R. Christie, M. E. Dolan, H. J. Drabkin, D. P. Hill, L. Ni, D. Sitnikov, M. A. Harris, S. G. Oliver, K. Rutherford, V. Wood, J. Hayles, J. Bahler, A. Lock, E. R. Bolton, J. De Pons, M. Dwinell, G. T. Hayman, S. J. Laulederkind, M. Shimoyama, M. Tutaj, S. J. Wang, P. D'Eustachio, L. Matthews, J. P. Balhoff, S. A. Aleksander, G. Binkley, B. L. Dunn, J. M. Cherry, S. R. Engel, F. Gondwe, K. Karra, K. A. MacPherson, S. R. Miyasato, R. S. Nash, P. C. Ng, T. K. Sheppard, A. Shrivatsav Vp, M. Simison, M. S. Skrzypek, S. Weng, E. D. Wong, M. Feuermann, P. Gaudet, E. Bakker, T. Z. Berardini, L. Reiser, S. Subramaniam, E. Huala, C. Arighi, A. Auchincloss, K. Axelsen, G. P. Argoud, A. Bateman, B. Bely, M. C. Blatter, E. Boutet, L. Breuza, A. Bridge, R. Britto, H. Bye-A-Jee, C. Casals-Casas, E. Coudert, A. Estreicher, L. Famiglietti, P. Garmiri, G. Georghiou, A. Gos, N. Gruaz-Gumowski, E. Hatton-Ellis, U. Hinz, C. Hulo, A. Ignatchenko, F. Jungo, G. Keller, K. Laiho, P. Lemercier, D. Lieberherr, Y. Lussi, A. Mac-Dougall, M. Magrane, M. J. Martin, P. Masson, D. A. Natale, N. N. Hyka, I. Pedruzzi, K. Pichler, S. Poux, C. Rivoire, M. Rodriguez-Lopez, T. Sawford, E. Speretta, A. Shypitsyna, A. Stutz, S. Sundaram, M. Tognolli, N. Tyagi, K. Warner, R. Zaru, C. Wu, J. Chan, J. Cho, S. Gao, C. Grove, M. C. Harrison, K. Howe, R. Lee, J. Mendel, H. M. Muller, D. Raciti, K. Van Auken, M. Berriman, L. Stein, P. W. Sternberg, D. Howe, S. Toro, M. Westerfield, *Nucleic Acids Research* **2019**, *47*, D330–D338.

[363] H. Mi, A. Muruganujan, D. Ebert, X. Huang, P. D. Thomas, *Nucleic Acids Research* **2019**, *47*, D419–D426.

[364] A. B. Keenan, D. Torre, A. Lachmann, A. K. Leong, M. L. Wojciechowicz, V. Utti, K. M. Jagodnik, E. Kropiwnicki, Z. Wang, A. Ma'ayan, *Nucleic Acids Research* **2019**, *47*, W212–W224.

[365] H. B. Ruan, *Journal of Molecular Cell Biology* **2020**, *12*, 775–784.

[366] S. L. Wolock, R. Lopez, A. M. Klein, *Cell Systems* **2019**, *8*, 281–291.e9.

[367] A. Gayoso, J. Shor, A. J. Carr, R. Sharma, D. Pe'er, JonathanShor/DoubletDetection: HOTFIX: Correct setup.py installation, **2019**.

[368] A. Raj, P. van den Bogaard, S. A. Rifkin, A. van Oudenaarden, S. Tyagi, *Nature Methods* **2008**, *5*, 877–879.

[369] A. D. Herbert, A. M. Carr, E. Hoffmann, M. Lichten, *PLoS ONE* **2014**, *9*, e114749.

[370] L. Scrucca, M. Fop, T. B. Murphy, A. E. Raftery, *R Journal* **2016**, *8*, 289–317.

[371] P. Tong, Y. Chen, X. Su, K. R. Coombes, *Bioinformatics* **2013**, *29*, 605–613.

[372] A. T. Lun, S. Riesenfeld, T. Andrews, T. P. Dao, T. Gomes, J. C. Marioni, *Genome Biology* **2019**, *20*, 63.

[373] S. Urs, C. Smith, B. Campbell, A. M. Saxton, J. Taylor, B. Zhang, J. Snoddy, B. Jones Voy, N. Moustaid-Moussa, *The Journal of nutrition* **2004**, *134*, 762–770.

[374] L. Satish, J. M. Krill-Burger, P. H. Gallo, S. Des Etages, F. Liu, B. J. Philips, S. Ravuri, K. G. Marra, W. A. LaFramboise, S. Kathju, et al., *BMC Medical Genomics* **2015**, *8*, 1–12.

[375] M. A. Ambele, C. Dessels, C. Durandt, M. S. Pepper, *Stem Cell Research* **2016**, *16*, 725–734.

[376] J. Söhle, N. Machuy, E. Smailbegovic, U. Holtzmann, E. Grönniger, H. Wenck, F. Stäb, M. Winnefeld, *PloS one* **2012**, *7*, e31193.

[377] J. Chen, Y. Lu, M. Tian, Q. Huang, *Journal of molecular endocrinology* **2019**, *62*, R239–R253.

[378] W. Fan, T. Imamura, N. Sonoda, D. D. Sears, D. Patsouris, J. J. Kim, J. M. Olefsky, *Journal of biological chemistry* **2009**, *284*, 12188–12197.

[379] O. A. MacDougald, M. D. Lane, *Annual review of biochemistry* **1995**, *64*, 345–373.

[380] S. B. Spurgin et al., **2016**.

[381] M. Ullah, M. Sittinger, J. Ringe, *Matrix Biology* **2013**, *32*, 452–465.

[382] E. Morandi, R. Verstappen, M. Zwierzina, S. Geley, G. Pierer, C. Ploner, *Scientific reports* **2016**, *6*, 1–14.

[383] S. Kim, C. Ahn, N. Bong, S. Choe, D. K. Lee, *PLoS One* **2015**, *10*, e0120073.

[384] J. A. Côté, J. Lessard, M. Pelletier, S. Marceau, O. Lescelleur, J. Fradette, A. Tchernof, *FEBS Open Bio* **2017**, *7*, 1092–1101.

[385] S. Karaman, M. Hollmén, S.-Y. Yoon, H. F. Alkan, K. Alitalo, C. Wolfrum, M. Detmar, *Scientific reports* **2016**, *6*, 1–12.

[386] E. C. Mariman, P. Wang, *Cellular and molecular life sciences* **2010**, *67*, 1277–1292.

[387] I. Nakajima, S. Muroya, R.-I. Tanabe, K. Chikuni, *Biology of the Cell* **2002**, *94*, 197–203.

[388] B. M. Spiegelman, C. A. Ginty, *Cell* **1983**, *35*, 657–666.

[389] L. Mor-Yossef Moldovan, M. Lustig, A. Naftaly, M. Mardamshina, T. Geiger, A. Gefen, D. Benayahu, *Journal of cellular physiology* **2019**, *234*, 3850–3863.

[390] M. Al Hasan, P. E. Martin, X. Shu, S. Patterson, C. Bartholomew, *Biomolecules* **2021**, *11*, 156.

[391] B. M. Spiegelman, S. R. Farmer, *Cell* **1982**, *29*, 53–60.

[392] W. Yang, X. Guo, S. Thein, F. Xu, S. Sugii, P. W. Baas, G. K. Radda, W. Han, *Biochemical Journal* **2013**, *449*, 605–612.

[393] J. Lilla, D. Stickens, Z. Werb, *The American journal of pathology* **2002**, *160*, 1551.

[394] C. H. del Pozo, G. Vesperinas-Garcıa, M.-Á. Rubio, R. Corripio-Sánchez, A. J. Torres-Garcıa, M.-J. Obregon, R. M. Calvo, *Biochimica et Biophysica Acta (BBA)-Molecular and Cell Biology of Lipids* **2011**, *1811*, 1194–1200.

[395] H. Xu, Y. Yang, L. Fan, L. Deng, J. Fan, D. Li, H. Li, R. C. Zhao, *Stem cell research & therapy* **2021**, *12*, 1–12.

[396] J. R. Schultz, H. Tu, A. Luk, J. J. Repa, J. C. Medina, L. Li, S. Schwendner, S. Wang, M. Thoolen, D. J. Mangelsdorf, et al., *Genes & development* **2000**, *14*, 2831–2838.

[397] A. G. Cristancho, M. A. Lazar, *Nature reviews Molecular cell biology* **2011**, *12*, 722–734.

[398] W. Shao, P. J. Espenshade, *Cell metabolism* **2012**, *16*, 414–419.

[399] A. Talebi, J. Dehairs, F. Rambow, A. Rogiers, D. Nittner, R. Derua, F. Vanderhoydonc, J. A. Duarte, F. Bosisio, K. Van den Eynde, et al., *Nature communications* **2018**, *9*, 1–11.

[400] S. Herzig, R. J. Shaw, *Nature reviews Molecular cell biology* **2018**, *19*, 121–135.

[401] M. Ahmadian, M. J. Abbott, T. Tang, C. S. Hudak, Y. Kim, M. Bruss, M. K. Hellerstein, H.-Y. Lee, V. T. Samuel, G. I. Shulman, et al., *Cell metabolism* **2011**, *13*, 739–748.

[402] N. Martinez-Lopez, D. Athonvarangkul, S. Sahu, L. Coletto, H. Zong, C. C. Bastie, J. E. Pessin, G. J. Schwartz, R. Singh, *EMBO reports* **2013**, *14*, 795–803.

[403] N. Minsky, R. G. Roeder, *Proceedings of the National Academy of Sciences* **2015**, *112*, E5669–E5678.

[404] L. Xu, X. Ma, A. Bagattin, E. Mueller, *Cell death & disease* **2016**, *7*, e2102–e2102.

[405] M. Cairo, J. Villarroya, R. Cereijo, L. Campderros, M. Giralt, F. Villarroya, *International journal of obesity* **2016**, *40*, 1591–1599.

[406] D. Kim, J.-H. Kim, Y.-H. Kang, J. S. Kim, S.-C. Yun, S.-W. Kang, Y. Song, *International journal of molecular sciences* **2019**, *20*, 3520.

[407] Y. Shi, F. Long, *Elife* **2017**, *6*, e31649.

[408] C. Fontaine, W. Cousin, M. Plaisant, C. Dani, P. Peraldi, *Stem cells* **2008**, *26*, 1037–1046.

[409] P. Bi, T. Shan, W. Liu, F. Yue, X. Yang, X.-R. Liang, J. Wang, J. Li, N. Carlesso, X. Liu, et al., *Nature medicine* **2014**, *20*, 911–918.

[410] T. Shan, J. Liu, W. Wu, Z. Xu, Y. Wang, *Journal of cellular physiology* **2017**, *232*, 1258–1261.

[411] Y. Ishihara, M. Tsuji, C. F. Vogel, *Archives of biochemistry and biophysics* **2018**, *642*, 75–80.

[412] D. Hrckulak, L. Janeckova, L. Lanikova, V. Kriz, M. Horazna, O. Babosova, M. Vojtechova, K. Galuskova, E. Sloncova, V. Korinek, *Genes* **2018**, *9*, 439.

[413] L. Choy, J. Skillington, R. Derynck, *Journal of Cell Biology* **2000**, *149*, 667–682.

[414] H. Lee, H. J. Kim, Y. J. Lee, M.-Y. Lee, H. Choi, H. Lee, J.-w. Kim, *PLoS One* **2012**, *7*, e52474.

[415] M. Li, Z. Liu, Z. Zhang, G. Liu, S. Sun, C. Sun, *Biological chemistry* **2015**, *396*, 235–244.

[416] A. R. Angueira, S. N. Shapira, J. Ishibashi, S. Sampat, J. Sostre-Colón, M. J. Emmett, P. M. Titchenell, M. A. Lazar, H.-W. Lim, P. Seale, *Cell reports* **2020**, *30*, 2869–2878.

[417] K.-i. Yamamoto, M. Sakaguchi, R. J. Medina, A. Niida, Y. Sakaguchi, M. Miyazaki, K. Kataoka, N.-h. Huh, *Biochemical and biophysical research communications* **2010**, *400*, 175–180.

[418] B. K. Sharma, M. Patil, A. Satyanarayana, *Journal of cellular physiology* **2014**, *229*, 1901–1907.

[419] I. G. Vonhögen, H. El Azzouzi, S. Olieslagers, A. Vasilevich, J. de Boer, F. J. Tinahones, P. A. da Costa Martins, L. J. de Windt, M. Murri, *Cells* **2020**, *9*, 1056.

[420] D. Lin, T.-H. Chun, L. Kang, *Biochemical pharmacology* **2016**, *119*, 8–16.

[421] F. J. Ruiz-Ojeda, A. Méndez-Gutiérrez, C. M. Aguilera, J. Plaza-Dıaz, *International journal of molecular sciences* **2019**, *20*, 4888.

[422] Q. Zhou, Z. Fu, Y. Gong, V. P. Seshachalam, J. Li, Y. Ma, H. Liang, W. Guan, S. Lin, S. Ghosh, et al., *Obesity* **2020**, *28*, 2153–2162.

[423] H. L. Fisk, C. E. Childs, E. A. Miles, R. Ayres, P. S. Noakes, C. Paras-Chavez, O. Kuda, J. Kopeck, E. Antoun, K. A. Lillycrop, et al., *Clinical Science* **2021**, *135*, 185–200.

[424] E. Arner, N. Mejhert, A. Kulyté, P. J. Balwierz, M. Pachkov, M. Cormont, S. Lorente-Cebrián, A. Ehrlund, J. Laurencikiene, P. Hedén, et al., *Diabetes* **2012**, *61*, 1986–1993.

[425] S. G. Dubois, L. K. Heilbronn, S. R. Smith, J. B. Albu, D. E. Kelley, E. Ravussin, L. A. A. R. Group, *Obesity* **2006**, *14*, 1543–1552.

[426] S. T. Nadler, J. P. Stoehr, K. L. Schueler, G. Tanimoto, B. S. Yandell, A. D. Attie, *Proceedings of the National Academy of Sciences* **2000**, *97*, 11371–11376.

[427] J. Jo, O. Gavrilova, S. Pack, W. Jou, S. Mullen, A. E. Sumner, S. W. Cushman, V. Periwal, *PLoS computational biology* **2009**, *5*, e1000324.

