# Multi-Dimensional Genomic Characterization of Adipose Tissue Across Depot, Disease and Development

Jinchu Vijay

Department of Human Genetics

Faculty of Medicine

McGill University, Montreal.

September 2021

A thesis submitted to McGill University in partial fulfillment of the requirements of degree of

Doctor of Philosophy

# Table of Contents

## Abstract

Obesity and associated metabolic complications continue to be a serious health concern around the globe. Among the different organ systems involved in the development of obesity, adipose tissue across distinct locations of the human body is known to play a key role in energy homeostasis. Despite the common function of fat storage, adipose tissue in different anatomical locations known as depots is associated with varied disease status. In addition, adipose tissue (AT) is also made up of several cell types including progenitors, immune cells, endothelial cells and smooth muscle cells. We used single-cell sequencing technologies to characterize the complex cell populations in AT and their role in metabolic disease risk. First, we investigated the association of cell populations in subcutaneous (SAT) and intra-abdominal visceral adipose tissue (VAT) from extreme obese adults with or without type 2 diabetes (T2D) using single-cell RNA sequencing (scRNASeq). We were able to profile different progenitors, immune cells and endothelial cells present in the two depots as well as contrasting their association with T2D with the help of gene expression patterns. We then further expanded these efforts to pediatric tissue characterization across lean and obese children as well as by including single nuclei open chromatin ATAC-Seq data (snATAC-Seq) combined with scRNA-Seq. We were able to confirm and validate our findings from adult study and explored dynamics of these cell population in response to obesity. Adipose resident immune cells showed age, tissue and disease dependency with a low proportion of adipose-infiltrating immune cells in young individuals compared to adults. In conclusion, we have shown that high-throughput single-cell approaches applied to deep phenotyped cohorts and carefully prepared tissue samples allow the identification of dysfunctional cellular and molecular machinery underpinning complex diseases such as obesity and type 2 diabetes.

## Résumé

L'obésité et les complications métaboliques qui y sont associées continuent d'être un grave problème de santé à travers le monde. Parmi les différents systèmes organiques impliqués dans le développement de l'obésité, le tissu adipeux (TA) distribué à différents endroits du corps humain est connu pour jouer un rôle clé dans l'homéostasie énergétique. En plus de sa fonction commune de stockage des graisses, ce tissu, aussi connus sous le nom de dépôt dans différents emplacements anatomiques, est associé à des états pathologiques variés. Le TA est composé de plusieurs types de cellules, notamment des progéniteurs, des cellules immunitaires, des cellules endothéliales et des cellules musculaires lisses. Nous avons utilisé des technologies de séquençage unicellulaire pour caractériser les populations cellulaires complexes dans le TA et leur rôle dans le risque de maladie métabolique. Tout d'abord, nous avons étudié l'association des populations cellulaires dans le tissu adipeux viscéral, sous-cutané et intra-abdominal d'adultes obèses extrêmes avec ou sans diabète de type 2 (DT2) en utilisant le séquençage d'ARN unicellulaire (scRNASeq). Nous avons pu profiler différents progéniteurs, cellules immunitaires et cellules endothéliales présentent dans les deux dépôts ainsi que contraster leur association avec le DT2 à l'aide de modèles d'expression génique. Nous avons ensuite étendu ces efforts à la caractérisation des tissus pédiatriques chez les enfants maigres et obèses, et avons ajouté des données de profilage unicellulaire de la chromatine ouverte (snATAC-Seq) combinées avec scRNA-Seq. Nous avons pu confirmer et valider nos résultats antérieurs chez l'adulte et exploré la dynamique de ces populations de cellules en réponse à l'obésité. L'analyse des cellules immunitaires a démontré une association à l'âge, aux tissus et à la maladie avec une faible proportion de cellules immunitaires infiltrant les tissus adipeux chez les jeunes par rapport aux adultes. En conclusion, nous avons

montré que les approches unicellulaires à haut débit appliquées à des cohortes phénotypées profondes et à des échantillons de tissus soigneusement préparés permettent l'identification de la machinerie cellulaire et moléculaire dysfonctionnelle à la base de maladies complexes telles que l'obésité et le diabète de type 2.

# List of Abbreviations

| | |
|---|---|
| AT | Adipose Tissue |
| ATAC-Seq | Assay for Transposase Accessible Chromatin sequencing |
| ATM | Adipose Tissue Macrophages |
| BAT | Brown Adipose Tissue |
| BMI | Body Mass Index |
| C/EBPs | CCAT-Enhancer Binding Proteins |
| CCA | Canonical Correlation Analysis |
| ChIP-Seq | Chromatin Immunoprecipitation assay with Sequencing |
| CIHR | Canadian Institutes of Health Research |
| CNS | Central Nervous System |
| CVD | Cardiovascular Diseases |
| DAVID | Database for Annotation, Visualization and Integrated Discovery |
| DNA | Deoxyribonucleic Acid |
| ENCODE | Encyclopedia of DNA Elements |
| eQTL | Expression Quantitative Trait Locus |
| FACS | Fluorescent Activated Cell Sorting |
| FRCBS | Finnish Red Cross Blood Service |
| GEM | Gel in Emulsion |
| GEO | Gene Expression Omnibus |
| GIANT | Genetic Investigation of ANthropometric Traits consortium |
| GTEx | Genotype-Tissue Expression Consortium |

| | |
|---|---|
| GWAS | Genome Wide Association Study |
| HCA | Human Cell Atlas |
| HSC | Hematopoietic Stem Cells |
| HuBMAP | Human Biomolecular Atlas Program |
| IUCPQ | Institut universitaire de cardiologie et de pneumologie de Québec |
| KRH | Krebs-Ringer-Henseleit |
| LPL | Lipoprotein Lipase |
| LYVE-1 | Lymphatic Vessel Endothelial Hyaluronan Receptor 1 |
| MARS-Seq | Massively Parallel RNA Single Cell Sequencing |
| MCC-Seq | MethylC-Capture Sequencing |
| METSIM | Metabolic Syndrome in Men |
| mQTL | Methylation Quantitative Trait Locus |
| MSC | Mesenchymal Stem Cells |
| MuTHER | Multiple Tissue Human Expression Resource |
| NIHR | National Institute for Health Research |
| NK cells | Natural Killer Cells |
| PBMC | Peripheral Blood Mononuclear Cell |
| PCA | Principal Component Analysis |
| PCR | Polymerase Chain Reaction |
| PGM | Pre-Adipocyte Growth Medium |
| PPARG | Peroxisome proliferator activated receptor gamma |
| QC | Quality Control |
| RIN | RNA Integrity Number |

| | |
|---|---|
| RNA | Ribonucleic Acid |
| SAT | Subcutaneous Adipose Tissue |
| scRNA-Seq | Single Cell RNA sequencing |
| SMART-Seq | Switch Mechanism at the 5′ End of RNA Templates |
| snATAC-Seq | Single Nuclei ATAC sequencing |
| SNPs | Single Nucleotide Polymorphisms |
| SVF | Stromal Vascular Fraction |
| t-SNE | t-Distributed Stochastic Neighbor Embedding |
| T2D | Type 2 Diabetes |
| TFBM | Transcription Factor Binding Motif |
| TF-IDF | Term Frequency-Inverse Document Frequency |
| TNF-a | Tumor Necrosis Factor Alpha |
| Treg | Regulatory T cells |
| TSS | Transcriptional Start Site |
| TZD | Thiazolidinediones |
| UMAP | Uniform Manifold Approximation and Projection |
| UMI | Unique Molecular Identifier |
| VAT | Visceral Adipose Tissue |
| WAT | White Adipose Tissue |
| WHO | World Health Organization |
| WHR | Waist to Hip circumference Ratio |
| WT1 | Wilms tumor gene |

List of Figures

**CHAPTER2**: **Cataloguing Human Adipose Tissue Across Depot and Disease at Single Cell Resolution**

Main Figures



Supplementary Figures

**CHAPTER3**: **Adipose Tissue Transcriptome and Epigenome Characterization During Development and in Health and Disease**

Main Figures

List of Tables

**CHAPTER2**: **Single-cell analysis of human adipose tissue identifies depot- and disease-specific cell types**

Main Tables

**CHAPTER3**: **Adipose Tissue Transcriptome and Epigenome Characterization During Development and in Health and Disease**

Main Tables



Supplementary Tables

Acknowledgments:

I would like to thank and acknowledge everyone who has helped me throughout my research work to bring this thesis to a complete picture. I thank my supervisor Dr. Elin Grundberg for guiding and supporting me the last 6 years. I thank you for believing in me and giving me the opportunity to work on an amazing project. I am grateful for the freedom and support she has provided and being patient with me in providing time and support. I thank you for mentoring me to understand how to approach a research problem, for the career guidance and all your support which helped me to bring my PhD to a finishing point. Next, I would like to thank my co-supervisor Dr. Guillaume Bourque for accepting me as his student and taking time to guide me in time of necessity. I also thank you for all your help with my thesis and my career explorations. I thank my supervisory committee members Drs. Hamed S. Najafabadi and Claudia Kleinman for always providing constructive feedbacks on my projects. I would also like to extend my thanks to Dr. Tomi Pastinen for taking time to reviewing my manuscripts and helping me to structure them more effectively. I extend my special appreciation to all my team members Albena Pramatarova, Warren Cheung, Xiaojian Shao, Marie-Michelle Simon, Elodie Boulier, Tony Kwan, Bing Ge and Fiona Allum for all the support and creating a friendly work environment. I also thank all our collaborators, a special note to Dr. André Tchernof who played a pivotal role in bringing my projects to completion by sharing their resources and expertise. I am also thankful to the Department of Human Genetics for the differential fee waiver, the travel awards, and the Excellence award. Finally, I would like to thank my parents, Anchu, Vijay, Isha and Megha for their continuous support and understanding throughout my PhD journey.

# Preface

**Contribution to Original Knowledge**

This doctoral thesis is aimed at characterizing adipose tissue cell population at single cell level and to ascertain their significance in development of metabolic disorders. To attain this goal, we studied adipose tissue from adults and children and the detailed description of our attempts are included as two research chapters in this thesis. We leveraged transcriptome and open chromatin regions of different cell population to study them in detail.

In the first project described in Chapter2, we characterized the adipose tissue from two different depots of human body- SAT and intra-abdominal VAT. Here we focused only on extremely obese adults discordant for T2D phenotype. We were able to identify two different population of progenitors in VAT which showed difference in gene expression signatures related to the disease status. We hypothesized that accumulation of one of the progenitor types marked by the expression of mesothelin (*MSLN*) is protective against metabolic disorders. This idea is further strengthened by presence of signals like fibrosis and expression of disease associated gene expression in the progenitor subset expressing adipsin (*CFD*). We were also able to profile many immune cell subtypes residing in adipose tissue in a depot and T2D-specific manner. For instance, we confirmed the notion that macrophages are residing to a larger extent in VAT compared to SAT including cells at distinct stages of polarization – a pattern that was further enhanced when the analysis was restricted to cells derived from T2D subjects. We also identified a subset of T cells enriched with metallothionein genes showing T cell exhaustion in obesity. Finally, we identified

the presence of multiple endothelial cell types including adipose tissue specific microvascular endothelial cells expressing *FABP4* and *CD36.*

In the second project, we included VAT samples from lean and obese children expanding our interrogating by including snATAC-Seq along with scRNA-Seq. These data confirmed the distinct relationships of dysfunctional and protective adipocyte progenitors with obesity identified in adults also in young individuals: *CFD*-expressing progenitor cells were enriched in children with obesity (34%) compared to healthy individuals (15%). *MSLN*-expressing cells were found to be present in healthy tissue from lean children (41%) compared to obese children (9%) and possess the signature of a brown-like progenitor population. Further analysis also showed snATAC-Seq peaks defining *MSLN* clusters being enriched for binding sites for TEAD-family of transcription factors in comparison with *CFD* expressing clusters. In addition, VAT resident immune cells showed age, tissue and disease dependency with a low proportion of adipose-infiltrating immune cells in young individuals compared to adults.

**Format of the Thesis**

This thesis follows manuscript-based thesis style with five chapters. The first chapter covers a brief background of adipose tissue biology and single cell technologies that are relevant to the research interest of this thesis. Followed by the background information, chapter 1 also includes the rationale, hypothesis, and specific aims of my research. Chapter 2 and 3 covers the original research projects of this doctoral thesis. Chapter 2 is published in Nature Metabolism and draft manuscript is included in chapter 3. General discussions on the results of the research chapters are incorporated in chapter 4. Chapter 5 comprise of conclusions and future directions which we ascertained from the research projects in chapter 2 and 3. Lastly, the appendix includes the list of other publications which I contributed during my PhD and additional supplementary tables and figures from chapter 2 and 3.

**Contribution of Authors**

The chapters described in this thesis was completed under the supervision of Dr. Elin Grundberg. Chapter 2 was published in Nature Metabolism in December 2019. The manuscript was authored by Jinchu Vijay, Marie-Frédérique Gauthier, Rebecca L Biswell, Daniel A Louiselle, Jeffrey J Johnston, Warren A Cheung, Bradley Belden, Albena Pramatarova, Laurent Biertho, Margaret Gibson, Marie-Michelle Simon, Haig Djambazian, The Multiple Tissue Human Expression Resource Consortium, Alfredo Staffa, Guillaume Bourque, Anita Laitinen, Johanna Nystedt, Marie-Claude Vohl, Jason D Fraser, Tomi Pastinen, André Tchernof and Elin Grundberg. E.G. conceived the study. E.G., A.T., T.P. and M.C.V. designed experiments. A.T., J.F. and M.F.G prepared and/or provided the clinical samples. L.B. and B.B. managed clinical aspects of the study. A.P., R.B., M.G., D.A.L. and M.M.S performed the sequencing experiments. W.A.C, J.J., H. D., A. S. and G.B. provided bioinformatics support. A.L and J.N. provided MSC cell lines and culture protocols. J.V. analyzed the data, interpreted the results and drafted the manuscript with input from E.G. All authors reviewed and contributed feedback on the final manuscript.

Chapter 3 was authored by Jinchu Vijay, Warren A Cheung, Byunggil Yoo, Guillaume Bourque, Jason D Fraser, Tomi Pastinen, Elin Grundberg. E.G conceived the study. E.G and T.P designed experiments. J.D.F provided the clinical samples. W.A.C, B. Y and G.B provided bioinformatics support. J.V. analyzed the data, interpreted the results and drafted the manuscript with input from E.G.