[428] L. A. Muir, C. K. Neeley, K. A. Meyer, N. A. Baker, A. M. Brosius, A. R. Washabaugh, O. A. Varban, J. F. Finks, B. F. Zamarron, C. G. Flesher, et al., *Obesity* **2016**, *24*, 597–605.

[429] T. Song, S. Kuang, *Clinical Science* **2019**, *133*, 2107–2119.

[430] A. D. Hildreth, F. Ma, Y. Y. Wong, R. Sun, M. Pellegrini, T. E. O'Sullivan, *Nature Immunology* **2021**, *22*, 639–653.

[431] A. E. Locke, B. Kahali, S. I. Berndt, A. E. Justice, T. H. Pers, F. R. Day, C. Powell, S. Vedantam, M. L. Buchkovich, J. Yang, et al., *Nature* **2015**, *518*, 197–206.

[432] D. Shungin, T. W. Winkler, D. C. Croteau-Chonka, T. Ferreira, A. E. Locke, R. Mägi, R. J. Strawbridge, T. H. Pers, K. Fischer, A. E. Justice, et al., *Nature* **2015**, *518*, 187–196.

[433] M. Litviňuková, C. Talavera-López, H. Maatz, D. Reichart, C. L. Worth, E. L. Lindberg, M. Kanda, K. Polanski, M. Heinig, M. Lee, et al., *Nature* **2020**, *588*, 466–472.

[434] C. Hepler, L. Vishvanath, R. K. Gupta, **2017**, *31*, 127–140.

Appendix A

Supplementary Information related to Chapter 3

A.1 Supplementary Figures

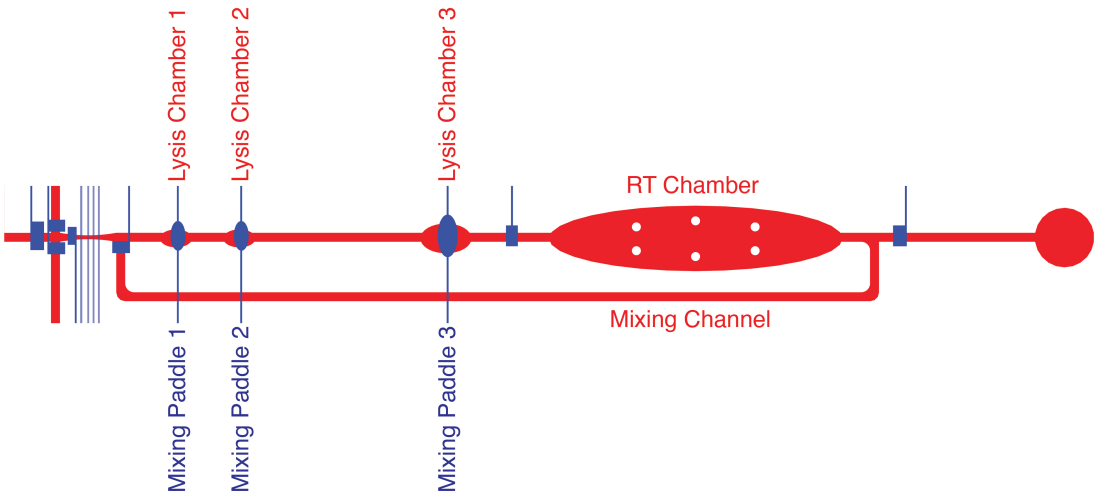

Figure A.1: **Detailed schematic of a single reaction lane on the µCB-seq device**. The lysis module has 3 reaction chambers and the RT module has 1 reaction chamber connected to the mixing channel. Both lysis and RT modules are separated from each other by the two reagent valves. RT primers with known barcode sequences are spotted in the Lysis Chamber 3 of each reaction lane. Positioned atop each of the reaction chambers in the lysis module are mixing paddles, which are actuated to resuspend barcoded RT primers in lysis buffer and circulate the relatively viscous RT mix throughout the mixing channel.

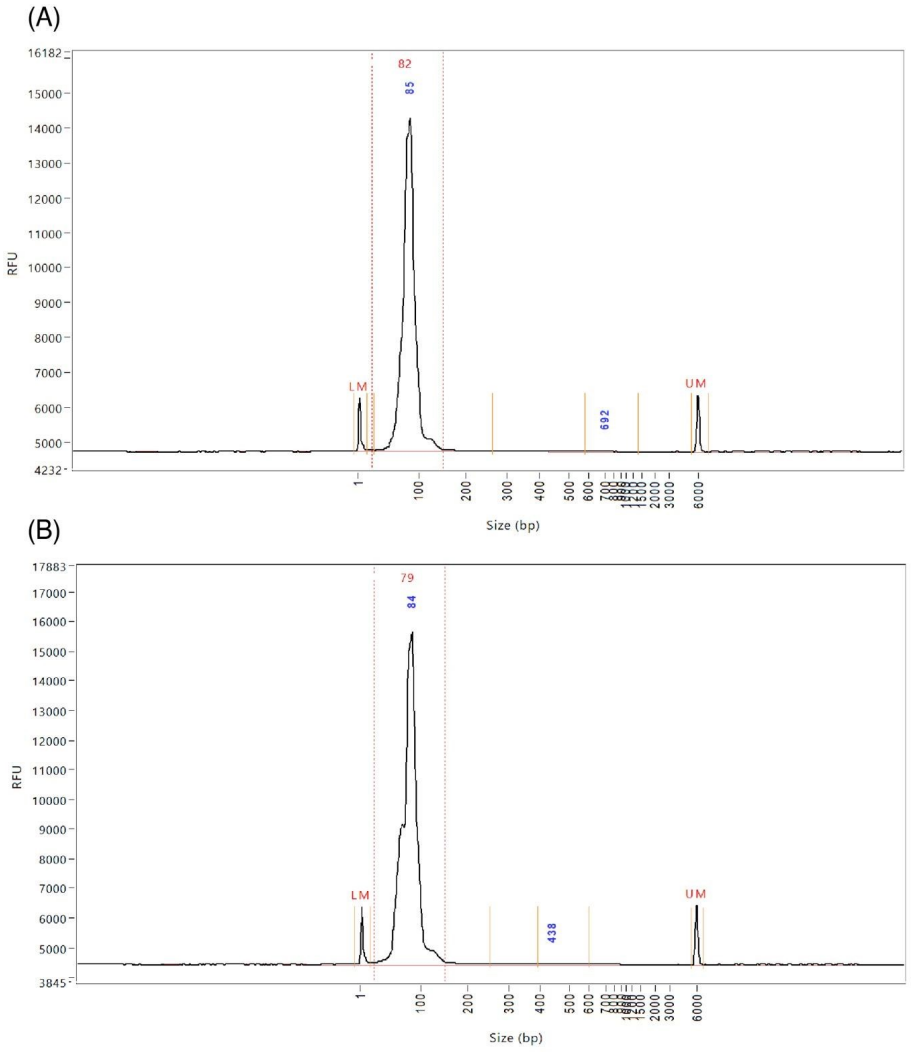

Figure A.2: **Validation of intact RT primer recovery from a PDMS slab after baking.** Fragment analysis size distribution traces for barcoded primers that were suspended in nuclease-free water at RT and (A) left in the original tube or (B) spotted on PDMS, dried, baked at 80 °C and recovered by resuspending in nuclease-free water.

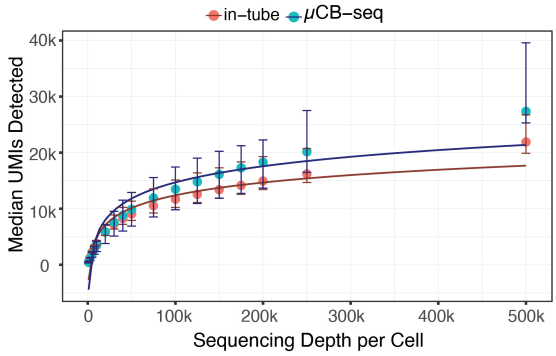

Figure A.3: Median UMIs detected for downsampled read depth across single HEK cells sequenced using µCB-seq (n = 16) and mcSCRB-seq in-tube (n = 16). Error bars indicate the interquartile range.

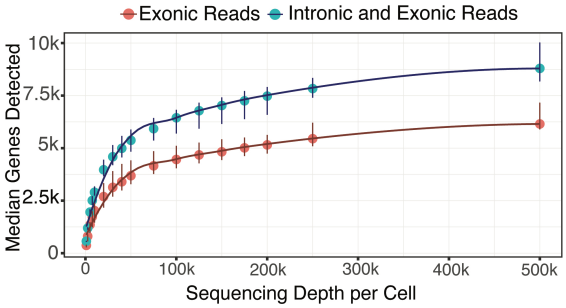

Figure A.4: Median genes detected using only exonic or both exonic and intronic reads for downsampled read depths across single HEK cells (n=16) sequenced using µCB-seq. Error bars indicate the interquartile range.

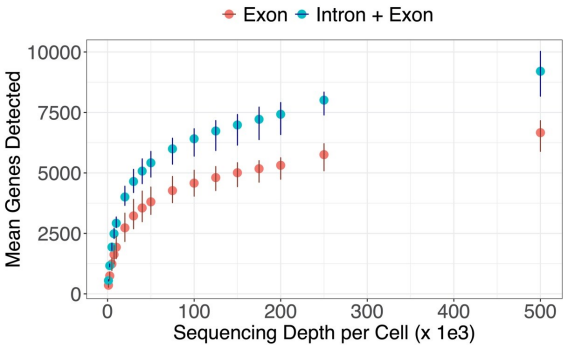

Figure A.5: Mean genes detected using only exonic or both exonic and intronic reads for downsampled read depths across single HEK cells (n=16) sequenced using µCB-seq. Error bars indicate the interquartile range.

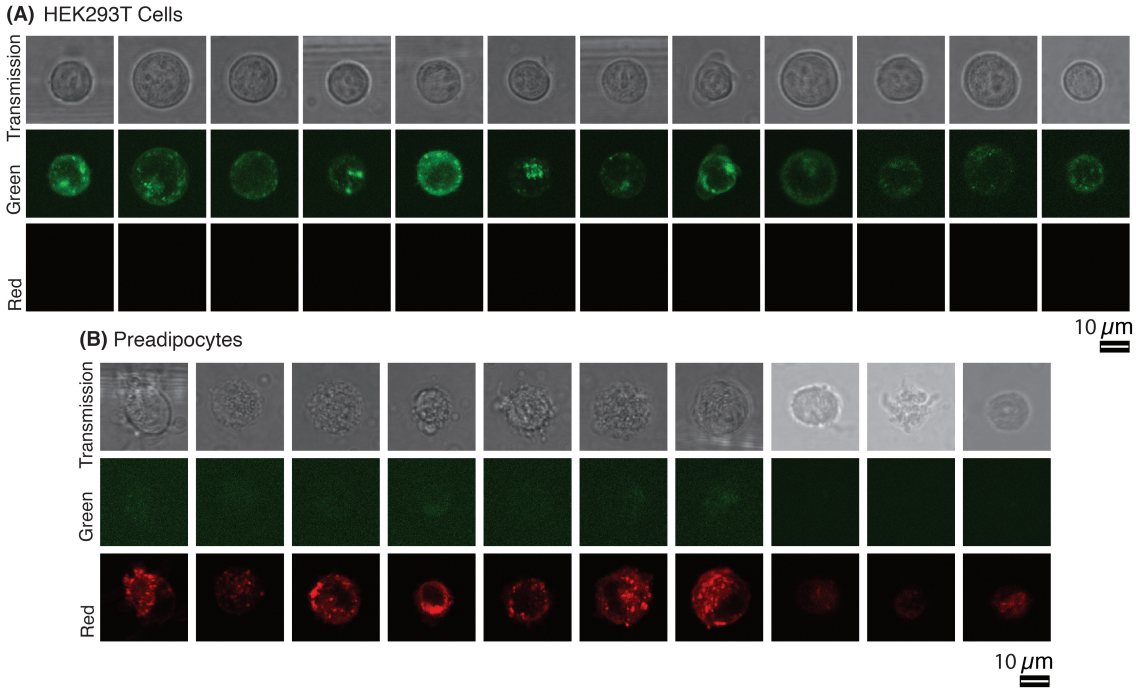

Figure A.6: Scanning transmission and two-channel fluorescent confocal images of all (A) HEK293T cells and (B) Preadipocytes stained using CellBrite Green and Red dye respectively.

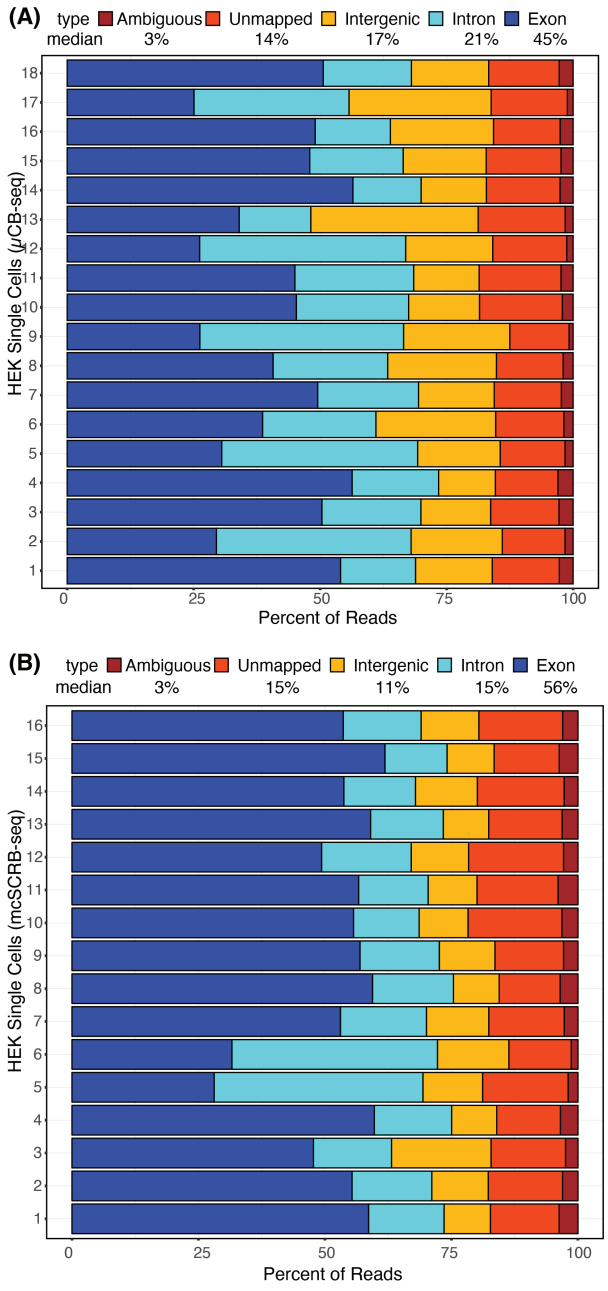

Figure A.7: Mapping Statistics for single HEK Cells sequenced using (A) µCB-seq and (B) mcSCRB-seq in-tube.