# Chapter 1: Introduction

Obesity and associated metabolic disorders continue to be a serious health concern around the globe. Obesity can be broadly explained as an imbalance between calorie intake and expenditure by the body. As per the estimates of World Health Organization (WHO) in 2016, 1.9 billion adults were found to be overweight, and 650 million adults were found to be obese worldwide[1]. Recent WHO reports also show that 5.6% of children under the age of 5 were identified to be overweight worldwide. Comparing the estimates from the year 2000 (4.9%), child obesity has a positive trend in line with global obesity trends seen in previous studies[2,3]. Obesity in turns leads to cardiac stress, alteration in pulmonary function, degeneration of cartilages, impairment of immune function and insulin resistance[4]. It is often measured using the ratio of weight-for-height of a person, termed as body mass index (BMI) and an adult with a BMI above 30 $kg/m^2$ is considered as obese[1].

Metabolic syndrome is a collective term given to the disorders that arise because of obesity. As per WHO, glucose intolerance is considered as the key component of metabolic syndrome. However, glucose intolerance is categorized as metabolic syndrome only if it is paired with two other comorbidities which include obesity, dyslipidemia, hypertension and microalbuminuria[5]. Though metabolic syndrome is prevalent among adults, considerable proportion of obese children also develops the disorder. It also increases the risk of development of cardiovascular diseases in later stages[3]. Framingham risk score is a widely used method to assess the 10-year risk for development of cardiovascular disease in the presence of metabolic disorders[6]. Though BMI gives an overall measure of obesity, fat accumulation at different regions of human body termed as depots plays a critical role in contributing risks to the development of metabolic syndromes.

## 1.1 Adipose Tissue Biology

### 1.1.1 Heterogenous Distribution of Fat in Human Body

In 1947, Vague[7] reported body fat accumulation shows two different patterns which he described as android and gynoid obesity. Android obesity is characterized by upper abdominal fat accumulation, prevalent in males and often associated with metabolic disorders, whereas gynoid obesity is marked by fat in peripheral regions and common in females. Further, Waist to Hip circumference Ratio (WHR) was found to be associated with development of cardio metabolic risks in many following studies[8]. The central obesity or android obesity is contributed by two main depots of fat tissue - subcutaneous adipose tissue (SAT) and intraabdominal visceral adipose tissue (VAT). Visceral adipose tissue also shows anatomical, cellular and molecular difference from subcutaneous adipose tissue. For instance, VAT is marked by small adipocytes, increased concentrations of glucocorticoid and androgen receptors, more vascularity and varied adipokine profile[9]. VAT is also drained by portal veins providing direct link to liver which is a vital organ associated with development of obesity. Added to the fact that the increase in WHR was driven by an increase in VAT volume, it was identified as the fat depot that is associated with development of cardio metabolic diseases[10]. Based on the occurrence of metabolic syndromes in different BMI range, patients are usually classified into two groups. Patients in normal weight range but with metabolic syndrome were usually referred as metabolically obese and obese individuals with normal metabolic status were referred as metabolically healthy obese. The VAT volume was identified to be factor regulating the disease status. The etiology of VAT disproportion can be pointed to a variety of factors including age, sex, genetics, ethnicity and hormones[11].

Methodological difficulties to access the tissue deep in the body cavity makes studying VAT challenging.

### 1.1.2 Sex Bias in Fat Accumulation

Adiposity in men is mostly contributed by VAT characterized by fat accumulation in the upper body. On the contrary, subcutaneous adipose tissue accumulated in lower body tend to contribute to obesity in women. After menopause, a shift in affinity towards adipose accumulation was seen in women resulting in more visceral adipose volume. Further, sex hormones were known to play a critical role in obesity pathology. Lipoprotein lipase (*LPL*) secreted by adipocytes plays a significant role in hydrolyzing triglycerides in lipoproteins and releasing free fatty acids. *LPL* was found to be expressed in a sex specific pattern. It is primarily expressed in subcutaneous adipose tissue in women and visceral adipose tissue in men[12]. In addition, circulating testosterone levels were found to be inversely correlated with *LPL* activity[13]. On the other hand, estrogen was found be protective against accumulating visceral adipose volume[14]. Development of insulin resistance also shows more prevalence in males than females. Studies show estrogen play a key role in glucose homeostasis in women and thereby preventing diabetes[15].

### 1.1.3 Adipose Tissue Expansion

Expansion of adipose tissue, as seen in obesity, can be due to increase in size of adipocytes known as hypertrophy or due to differentiation of new adipocytes known as hyperplasia[16,17]. Adipocytes in adults were found to be static and main mechanism of adipose expansion was hypertrophy[16]. Hypertrophy alters normal functioning of adipose tissue by several ways. It includes increased

inflammation, altered adipokine secretion and promoting hypoxia which in turn leads to the development of metabolic syndromes[18].

In hyperplasia, adipocytes were known to be differentiated from mesenchymal stem cells (MSC), but the complete mechanism is unknown. MSCs are present in different regions of the human body, including adipose tissue. Multipotent MSCs can differentiate into adipocytes, osteocytes, myocytes, and chondrocytes. Adipogenesis starts with commitment of MSCs to proliferative fibroblasts like pre adipocytes. Pre adipocytes remains in a state of growth arrested for extended periods due to contact inhibition from the neighboring cells. In the presence of a stimulus, they undergo mitotic clonal expansion and terminal differentiation to mature lipid laden adipocytes[19]. A cascade of transcription factors is known to mediate adipogenesis. Among them, Peroxisome proliferator- activated receptor γ (PPARγ) and CCAT-enhancer binding proteins (C/EBPs) are studied extensively[20]. However, new transcription factors, genes and regulatory mechanisms involved in adipogenesis are still actively being discovered.[21,22]

Adding another layer of complexity, the embryonic origins of different adipose depots are yet to be understood. Chau et. al. identified subcutaneous and visceral adipose tissue have different origins[23]. Transcriptome study of subcutaneous and visceral adipose tissue showed the expression of Wilms tumor gene, *WT1* exclusively in visceral adipose tissue. Further, lineage tracing studies showed visceral adipose tissue has a group of progenitors that may have a development origin from mesothelium. Mesothelium forms a lining of internal organs such as heart, lungs, and intestines. Interestingly, a layer of mesothelium was also found covering the visceral adipose depot. The authors hypothesized that a some of the visceral adipocyte progenitors were contributed

by the mesothelial layer. This was further confirmed in several ways, including the observation of mesothelial marker *UPK3B* in visceral adipose tissue[24].

### 1.1.4 Thermogenic Adipose Tissue

Energy storing adipose tissue found in different depots of human body are termed as white adipose tissue (WAT). Thermogenic adipose tissue referred to as brown adipose tissue (BAT) have gained attention of researchers for decades due its capacity to breakdown available substrates including glucose and free fatty acids into heat[25]. BAT was observed in hibernating animals as a protective mechanism against cold temperatures through non-shivering thermogenesis[26]. This is facilitated by oxidative phosphorylation using Uncoupling protein 1 (*UCP1*)[27]. As a result, BAT is rich in mitochondria compared to WAT. BAT was believed to be absent in adult humans. However, recent studies show presence of BAT in adults and their abundance is inversely correlated with BMI[28] and insulin resistance[25].

Thermogenic adipose tissue can also be developed from WAT when there is a stimulus like cold or β3- adrenergic agonists. These cell types are termed beige or brite adipocytes. There have also been evidence suggesting development of beige fat from precursor cells. Beige fat has the capacity to transform back to WAT or induced again based on the presence of stimulus[29]. Brown adipocytes are known to develop from precursor cells that show equivalent properties as skeletal muscle precursors. On the other hand, beige adipocyte precursors show similarity with white adipocyte precursors. Irrespective of the precursors, *PRDM16* is the transcription factor that is known to play major role in development of both brown and beige adipocytes[30].

### 1.1.5 Infiltration of Immune Cells

Adipose tissue is made up of adipocytes, progenitor cells, endothelial cells, smooth muscle cells and immune cells. The non-adipocyte cell groups in the adipose tissue are known as stromal vascular fraction (SVF). Immune population in adipose tissue include macrophages, T cells, B cells, neutrophils, eosinophils, and mast cells. Adipose tissue also secretes several pro-inflammatory and anti-inflammatory cytokines and chemokines. Increase in macrophage population has been associated with increase in BMI and obesity[31]. Immune cells are known to form crown like structures called milky spots around dying adipocytes weeks after high fat diet.[32]

Development of insulin resistance is directly associated with the level of inflammation of adipose tissue and is marked by Tumor Necrosis Factor (TNF-α). Macrophages are identified as the main source of TNF-α and an increase in macrophage population is observed with development of insulin resistance[31]. Further, macrophages are also found to polarize to proinflammatory M1 phenotype from anti-inflammatory M2 cells[33]. Apart from macrophages, Regulatory T cells (Treg) and B cells are known to play a leading role in adipose inflammation. For instance, Treg population was significantly reduced with onset of obesity as well as linked to insulin sensitivity through the secretion of peroxisome proliferator activated receptor γ (*PPARG*)[34]. B cells promotes inflammation and insulin resistance. However, the exact mechanism is not known.

## 1.2 Genome-wide Studies

### 1.2.1 Genetic, Epigenetic and Transcriptomic Studies

Obesity is a complex trait meaning the disease is caused by both genetic and environmental factors. A recent study showed that the heritability of BMI is around 30 – 40%[35]. Genome wide association

studies (GWAS) have identified numerous genetic regions (i.e., loci) associated with phenotypes linked to obesity and metabolic disorders. For instance, Locke et al. identified 97 loci associated with BMI using 339,224 individuals from Genetic Investigation of ANthropometric Traits consortium (GIANT) meta-analysis. However, most of the loci were found to be involved in Central Nervous System (CNS) functions further suggesting involvement of the brain in the development of obesity[36]. On the other hand, loci that were found to be associated with fat distribution (e.g., waist to hip ratio) were more likely to be linked to adipose tissue function, confirming differential fat accumulation in depots were driven by adipose tissue[37]. The widely used system for annotating GWAS loci is to link them to nearby genes to predict function. However, comprehensive, integrative studies are often required to characterize the Single Nucleotide Polymorphisms (SNPs) as explained in the classical example of *FTO* gene association with obesity. Chromosome conformation studies showed different SNPs in vicinity of *FTO* disrupts the functions of distant genes including *IRX3, IRX5, FTO* and *RPGR1P1L* rather than associating the pathogenesis to *FTO*[38]. Another example is the GWAS loci associated with leptin-melanocortin pathway. Results from studies on rare and severe early-onset obesity (so called monogenic from of obesity) have identified SNPs associated with multiple genes each with a moderate risk score but collectively affecting leptin – melanocortin pathway which is involved in energy intake[39,40].

Environmental factors underlying complex disease risk such as obesity can be associated with epigenetic changes. DNA methylation patterns have been studied extensively to identify epigenetic links to obesity and metabolic traits. Mendelson et. al. studied methylation patterns associated with BMI and cardiometabolic diseases using Framingham heart study cohort and Lothian birth control

cohorts. 83 methylation sites associated with BMI was found to be replicated in other cohorts confirming that the loci are highly conserved for disease association. Using tissue specific expression, they also characterized a novel locus near *SREBF1* to associated with BMI and coronary artery disease[41]. As DNA methylation pattern is known to be tissue specific and inform about regulatory regions such as promoters and enhancers throughout the genome, our team recently implemented a capture approach called MethylC-Capture Sequencing (MCC-Seq) to study the methylome in specific tissue and disease context using user defined regions. MCC-Seq is a versatile tool providing cost effective solution as well as defining target genome regions specific to particular tissue.[42] Apart from methylation, identification of open chromatin regions using ATAC-Seq (assay of transposase-accessible chromatin with sequencing), regions of post translational histone marks using ChIP-Seq (chromatin immunoprecipitation assay with sequencing) and chromosome conformation using Hi-C also gained interest.

### 1.2.2 Large Cohort Reference Studies

Numerous population-based tissue resources have been established to improve our insight into the etiology of metabolic diseases. I would like to highlight three such studies – MuTHER, METSIM and GTEx which we used for validation of the results described in chapter 2 and chapter 3.

Multiple Tissue Human Expression Resource (MuTHER) is an initiative to understand genetic and non-genetic factors associated with epigenetic and gene expression variation and subsequent links to metabolic disease risk using data collected from ~800 female twins. Tissue samples included lymphocytes, subcutaneous adipose tissue, muscle, and skin. Joint analysis of methylation, gene expression and genotype information from monozygotic and dizygotic twins helped to characterize

functional CpGs in the population. Methylation patterns were found to be invariable across healthy individuals in regulatory regions such as hypomethylated promoters while gene body and intergenic regions accounted for most of variability. Variations in cis-eQTL regions were also found to have ~40% heritability rate. The study using MuTHER data emphasize the importance of considering common and rare SNPs as well as both cis and trans regulatory regions to understand the global picture of disease susceptibility.[43,44]

Metabolic Syndrome in Men (METSIM) is a population wide cohort of ~10,000 Finnish men. The study focused on identifying genomic factors associated with Type 2 Diabetes (T2D) and cardiovascular diseases (CVD). The METSIM dataset covers exome, whole genome, methylome, gut microbiome and transcriptome with a cross sectional 5 year follow up study (N = ~6,500) to assess the development of T2D and CVD. For the transcriptome profiling, subcutaneous adipose tissue from 1,410 participants was studied either using Affymetrix U219 microarray or bulk RNA-Seq. METSIM initiative helped to gain many novel insights into disease etiology. Numerous GWAS and eQTL loci were discovered associated with T2D, CVD, insulin resistance and adiposity[45].

The Genotype Tissue Expression (GTEx) project is an attempt to catalog genetic variation and gene expression among 54 different tissues of human body, thereby providing a tissue specific reference dataset to the research community. Samples were obtained from postmortem and current version of GTEx (V8) include nearly 1,000 individuals. Adipose tissue from both subcutaneous and visceral depots were included in the study. Gene expression data was studied using RNA-Seq.

The GTEx analysis results and methods used were publicly available through its portal (https://www.gtexportal.org/home/) and the complete data is available through dbGaP.

These population-based studies provide tremendous resource for researchers. However, most of the resources uses bulk sequencing or microarray profiling data which provides an average measurement of all cells in a biological sample. Single cell sequencing has proven to be robust in characterizing individual cells in complex cell population at multiple genomic levels accessing constituent genome, transcriptome, epigenome, metabolome, and proteome. It is often applied to identify heterogenous cells in complex tissues, tumor samples and to identify developmental trajectory of the cells. Single cell RNA sequencing and single nuclei ATAC sequencing strategies are reviewed in the following sections with focus on 10X Genomics protocols considering its relevance to the research chapters 3 and 4.

## 1.3 Single cell Sequencing

### 1.3.1 Single cell RNA Sequencing (scRNA-Seq)

Researchers attempted to study single cells long before the development of current high throughput microfluidic systems[46-48]. Single cell technologies have achieved a tremendous growth in technological front as well as acceptance among scientific community with the expansion of droplet based high throughput systems. One of the earliest references of single cell transcriptome sequencing was from Tang et al. in 2009. The authors modified existing transcriptome amplification method to sequence a mouse blastomere. The technique involved manual selection of single cells using pipetting under microscope[49]. Irrespective of the technique used, once the

single cells were isolated, the nucleic acid (DNA or RNA) is subjected to amplification and further treatments depending on the sequencing requirements such as RNA-Seq or ATAC-Seq.

Capturing of single cells from a biological sample is a challenging task and many methods were developed over time. Fluorescent activated cell sorting (FACS) is one of the widely accepted methods to isolate substantial number of cells based on cell surface markers[50]. FACS helps to sort similar cell types, but methods of choice from serial dilution to droplet based single cell capture techniques should be used to attain single cell resolution. The other traditional methods for single cell isolation include micromanipulation, immunomagnetic separation, laser capture microdissection, array-based methods, and microfluidic platforms. Micromanipulation involves manual intervention to separate single cells from tissue using microscopes and specialized micro dissectors[48]. The method is laborious and prone to high error rates. Immunomagnetic separation uses magnetic beads coated with antigens specific for the cell surface antibody to isolate cells of interest[51]. Laser pulses were used to disintegrate the tissue slices fixed on specialized membranes in laser capture microdissection[52]. Though this method provides spatial information of the cells, excessive cost and low accuracy are major drawbacks[53]. The first implementation of array-based technology is called as CytoSeq. In this method, cells in suspension are loaded to custom designed array with up to 100,000 microwells. The microwells are loaded with magnetic beads which are functionalized with oligonucleotides for PCR priming sequences, cell identity, molecular index for transcripts and oligo dT sequence. The cells are captured at a rate of one in ten wells. However, the limitations include excessive cost and restricted capture of pre-defined set of genes[54]. The latest and one of the most widely used techniques is microfluidics through droplet based single cell capture. In this method, single cells are captured in immiscible liquid emulsion by passing the

cells through micro tubes. Ideally, each droplet provides isolated reaction environment for single cell for further manipulation. Though the droplet-based technologies are robust to identify rare cell populations, enriching cells of interest using FACS before single cell capture is often followed to study the minority cell population[55]. Two pioneering methods that are developed independently using microfluidics are InDrop and Drop-Seq. InDrop used hydrogel beads coated with oligonucleotides made up of photocleavable linker, T7 promoter, PCR primer, cell barcode, Unique Molecular Identifiers (UMIs) and poly dT tail. The hydrogels, cells, reaction mixture (lysis buffer and reverse transcription reagents) and oil are loaded separately to a microfluidic device. The single cells are captured in droplets along with hydrogels and reaction mixture for further processing[56]. Drop-Seq used similar design with a few differences. The authors used hard resin beads functionalized with common PCR amplification sequence, cell barcode, UMI and a poly dT sequence. Also, photoactivation is not required in Drop-Seq as like in InDrop[57].

Droplet based technology is successfully implemented commercially by 10X Genomics and it is one of the widely used method for single cell RNA sequencing (scRNA-Seq). They use gel beads to capture mRNA from the cells and are termed as gel in emulsion (GEMs). The gel beads are functionalized with oligonucleotides comprising of sequencing adapter and primers, 10X barcode, Unique molecular identifier (UMI) and poly dT primer (Figure 1.3.1 A). The 10X barcode is 16 bp (base pairs) in length, designed to identify each gel bead based on ~750,000 sequences designed by 10X Genomics. UMI is a 12bp random sequence to identify each transcript in the cell. The last poly dT primer is for capturing each transcript by binding poly A tails. The poly dT primer can also be replaced with capture sequences to sequence particular region in the transcript. The gel beads and the cells are then passed through chromium next GEM chip to generate GEMs by

capturing single cells (Figure 1.3.1 B). Inside GEMs, the cells are lysed, and poly A tailed mRNA are captured by the oligos released from the Gel beads. The mRNA is then reverse transcribed to full length cDNA (complementary DNA) and then sheared enzymatically. The cDNA molecule with the oligos is selected and sequenced as per standard sequencing protocol[58,59].

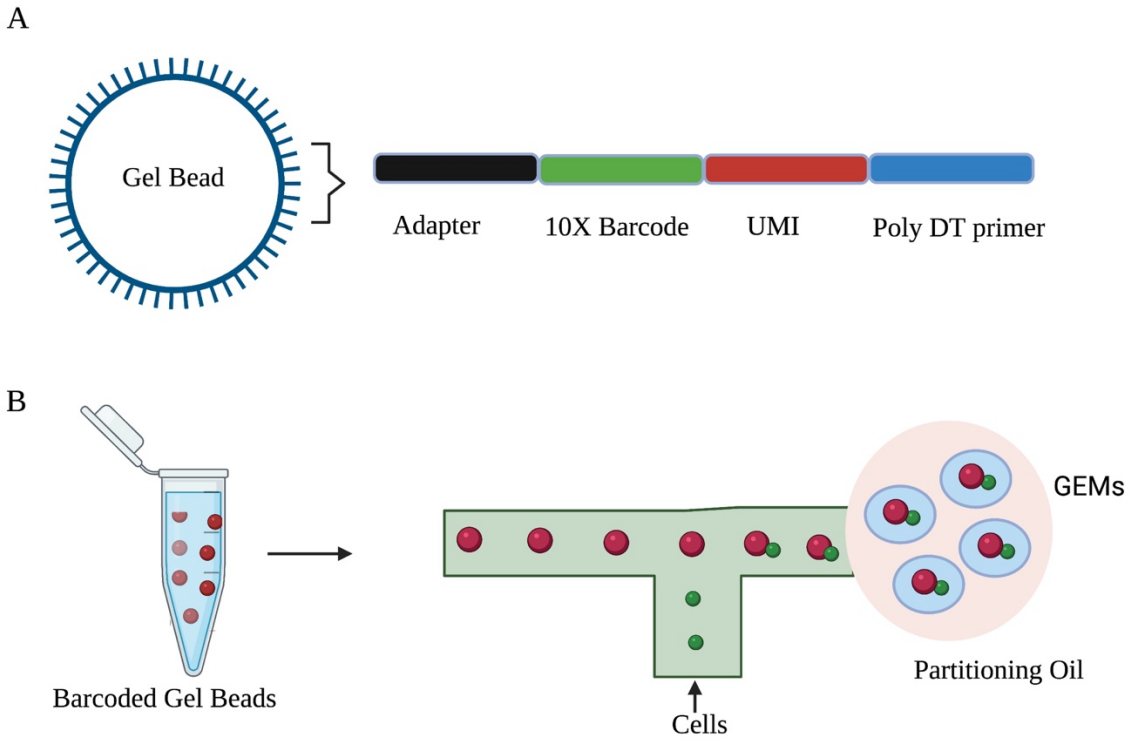



Figure 1.3.1 Experiment setup of 10X Genomics scRNA-Seq *. (A) The Gel bead is coated with oligonucleotides which has sequencing adapters, 10X barcode, UMI and the capture sequence. (B) The gel beads and cells are brought together in an oil medium to form GEMs.

* Created using Biorender by adapting data from 10X Genomics

In addition to 10X Genomics, there are a variety of methodologies available to interrogate the transcriptome at single cell resolution which includes MARS-Seq, SMART-Seq2 and Fluidigm C1 system. SMART–Seq2 provides the sequence of full-length transcriptome ensuring better sensitivity and in-depth analysis of transcripts of interest. However, choosing the methodology

depends on the biological question. For a broad analysis such as to characterize the cell population in complex tissues, molecular tag-based methods such as 10X Genomics, MARS-Seq and Fluidigm C1 system have proven to be sufficient. MARS-Seq has gained acceptance with ease of use of FACS and illumina sequencing protocol. In the recent improvement MARS-Seq2.0, authors claim to attain extremely low levels of doublets (reviewed in section 1.3.2) in single cell capture. Single cells captured using MARS-seq can be stored for long term which is an added advantage against 10X Genomics protocol[60]. Fluidigm C1 system uses array-based technology, and it is restricted to uniform cell size and shape which makes it a less favorable choice. On the other hand, low turnaround time for large number of cells, relative low cost and UMI based transcript capture become an advantage for 10X genomics[61].

### 1.3.2 Data Analysis Methods for scRNA-Seq

The primary processing of scRNA-Seq data from 10X Genomics is done by Cell Ranger pipeline[58]. It includes demultiplexing, quality check, alignment, visualization and aggregating the similar libraries. The expression matrix is generated through this step. It quantifies the expression of each gene for each cell, and it forms the input for further downstream analysis. Among the other result files, Cell Ranger also provides a html document summarizing the results of analysis which gives a general idea about the data.

Several software are available for downstream processing depending on the analysis requirement starting with quality check and filtering. The single cell capture using microfluidics generates empty droplets. However, they may contain cell free RNA called as ambient RNA giving a false impression of presence of cell[58]. The filtering criteria for empty droplets assumes that empty

droplets have a small number of transcripts quantified as UMI count in comparison to regular cells. The cells with low UMI compared to the average quantification of the cells in the sample are removed to overcome this issue. Another factor of consideration is the identification of dying cells. The dying cells could be identified by the presence of low gene count and high count for mitochondrial genes[62]. The general practice is to remove cells with high mitochondrial content (greater than 5% of the total mRNA captured). However, the filtering should be done taking in to account the biology of the sample analyzed. For instance, liver and adipose tissue have high mitochondrial activity, and increase in mitochondrial gene expression relates to biological property of tissue rather than cell death.

The single cell capture techniques have been improved tremendously from the day of inception. But one of the challenges that still require attention is the capture of two or more cells in a droplet termed as doublet or multiplets respectively. Rough estimate from 10X is ~8.0% doublets in an experiment to recover ~10,000 cells when ~16,500 cells are loaded as input. Different approaches are followed to remove the doublets computationally[63]. The cells with more than double the amount of average UMI count or gene count is expected to be doublets and they are removed from analysis. The R package, DoubletFinder tries to identify doublets by random simulation of the artificial doublets from the data analyzed. These artificial doublets are compared with the sample to identify doublets in the data[64]. Another approach is using genetic deconvolution as implemented in demuxlet. Demuxlet allows multiplexing of samples and then deconvoluting the data using genotype information of each sample[65]. The approach is beneficial not only to identify multiplets, as multiplexing also help to overcome batch effects and helps in reducing costs.

Once the superior quality cells are selected, the data can be corrected to overcome known technical bias introduced by varied sequencing depth and batch effects. Single cell data is inherent of specific challenges distinct from bulk RNA-Seq data. Input mRNA is low as it comes from single cell and depends on the abundance of cell population. Further, the quantification of the transcripts can be due to biological signals as well as technical limitations for capturing the mRNA. Added to that, the mRNA capture is also influenced by cell state i.e., stage of cell cycle and batch effects[66]. Normalization of the data can be achieved through log transformation as implemented in Seurat pipeline[67] or using specialized methods like SCnorm[68] or scran[69]. The cells tend to segregate based on the cell cycle. However, cell cycle effects often mask the true variation that the experiment is designed to unveil. Hence regressing the cell cycle effects is advised when necessary. Seurat has in built option to regress cell cycle effects. The removal batch effects require special methods due to rapidly changing scRNA-Seq sequencing and data analysis techniques. Canonical correlation analysis (CCA)[70] by Seurat is one of the widely used methods. CCA tries to align two cell population based on common source of variation. CCA is used to find the similar cell clusters between two sample sets. On the other hand, Conos[71] is designed to remove variation among samples in a data set.

Dimensionality reduction and clustering are effective ways to visualize the cell population captured in scRNA-Seq data analysis. Top variable genes (~ 2,500) are used for this purpose[67]. The common dimensionality reduction method used is principal component analysis (PCA)[72]. Number of components to be used dependent on the complexity of the data. Often it is calculated based on variation in standard deviation of the components using an elbow plot. Visualization methods such as Uniform Manifold Approximation and Projection for Dimension Reduction

(UMAP) can help to view the clustering[73]. Clusters can then be annotated and processed further depending on the aim of the experiment. The data can be explored to identify novel cell clusters, similar clusters with differential gene expression pattern, trajectory of development, cell – cell communication, etc[74-76].

### 1.3.3 Single Nuclei ATAC Sequencing (snATAC-Seq)

The chromatin organization starts with the nucleosomes[77,78]. A nucleosome is a stricture formed by 8 histone proteins (2 sets of H2A, H2B, H3 and H4) and 147 bp length double stranded DNA[79]. Each nucleosome is connected by variable length linker DNA and H1 histone protein. The nucleosomes are then compacted to 30-nm wide chromatin fibers. DNA replication and transcription is dependent on accessibility of DNA from this compact structure and the accessibility is regulated by numerous factors including tissue and disease specificity[79,80]. ATAC-Seq is a robust technique to study open chromatin regions which in turn helps us to identify cis and trans regulatory regions. Briefly, the DNA from the study sample is treated with hyperactive Tn5 transposase which bind to the open chromatin regions and cleaves them. These extracted DNA fragments are then sequenced using standard sequencing protocol[81].

Due to the potential of understanding functional epigenetic regions in the genome, ATAC-Seq is implemented at single cell level which is termed as single nuclei ATAC-Seq (snATAC-Seq)[82,83]. snATAC-Seq profiling is found to be efficient in identifying different cell population in the sample as well as identifying the regulatory regions including promoters and enhancers. It also has the potential to characterize similar cells with distinct transcriptome signatures due to intrinsic changes such as cell cycle effects[84]. Commercially, snATAC-Seq can be implemented using sci-ATAC-

Seq, Fluidigm C1 or 10X Genomics protocol depending on the choice of single nuclei capture techniques. 10X Genomics uses similar capture technique using GEMs as that of scRNA-Seq. The oligonucleotide in the gel bead is made up of 29 bp of sequencing adapters, 16 bp of 10X barcode and 14 bp of primer to the read 1N. GEMs are created from the gel beads and Tn5 transposase treated nuclei. Within the GEMs, the oligonucleotides attach to transposed DNA and activates linear amplification. Then, the DNA is released from the emulsion, PCR amplified and sequenced using established Illumina protocol[85].