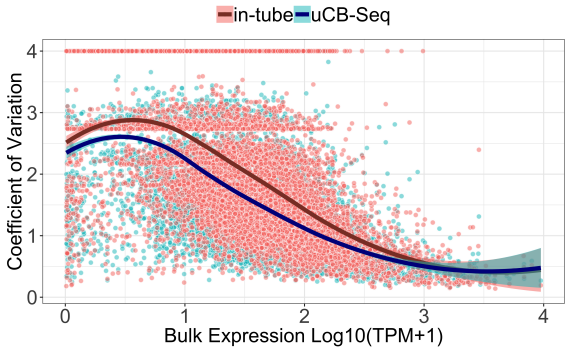

Figure A.8: **µCB-seq and in-tube mcSCRB-seq protocol have comparable precision.** The coefficient of variation for each gene (SD normalized by the mean) is plotted against its bulk expression for HEK cells sequenced using µCB-seq (n=16) and mcSCRB-seq in-tube (n=16). HEK Cells were sequenced to a depth of 200,000 reads and bulk RNA-seq library was prepared using 1 µg HEK total RNA sequenced to a depth of 63 million reads. CV was calculated for all common genes detected in bulk RNA-seq, µCB-seq and mcSCRB-seq libraries. The highlighted region displays the 95% confidence interval around the smooth fit as determined by loess regression.

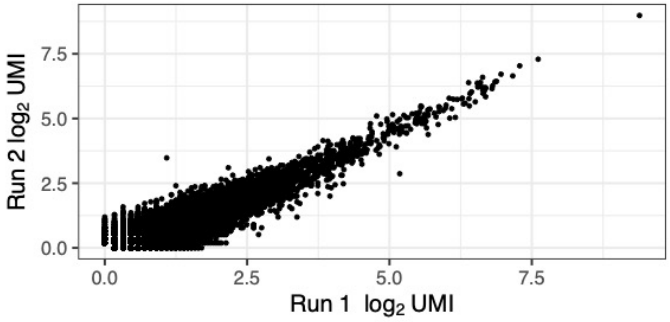

Figure A.9: **Pairwise correlation of the mean UMI transcript counts for two µCB-seq HEK293T single cell transcriptome sequencing experiments.** Each dot represents the log-transformed mean UMI counts for a given transcript for all cells at a depth of 250,000 reads per cell. Data from 8 and 7 cells are shown for Run1 and Run2 respectively.

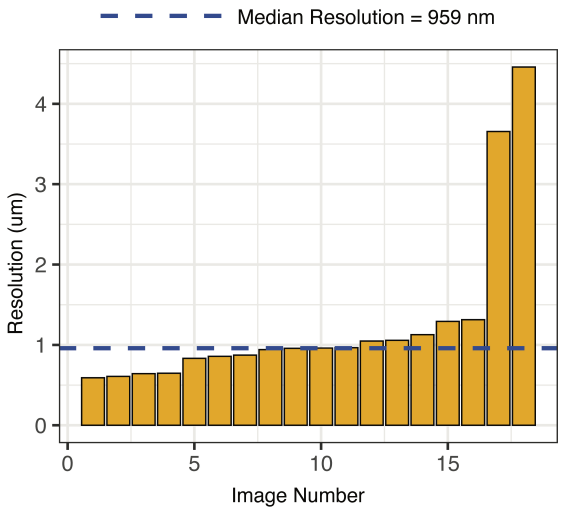

Figure A.10: **Spatial resolution in confocal fluorescent images of HEK cells and Preadipocytes.** The blue dashed line indicates the median resolution across 18 images. Detailed image analysis steps are explained in the Materials and Methods section.

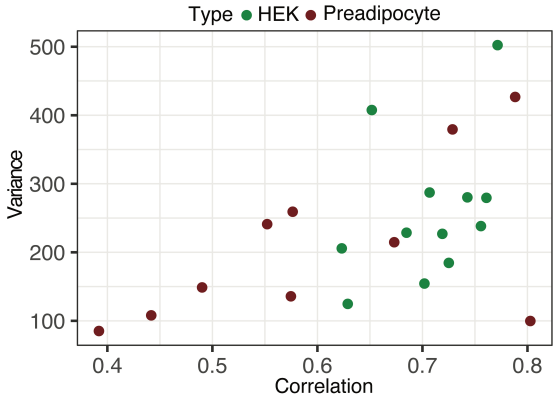

Figure A.11: Annotation of HEK293T cells and Preadipocytes in a 2-Dimensional Correlation vs Variance space as quantified for grayscale intensities in the scanning transmission images. Detailed image analysis steps are explained in the Materials and Methods section.

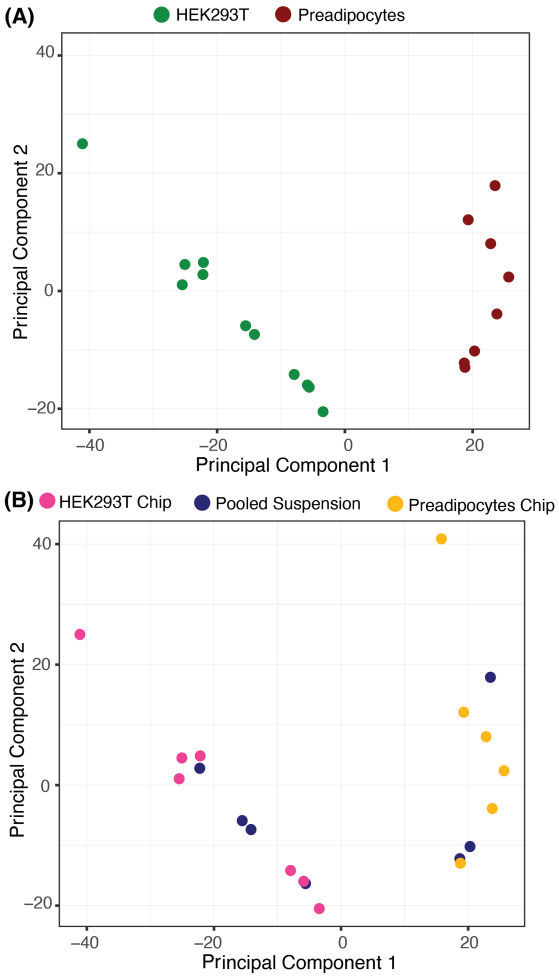

Figure A.12: Annotation of HEK293T cells and Preadipocytes in Principal Component Space based on (A) Cell-clusters identified using unsupervised hierarchical clustering in the PCA space and (B) µCB-seq devices on which cells were processed. Device 1 processed just HEKs (n=7), Device 2 processed a mix of both HEKs (n=4) and preadipocytes (n=3), and Device 3 processed just Preadipocytes (n=6)

A.2 Supplementary Tables

Table A.1: RT Primers with known barcode sequences used in µCB-seq. Barcodes bc1-bc10 were used for experiments on HEK293T Total RNA and HEK293T single cells (Fig. 3.3 and Fig. 3.4), whereas underlined barcodes were used for the imaging and sequencing of HEK293T cells and Preadipocytes (Fig. 3.5). The underlined subset of ten barcodes was selected to ensure sequence diversity at every barcode base for optimal next-generation sequencing performance without PhiX spike-ins.

Barcode Number	Barcode Sequence
bc1	TCACAGCA
bc2	GTAGCACT
bc3	ATAGCGTC
bc4	CTAGCTGA
bc5	CTACGACA
bc6	GTACGCAT
bc7	ACATGCGT
bc8	GCATGTAC
bc9	ATACGTGC
bc10	GCAGTATC
bc13	TGCTACAG
bc15	CGCTATGA
bc26	ATGCACGT
bc40	TATGCACG
bc47	CATCGTGA
bc82	CCAGTTAG
bc92	GGCATTGT

Table A.2: Sequencing summary statistics for all 10 pg HEK total RNA samples processed on two µCB-seq devices and analyzed as presented in Figure 3.3 of this Chapter 3. All libraries were sequenced in a single batch using the Illumina MiniSeq sequencing platform. Total reads for all 20 libraries are 1,358,764 with an average sequencing depth of 67,938 per library.

Barcode	**Chip #**	**Sequencing Depth**
ACTGCGT	1	76,710
ATCGTGC	1	111,727
ATGCGTC	1	41,827
CTCGACA	1	60,085
CTGCTGA	1	38,561
GCGTATC	1	84,913
GCTGTAC	1	93,049
GTCGCAT	1	81,865
GTGCACT	1	44,354
TCCAGCA	1	68,052
ACTGCGT	2	64,073
ATCGTGC	2	64,705
ATGCGTC	2	50,154
CTCGACA	2	45,715
CTGCTGA	2	60,891
GCGTATC	2	51,237
GCTGTAC	2	66,989
GTCGCAT	2	92,830
GTGCACT	2	69,444
TCCAGCA	2	91,583

Table A.3: Sequencing summary statistics for all single HEK cells processed on two μCB-seq devices and analyzed as presented in Figure 3.4 of this manuscript. All libraries were sequenced in a single batch using the Illumina MiniSeq sequencing platform. Total reads for all libraries combined are 8,908,444 with an average sequencing depth of 494,914 per cell.

Barcode	**Chip #**	**Sequencing Depth**
ACATGCGT	1	498,991
ATACGTGC	1	747,079
ATAGCGTC	1	269,734
CTAGCTGA	1	307,760
GCAGTATC	1	388,638
GCATGTAC	1	1,075,768
GTACGCAT	1	223,584
GTAGCACT	1	446,942
TCACAGCA	1	339,732
ACATGCGT	2	445,705
ATACGTGC	2	843,948
ATAGCGTC	2	227,311
CTACGACA	2	378,319
CTAGCTGA	2	244,798
GCATGTAC	2	668,187
GTACGCAT	2	570,252
GTAGCACT	2	620,452
TCACAGCA	2	611,244

Table A.4: Sequencing summary statistics for all single HEK cells and Preadipocytes processed on three µCB-seq devices and analyzed as presented in Figure 3.5 of this manuscript. All libraries were sequenced in a single batch using the Illumina MiniSeq sequencing platform. Total reads for all libraries combined are 6,925,205 with an average sequencing depth of 346,260 per cell.

Barcode	**Chip #**	**Sequencing Depth**
ATGCACGT	1	254,436
CATCGTGA	1	172,172
GGCATTGT	1	166,915
GTAGCACT	1	305,948
TATGCACG	1	254,321
TCACAGCA	1	371,975
TGCTACAG	1	687,106
CATCGTGA	2	129,908
CGCTATGA	2	553,155
GGCATTGT	2	330,301
GTAGCACT	2	318,773
TATGCACG	2	631,094
TCACAGCA	2	398,464
ATAGCGTC	3	319,417
CATCGTGA	3	595,996
CGCTATGA	3	173,339
GGCATTGT	3	426989
GTAGCACT	3	245332
TCACAGCA	3	346914
TGCTACAG	3	242650

Table A.5: Sequences of DNA primers used in both mcSCRB-seq in-tube experiments and on µCB-seq devices for off-chip library preparation. Same primer sequences as in mcSCRB-seq [13] are used in this work. /5Biosg/ indicates a 5' Biotin, * indicates a phosphorothioated nucleotide, and r indicates an RNA base.

Primer	**Sequence**
SINGV6	/5Biosg/ACACTCTTTCCCTACACGACGC
P5NEXTPT5	AATGATACGGCGACCACCGAGATCTAC ACTCTTTCCCTACAC GACGCTCTTCCG*A*T*C*T
E5V6 TSO	CGCACACTCTTTCCCTACACGACGCrGrGrG

A.3 Supplementary Notes

Imaging Chamber Volume Measurement for Trapping 10 pg of Total RNA

When measuring chamber volume for Total RNA experiments in the µCB-seq device (Fig. 3.3), we observed a discrepancy in height between the µCB-seq flow molds and the reaction chambers of the PDMS µCB-seq devices with actuated control valves. Flow molds were measured by Dektak profilometer, giving an imaging chamber height of 29 µm. When imaging the corresponding chamber on the µCB-seq device via Coherent anti-Stokes Raman spectroscopy (CARS), we recorded a chamber height of 53.5 µm. Profilometry was not feasible for the closed µCB-seq device, so we elected to conservatively use the CARS measurement at the risk of overestimating volume and loading less than 10 pg Total RNA into the µCB-seq device. To measure chamber volume, we pressurized the isolation valves on a µCB-seq device and acquired a z-stack of the resultant air-filled imaging chamber. Images were thresholded in ImageJ and manually outlined to record the cross-sectional area of each imaging chamber slice. The volume of the chamber was estimated by a Riemann sum to ensure that chamber volume erred on the larger side. The chamber volume measured by this method was 1.88 nL, which resulted in our conservative input concentration of 5.31 ng/µL Total RNA to ensure no more than 10 pg of RNA was processed in each lane of the µCB-seq device.

Comparison of µCB-seq and Fluidigm C1 performance

The Fluidigm C1 is a commercial microfluidic platform with integrated valves that uses SMART-seq for full-length transcript quantification in single cells [24]. Arguel et al. demonstrated the use of Fluidigm C1 to sequence single HEK293T cells with a 5' UMI tagging protocol [244]. We used this study to benchmark the current performance of µCB-seq against the Fluidigm C1, as both studies process HEK293T cells and implement UMI-based transcript counting. There are some limitations to this comparison, however. First, there may be differences in capture efficiency and bias between a 5' sequencing chemistry and the 3' sequencing chemistry used here. Additionally, Arguel et al. used the hg19 genome for alignment with STAR, whereas we used the GRCh38 genome with STAR. Finally, Arguel et al. used Dropseq Core [32] for UMI counting whereas we used zUMIs [235] and filtered for exons. Using 500k reads as an individual point of comparison, the protocol on the C1 detected a mean of 6,000 genes. Another published result using the standard SMARTer protocol on the C1 suggests similar gene detection level of 6,000 genes, although this was carried out on a different cell type (HCT116) [23]. Our µCB-seq protocol detected a mean of 6,663 genes when counting only exons, and 9,203 genes when counting exons and introns (Fig. A.5). This suggests that µCB-seq has similar or improved sensitivity compared to other protocols on the Fluidigm C1.