### 1.3.4 Analysis Methods for snATAC-Seq

The primary processing of the 10X snATAC-Seq data is done using Cell Ranger ATAC pipeline[85]. Briefly, the raw data is demultiplexed, filtered and aligned to the reference genome. PCR duplicates are removed and ATAC peaks are called based on the enrichment of sequencing fragments in a genomic region. Similar to scRNA-Seq data, cell by peak matrix is generated in this step which serves as the main peak quantification input for downstream processing. Signac[86] pipeline developed by the developers of Seurat can be used for further quality check and clustering. The quality check includes checking for nucleosome banding pattern in a data set. The ratio of fragments of mononucleosome (147 bp and 294 bp) to nucleosome-free (< 147 bp) fragments is calculated to check the distribution of nucleosome signal. Fragments representing transcriptional start site (TSS) can also help to determine the quality of data from the cell. Ideally, ATAC data is expected to show enrichment in regulatory regions which include TSS of genes[81]. Empty droplets, low sequencing depth and multiplets can be assessed by checking the number of fragments in each peak for a particular cell. Low fragment count may be due to empty droplets or low sequencing depth. On the other hand, higher number of fragments more than the double of the average

measurement is considered as a signal of multiplets. Similarly, cells with small number of total peaks should be removed as it could be a technical artefact. ENCODE project[87] identified the blacklisted regions that are prone to technical biases in sequencing data. Cells that have concentration of peaks in blacklisted regions are also removed from analysis. Batch effects can be removed computationally using similar methods to scRNA-Seq such as Harmony[88], Seurat or Conos. snATAC-Seq data is sparse capturing only 1-10% of open chromatin regions[89] and requires special methods for normalization. Signac uses term frequency-inverse document frequency (TF-IDF) normalization by transforming the data by giving increased weight to rare peaks. Once the data is normalized, variable peaks are selected, Latent Semantic indexing and singular value decomposition is performed. Clustering is usually done using k-means or Louvain algorithms followed by visualization using UMAP or t-SNE. Further analysis of the data includes identification of cluster specific peaks, differential peak analysis among clusters or biological conditions, identifying motifs that are enriched in the peaks, etc. It is a widespread practice to integrate scRNA-Seq data to annotate the clusters. Further, combined analysis of both data modalities will help to understand the epigenetic regions influencing changes in gene expression pattern[84]. The peaks could be annotated for functional significance such as promoters, enhancers or transcription factor binding sites using accepted methods of bulk ATAC-Seq data. Transcription factor binding sites are explored using existing databases like JASPER[90] or HOMER[91]. Co-accessible peaks can be calculated by using Cicero to understand the regulatory network of open chromatin regions[92].

### 1.3.5 Recent Efforts in Adipose Research Using Single Cell Methods

Numerous efforts to understand adipose tissue at single cell resolution have been published recently which are focused on adipocyte progenitors, tissue resident immune cells or both[93-96]. However, most of the studies relied on animal model systems such as mouse models owing to the difficulty in obtaining deep adipose tissue samples especially from healthy individuals for reference. One of the initial studies using scRNA-seq identified a CD142+ novel adipocyte progenitor population that regulates adipogenesis. These cell cluster termed as Aregs was found to be conserved in mouse and human samples[94]. Further, studies by Merrick et al. identified three different progenitor population – DDP4+ proliferative MSCs, CD54+ committed pre adipocytes and CD142+ Aregs[97]. DPP4+ MSCs has the potential to differentiate in to ICAM1+ preadipocytes as well as CD142+ Aregs. Studies on lineage tracing of cold induced brown adipocytes by Shamsi et al. identified alternative BAT origin from vascular smooth muscles[98]. Hildreth et al. profiled SVF from abdominal SAT using ~ 110,000 cells. They used scRNA-Seq to identify the clusters and then they applied FACS to study and validate their observations in detail. The study was mainly focused on tissue resident immune cells identifying 15 distinct cell population with differential abundance in lean and obese subjects[93].

Taking into consideration the importance of understanding tissues at single cell level, largescale consortia namely the Human Biomolecular Atlas Program (HuBMAP) and the Human Cell Atlas (HCA) were launched. HuBMAP is an initiative from National Institute of Health (NIH) common fund predicted in 2019 to span for seven years. The organs studied in the initial launch included kidney, spleen, lymph nodes, small intestine, large intestine, thymus, and heart. In addition to single cell transcriptome and open chromatin assays, HuBMAP also targets to resolve spatial conformation using RNA, protein, metabolites, and lipids. The data generated through various

collaborations will be made accessible to the research community[99]. HCA is an international initiative aimed at providing a public reference set of gene expression profile for different tissue, organ and organ system in human body. The current version of HCA portal (DCP 2.0) has data from 13.8 million cells from different human tissues[100].

### 1.4 Rationale, Hypothesis, and Objectives

Adipose tissue contains diverse cellular populations at various depots of human body, which include adipocytes, progenitors, immune cells, endothelial cells, and smooth muscle cells. Abundance and function of these cells differ in individuals with or without obesity and metabolic disorders[93]. Despite the recent attempts, we lack comprehensive understanding of adipose tissue at single cell level. We hypothesized that investigating adipose tissue using a multidimensional approach, focusing on in depth analysis of its pure cellular subpopulations from carefully selected study population will help to understand molecular mechanisms involved in obesity and associated metabolic complications during the lifespan. We aim to provide novel insight into metabolic disease etiology by unraveling transcriptomic and epigenetic signatures and mechanisms underlying the risk of obesity and obesity induced T2D at single cell resolution.

The focus of chapter 2 was to catalogue the different cellular population in adipose tissue derived from obese (adult) individuals discordant for metabolic complications and across depots. SVF from multiple adipose depots (i.e., SAT and VAT) of obese adults was interrogated at single cell resolution to identify gene expression patterns in cellular sub-clusters that were linked to adipose tissue depots and T2D. We then expanded our efforts in chapter 3 by studying SVF obtained from VAT of children to understand early developmental markers associated with obesity. We included an additional layer of interrogation using snATAC-Seq in addition to scRNA-Seq to identify regulatory elements specific to each cell population and disease status. With the in-depth reference of cell population from chapter 2[101] we aimed to understand the changes in individual cell population among lean and obese individuals in chapter3.

# CHAPTER 2: Cataloguing Human Adipose Tissue Across Depot and Disease at Single Cell Resolution

## 2.1 Bridging Statement Between Chapter 1 and 2

The current understanding of adipose biology in the etiology of obesity is based on 1) bulk sequencing approaches in tissue samples, 2) studies using model organisms or 3) matured adipocytes only. In this research chapter, we aimed to extend these efforts and characterize human adipose tissue at single-cell resolution and across depots and disease status using contemporary scRNA-Seq approaches. To accomplish this, we enrolled obese adults undergoing bariatric surgery who were carefully selected based on metabolic phenotypes to obtain a matching group of individuals based on their BMI with and without T2D. We obtained AT collected during the bariatric surgery procedure from the two different fat compartments of interest: greater omentum corresponding to VAT and abdominal subcutaneous fat compartment (SAT). The main goal of the project was to characterize the tissue composition in AT and contrast cellular and transcriptomic signatures across cell type, tissue depot and T2D status. We categorized these signatures based on the known AT cellular populations such as adipose progenitors and stem cells, immune cells, and the endothelium and captured significant variability within each group with evidence of novel, latent cell subpopulations in a depot-specific manner. We used T2D status to cellular and gene expression pattern that contribute to insulin resistance and to identify molecular signatures contributing to metabolic healthy as well as unhealthy obesity. Our findings were further validated using data from MuTHER, METSIM and GTEx study cohorts. To concise, we aimed to create a comprehensive catalogue of adipose tissue resident cell population and their gene expression identities based on the depot and T2D status.

## 2.2 Title, Authors and Affiliations

Single-cell analysis of human adipose tissue identifies depot and disease specific cell types

Jinchu Vijay[1,2], Marie-Frédérique Gauthier[3], Rebecca L Biswell[4], Daniel A Louiselle[4], Jeffrey J Johnston[4], Warren A Cheung[4], Bradley Belden[4], Albena Pramatarova[2], Laurent Biertho[3], Margaret Gibson[4], Marie-Michelle Simon[2], Haig Djambazian[2], The Multiple Tissue Human Expression Resource Consortium[^], Alfredo Staffa[2], Guillaume Bourque[1,2], Anita Laitinen[5], Johanna Nystedt[5], Marie-Claude Vohl[6], Jason D Fraser[7], Tomi Pastinen[4], André Tchernof[3,*], Elin Grundberg[4,*]

[1]Department of Human Genetics, McGill University, Montreal, Québec, Canada

[2]McGill University and Genome Quebec Innovation Centre, Montreal, Québec, Canada

[3]Québec Heart and Lung Institute, Université Laval, Québec, QC, Canada

[4]Center for Pediatric Genomic Medicine, Children's Mercy Kansas City, MO, USA

[5] Finnish Red Cross Blood Service, Kivihaantie 7, FI-00310 Helsinki, Finland.

[6]Institute of Nutrition and Functional Foods (INAF), Université Laval, Québec, Québec, Canada

[7]Department of Surgery, Children's Mercy Kansas City, MO, USA

[^] Full list of members and affiliations appear in the Acknowledgements

[*]Correspondence should be addressed to A.T. (andre.tchernof@criucpq.ulaval.ca) or E.G. (egrundberg@cmh.edu)

Published in Nature Metabolism 2019 December ;2(1):97-109. doi: 10.1038/s42255-019-0152-6.

## 2.3 Abstract

The complex relationship between metabolic disease risk and body fat distribution in humans involves cellular characteristics which are specific to body fat compartments. Here we show depot-specific differences in the stromal vascular fraction of visceral and subcutaneous adipose tissue by performing single-cell RNA sequencing of tissue specimen from obese individuals. We characterize multiple immune cells, endothelial cells, fibroblasts, adipose and hematopoietic stem cell progenitors. Subpopulations of adipose-resident immune cells are metabolically active and associated with metabolic disease status and those include a population of potential dysfunctional CD8+ T cells expressing metallothioneins. We identify multiple types of adipocyte progenitors that are common across depots, including a subtype enriched in individuals with type 2 diabetes. Depot-specific analysis reveals a class of adipocyte progenitors unique to visceral adipose tissue, which shares common features with beige preadipocytes. Our human single-cell transcriptome atlas across fat depots provides a resource to dissect functional genomics of metabolic disease.

## 2.4 Introduction

White adipose tissue (WAT) and its endocrine activities are known to be implicated in the development of obesity and associated metabolic disorders. Specifically, the risk increases with increase in abdominal obesity contributed by excessive visceral adipose tissue (VAT)[11] – a linear relationship that is not seen with abdominal subcutaneous adipose tissue (SAT)[102]. Susceptibility to obesity-related cardiovascular and metabolic disorders has also been linked with the increase in adipose volume resulting from enlargement of tissue resident adipocytes (i.e. hypertrophy)[18]. On the other hand, adipocyte expansion by recruiting new progenitors (hyperplasia) is often

considered as a protective mechanism from the metabolic standpoint[103]. Studies have also shown that adipose tissue dysfunction leading to insulin resistant type 2 diabetes (T2D) is marked by inflammation, fibrosis and / or lipodystrophy[104] which emphasizes the importance of adipose-infiltrating immune cell populations in modulating and developing metabolic disorders. For instance, M1 macrophages, mast cells, B-2 cells, CD8+ T cells and IFN-γ+ Th1 cells were seen to be increased in adipose tissue of individuals with obesity compared with those who were normal weight and the reverse pattern was observed in M2 macrophages, eosinophils, Treg, iNKT, B1 and γδ T cells[105]. These adipose tissue resident immune cells have also been shown to create a microenvironment that can inhibit adipocyte progenitor differentiation to lipid-storing adipocytes[106]. However, despite extensive work on characterizing various cell subpopulation in adipose tissue, the complete human non-adipocyte fraction also known as the stromal vascular fraction (SVF) has not been profiled across depots in an unbiased manner. Given the multitude of factors affecting adipose tissue function, a thorough understanding of the cell types involved, and their specific gene expression pattern is essential. The advent of single-cell transcriptomic approaches in the past years have made it possible to use these technologies to determine cellular heterogeneity and functional states at the single-cell level with high reproducibility and sensitivity[107]. Current high-throughput microfluidics techniques are capturing thousands of cells from each sample simultaneously for gene expression profiling and together with new algorithms for clustering, visualization, and modeling this allows for high-powered analysis of disease-targeted tissue samples for efficient cataloging of cellular composition and the role in disease risk. Recent studies utilizing single-cell RNA sequencing (scRNA-Seq) in adipose tissue from mouse models have identified a subset of adipocyte progenitors that regulates adipocyte differentiation[94]

as well as the presence of a novel type of inflammatory progenitors residing in the visceral fat depot of the mice[108]. Similar strategies in human adipose samples have not been applied to date.

We present a high-throughput single-cell expression profiling study of human adipose tissue including 25 samples derived from multiple depots of individuals with obesity. We provide a rich catalog of cell types residing in adipose tissue including both latent and common cell populations. We characterize and validate distinct cell types that are metabolically active, specific to each depot or correlate with metabolic disease status.

## 2.5 Results

### 2.5.1 Characterization of SVF across multiple adipose depots

We generated scRNA-Seq data from 25 adipose samples (12 VAT and 13 SAT) derived from 14 individuals undergoing bariatric surgery (Supplementary Table 1, Supplementary Figure 1, Methods). All samples were matched for age and BMI but differed based on fasting glycemia as an indication of T2D (Table 1). We annotated the clusters using marker genes (Supplementary Table 2-3) which resulted in three groups of cells: adipocyte progenitors and stem cells (P1-P7), immune cells (I1-I7) and endothelial cells (E1-E3) (Figue 1). The proportion of the cell types based on individual average was 55%, 37% and 8% for progenitors, immune and endothelial cells, respectively. This distribution was similar for the merged sample average and corresponded to 60%, 34% and 6% which are in line with classical fluorescence-activated cell sorting (FACS) experiments [109]. To avoid subjectivity and to add strength to analyses, we also performed reference-based single-cell annotation (Methods) which confirmed our broad clusters of cell populations (Supplementary Table 4). Finally, as the distribution of female and male samples was

slightly skewed (Table 1) we performed in-depth sex-specific expression analysis of the marker genes in 245 (N=165 male and N=80 female) VAT and SAT samples, respectively, from the GTEx Consortium (Supplementary Table 5).

2.5.2 Multiple types of endothelial cells reside in adipose tissue

We identified three types of endothelial cells (E1-E3, Figure 1) all showing selective expression for *GNG11* and *SEPW1*. E1 and E2, were found to be relatively similar sharing nearly half of the top expressed genes including *SPARCL1, FABP4* and *IFI27*. However, comparative analysis between E1 and E2 cells showed the latter group expressing classical endothelial markers such as *ACKR1, SELE, TM4SF1, VCAM1, TMEM173, PLVAP, ICAM1, PECAM1, VWF, ADAMTS9* and *TFPI*. On the other hand, E1 cells had pronounced expression of *FABP4, LGALS1, RBP7, GPX3* and *CD36* which are known to be expressed in microvascular endothelial cells of adipose tissue involved in the fatty acid handling machinery[110]. To characterize E3, we performed differential gene expression analysis between E3 and the combined population of E1 and E2 cells, respectively, (Supplementary Table 6) and found *LYVE1* among the top expressed genes. *LYVE1* is a marker of lymphatic endothelial cells [111] and thus these results indicate the presence of lymphatic vasculature in SVF with 78% of its cell population from VAT samples.

2.5.3 Characterization of Immune cells and their link to adipose tissue inflammation

We found that 34% of cells from our merged clusters (Figure 1, I1-I7; Supplementary Table 4) expressed markers of different immune cell populations. To facilitate the classification of the cell types we subset the immune cell clusters and identified 14 new clusters, labelled as IS1 - IS14, where all groups were present across depots (Figure 2a-b).

A large proportion (40%) of the immune cells (IS1, IS4, IS6, IS8) were clustered closely together and showed gene expression signatures of NK / T cells. IS1 exhibited selective expression of IL7R which is known to be expressed in naive T cells[112]. However, unsupervised cell type definition pointed towards a mix of naïve and memory CD4+ and CD8+ T cells (Supplementary Table 7). Manual annotation of IS4 cells identified specific expression of *GNLY, NKG7, FGFBP2, GZMB, GZMH* and *CTSW* which points toward these cells being NK cells (Supplementary Table 8). Indeed, this was supported by unsupervised annotation which showed 68% of IS4 cells being correlated with the NK cell reference. A recent report found that obesity is associated with reprogramming of blood-derived NK cells with upregulation of genes involved in lipid metabolism including *CD36*[113]. However, the IS4 NK cells identified here did not express any genes involved in lipid metabolism pointing towards different obesity-related effect on NK cells dependent on resident tissue.

IS6 showed increased expression of *CCL5* and *IL32*, indicating that these cells may be activated T cells [114] which was confirmed by the unsupervised annotation revealing 68% of the cells being associated with memory CD8+ T cells. While large proportion (59%) of the IS8 cells were, similar to IS6 cells, mapped as memory CD8+ T cells in our unsupervised annotation we noted in the manual curation that they showed a unique pattern of high expression of metallothionein genes – *MT1E, MT1F, MT1G, MT1X* and *MT2A* (Supplementary Table 8, Figure 2c). We followed-up this finding by performing additional scRNA-Seq profiling on CD34- sorted SVF cells and found evidence of the presence of metallothionein-rich T cells also in the validation sample (Cluster 7; Supplementary Table 9; Figure 2d) where both discovery and validation samples specifically

showed *MT1F* and *MT1G* to be uniquely expressed by this T-cell subpopulation (Figure 2c-d). Similar to what was recently shown for circulating immune cells[113], these potential adipose-resident dysfunctional T cells may be induced by the obesity and thus enriched in our scRNA study. To test this, we used bulk gene expression data from two large SAT population-based collections including 1) 776 female samples from the MuTHER resource and 2) 770 male samples from the METSIM Study where both studies have in addition to tissue collection, study subjects deeply phenotyped for obesity-related traits [115,116]. Here, we used BMI and DXA-derived % Trunk fat (PTF) in association with the expression of *MT1E, MT1F, MT1G, MT1X* and *MT2A* (Supplementary Table 10). Encouragingly, we found consistent strong positive associations between the obesity-traits and expression of *MT1F, MT1G* and *MT2A* whereas *MT1E* and *MT1X* showed weaker or absence of significant association in line with our observations of cell-specific expression of these two genes. Finally, we found that all metallothionein genes are expressed in adipose tissue irrespective of tissue depot with no sex-specific expression pattern (Supplementary Figure 2, Supplementary Table 5).

Our manual and/or unsupervised annotation mapped clusters IS2, IS3, IS7, IS9 and IS12 to adipose tissue macrophages (ATM) with the proportion of cells expressing the classical macrophage marker *CD68* ranging from 19 to 52% (Extended Data 1a) with IS2 having the highest (52%) and IS9 the lowest (19%). IS2 cells were also enriched with genes that are involved in lipid metabolism during obesity (Supplementary Table 8) including *LIPA, LPL, CD36* and *FABP4* – findings in line with the notion that obesity activates a non-classical inflammatory phenotype of ATM involving lipid accumulation and trafficking[117]. In addition, we find these metabolically active ATMs (i.e. IS2 cells) to express *CD9* more pronounced than the other ATMs (Extended Data 1a). We

performed cellular phenotyping by immunohistochemistry and validated the co-expression of *CD68* and *CD9* in obese adipose tissue (Figure 2e-f) as well as confirmed [115,116] the linear relationship of *CD9* expression and BMI in the MuTHER and METSIM SAT cohorts, respectively (MuTHER BMI: p-value=2.2E-13, beta=0.011; METSIM BMI: p-value=9.4E-28, beta=0.379).

IS12 and IS3 clusters showed a similar signature as IS2 but where IS3 cells had increased expression for inflammatory genes *CXCL3, CXCL2, IL1B, CCL3* and *CXCL8* (Extended Data 1b). On the other hand, IS9 cells are marked by the expression of *FOLR2* and *KLF4* which are known signatures of M2 macrophages [118,119]. Finally, although our unsupervised annotation indicated IS12 cells as macrophages we could not distinguish them further.

Our unsupervised annotation identified IS5 and IS10 as classical CD16- monocytes with high expression of *S100A8, S100A9* and *S100A12* in IS10. However, further literature mining showed that IS5 cells are rather dendritic cells due to their high expression of HLA genes including *HLA-DPB1, HLA-DQA1, HLA-DPA1, HLA-DRA* and *HLA-DQB1*. IS13 cells possessed signatures of a newly identified subtype of dendritic cells [120] with high expression of *LST1, SERPINA1, AIF1* and *FCGR3A*.

Finally, we were able to link the IS11 cluster to B cells due to the expression of *IGKC, JCHAIN, CD79A* and *CD37*. Remaining cluster (IS14) showed mixed signatures in both annotation strategies and thus was not explored further.

2.5.4 Identification of adipocyte progenitors derived from different adipose tissue depots

Next, we characterized our progenitor clusters (P1-P7) by studying the expression pattern of cell surface makers used in FACS experiment of SVF progenitors[121] and noted that all clusters appeared CD45− CD34+ CD31− (Figure 3a) with a profound signature in P1-P6. We also noted that five of the clusters (P2, P4-P7) were identified to express *CFD*[122] encoding the adipokine Adipsin which is known as a marker of adipocyte differentiation (Figure 3b) and thus the range of *CFD* expression observed across the clusters may be linked to different stages of adipogenesis. To validate this, we used in vitro differentiation studies of human mesenchymal stem cells (MSC) and identified a 5.6 log2 fold induction between undifferentiated MSCs and the first stage of adipogenic induction (ACI vs AI, Adjusted p-value=1.19E-26, Figure 3c). We further noted that *CFD* expression is additionally increased at later stages of adipogenesis (AI vs AD1, log2fold=1.3, Adjusted p-value=0.012; AI vs AD2, log2fold=1.2, Adjusted p-value=0.036). Incorporating this information with our progenitor signature indicated that P2 and P7 include more mature preadipocytes compared to cells within clusters P4-P6. We could further validate this by unsupervised annotation which linked on average 88% of the cells in P2 and P7 to Adipocytes compared to only on average 70% of cells in P4-P6 (Supplementary Table 11).

We then compared the seven progenitor clusters based on tissue type and found a striking depot-specific pattern (Figure 4a) where three of the clusters (P2, P6, P7) were mainly composed of cells from SAT whereas P1, P3, P5 included mostly cells from VAT and only one cluster of progenitors (P4) included a mixed population of SAT- and VAT-derived cells. For instance, the VAT-specific clusters P1 and P3 exhibited pronounced expression (average log fold change > 1.5) of omentin (*ITLN1*)[123] and mesothelin (*MSLN*). Omentin is a VAT-specific adipocytokine serving as a

biomarker for metabolic diseases[124] and together with *MSLN* also a marker of mesothelial cells[125] in line with findings that VAT develops from the mesothelium[23]. Additionally, we identified 16 other genes characterizing P1 and P3 (average log fold change >1, Supplementary Table 2) which we followed-up with bulk RNA-Seq and found 11/16 (68%) genes being differentially expressed (log fold change>=1 and Adjusted p-value<0.05) (Supplementary Table 12, Figure 4b).

To study these cell populations in more detail, we created two subsets for re-clustering including 1) SAT- (N=5,458) and VAT-derived (N=9,847) progenitor cells, respectively.

2.5.5 Subcutaneous adipocyte progenitors and their role in metabolic conditions

The clustering of the adipocyte progenitors identified five groups (SP1-SP5, Figure 5a-5b) all expressing *CFD*. SP1 and SP3 cells expressed pre-adipocyte/adipose stem cell markers (e.g., *MGP*, *APOD, CXCL14, WISP2*) whereas SP2 showed signatures of a more mature adipocyte progenitor cell (e.g., *APOE, FABP4, CEBPB* and *CD36*). The SP4 group contained cells expressing genes involved in fibrosis and extra cellular matrix accumulation including *COL3A1, COL6A3, COL1A1* and *COL6A1*. This fibroblast signature was validated by unsupervised annotation where 69% of the cells were mapped as fibroblasts compared to SP1-SP3 with over 80% of the cells mapped as adipocytes (Supplementary Table 13). SP5 cells expressed high levels of inflammatory markers (e.g., *CCL5, CD3E, IL7R* and *IL32*) including *CD45*/ *PTPRC* (Figure 5c) indicating that SP5 cells represent hematopoietic stem cell (HSC) progenitors.

We then investigated whether the proportion of adipocyte progenitor clusters specific to SAT in each individual sample was associated with T2D status. We correlated fasting glucose levels with

all the SP cell proportions and found the abundance of SP1 to be significantly correlated with glucose levels (Pearson r=0.56, p-value=0.046, Figure 5d, Supplementary Table 14). We followed-up this finding in multiple ways and provide further evidence of the link between SP1 cells and T2D. First, we used top differently expressed genes in this group (SP1, N=43, Adjusted p-value < 0.01) compared to all other cell types (SP2-SP5; Supplementary Table 15) for disease annotation (Methods) and found T2D as the top annotated disease (Supplementary Tables S16, p-value=3.21E-6). Secondly, we used the MuTHER resource [115] followed by the METSIM Study[116] for transcription-wide association analysis of multiple T2D-related traits. We found the SP1 gene set (N=43) to be significantly enriched for association to the T2D traits in both studies (Supplementary Table 17) corresponding to fold changes between 2.5 and 9.4 dependent on the trait and study (Fisher's p-value = 0.06 – 7.96E-08, Supplementary Table 18). Specifically, the female MuTHER Study included enriched association for 44% of the SP1 genes using strict Bonferroni corrected study-wise p-value threshold of p<2.53E-6. Replication of these genes in the male METSIM Study showed strong correlation of effect sizes across T2D traits (r=0.92, Supplementary Figure 3).

Finally, we went back to our single-cell data set and performed differential expression analysis between SP cells from T2D vs non-T2D samples as defined by their fasting glycemia restricting to the 19 validated SP1-specific genes (Supplementary Table 17-18). Here, we found 53% of the genes (10/19) being also differently expressed (Bonferroni p-value < 0.05/19, Figure 5e) in cells based on T2D status with the same direction of effect as seen in MuTHER (Supplementary Table 19). These genes include *GPX3, WISP2, ATF3, EIF1, MGP, TMEM176B, CXCL12, CFH, C1S* and *ADH1B*[126]. The top differentially expressed gene was *GPX3* with significantly higher

proportion of cells expressing the gene in non-T2D samples (73%) versus in T2D samples (66%) (Seurat DGE, p-value=1.5.9E-14). In line with these findings, we found *GPX3* to be negatively correlated with Homeostatic Model Assessment for Insulin Resistance (HOMA-IR) in both the MuTHER (beta=-0.084, p-value=1.03E-11) and METSIM cohort (beta=-0.147, p-value=4.46E-5), respectively.

We also noted *WISP2* among those top 10 genes with differential expression between T2D and non-T2D cells (p-value=3.1E-5) but as opposed to *GPX3, WISP2* seems to have a negative effect on T2D status. We found 72% of T2D cells expressing the genes versus only 52% of the non-T2D cells. Again, we validated this pattern in the two large MuTHER and METSIM adipose tissue cohorts with a positive correlation with HOMA-IR (beta=0.115, p-value=2.2E-20 and beta=0.132, p-value=2.3E-4). Similar to *WISP2*, we found that T2D cells expressed *ATF3* to a larger extent (51%) compared to non-T2D cells (29%) and *ATF3* was positively correlated to HOMA-IR in MuTHER (beta=0.079, p-value=2.9E-21) and METSIM (beta=0.102, p-value=4.4E-3), respectively.

2.5.6 Validation of depot independent adipocyte progenitors in CD34+ sorted cells

Clustering of the VAT-derived progenitor cells resulted in six clusters (Figure 6a, Supplementary Table 20) with cells expressing *CFD* (referred to as VPC and correspond to VP4-6) versus *MSLN* (referred to as VPM and correspond to VP1-3) were clustered separately (Figure 6b). The VPCs (VP4-6) resembled the progenitor types identified in SAT and thus may represent adipocyte progenitors that are present across depots. Specifically, VP4 included cells enriched for genes such as *APOD, CXCL14, DPT, GPX3, MGP, EIF1, C1S* and *ADH1B* and thus with a similar expression

pattern as the T2D-linked SP1 cells. VP5 cells show an enrichment for *MFAP5* which in turn is involved in adipose tissue remodeling in individuals with obesity[127]. Another notable gene expressed in cluster VP5 was *S100A4*, a biomarker for inhibition of adipogenesis, associated with reduced obesity and inflammation in rodents [128]. In all, the expression signature of VP5 was similar to the fibrotic signatures of the cells grouped as SP4. Finally, VP6 is marked by the expression of *TYROBP, HLA-DRA* and *CD74* and in general have the inflammatory signature identified for the HSC progenitors in SP5. Taken together, the combined results from our SAT and VAT single-cell data indicate presence of three depot-independent cell types: T2D-associated adipocyte progenitors, fibroblasts and HSC, respectively. To provide further support to these observations, we generated scRNA-Seq profiles of CD34+ sorted SVF cells (Methods). The clustering of these cells resulted in thirteen groups (0-12, Supplementary Figure 4, Supplementary Table 21) with clear evidence for overlapping fibrotic and HSC properties of two of the clusters. Specifically, we linked cluster 10 (134 cells) to HSC progenitors which corresponded to 2% of the complete cell population concurring with the proportion of cells in the SP5 cluster (3%) from the deep sequenced sample when restricting to progenitors only. The expression signature of this cluster also resembles that of SP5 with *PTPRC, CCL5* and *IL32* among the top genes distinguishing these CD34+ progenitors from the other. Similarly, we found clusters 5 and 7 in the validation sample to have similar properties as the fibrotic SP4 cluster including expression of *FBN1, PI16* and *IGFBP6*.