With regards to sample-to-sample variability, Arguel et al. compute pairwise correlations across different microfluidic runs (with 37, 47 or 74 cells). For each device run, log-

transformed mean UMI counts across all the cells were used as input for correlation analysis, giving R = 0.92, 0.98, and 0.93. Our µCB-seq device has a slightly lower but comparable correlation value of R = 0.91 (p-value <.05) between device runs with 8 and 7 cells (Fig. A.9). This slightly lower correlation value is expected due to the 4- to 9-fold higher number of cells averaged in the C1 experiments as compared to our µCB-seq experiments.

Appendix B

Supplementary Information related to Chapter 4

B.1 Supplementary Figures

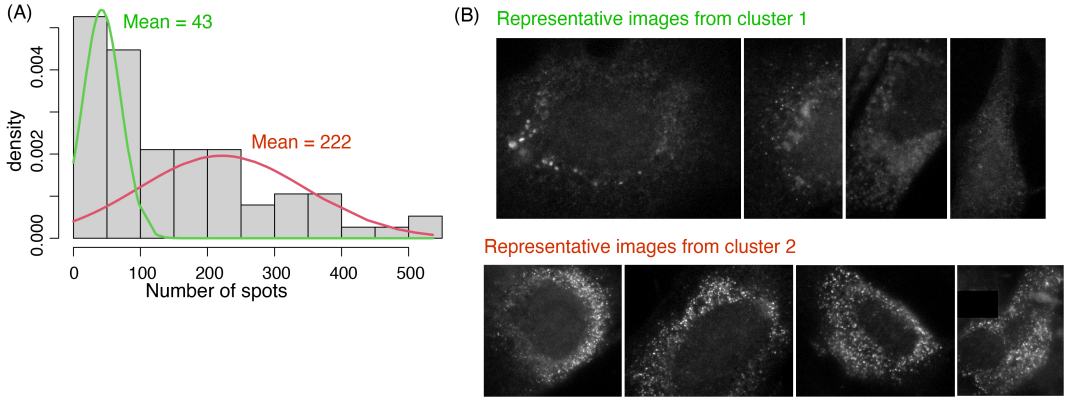

Figure B.1: Validation of scRNA-seq markers for recovering cell-type heterogeneity in brown preadipocytes using smFISH. (A) Distribution of number of MMP1 mRNA spots per cell in brown preadipocytes. Overlaid gaussian distributions represent the 2-component fit identified using Gaussian finite mixture model fitting. (B) 4 representative images of cells from cluster 1 (mean = 43) and cluster 2 (mean = 222). Representative images are cells within +/− 7 transcript counts from the mean.

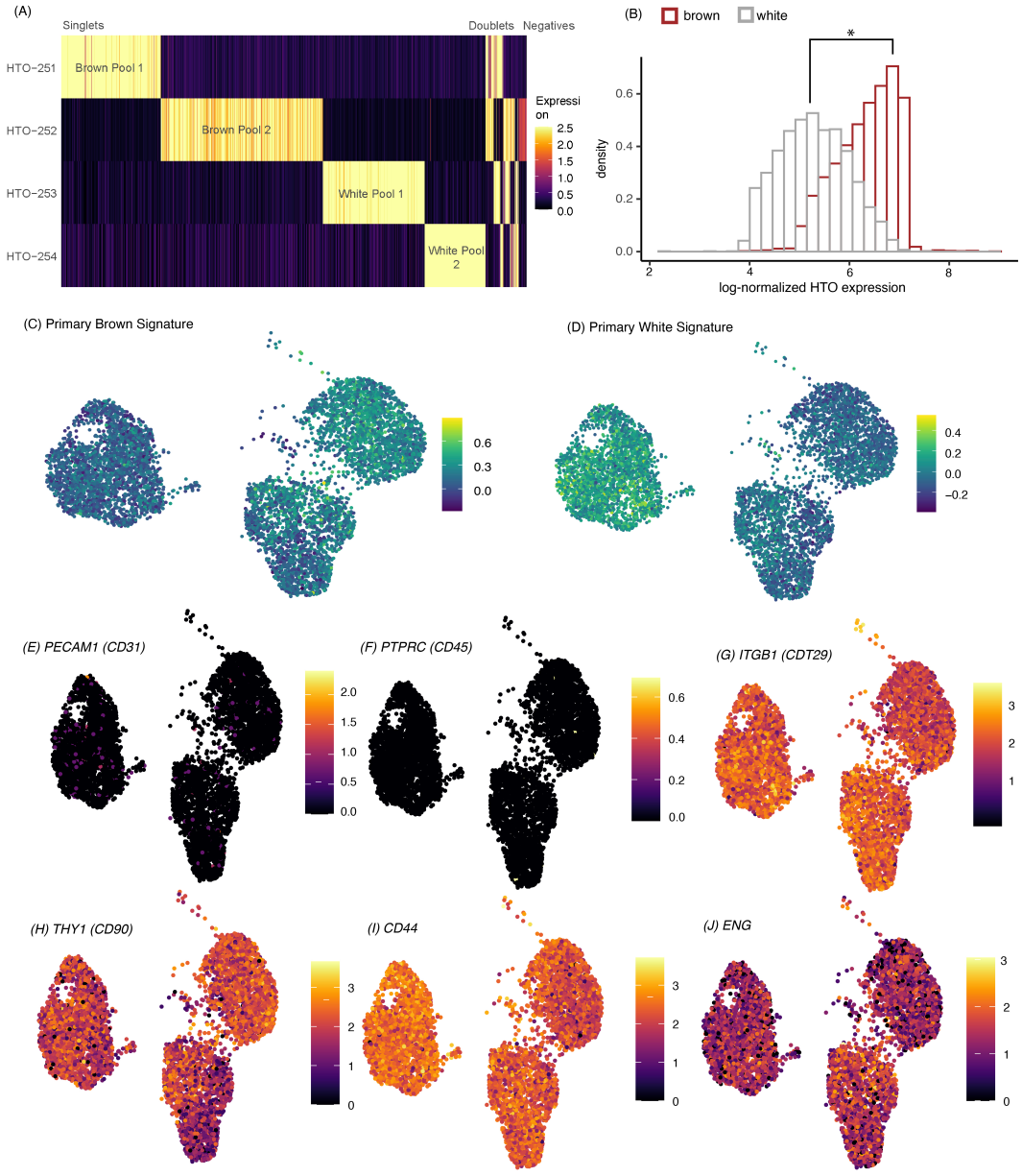

Figure B.2: **Analysis of white and brown preadipocyte scRNA-seq dataset** (A) Log-normalized expression of four hashtag antibodies used for multiplexing of white and brown preadipocytes (whole-cells). (B) Distribution of normalized hashtag antibody expression in white and brown preadipocytes identified as singlets. Statistical testing was performed using a two-sided t-test. (C) to (D) Heatmap of transcriptional signature scores for white preadipocyte (C) and brown preadipocyte (D). Original signatures were defined using primary white and brown preadipocytes isolated from the same anatomical region as the *in vitro* model system used in our study. (E) to (J) Expression profiles of marker genes in scRNA-seq dataset.

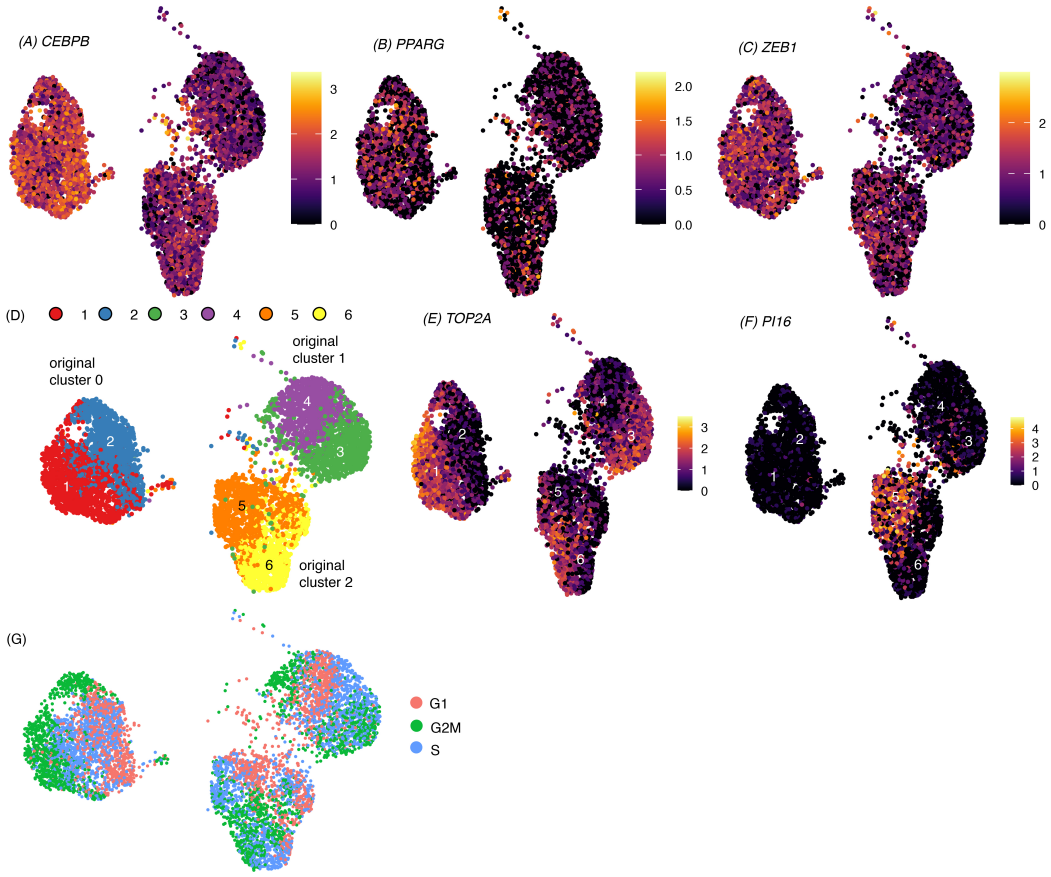

Figure B.3: **Analysis of white and brown preadipocyte scRNA-seq dataset** (A) to (C) Expression profiles of marker genes in scRNA-seq dataset. (D) Sub-clusters identified for each of the original cluster 0, 1, and 2 in Fig. 4.1A. (E) Expression profile of mitotic cell marker TOP2A, one of the top marker genes during sub-clustering of original clusters 0, and 1 in scRNA-seq dataset. (F) Expression profile of adipocyte progenitor marker PI16, the top marker gene during sub-clustering of original cluster 2 in scRNA-seq dataset. (G) Cells annotated by cell-cycle phase as calculated using Seurat

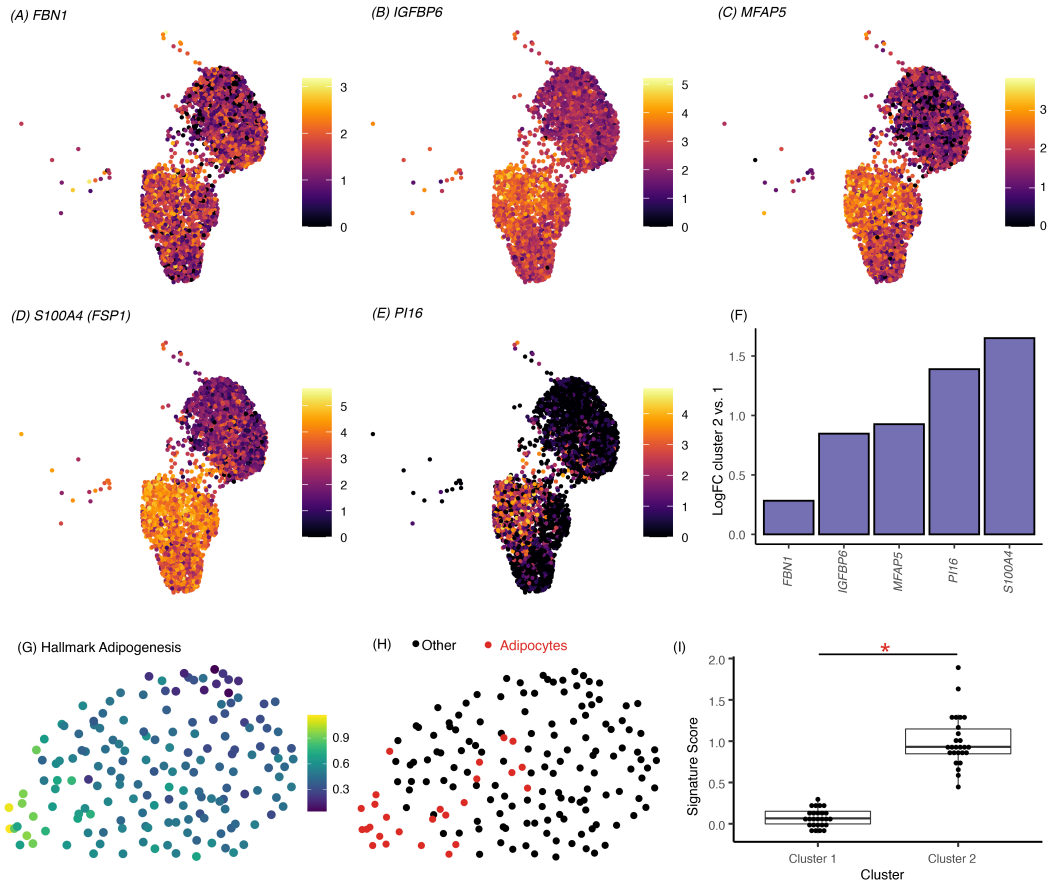

Figure B.4: Differential expression in brown preadipocyte scRNA-seq dataset between cluster 2 and cluster 1. (A) to (E) Expression profile of marker genes for Fsp1+ fibroblasts identified in [307] (F) Log fold change values of the marker genes as calculated using cluster 2 vs cluster 1 differential expression test. All genes were significantly enriched in cluster 2 with FDR < 0.05. Also see Supplemental Table 2C. (G) Heatmap of Hallmark Adipogenesis signature defined in MSig database. The signature consists of genes up regulated during adipocyte differentiation. (H) Cells identified as mature adipocytes after unsupervised clustering in Seurat (I) Boxplot of transcriptional signature scores in mature adipocytes (highlighted in red in panel H). Signatures were defined for cluster 1 and cluster 2 cells using scRNA-seq dataset (see Supplemental Table 2C). Statistical testing was performed using two-sided Mann-Whitney U-test.

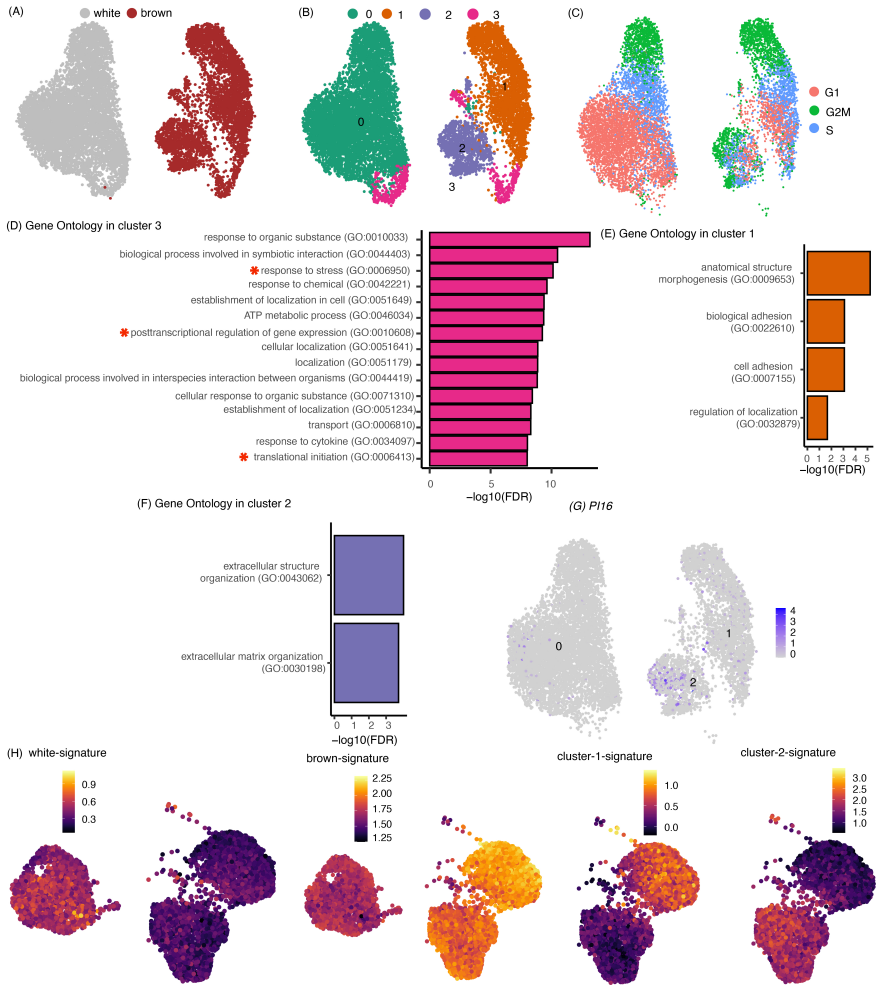

Figure B.5: **Unsupervised clustering of white and brown preadipocytes snRNA-seq dataset** (A) and (B) UMAP visualization of white and brown preadipocytes annotated either manually to reflect the sample of origin (A) or based on unsupervised clustering (B). (C) Cells annotated by cell-cycle phase as calculated using Seurat (D) Top gene ontology biological processes (BP) terms enriched in cluster 3 based on a cluster 3 vs. all DE test. Marked in red is the enrichment of BP terms because of stress response genes (response to stress), mitochondrial genes (ATP metabolic process), and ribosomal mRNA genes (translational initiation). Enrichment of mitochondrial and ribosomal mRNA genes indicates the presence of cellular background RNA contamination (see Fig. B.8C). (E) and (F) Top 10 gene ontology terms in brown cluster 1 (E) and cluster 2 (F) in scRNA-seq dataset (Fig. 4.1C) that are also enriched in respective clusters in snRNA-seq dataset. (G) Expression profile of adipocyte progenitor marker PI16 in snRNA-seq dataset. Also see Fig. B.3F. (H) Heatmap of transcriptional signature scores for white preadipocyte (white), brown preadipocyte (brown), brown preadipocyte cluster 1 (one), and brown preadipocyte cluster 2 (two) as plotted on the UMAP visualization of scRNA-seq data. Signatures were defined using snRNA-seq data using white vs brown, or cluster-1 vs cluster-2 differential expression testing

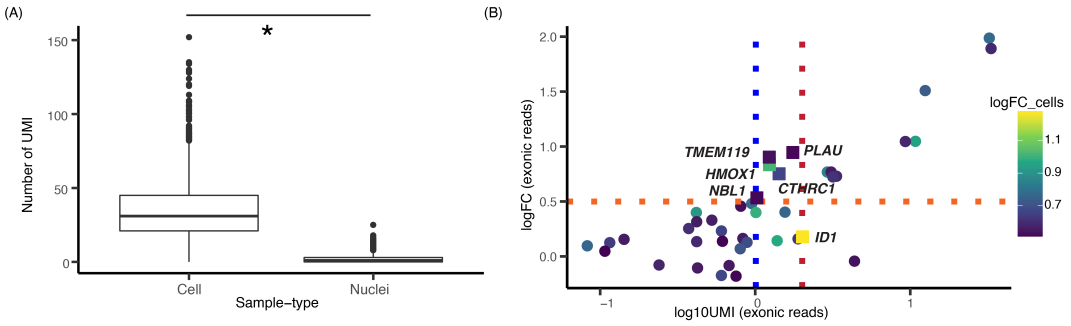

Figure B.6: **Investigating lack of ID1 DE in white nuclei over brown nuclei.** (A) Boxplot of number of ID1 UMIs detected in each cell or nuclei isolated from white preadipocyte. (B) Log-fold-change vs log-UMI counts in white nuclei when using only exonic reads, where each dot represents a white-preadipocyte-enriched gene (white vs brown DE test) detected using scRNA-seq dataset (Fig. 4.1A). Horizontal dotted line indicates logFC cutoff value of 0.5 used as a threshold for DE testing. All genes had a logFC > 0.5 in scRNA-seq dataset. Vertical blue dotted line indicates smallest mean UMI count at which a gene was detected to be differentially expressed. Vertical red dotted line indicates the mean UMI count for ID1 gene. ID1 gene is marked with a square, along with genes TMEM119, PLAU, HMOX1, NBL1, and CTHRC1.

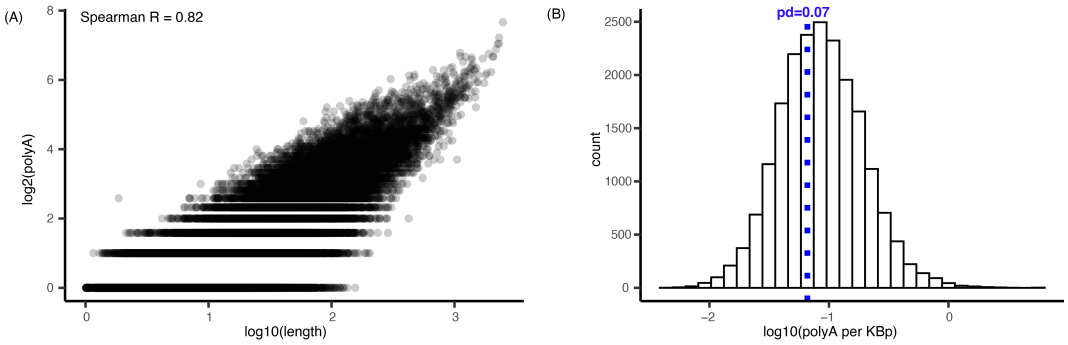

Figure B.7: **Estimating polyA-tract density per Kbp in the genic region.** (A) Scatter plot of total number of polyA-tracts (greater than 15−bp) plotted against gene length. Each dot is a gene in the GRCh38−2020A reference from cellranger analysis pipeline. (B) Distribution of mean number of polyA-tracts per Kbp for each gene in panel A. Blue dotted line indicates mean number of poly−A tracts per Kbp across all genes and is used to estimate pd=0.07. See Note SB.2 for details on normalization strategy.

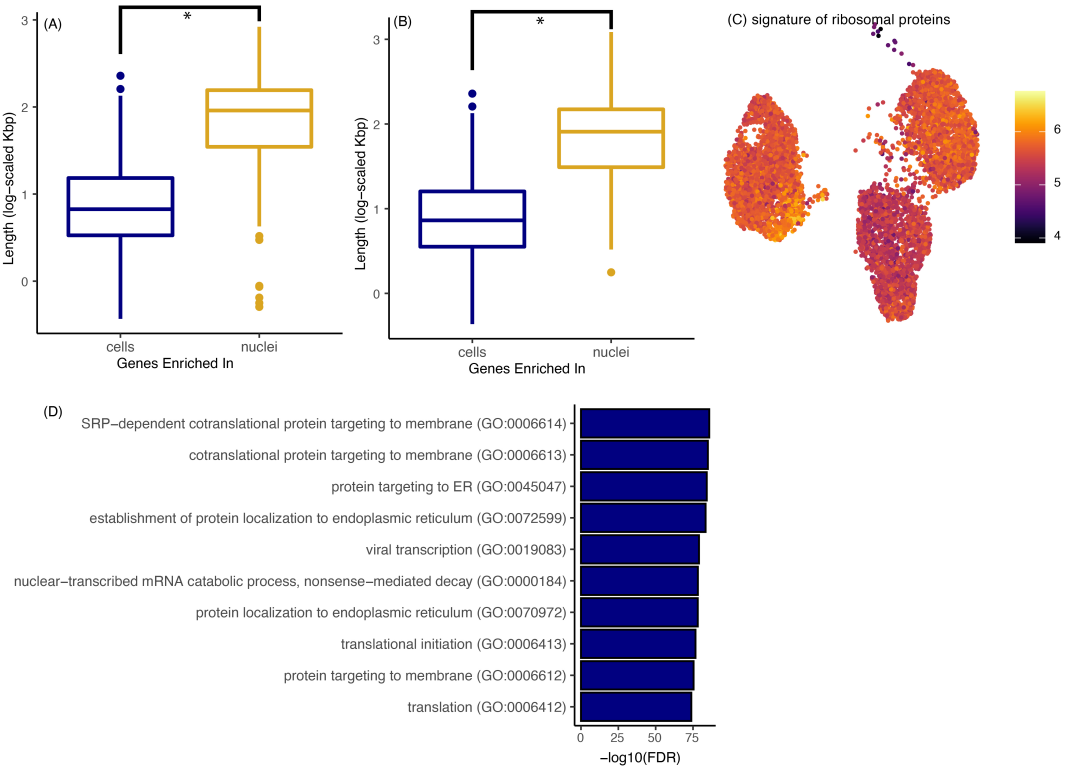

Figure B.8: **Gene length-associated detection bias in snRNA-seq.** (A) Distribution of gene length for genes enriched in cells (in blue) and nuclei (in yellow) with log fold-change > 1 and FDR < 0.05 including both intronic and exonic reads. Intronic UMI-count matrix was normalized to correct for gene length bias in both cells and nuclei (see Note B.2). (B) Distribution of gene length for genes enriched in cells (in blue) and nuclei (in yellow) with log fold-change > 1 and FDR < 0.05 using only exonic reads. (C) Heatmap of transcriptional signature score defined using top 100 genes enriched in cells vs. nuclei in white preadipocytes based on log fold-change values after normalization. The scores are plotted on the 2D UMAP visualization of scRNA-seq preadipocyte data. (D) Top 10 gene ontology terms enriched in white cells as compared to white nuclei based on differential expression after normalization.

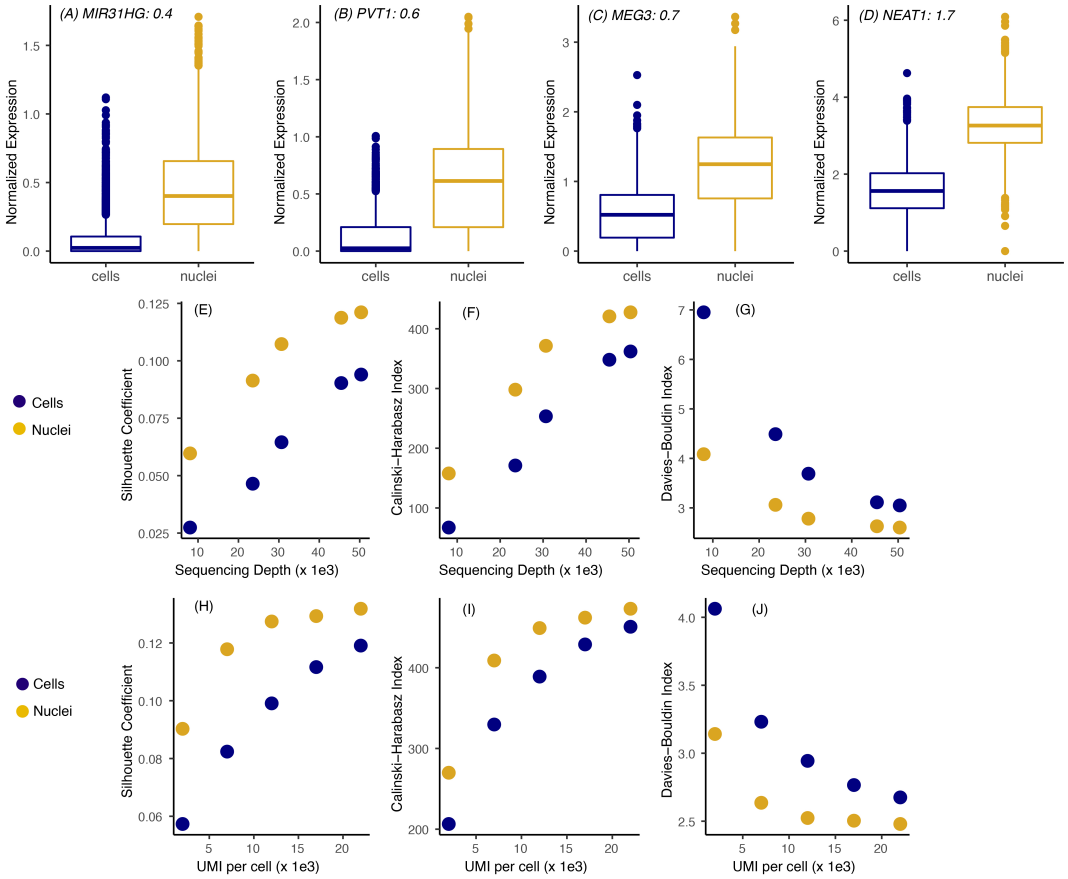

Figure B.9: **Enrichment of lncRNAs in the nuclear transcriptome.** (A) to (D) Expression of adipogenic regulatory lncRNAs in brown nuclei over brown whole cells. Black text indicates logFC value for brown nuclei vs. brown cells DE test with FDR < 0.05 after normalization. (E) to (G) Cluster separation resolution quantification between brown cluster 2 vs cluster 1 in scRNA-seq and snRNA-seq dataset. Only lncRNAs were considered for PCA manifold generation. Both datasets were subsampled to have the same number of cells/nuclei and same number of mean transcriptome mapped reads. (H) to (J) Similar analysis as panel (E) to (G) but normalization was performed to have the same number of UMI counts per cell/nuclei. A higher Silhouette coefficient and Calinski Harabasz and a lower Davies Bouldin index indicate superior cluster separation performance.