2.5.7 Visceral specific progenitor cells have different origin and mitochondrial activity

In contrast to VPC, VPM appeared to have unique properties specific to the visceral depot with signature of a mesothelial origin as discussed above. Apart from specific expression of *MSLN*, the VPM progenitor cells also expressed the mesothelial adipocyte markers *UPK3B* and *WT1*

(Extended Data 2). Comparing VPC and VPM cells, we also noted clear differences in mitochondrial gene expression. Specifically, we found that the VPM had comparatively high mitochondrial expression (i.e., 6-20%) whereas VPC showed a consistent pattern of cells with low (i.e., <=5%) expression of mitochondrial genes (Figure 6c). As it is well known that brown adipocytes contain higher number of mitochondria than white adipocytes, this intriguing finding prompted us to test whether the VPM cells constitute progenitors with differentiation potential towards beige adipocytes and thus potentially have a more protective function against obesity and insulin resistance. To test this, we first correlated VPM proportion with fasting glucose levels and found a significant negative correlation (Pearson r=-0.64, p-value=0.025, Figure 6d). We also noted that VPM expressed *PLA2G2A* (VP1; Adjusted p-value= 5.84E-109) which was recently shown to activate mitochondrial uncoupling in brown adipose, and in line with our observations here, provided protection from the deleterious effects of high fat diets in mice[129]. In addition, the mitochondrial marker gene *SOD2* (VP3; Adjusted p-value=2.72E-245) known to be upregulated in beige and brown AT[130] compared to white AT was found to be specific to the VPM[131] cells.

Next, we investigated whether this beige adipose signature found among a subset of the VAT progenitor cells (i.e., VPM) could be validated in independent datasets of purified VAT and SAT - derived mature adipocytes from the same clinical study population (Supplementary Table 22). Indeed, we noted striking overrepresentation of not only the mesothelial markers *MSLN* ($log_2$fold=1.48, Adjusted p-value=1.7E-3), *WT1* ($log_2$fold=2.73, Adjusted p-value=8.1E-10) and *UPK3B* ($log_2$fold=2.54, Adjusted p-value=8.3E-9) but also the classical markers for beige and brown adipocytes *EBF2* ($log_2$fold=2.39, Adjusted p-value=2.1E-20), *PRDM16* ($log_2$fold=1.44, Adjusted p-value=9.9E-5) and *UCP1* ($log_2$fold=2.20, Adjusted p-value=2.3E-7). However,

progenitor or mature brown adipocyte markers *MYF5* ($log_2$fold=0.04, Adjusted p-value=0.96) or *ZIC1* ($log_2$fold=-0.01, Adjusted p-value=0.98) were not differentially expressed, suggesting that the VPM cells are not of the brown adipocyte lineage. We then used bulk SVF and whole tissue from VAT and SAT to validate the induction of *UCP1* during differentiation to mature adipocytes whereas the expression of *MSLN* and *WT1* peaks in progenitors (Figure 6e). In addition, we noted a strong negative correlation in VAT between *UCP1* and both mesothelial markers (*MSLN* vs *UCP1*, r=-0.72, p-value=0.018 and *WT1* vs *UCP1*, r=-0.72, p-value=0.019; Supplementary Figure 6). We also noted a significant positive correlation between *UCP1* and *IRX3* in VAT (r=0.85, p-value=1.71E-3) in line with recent reports showing the induction of *IRX3* in the browning process of adipocytes[132] (Supplementary Figure 6). Lastly, in line with our observations of marker genes characterizing VPM, we observed a trend of *SOD2* expression to be positively correlated with the expression of *UCP1* (r=0.56, p=0.09) further indicating browning signatures of a subset of VPM-derived adipocytes residing in the visceral depot only.

These several lines of evidence indicate that there are UCP1+ cells originating from *MSLN* / *WT1* expressing progenitors in VAT only where the expression of *UCP1* is peaking in mature adipocytes that underwent browning. This may explain why we failed to detect any *UCP1* expressing cells in our large single-cell data of VPM progenitors from individuals with obesity. However, as we observed a negative correlation between proportions of VPM cells and metabolic disease status (Figure 6d) we hypothesized that *UCP1*-expressing preadipocytes derived from VPM progenitors may be more pronounced in healthy tissue early in development. To address this, we accessed VAT from a healthy young donor. We subset CD34+ cells for clustering (Figure 7) and confirmed the high abundance of *MSLN* / *WT1* progenitors in SVF derived from VAT.

Encouragingly, we also identified a distinct cluster of cells among the VPM that not only expressed *UCP1* (Figure 7c) but also had the highest mitochondrial content (Figure 7d). Finally, we also noted *SOD2* among the top genes with differential expression between *UCP1*-expressing versus non-UCP1 VPMs (Adjusted p-value=7.02E-93) with 95% of UCP1+ cells expressing the gene versus only 75% of the *UCP1*- VPMs. On the contrary, as shown above, UCP1+ cells had significantly lower expression of *WT1* (Adjusted p-value= 4.24E-08) with only 22% of UCP1+ cells expressing the gene versus 56% of the UCP1- VPMs (Supplementary Table 23).

**2.6 Discussion**

We have performed a large, unbiased assessment of the cellular landscape in the non-adipocyte fraction of human adipose tissue (known as the SVF). The corresponding mature adipocyte fraction per depot was studied in parallel by bulk sequencing approaches due to current technical limitations in profiling large and lipid-rich cells at single-cell resolution. Study subjects were selected based on obesity-related metabolic phenotypes including presence or absence of type 2 diabetes (T2D), which allowed us to incorporate differences in cell populations not only dependent on depot but also based on disease status.

Our initial classification resulted in three broad categories of cell types including progenitors or stem cells, immune cells and endothelial cells. The latter cell type represented ~8% of all cells but with distinctive signatures dividing the group further into subpopulations. Although in minority, it is well established that endothelial cells play an important role in adipose tissue inflammation where obesity induced T2D has been associated with profound dysfunction of endothelial cells[133]. Our largest population of endothelial cells were characterized as fatty acid handling microvascular

endothelial cells due to pronounced expression of genes encoding fatty acid transport and binding proteins, respectively. These results are in line with recent work showing the importance of endothelial fatty acid uptake in adipose tissue especially during obesity-induced tissue inflammation[110]. Another subpopulation was identified as lymphatic-derived due to expression of the lymphatic vessel endothelial hyaluronan receptor (*LYVE1*) gene. This is of interest as recent efforts have shown the importance of a crosstalk between lymphatic vessels and adipose tissue, and that lymphatic dysfunction are linked to metabolic diseases [134]. We extend these efforts and show that the lymphatic vasculature is depot specific as the majority of these cells were derived from visceral samples.

Our characterization of adipose-resident immune cells, representing ~30% of the cells, confirmed the complex nature of obesity-induced adipose tissue inflammation. We identified 14 different immune cell types including T cells, B cells, NK cells, dendritic cells, monocytes and macrophages. T cells, and particular the CD8+ subset, are known to play an important role in adipose inflammation where the obese adipose tissue activates these cells which then recruit and activate adipose-resident macrophages[135]. We identified activated memory CD8+ T cells expressing *CCL5* which is known to be positively correlated with obesity[136] where increased expression associate with increased BMI. Interestingly, we found a subpopulation of these CD8+ T cells expressing metallothionein with unique clustering pattern compared to the activated memory CD8+ cells and with indication of an association to obesity-related traits. Specifically, using two large population-based adipose tissue resources we found a strong association of the expression of the cell-specific metallothionein genes *MT2A, MT1F* and *MT1G* and obesity phenotypes indicating a role of these cells in adipose dysfunction. This T cell expression profile

of metallothionein genes was recently linked to a novel type of dysfunctional CD8+ T cell population identified in single-cell studies of tumor-infiltrating lymphocytes[137] where targeted deletion of the metallothionein resulted in loss of T cell dysfunction. Similar to Singer et al, we also found that these metallothionein genes are expressed only in the potential dysfunctional adipose-resident T cells but not in the activated memory CD8+ T cells, and as such break new grounds in possibly defining the role of T cells in adipose inflammation and potentially insulin resistance.

As mentioned above, macrophages are activated by T cells and infiltrate the adipose tissue contributing to the obesity-related tissue inflammation and subsequently insulin resistance. Adipose resident macrophages play different key role in energy metabolism, clearance of dead adipocytes and removing extracellular lipids[138] which are found to be altered during the development of obesity and insulin resistance. We observed that with high expression of genes involved in lipid metabolism during obesity, our largest macrophage population likely represent the so called metabolically active ATMs which have been associated with a beneficial role in obesity including promoting dead adipocyte clearance[139]. We further showed that they have selective expression of *CD9* which points towards them representing a distinct type of obesity-related ATMs.

We identified multiple clusters of progenitor cells with clear expression signatures dependent on cell source (i.e., subcutaneous or visceral). They were further characterized to be at different stages of adipogenesis which was marked with varying degree of *CFD* expression. *CFD* has been identified as a key player in adipogenesis where knockdown of the gene inhibits lipid accumulation

and expression of adipocyte markers during adipocyte differentiation[140]. On the other hand, overexpression of *CFD* promotes adipocyte differentiation. We confirmed the association of *CFD* expression and adipocyte differentiation with validation experiments using mesenchymal stem cells that were monitored during adipogenesis. These stem cells were derived from bone-marrow and cultured in vitro and thus may possess different properties than adipose-derived stem cells, however it was recently shown that they do have similar adipogenic differential potential[141]. We also show clear distinction of the subpopulations beyond what has been shown before[142]. We identified a subtype of premature adipocytes that associated with T2D where individuals with high glucose levels had higher abundance of this cell type than those with normal glucose levels. We noted that this T2D-linked cell type was characterized by genes within the PPARγ pathway, some with protective role against insulin resistance. One of those key protective genes was *GPX3* where significantly higher expression was identified in cells derived from nondiabetics. It is known that T2D-associated oxidative stress can be reversed by PPARγ-mediated antioxidant regulation mediated through the expression of *GPX3*[143]. Furthermore, PPARγ agonists such as Thiazolidinediones (TZDs), commonly used to treat T2D, although not used in our samples, induce these antioxidant effect with increased expression of *GPX3* as a result. On the contrary, we also found genes expressed by these T2D-associated cells with seemingly negative effect on T2D status including *WISP2* and *ATF3*. The expression of these genes was significantly higher in T2D cells as well as a larger proportion of T2D cells expressing the genes compared to non-T2D cells. *WISP2* is an adipokine and while one of its key functions is to regulate adipogenic commitment and PPARγ activation, our results confirm previous efforts showing it can also induce insulin resistance likely through the inhibition of adipocyte recruitment and promotion of hypertrophic obesity[144]. *ATF3* inhibits PPARγ-mediated transactivation[145] and hepatic *ATF3* expression is

associated with T2D[146] and as such *ATF3*, similar to *GPX3*, may serve as a potential therapeutic target for the reduction of insulin resistance.

Finally, while our depot-specific analysis confirmed earlier efforts showing that visceral-derived adipocyte progenitors originate from the mesothelium [23], we provide additional support to the notion that a subset of these are also likely inducible by responding to environmental stimuli and may be differentiating into beige preadipocytes. We noted that for instance the visceral-specific VPM cells, while being CD34+/WT1+ they did not express *MYF5*, suggesting that rather than being classical brown preadipocytes, they are more likely induced and could be consistent with so called beige adipocyte progenitors[147]. Although the physiological significance of such beige adipocyte progenitors in visceral adipose of individuals with morbid obesity is uncertain, they may possibly represent a compensatory mechanism to reduce the adverse effects of obesity. This hypothesis is supported by our finding of overrepresentation of this particular cell population in non-insulin resistant as well as healthy subjects.

We acknowledge that a limitation of our study is the focus of individuals with obesity where further studies are needed to contrast cellular architecture in adipose tissue derived from healthy individuals. However, as there are clear depot and region-specific differences in cellular distributions, single-cell profiling spanning healthy and obese individuals would require a surgical procedure common in the general population where tissue resection can be obtained for research purpose.

In conclusion, we show the richness of the cellular contribution in human adipose tissue with multiple types of subpopulations of immune cells, endothelial cells, fibroblasts and progenitors. Interrogation of the non-adipocyte fraction of adipose tissue at single cell level suggests that development of obesity and T2D could be a joint action of subpopulation of different cell types residing in adipose tissue with altered function.

## 2.7 Methods

Study Subjects

Participants were selected through an institutionally approved biobank infrastructure with ongoing recruitment at the Quebec Heart and Lung Institute (IUCPQ; Université Laval, Quebec City, Canada). Inclusion and exclusion criteria were those related to bariatric surgery. Specifically, men and women who had a BMI≥40kg/$m^2$ or ≥35kg/$m^2$ with major comorbidities, who required surgery and who met the NIH Guidelines for bariatric surgery were invited to participate. Exclusion criteria were general contra-indications for bariatric surgery, a BMI<35 kg/$m^2$, age under 18 or over 60 years, respectively, abnormal bowel habits including irritable bowel syndrome, pregnancy, cirrhosis, inflammatory bowel disease and previous bariatric surgery. Participants were further selected based on i) the presence of a signed consent form and ii) availability of type 2 diabetes (T2D) status where subgroups (i.e., with or without T2D) were matched for BMI, age as well as males:females ratio as much as possible.