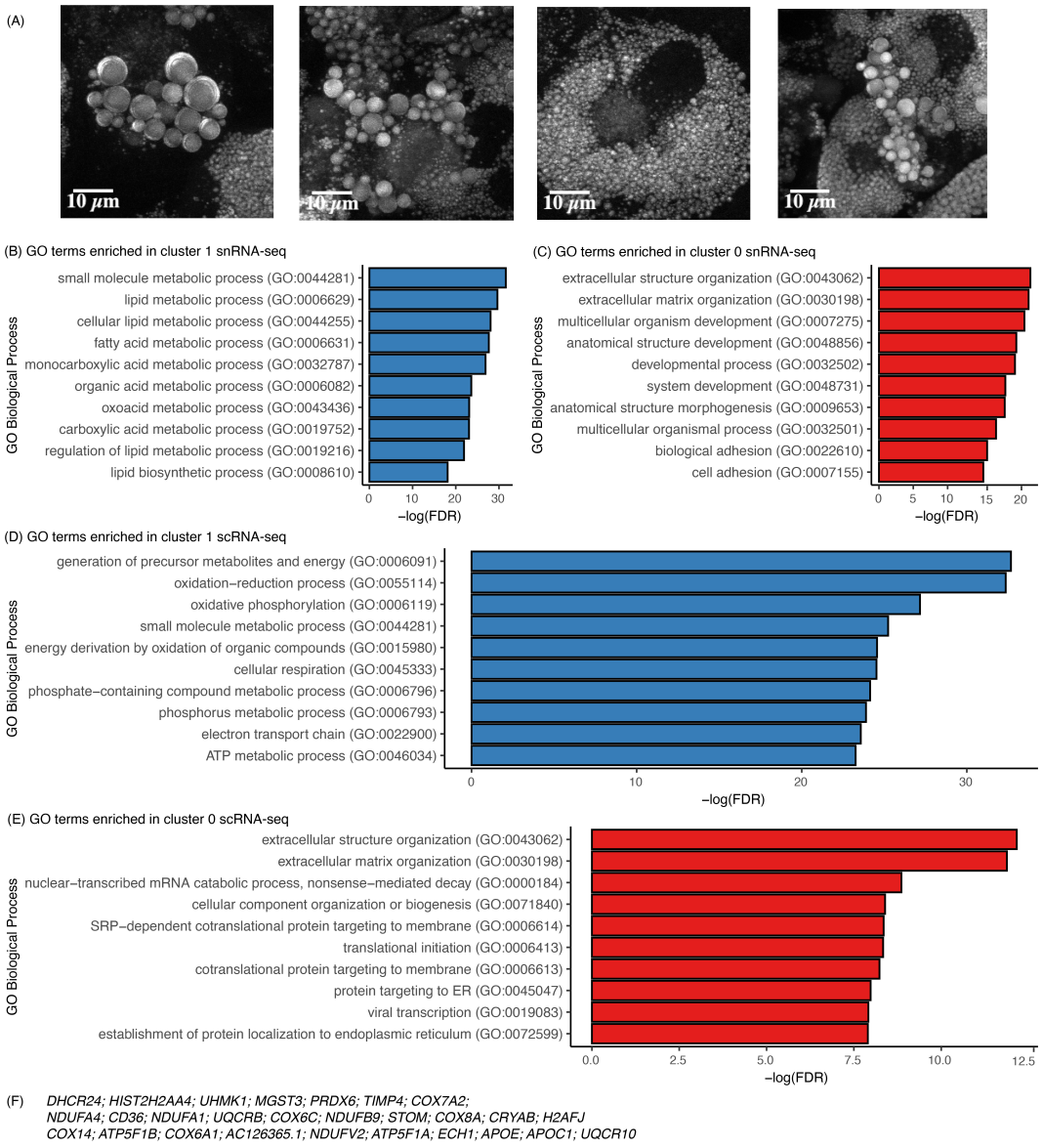

Figure B.10: **Comparative analysis of nuclear and whole-cell transcriptome at mature adipocyte stage** (A) Coherent anti-stokes Raman imaging of human white preadipocytes differentiated for 20 days using a chemical adipogenic induction cocktail. The images were acquired at 2845 cm^{-1} wavenumber, which corresponds to the CH_3 peak present in lipids. Z-stacked images were acquired and the maximum intensity projection for each pixel was plotted. (B) and (C) Top 10 gene ontology terms enriched in cluster 1 (panel B) and cluster 0 (panel C) in snRNA-seq dataset. (D) and (E) Top 10 gene ontology terms enriched in cluster 1 (panel D) and cluster 0 (panel E) in scRNA-seq dataset. (F) List of 27 genes differentially enriched in cluster 1 (mature adipocytes) in scRNA-seq dataset but not differentially enriched in cluster 1 of snRNA-seq dataset

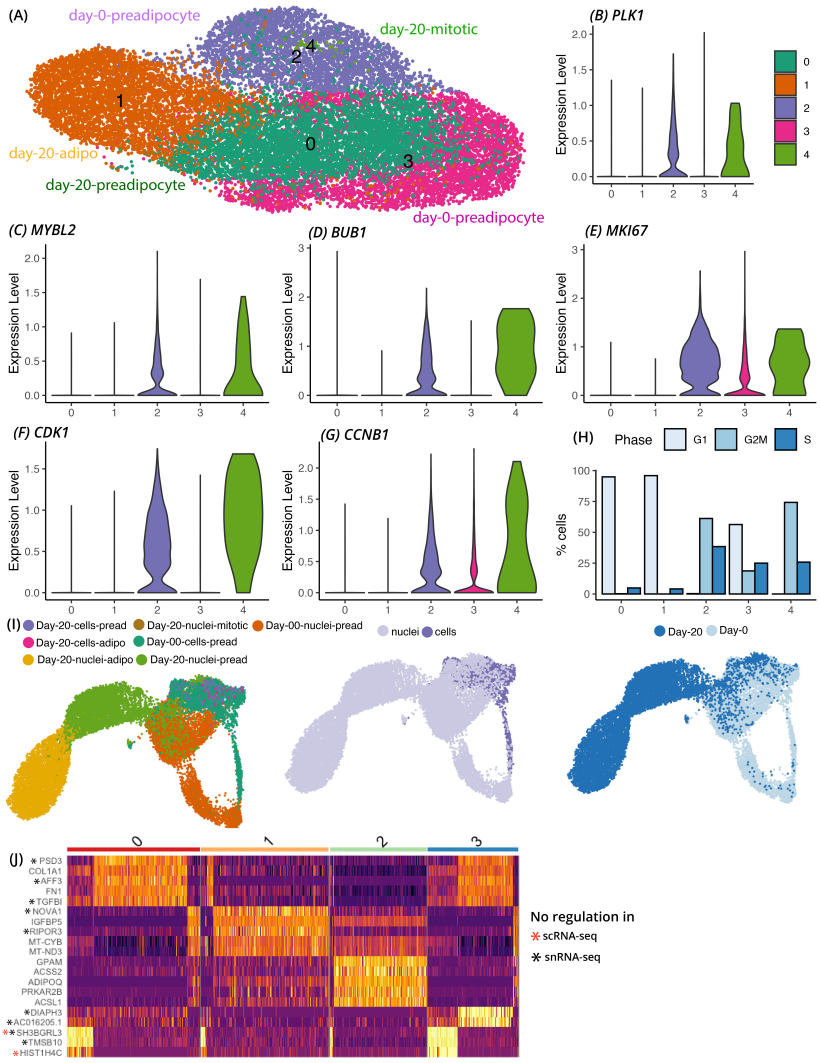

Figure B.11: **Proliferating vs growth arrested cells in snRNA-seq and scRNA-seq white preadipocyte dataset.** (A) Supervised clustering of integrated scRNA-seq and snRNA-seq white preadipocyte (day-0) and white adipocyte (day-20) dataset. See Note B.2 for details regarding clustering scheme. (B) to (G) Violin plots of common proliferation and mitosis marker genes in clusters identified in panel (A). (H) Bar plot of distribution of cell cycle phase assignment in the clusters identified in panel (A). Y-axis plots the percent of cells belonging to different cell cycle phase for every cluster. See Note B.2 for details regarding cell cycle phase assignment. (I) UMAP visualization of integrated white preadipocyte day-0 and day-20, scRNA-seq and snRNA-seq datasets. Cells are annotated by original clusters (left panel), sequencing technique (middle panel), and harvestation day (right panel). Integration was performed using Seurat v3. (J) Heat map of top 5 marker genes for each cluster identified using Seurat (Fig. 4.7C right-most panel), with genes as rows, and cells as columns. The color bar on top represents cluster assignment. All genes were differentially expressed in both scRNA-seq and snRNA-seq datasets, except for the ones marked in red or black (see Methods).

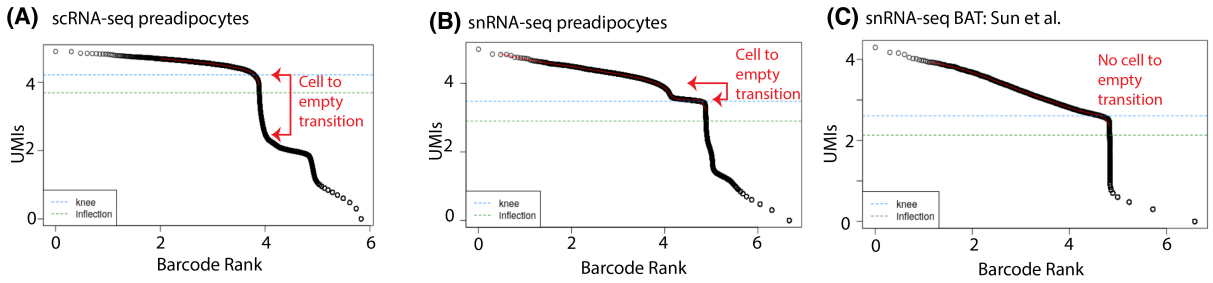

Figure B.12: **Background mRNA levels in scRNA-seq and snRNA-seq libraries**(A) Elbow plot for scRNA-seq dataset of human preadipocytes. On x-axis are barcodes ranked by their UMI counts (y-axis). Both X and Y axes are log10-transformed (B) Same plot as (A) but for snRNA-seq dataset from same cell-types(C) Same plot as (A) but for a publicly available snRNA-seq dataset [273]. The red line marks the transition region from droplets containing cells to empty droplets.

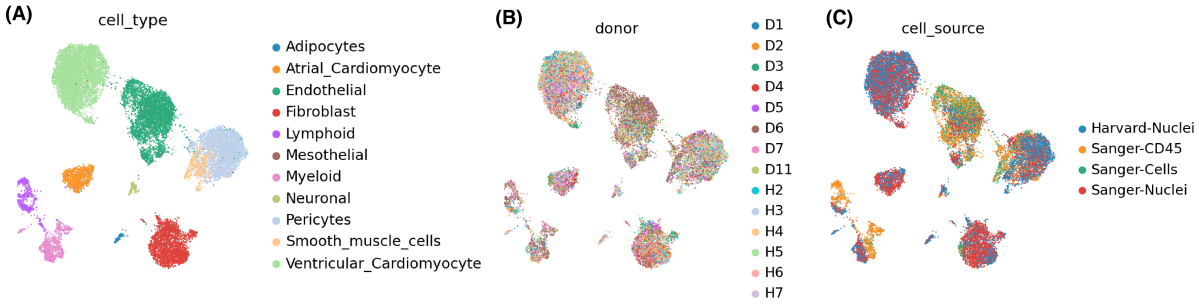

Figure B.13: **Integration of *in vivo* derived scRNA-seq and snRNA-seq datasets with scvi-tools** UMAP visualization of human heart cell atlas dataset [433] colored by (A) cell-type classification (B) donor classification and (C) technique classification. Sanger-nuclei and Harvard-nuclei are snRNA-seq datasets and Sanger-CD45 and Sanger-cells are scRNA-seq dataset. Integration was performed by Adam Gayoso.

B.2 Supplementary Notes

Validation of scRNA-seq marker genes in recovering brown preadipocyte heterogeneity using smFISH

scRNA-seq of brown preadipocytes revealed existence of two distinct cell-types (Fig. 4.1B) marked by differential expression of several genes. We used single-molecule fluorescent in situ hybridization (smFISH) imaging to validate the differential expression of these genes in situ. Specifically, we targeted cluster-2 enriched gene MMP1 for smFISH by designing short oligonucleotide probes complementary to the coding region of this gene (see Methods). Quantitative spot counting analysis, followed by gaussian mixture model fitting identified a 2-component bimodal distribution as the best fit (Fig. B.1, mean cluster 1 = 43 transcripts/cell, mean cluster 2 = 222 transcripts/cell), thereby corroborating the observation of two types of brown preadipocytes in our system. We used a likelihood ratio test of a 1-component fit against 2-component fit to determine which model best fits the data. This test yielded a p-value of 0.002 for the goodness of fit assessment between the two models, suggesting that the 2-component model is a more accurate approximation than the 1-component model for describing the parametric space of the observed distribution. We then used a 2-component negative binomial model to fit the count distributions, which is commonly used in single-cell RNA-seq measurements [360]. The negative binomial model yielded similar cluster distribution, with mean transcripts per cell of 45 for cluster 1 and 174 for cluster 2, suggesting that our findings are independent of the distribution used.