For the discovery cohort used for single-cell RNA sequencing, we identified 14 individuals (4 males and 10 females) using the above described inclusion criteria where five individuals were diagnosed with type 2 diabetes and nine individuals without type 2 diabetes. Detailed clinical

information is provided in Table 1. For single- cell expression validation, two additional individuals were identified using the same selection criteria including one 35-year-old male without type 2 diabetes and with BMI of 59 kg/$m^2$ and one 42-year-old female with type 2 diabetes with BMI 44.8 kg/$m^2$

For validation studies using bulk SVF and adipocyte samples, we identified 13 independent individuals from the same study population. This sample set included 12 females and one male with a BMI range from 39.7 to 59.2 kg/$m^2$ and age values ranging from 29 to 60 years. Six of these individuals were diagnosed with type 2 diabetes and six were not. Similarly, for validation studies using bulk adipose tissue samples, we identified another set of 10 independent individuals with BMI 35.5 to 71.8 kg/m2, age between 42 and 49 years and the male-to-female ratio as 1:1.

For all sample sets, adipose tissue was collected during the bariatric surgery from two different fat compartments: greater omentum corresponding to VAT and abdominal subcutaneous fat compartment (SAT). Informed consent was obtained from each participant through the management of the framework of the Quebec Heart and Lung Institute Obesity Biobank. The protocol was approved by the Research Ethics Committee of the IUCPQ (Protocol # 21320).

Sample Preparation

SVF and mature adipocytes were obtained as follows: adipose tissue was digested within 30 min of collection with collagenase according to a modification of the Robdell method[148]. Briefly, adipose tissue samples were digested with collagenase type 1 in Krebs-Ringer-Henseleit (KRH) buffer for 45 minutes at 37°C. Cell suspensions containing mature adipocytes and SVF were then

filtered with a nylon mesh and washed 3 times with KRH buffer. The nature of the buoyancy adipocytes allows them to float to the surface. Mature adipocytes were aliquoted and the remaining solution containing the SVF was centrifuged 1500 rpm for 5 minutes. The pellet was washed with pre-adipocyte growth medium (PGM) (DMEM-F12 supplemented with 10% calf serum, 1% penicillin-streptomycin, 17µM pantothenic acid, 33µM biotin, 100µM ascorbic acid and 2,5 µm/ml amphoB) followed by a second centrifugation. SVF cells were then cryopreserved using freezing medium (PGM supplemented with 40% FBS and 10% DMSO). The medium was added to the pellet and was frozen with a temperature gradient (-1 °C/ minutes) and stored in liquid nitrogen until analysis. Whole adipose tissue samples were following collection quickly frozen in liquid nitrogen and stored until analysis.

Isolation of CD34+/CD34- SVF cells by fluorescence activated cell sorting

Cells from the SVF were thawed and suspended in PBS-0.1% BSA. They were centrifuged at 1500 rpm for 5 minutes. Erythrocyte lysis buffer was added for 5 minutes to the suspension to get rid of red blood cells. The suspension was centrifuged again at 1500 rpm for 5 minutes. Primary antibody (CD34, PE, eBioscience) was added to the samples and incubated for 60 minutes at 4°C in the dark. PBS-0.1% BSA was added to wash the suspension followed by 80 µm filtration with nylon mesh to remove debris. The suspension was then transfer to a 10ml polypropylene tube, centrifuged at 1500 rpm for 5 minutes and the pellet was suspended with PBS-0.1% BSA. Viability staining solution has been added 15 minutes before the sorting process. Viability control was performed with live/dead cells to ascertain that the staining solution worked well and for gate adjustment. OneComp Ebeads control was also used to provide positive/negative control for the antibody. The samples were sorted with the BD FACSAria II (BD, San Diego, California, USA)

with the following parameters: Cell size, graininess, viability, singlet cell and CD34+/CD34-. Dead and non-singlet cells were automatically discarded. Cells that were positive for CD34 and negative for CD34 were collected in two distinct tubes (Supplementary Figure 7). After sorting, cells were centrifuged at 1500 rpm for 5 minutes and were conserved in freezing medium as described above and stored at -80 °C until further analysis.

MSC culture and differentiation

Bone-marrow derived mesenchymal stromal cells (LY-MSCs) from eight healthy 19-27 years old female donors were provided by the GMP facility of the Advanced Cell Therapy Centre, Finnish Red Cross Blood Service (FRCBS), Helsinki, Finland. The cells were thawed, expanded and differentiated into adipocytes following the FRCBS protocol previously published[149]. Briefly, in preparation for adipogenic differentiation cells at passage #3 were plated in 10 cm cell culture plates and grown in expansion media containing low glucose DMEM (Thermo Fisher Scientific), 100 U/ml penicillin /100 μg/ml streptomycin (Thermo Fisher Scientific) and 10% Stemulate (Pooled Human Platelet Lysate culture media supplement, Cook Regentec) until they reached 70%-90% confluence (Timepoint 1), at which point the media was changed to adipogenic basal medium containing alpha-MEM Glutamax, 10 % FBS, 20 mM HEPES, 100 U/ml penicillin and 100 μg/ml streptomycin (Thermo Fisher Scientific). Half of the culture plates were used as controls and maintained in the adipogenic basal medium until harvesting (AC samples). The other half was subjected to adipogenic differentiation (AD samples). More specifically the cells were first incubated for 3-4 days in adipogenic basal media supplemented with the induction cocktail of 0.1 mM indomethacin (Sigma), 0.5 μg/ml insulin, 0.2 mM 3-isobutyl-1-methylxanthine (IBMX), and 0.4 μg/ml dexamethasone (Preadipocyte Differentiation Medium Supplement Pack, PromoCell)

(Timepoint 2), then the media was changed to the terminal differentiation cocktail containing 0.1 mM indomethacin (Sigma), 0.5 μg/ml insulin and 3 μg/ml Ciglitazone (PromoCell). The media of all plates was changed twice a week and the differentiation was allowed to proceed for 3 weeks for Timepoint 3 or 4 weeks for Timepoint 4. For each donor samples were harvested at Timepoint 1 (ND – non-differentiated cells), Timepoint 2 (AIC and AID: Adipocyte induced control without drugs in the media or differentiated with induction drugs), Timepoint 3 (AC1, AD1: Adipocyte control without drugs in the media or differentiated with drugs) and Timepoint 4 (AC2, AD2: Adipocyte control without drugs in the media or differentiated with drugs).

Single cell RNA sequencing

Cells from the SVF were thawed and serially diluted in DMEM/F12 supplemented with 10% fetal bovine serum and 1% penicillin-streptomycin in 15ml conical tube to a volume of 14ml. They were centrifuged at 300 rcf for 10 minutes and the supernatant was removed and discarded. The pelleted cells were suspended in an adequate amount of DMEM/F12 supplemented with 10% fetal bovine serum and 1% penicillin-streptomycin and transferred to a 1.5ml microfuge tube for counting. 10 μl of the suspension was used for a cell count and viability assay on a Countess II Automated Cell Counter (Invitrogen) to determine the volume of suspension to be used in the 10x Chromium Single Cell Library protocol. The desired number of viable cells to be loaded onto the Chromium Single Cell A Chip (10x Genomics) was 12,000 cells. After cell capture, the remaining cells were centrifuged at 300 rcf for 10 minutes and cryopreserved using freezing medium (Recovery Cell Culture Freezing Medium, Thermo Fisher Scientific). The cells were suspended in freezing medium and frozen with a temperature gradient of -1ºC per minute. Cell capture, cDNA amplification, and library preparation were performed using the Chromium Single Cell 3' Library

& Gel Bead Kit v2 (10x Genomics), according to the manufacturer's protocol. The final libraries were analyzed using a TapeStation to determine library size and a Qubit Broad Range dsDNA assay to determine library concentration and sequenced using Illumina HiSeq at high depth (mean 171,560 reads/cell) capturing on average 1,527 cells/sample.

Bulk RNA Sequencing

In total of 0.5 to 3 million cells were re-suspended in 500 uL TRIzol Reagent and total RNA was extracted using the miRNeasy Mini Kit (Qiagen) according to the manufacturer's protocol. RNA library preparations were carried out on 500 ng of RNA with RNA integrity number (RIN)>7 using the Illumina TruSeq Stranded Total RNA Sample preparation kit, according to manufacturer's protocol. Final libraries were analyzed on a Bioanalyzer and sequenced on the Illumina HiSeq 2500 (pair-ended 100 - 150 bp sequences). On average 61,222,621, 49,175,966, 66,301,281 and 84,454,561 paired-end reads were sequenced in adipocytes, SVF, adipose tissue and MSCs respectively.

QC and clustering of single cell data from SVF

Initial analysis of the single cell libraries was done using cellranger v.2.1.0 from 10X Genomics. GRCh38 was used as the genome reference. The analysis was started with 26 samples. All sequenced samples were aggregated using cellranger aggr pipeline where read depth difference between libraries was normalized by subsampling higher depth libraries. Seurat[70] was used for further analysis of clustering, dimensionality reduction and differential gene expression analysis. For filtering, we used only cells with a minimum gene count of 200 and a maximum of 2500 genes/cell based on their distribution in the sample. We also removed cells with unique molecular

identifiers (UMIs) less than 200 and greater than 13,750 per cell. Cells with percentage of mitochondrial gene expression greater than 20 was also removed. However, clustering showed that the majority of the cells (90%) from one sample were clustered together apart from the clusters of all the other samples. The sample was removed the sample and analysis was repeated.

The final input data to Seurat comprised of 38,170 cells and after filtering we obtained 26,350 cells. 11,111 cells were removed because of low gene / UMI count. We regressed out the variability due to UMI distribution, mitochondrial gene expression and the difference between G2M and S phase scores based on the gene expression of cell cycle genes [150]. Top 23 dimensions were used to generate final clusters using principal component analysis (PCA) and graph based clustering. Top 100 genes in each cluster were identified in comparison with all other cells using the function – 'FindAllMarkers' in Seurat while keeping a cutoff of padj <= 0.01. Genes that are expressed to a minimum of 25% of cells in either of the test population is considered for analysis. Clusters were annotated using overlapping known marker genes among the cluster specific genes. Differential gene expression analysis between cell population was done using Wilcoxon rank sum test in 'FindMarkers' function. The raw expression data of the filtered cells used in clustering of Seurat was used for generating annotation using SingleR with default parameters[151]. The annotation for each cell is then overlapped with Seurat cluster identities to identify the cell type of each cluster. The subset for the VAT progenitors and immune clusters were created from the clustering results using 'SubsetData' function of Seurat. Top 10 and 15 dimensions were used for clustering VAT progenitors and Immune cells respectively using highly variable genes in the corresponding data.

The sample with highest number of cells captured (n=6,672) is randomly subsampled to 1,500 cells to match the cell counts of remaining samples. The clustering is repeated using the same settings as that of initial analysis. A subset of SAT progenitors (P2, P4, P6, P7) was created. Cells from VAT that were included in the subset was removed and we obtained 2,705 that are from SAT progenitors. Top variable genes were identified and clustering was done using top 10 dimensions. The cluster specific genes were used for functional annotation using DAVID Bioinformatics Resources 6.8 [152,153].

Differential gene expression analysis using bulk RNA

Raw reads were trimmed for quality (phred33 ≥30) and length (n≥32), and Illumina adapters were clipped off using Trimmomatic v.0.35[154]. Filtered reads were then aligned to the GRCh38 human reference using STAR v.2.5.3a[155]. Raw read counts of genes were obtained using htseq-count v.0.6.0[156]. Differential gene expression analysis was done using DeSeq2 v.1.18.1[157].

The gene raw count table (V6p) of RNA-Seq data was downloaded from the GTEx web portal. The subset of data is then created selecting only tissues of interest based on the sample attribute file provided in the GTEx portal. Differential expression gene analysis is done using DeSeq2 v.1.18.1.

Detection of CD9+/CD68+ macrophages by histological immunofluorescence

Visceral and subcutaneous adipose tissue samples from four patients undergoing bariatric surgery were identified of which two were diagnosed with type 2 diabetes. The mean age and BMI were 47.3±11.7 years and 49.9±5.2 kg/m2 respectively. Adipose tissue slides were disembodied by xylene (2 x 10 min) followed by rehydration step using ethanol (2 x 5 min 100% and 2 x 5min

95%). Antigen retrieval was performed using citrate buffer 0.01M (0.1M citric acid with 0.1 sodium citrate mixed with water) for 5 min at 65°C and 10 min at 95°C. The final rehydration step was performed by immersion of the slides for 5 minutes in water and for 5 minutes in PBS. Non-specific sites were blocked for at least 45 minutes using PBS 1X, 0.1% BSA, 0.4% Triton-X100 and 10% goat serum. Primary antibodies, rabbit anti-human CD9 (PA5-11556, Invitrogen, California, USA) and mouse anti-human CD68 (ab955, Abcam, Cambridge, United Kingdom) were incubated overnight at 4°C in a humidified chamber. The slides were washed 3 times in PBS 1X for 5 minutes. Secondary antibodies, goat anti-rabbit Alexa488 (A11008, Invitrogen, California, USA) and goat anti-mouse Alexa594 (A11005, Invitrogen, California, USA) were incubated for 2 hours at room temperature in the dark. The slides were then washed 3 times with PBS 1X for 5 minutes. Vectashield mounting medium with DAPI (H-1200, Vectors Laboratories, California, USA) was used with coverslip and sealed with nail polish. Digital images were taken using Zeiss LSM800 confocal system (Zeiss, Oberkochen, Germany). Image were taken at 20x and 63x magnification. The slides were stored at 4°C in the dark and imaged less than 4 days after staining.