Validation of scRNA-seq marker genes in recovering brown preadipocyte heterogeneity using smFISH

In our scRNA-seq dataset, ID1 was the top differentially expressed (DE) gene in white preadipocyte over brown. However, in snRNA-seq dataset, ID1 was not DE in white nuclei over brown nuclei. Comparison of transcript abundance for ID1 across scRNA-seq and snRNA-seq in white preadipocytes revealed a significantly higher number of UMIs in scRNA-seq, even at a shallower sequencing depth (50,000 reads vs 75,000 reads; Fig. B.6A). To better understand the lack of DE of ID1 in single nuclei, we compared the transcript abundance (UMI count) in nuclei to the logFC in nuclei for all genes that were detected as DE in white preadipocytes using scRNA-seq (logFC>0.5). Among genes that were detected as DE in whole cells, any gene that had a transcript abundance below ≈ 1 UMI in nuclei, were not detected as significantly DE in snRNA-seq (logFC < 0.5, Fig. B.6B). Meanwhile, multiple genes (PLAU, TMEM119, HMOX1, CTHRC1, and NBL1) with lower nuclear transcript abundance than ID1, were significantly differentially expressed in white nuclei (logFC > 0.5, Fig. B.6B). Moreover, these genes had a smaller effect size (logFC enrichment) than ID1 in single cells. Together, these results suggest that ID1 is not differentially enriched in nuclei but is differentially enriched in the cytoplasm between white and brown preadipocytes. In

this analysis, we only considered exonic reads for this analysis to avoid transcript abundance inflation of long genes.

Validation of scRNA-seq marker genes in recovering brown preadipocyte heterogeneity using smFISH

Fluorescent-activated cell sorting (FACS) has been instrumental in identifying lineage-specific preadipocyte marker genes in mice [434]. However, markers identified in mice are not comprehensively selective for humans [286, 308]. We therefore sought to define a set of white-specific and brown-specific marker genes as well as a set of genes specifically expressed in cluster 1 and cluster 2. Using the identified list of differentially expressed genes (Supplemental Table 1B and 1C), we implemented stringent cutoff criteria with logFC $>$ .8 in each cell-type, minimum detection of 60% and maximum detection of 40% in the other cell-type, for classifying genes with highly enriched and specific expression as marker genes. On this basis, we recognized NTNG1, RPL39L, PGF, LAMA4, BAALC, HIP1, and HAS2 as markers of white preadipocytes; LIMCH1, LYPD1, RGS4, ITGBL1, CDH13, and COL4A2 as markers of brown preadipocytes in the human neck depot (Fig. 4.1B and Supplemental Table 1B). We also identified KRT18, LUZP2, DLGAP1, SBSPON, MAP3K7CL, and NRXN3 as markers of brown cluster 1; CTSK, BST2, and MOXD1 as markers of brown cluster 2 (Fig. 4.1B and Supplemental Table 1C).

Outline of normalization strategy to correct for gene-length-based detection bias arising from including intronic reads

Rationale for Normalization

Multiple recent studies have demonstrated internal hybridization of polyT RT-primer to intronic polyA stretches as the primary mechanism for the capture and detection of intronic reads. Assuming that all intronic reads are derived from such hybridization incidences, number of observed intronic UMIs for any gene g in a given nuclei can be estimated by the following equation:

$$p_i \times pA_g \times N_g = x_g \tag{B.1}$$

where p_i is the probability to capture an intronic read, pA_g are the number of polyA stretches in gene g and N_g is the true transcript abundance. Assuming that p_i is independent of gene g

$$N_g \propto x_g / pA_g \tag{B.2}$$

Estimating pA_g

pA_g can be modeled by the following equation, where pd is the number of polyA stretches per kilobase of the genic region in the human genome, and gl_g is the total length of the gene in kilobase, including introns and exons:

$$pA_g = pd \times gl_g \tag{B.3}$$

Since the polyT tail in 10x Chromium RT primer is 30-bp long, we assumed hybridization to occur between the polyT tail and a polyA stretch, if the polyA sequence is at least 15-bp long (50% of the polyT tail). We queried the GRCh38 human genome to get positions of all polyA tracts at least 15-bp long, without mismatch, and screened for overlaps between such polyA tracts and gene coordinates for all genes in the cellranger GRCh38-2020A reference (which includes lncRNAs). As expected, total number of polyA tracts were highly correlated with gene length (Spearman R = 0.82, p-value < 0.05, Fig. B.7A) for each gene. We also calculated mean number of polyA tracts per Kbp for each gene, and estimated pd as the mean number of polyA tracts per Kbp across all genes, including zeroes (Fig. B.7B). Following this analysis, we estimated pd to be equal to 0.07.

We retrieved gene coordinates, strand, and gene length information using the GRCh38 gene annotation file downloaded from Gencode (Release 32). The same GTF was used for cellranger analysis. Briefly, each gene was first summarized by setting 3rd column in the GTF to gene, followed by calculation of gene length by subtracting the 5th and 4th columns.

Normalization Strategy

Based on the equation above, we present a normalization framework to reduce the technical bias arising from comparisons of nuclear and cellular data upon inclusion of intronic reads. This normalization strategy is implemented on the count matrix generated using only intronic reads, for both scRNA-seq and snRNA-seq datasets, and provides a modified UMI count-abundance, taking gene-length into account, for each cell and nuclei, based on the following equations:

$$\bar{x_g} = \frac{x_g}{gl_g \times pd} \tag{B.4}$$

where x_g is the original UMI-count for gene g in a given cell/nuclei, and $\bar{x_g}$ is the modified UMI count after normalization for gene g in the same cell/nuclei. This modified intronic UMI-count is then added to the observed exonic UMI-count for each gene g in a given cell/nuclei and finally library-normalized as following:

$$\bar{z_g} = \log\left(\frac{\bar{x_g} + y_g}{N_i + N_e} \times e^4 + 1\right) \tag{B.5}$$

$$N_i = \sum_{g \in G} \bar{x_g} \tag{B.6}$$

$$N_e = \sum_{g \in G} y_g \tag{B.7}$$

where y_g is the original UMI-count for gene $g$using exonic reads, and $\bar{z_g}$ is the final log-normalized count used for downstream differential expression testing between cells and nuclei.

Proliferating vs growth arrested cells in snRNA-seq and scRNA-seq white preadipocyte dataset

As highlighted in Fig. 4.7B UMAP visualization, both scRNA-seq and snRNA-seq white preadipocyte dataset (day-0) were cleaved into two halves. We investigated the differences between these two halves by manually annotating clusters as following (Fig. B.11A):
Cluster 0: day-20-differentiating-preadipocyte nuclei and cells
Cluster 1: day-20-adipocyte nuclei and cells
Cluster 2: top half of cleaved day-0-preadipocyte nuclei and cells
Cluster 3: bottom half of cleaved day-0-preadipocyte nuclei and cells
Cluster 4: day-20-cluster-2-nuclei
We normalized the data using NormalizeData command in Seurat and plotted expression profiles of proliferation and mitotic marker genes PLK1, MYBL2, BUB1, MKI67, CDK1, and CCNB1 (Fig. B.11B to B.11G). As expected, cluster 0 and 1 which primarily comprised of day-20 cells had no expression of proliferation markers, which is in line with their growth arrested behavior post adipogenic induction. However, prior to differentiation (day-0-preadipocytes), cells undergo cell cycle progression, thereby explaining the positive expression of proliferation marker genes in cluster 3. Cluster 2 cells were perhaps preadipocytes that underwent growth arrest due to contact inhibition during cell culture. Notably, even after 20 days of differentiation, a very small number of cells (cluster 4) were still highly proliferating, suggesting that these cells could be preadipocytes that never underwent growth arrest. We also calculated cell cycle phase scores based on canonical markers using the CellCycle Scoring pipeline in Seurat and assigned either G1, G2M, or S Phase to each of these cells. As expected, most of the cells in clusters 0, 1, and 2 were in G1 phase as opposed to G2M and S phase in clusters 3 and 4 (Fig. B.11H).

Hashing of white and brown preadipocytes using oligo-conjugated hashtag antibodies

Cell hashing enables pooling of all samples prior to loading them onto a single 10X chromium controller lane, thereby enabling combined library preparation and sequencing of all samples to eliminate potential "batch" artifacts. During single-cell suspension preparation of

white and brown preadipocytes for downstream scRNA-seq, we split each preadipocyte type into two individual microcentrifuge tubes for a total of four working samples. Brown preadipocytes were labelled with Hashtag-A0251 and A0252 antibodies, and white preadipocytes with A0253 and A0254 antibodies (Supplemental Table 1A). By sequencing these hashtag antibodies alongside the cellular transcriptome, we assigned each cell to its sample of origin, and identified doublets originating from multiple samples. Hashtag-antibody library was counted using the CITE-seq-Count workflow (10.5281/zenodo.2585469) and demultiplexed using the Seurat function 'MULTIseqDemux'. Demultiplexing pipeline identified 143 negative barcodes, 532 doublet barcodes, and 6918 cell-containing barcodes. As expected, every cell-containing barcode had highly positive and specific expression of only a single hashtag antibody, every doublet had marked expression for a combination of two antibodies, and negatives had very low expression for all antibodies (Fig. B.2A). While hashtag UMI counts revealed enrichment of protein targets in brown preadipocytes (Fig. B.2B, two-sided t-test), there was sufficient detection in both cell-types to enable robust cellular demultiplexing (Fig. B.2A).

Labeling protocol: For staining cells with hashtag antibodies, we followed supplier's protocol. Briefly, cells were harvested from a single 100 mm cell-culture dish and suspended in 100 µl of cell staining buffer in 2 ml low bind tubes. 5 µl of Human TruStain FcX Fc Blocking reagent was then added, and cells were incubated for 10 minutes at 4°C. 0.5 µg of a unique Cell Hashing antibody was added to each tube and cells were incubate for 30 minutes at 4°C. Cells were washed with 1 mL of cell staining buffer for 3 times by centrifuging at 1200 rpm for 4 minutes at 4°C. Finally, cells were suspended in PBS and 0.04% BSA at 1000 cells/µL for downstream 10x sequencing.

Appendix C

Supplementary Information related to Chapter 5

C.1 Supplementary Figures

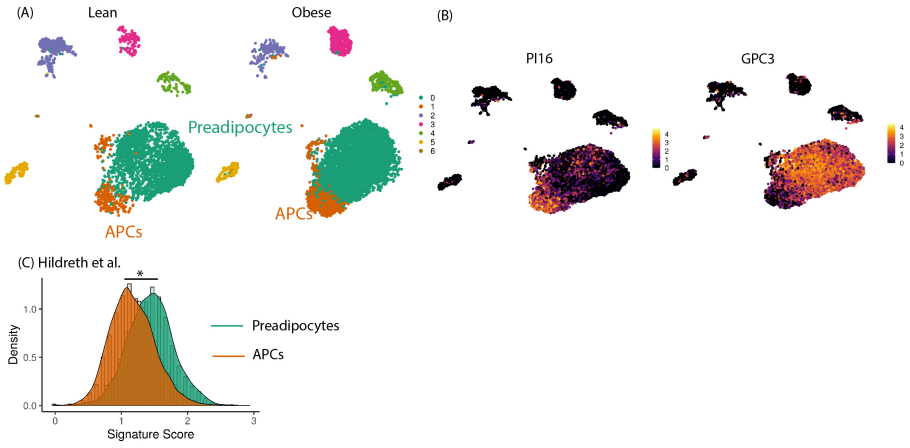

Figure C.1: **Application of adipogenic gene signatures in public scRNA-seq datasets** (A) UMAP of scRNA-seq dataset used in Hildreth et al. [430] study. (B) PI16 marks APCs and GPC3 marks preadipocytes. (C) Distribution of cell-maturity score between APCs and Preadipocytes in Hildreth et al. dataset

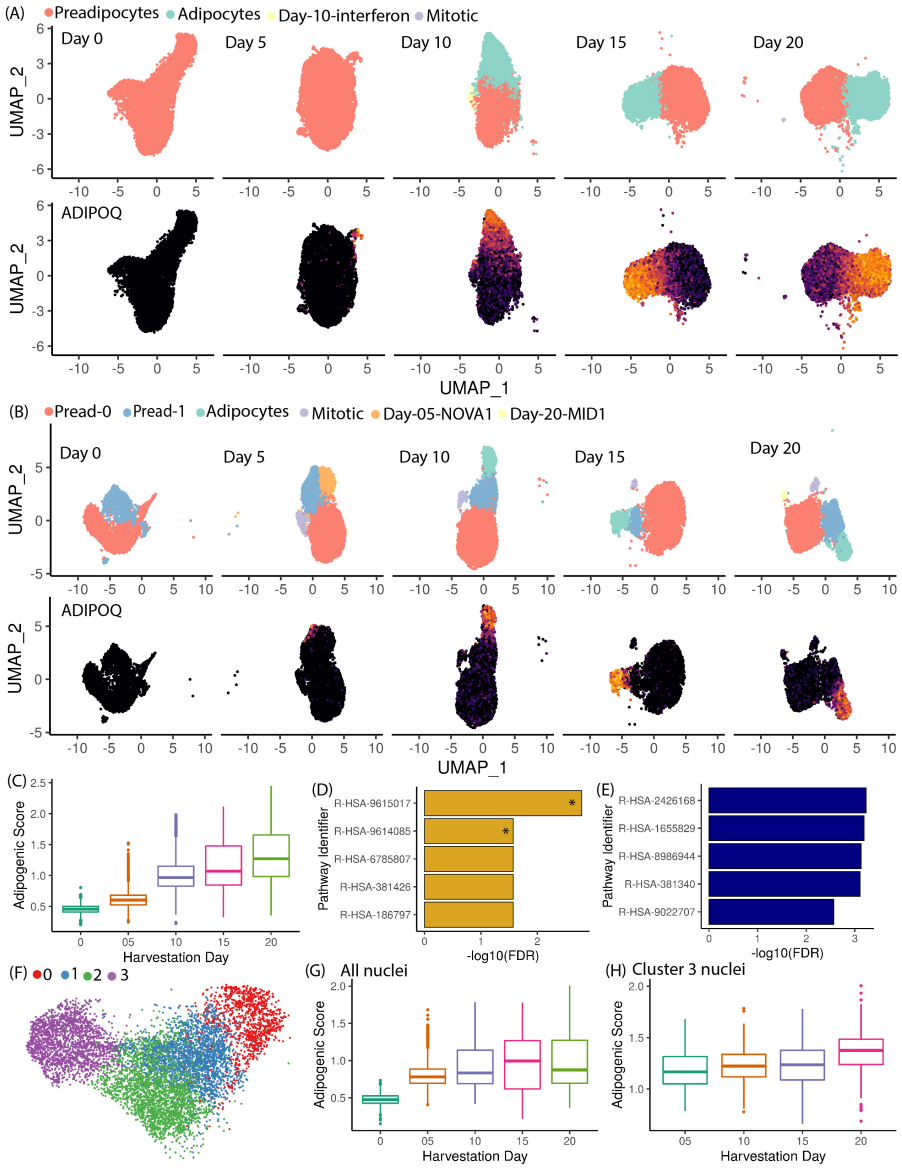

Figure C.2: (A) UMAP visualisation of differentiating white preadipocyte dataset from each day of harvestation colored using unsupervised clustering (top) or ADIPOQ expression (bottom). (B) Same plot as (A) but for differentiating brown preadipocyte dataset. (C) Adipogenic score for nuclei harvested from 5 time-points from differentiating white preadipocyte. (D) and (E) Top enriched pathways in non-adipogenic brown trajectory (D) and adipogenic brown trajectory (E). (F) Joint unsupervised clustering of all cells in the adipogenic brown trajectory (G) Adipogenic score for nuclei harvested from 5 time-points from differentiating brown preadipocyte. (H) Adipogenic score for nuclei harvested from 5 time-points from differentiating white preadipocyte. Only adipocytes were considered for this plot (cluster 3 in panel F).

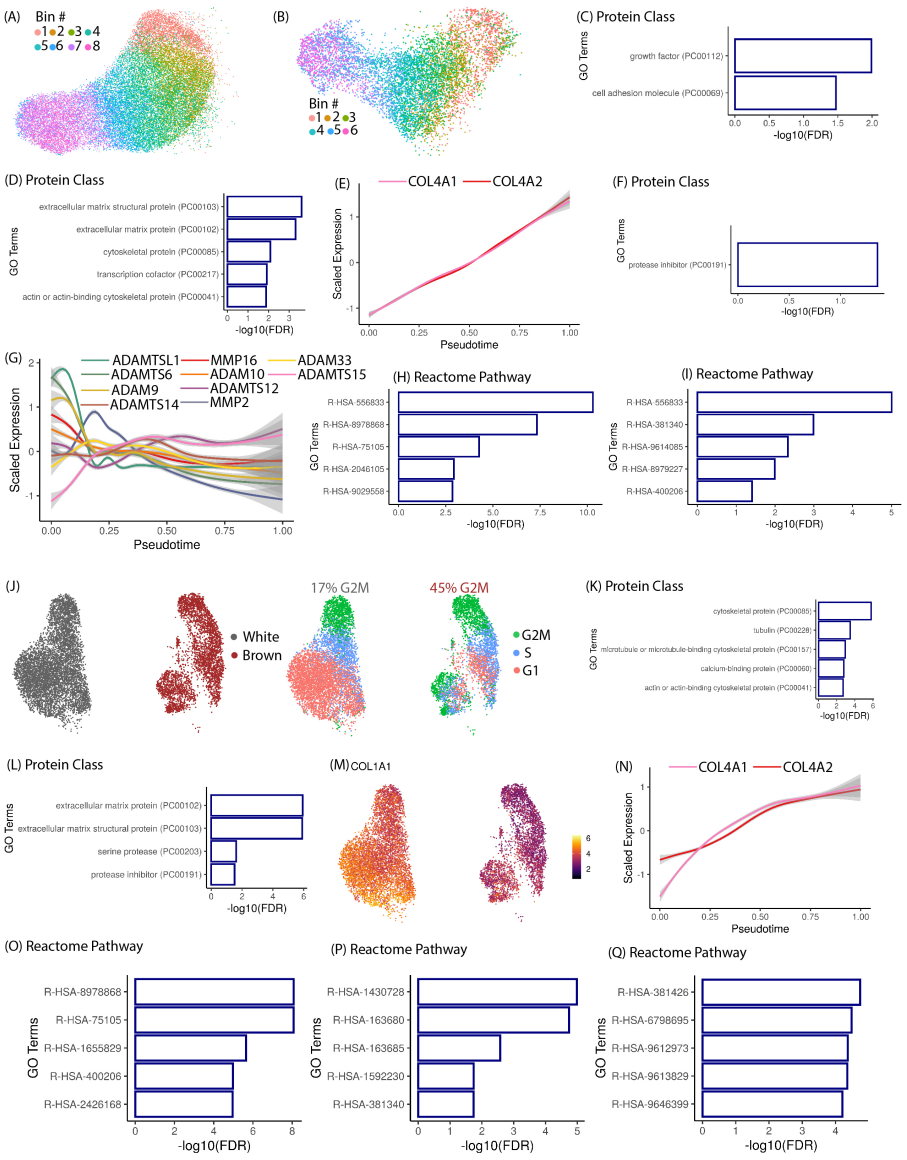

Figure C.3: **Pseudo-temporal ordering of white and brown preadipocytes** (A) and (B) Pseudo-temporal bins for white and brown dataset respectively. (C) GO terms for genes enriched in Group 1 in white dataset. (D) GO terms for genes enriched in Group 2 in white dataset. (E) Expression dynamics of ECM components in white dataset. (F) GO terms for genes enriched in Group 3 in white dataset (G) Expression dynamics of metalloproteases in white dataset. (H) GO terms for genes enriched in Group 4 in white dataset (I) GO terms for genes enriched in Group 5 in white dataset. (J) UMAP of white and brown nuclei from day 0, colored by lineage (left) and cell-cycle phase (right). (K) GO terms for genes enriched in Group 2 in brown dataset. (L) GO terms for genes enriched in Group 3 in brown dataset. (M) Expression of COL1A1 in white and brown nuclei harvested from day 0. (N) Expression dynamics of ECM components in brown dataset. (O) GO terms for genes enriched in Group 4 in brown dataset. (P) GO terms for genes enriched in Group 5 in brown dataset. (Q) GO terms enriched in genes exclusively regulated in brown dataset.

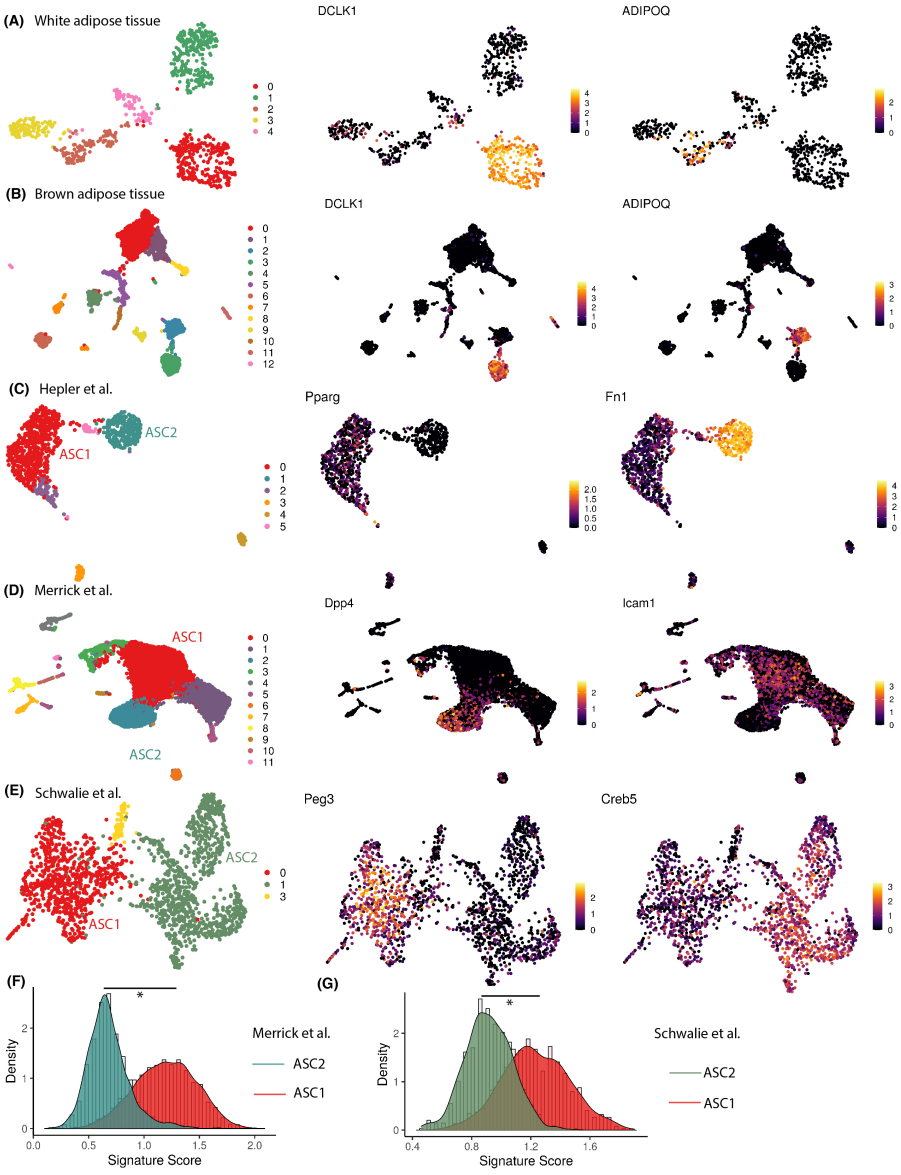

Figure C.4: **Application of adipogenic gene signatures in public scRNA-seq datasets** (A) and (B) UMAP of WAT (A) and BAT (B) datasets used in Sun et al. [273] study. DCLK1 marks preadipocyte population and ADIPOQ marks mature adipocyte population. (C) to (E) UMAP of WAT SVF used in Hepler et al. [347] (C), Merrick et al. [348] (D), and Schwalie et al. [349] (E) along with marker gene expression for ASC2 and ASC1 cell-types (F) Distribution of cell-maturity score between ASC2 and ASC1 cell-types in Merrick et al. and Schwalie et al. datasets.

C.2 Supplementary Notes

Pseudotemporal ordering of differentiating white and brown preadipocytes into mature adipocytes

After integration of snRNA-seq datasets from distinct time-points, a coarse cellular ordering was obtained for both white and brown fat development, with day-0 nuclei on one end, and day-20 nuclei on the other. In order to further refine this cellular-ordering for a higher resolution, we utilized the idea of pseudo-temporal analysis. Typically, pseudo-temporal analysis is performed using advanced bioinformatic algorithms that require prior information on the starting cell/cluster for ordering. Examples include Slingshot, which was identified as one of the most robust tools for pseudotemporal ordering by recent benchmarking investigations. However, Slingshot does not take into account prior biological information for ordering cells. Hence, multiple researchers have rather relied on expression values of biologically relevant genes (or highly variable genes) as proxy for cell-ordering. For adipogenesis, a list of such biologically relevant genes can be found as a molecular signature in the MSigDB (HallmarkAdipogenesis), and combined expression of these genes (or a signature "score") could be used as a proxy to order cells. Although, most genes which are part of the HallmarkAdipogenesis signature are identified via transcriptomic enrichment analyses in mature adipocytes (terminal state), as compared to preadipocytes (initial state). Consequently, expression of such genes loses resolution for ordering cells that are in middle stages of adipogenesis. Therefore, for ordering nuclei in our dataset, we developed a strategy which utilizes expression of genes that are monotonically increasing in expression with cellular differentiation, thereby providing a high dynamic range as well as pseudo-temporal resolution. Such monotonically increasing genes were identified via a consensus of Slingshot and HallmarkAdipogenesis ordering.

Specifically, differentiating white and brown nuclei were first ordered using both Slingshot, and HallmarkAdipogenesis signature score (calculated using Vision). Then, dynamically regulated genes were identified for each ordering (see Methods) and clustered based on their expression profiles. Genes that were monotonically increasing in both ordering strategies were then defined as a custom signature. Different signatures were defined specific to white and brown adipogenesis. Finally, each nuclei was assigned a score based on the expression of genes constituting these custom-defined signatures (using Vision), and this score was used as a proxy for pseudotime. Gene signatures can be found below.

White Signature: ADM APOE FABP5 MDH1 DCXR AOC3 PLIN2 MME BTG1 CD36 PDK4 FKBP5 SOX5 BCL6 ZBED3 SIK2 COL4A1 COL4A2 RPLP2 LMO4 AL845331.2 BMS1P14 SORT1 ACER3 HK2 ITGA7 SLC7A6 FZD4 TMEM135 PLA2G16 KCNIP2-AS1 KCNIP2 TNS1 DECR1 PPARGC1A PSMA1 G0S2 PNPLA2 FABP4 MLXIPL AQP7 ACADVL PALMD CYB5A PDE3B SAT1 ACACB CSAD MALAT1 DOCK11 PPARG FOXO1 CALCRL CHCHD10 ACO2 TOB2

Brown Signature: CYB5A COL4A1 CIDEC CD36 AC002066.1 CAV2 TMEM164 ACSL5 GIPR LBP NAMPT PDK4 ACSL4 FADS1 ACACA ME1 GHR DDIT4 LPCAT3

SREBF1 NR1H3 ABCA1 SOX5 NRCAM PDE3B PSMA1 BCL6 DECR1 SIK2 PLIN4 PALMD ACSL1 SAT1 TMEM135 SCD FABP4 PNPLA2 ACACB CSAD PPARG ELOVL5 TLE1 ACER3 UVRAG LINC01239 PTK2B MAPK10 SOS1 RHOBTB3 FKBP5 ZBTB16 EBF1 GBE1 AKR1C2