Validation of beige adipocyte signatures by single cell sequencing

Adipose tissue was collected from a 15-month-old boy undergoing inguinal hernia repair at Children's Mercy Kansas City. Consent was obtained through the management of the study protocol approved by the Institutional Review Board at Children's Mercy Kansas City (Protocol # 17110653). An easily accessible portion of the greater omentum at the lower end of the omental drape was identified and a portion was pulled out and a 2cm segment was amputed to allow for areas of cautery changes to be discarded. Specimen was passed off the field in a jar containing

saline. Adipose tissue was then prepared as described above to obtain SVF. The SVF pellet was treated with ACK lysis buffer at RT for 5 minutes, followed by another centrifugation step. The ACK lysis buffer was removed from the pellet and the cells were washed once more with HBSS+200uM adenosine+2%FBS and centrifuged one final time. Cells were counted to determine the total number of cells in suspension and viability. Cells were then cryopreserved using freezing medium (Gibco™ Recovery™ Cell Culture Freezing Medium, 10% DMSO). The medium was added to the pellet and was frozen with a temperature gradient (-1 °C/ minutes) and stored in liquid nitrogen until analysis. At the time of single-cell analysis, cells from the SVF were thawed and prepared as described above including cell count and viability assay on a Countess II Automated Cell Counter (Invitrogen) to determine the volume of suspension to be used in the 10x Chromium Single Cell Library protocol. The desired number of viable cells to be loaded onto two wells of a Chromium Single Cell A Chip (10x Genomics) was 8000 cells for a total of two captures with 4000 cells/capture. Cell capture, cDNA amplification, and library preparation were performed using the Chromium Single Cell 3' Library & Gel Bead Kit v2 (10x Genomics), according to the manufacturer's protocol. The final two libraries were analyzed using a TapeStation to determine library size and a Qubit Broad Range dsDNA assay to determine library concentration and sequenced on an Illumina NovaSeq6000 instrument. For QC and clustering, we followed same procedure used for the discovery samples to analyze the validation sample. However, for filtering, we used only cells with a minimum gene count of 200 and a maximum of 6000 genes/cell based on their distribution in the sample. We also removed cells with unique molecular identifiers (UMIs) greater than 50,000 per cell. Cells with percentage of mitochondrial gene expression greater than 40 was also removed. After filtering we used 2,779 cells for further clustering and analysis. Top 55 dimensions were used to generate final clusters using principal component

analysis (PCA) and graph-based clustering. The subset for the progenitors were created from the clustering results using 'SubsetData' function of Seurat. Top 50 dimensions were used for clustering VAT progenitors using highly variable genes in the corresponding data.

Gene expression analysis in population based cohorts

Associations between gene expression levels (IlluminaHT12) and phenotypes within the MuTHER cohort were modeled using a linear mixed effects model as described previously[115]. Briefly, the lmer function in the lme4 package, was fitted by maximum-likelihood. The linear mixed effects model was adjusted for age and experimental batch (fixed effects) and family relationship (twin-pairing) and zygosity (random effects). A likelihood ratio test was used to assess the significance of the phenotype effect. The p-value of the phenotype effect in each model was calculated from the Chi-square distribution with 1 degree of freedom using -2log (likelihood ratio) as the test statistic. Summary statistics from associations between gene expression levels and phenotypes within the METSIM cohort were obtained from METSIM Study[116].

## 2.8 Acknowledgements

This work was supported by a Canadian Institute of Health Research (CIHR) team grant awarded to E.G (EGM141898) and the CIHR funded Epigenome Mapping Centre at McGill University (EP1-120608) awarded to T.P. E.G. holds the Roberta D. Harding & William F. Bradley, Jr. Endowed Chair in Genomic Research and T.P. holds the Dee Lyons/Missouri Endowed Chair in Pediatric Genomic Medicine. A.T. is the director of a Research Chair in Bariatric and Metabolic Surgery. M.C.V. is the recipient of the Canada Research Chair in Genomics Applied to Nutrition and Metabolic Health (Tier 1). The MuTHER Study was funded by a program grant from the

Wellcome Trust (081917/Z/07/Z) and core funding for the Wellcome Trust Centre for Human Genetics (090532). The TwinsUK study was funded by the Wellcome Trust and European Community's Seventh Framework Programme (FP7/2007-2013). The TwinsUK study also receives support from the National Institute for Health Research (NIHR)- funded BioResource, Clinical Research Facility and Biomedical Research Centre based at Guy's and St Thomas' NHS Foundation Trust in partnership with King's College London. We further thank additional members of the MuTHER consortium for providing valuable data for this study; Kourosh R. Ahmadi, Chrysanthi Ainali, Amy Barrett, Veronique Bataille, Jordana T. Bell, Alfonso Buil, Emmanouil T. Dermitzakis, Antigone S. Dimas, Richard Durbin, Daniel Glass, Neelam Hassanali, Catherine Ingle, David Knowles, Maria Krestyaninova, Cecilia M. Lindgren, Christopher E. Lowe, Eshwar Meduri, Paola di Meglio, Josine L. Min, Stephen B. Montgomery, Frank O. Nestle, Alexandra C. Nica, James Nisbet, Stephen O'Rahilly, Leopold Parts, Simon Potter, Johanna Sandling, Magdalena Sekowska, So-Youn Shin, Kerrin S. Small, Nicole Soranzo, Gabriela Surdulescu, Mary E. Travers, Loukia Tsaprouni, Sophia Tsoka, Alicja Wilk, Tsun-Po Yang & Krina T. Zondervan. The authors also thank Matti Korhonen and Marja Ahti for valuable help in this study. The Genotype-Tissue Expression (GTEx) Project was supported by the Common Fund of the Office of the Director of the National Institutes of Health, and by NCI, NHGRI, NHLBI, NIDA, NIMH, and NINDS. The data used for the analyses described in this manuscript were obtained from the dbGaP accession number phs000424.v7.p2 on 07/10/2019.

**2.9 Additional Information**

Declaration of Interests

A.T. receives funding from Johnson & Johnson Medical Companies and Medtronic for research

unrelated to the present manuscript.

Data Availability

Raw data files from single-cell and bulk RNA-Sequencing used in this study are uploaded in GEO in a SuperSeries with accession number: GSE136230. The expression data from the MuTHER cohort have been deposited in the ArrayExpress, (https://www.ebi.ac.uk/arrayexpress/) with accession number E-TABM-1140.

## 2.10 Main Tables and Figures

2.10.1 Tables

Table 1: Characteristics of the study subjects. P values are calculated used two-sided t-test.

| | T2D | non-T2D | |
|---|---|---|---|
| N | 5 | 9 | |
| Male:Female (N) | 2:03 | 2:07 | |
| | mean ± SD | mean ± SD | P-value |
| Age (years) | 52.8±12.5 | 43.3±11.4 | 0.2 |
| BMI (kg/m2) | 41.0±6.0 | 43.5±7.2 | 0.49 |
| Fasting glycemia (mmol/L) | 10.7±1.3 | 5.57±0.5 | 4.97E-04 |
| Percentage of glycated hemoglobin | 0.08±0.005 | 0.05±0.002 | 1.61E-04 |

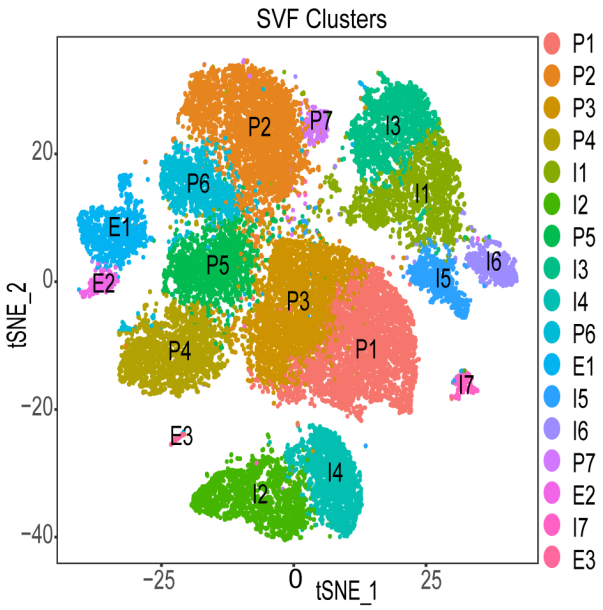



Figure 1: Identified cell populations in the non-adipocyte fraction of adipose tissue. Clustering results of 26,350 cells from the stromal vascular fraction (SVF) derived from 25 adipose samples that underwent single-cell RNA sequencing identifying 17 clusters. Cell populations were classified as Progenitors (P), Immune cells (I) and Endothelial cells (E) and labelled accordingly.

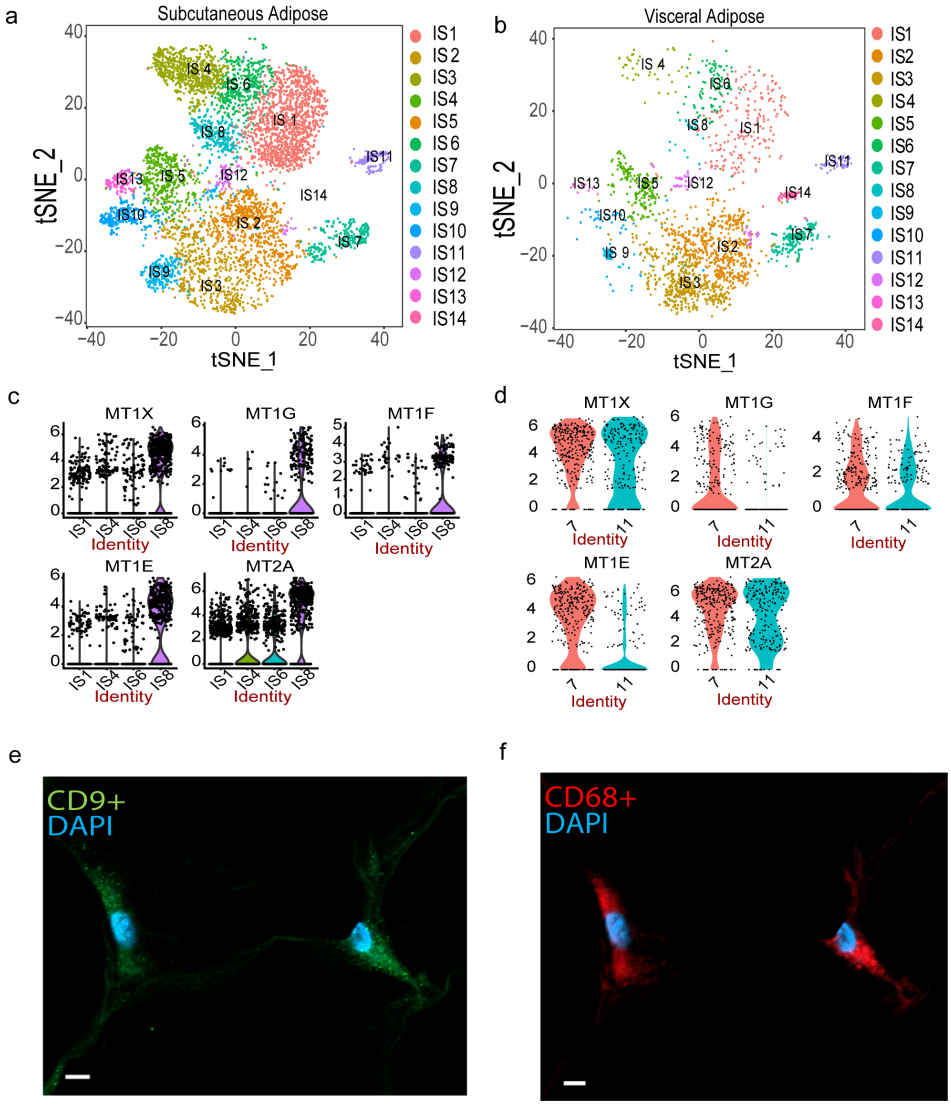



Figure 2: SVF-derived immune cells. Re-clustering of 9,025 CD34- cells representing immune cells in the stromal vascular fraction (SVF) of adipose tissue identified 14 cell types including T/NK cells (IS1, IS4, IS6, IS8), macrophages (IS2, IS3, IS7, IS9, IS12), dendritic cells (IS5, IS13), monocyte (IS10) and B-cells (IS11) in SAT (a) and VAT (b). Violin plots of expression density of Metallothionein genes across T / NK cell clusters in discovery (c) and CD34- validation samples (d) comprising 25 and 3 samples, respectively. The y axis indicate log transformed expression values and the width indicate number of cells expressing the particular gene. Immunohistochemistry in subcutaneous adipose derived from a 60 year-old woman with a BMI of

52.2 kg/m2 showing co-expression of *CD9* (e) and *CD68* (f) cells. The scale bar indicates 5 µm. The staining was done on 4 independent individuals to confirm.

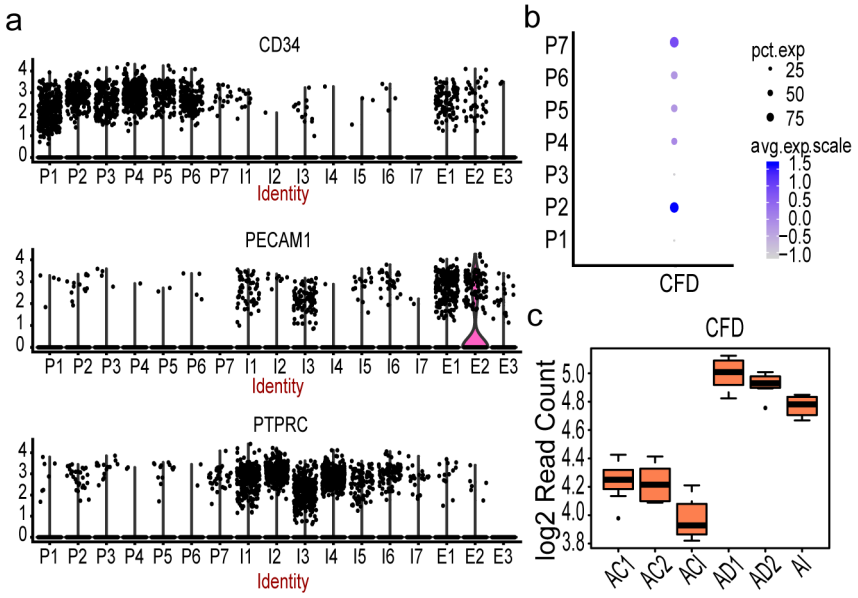



Figure 3: SVF-derived progenitor clusters. (a) Violin plots of log transformed expression density of *CD34* (upper panel), *CD31/PECAM1* (middle panel) and *CD45/PTPRC* (bottom panel) across all SVF clusters from 25 samples. The width of the violin plot indicates number of cells expressing the particular gene. (b) Dot plot of the expression of *CFD* across progenitor (P) clusters. The size of the dot corresponds to the percentage of cells expressing *CFD* in each cluster and the color represents the average *CFD* expression level (c) Box plot of *CFD* expression at 4 different time points of adipocyte differentiation from mesenchymal stem cells (MSCs, n= 8 individuals). AI corresponds to the first time point three days after culturing MSCs in the induction media. AD1 and AD2 are 1 and 2 weeks of differentiation in adipogenic media after the AI time point. ACI, AC1 and AC2 are corresponding control sets for AI, AD1 and AD2 without any adipocyte differentiation treatments. Each time point with corresponding control includes three independent MSC cultures and shown by average log 2 read count. The black line inside the boxplot represents

and median value and the size of the box is determined by the 25th and 75th percentile of the data. The wiskers of box plot represents the maximum and minimum values of the data shown.

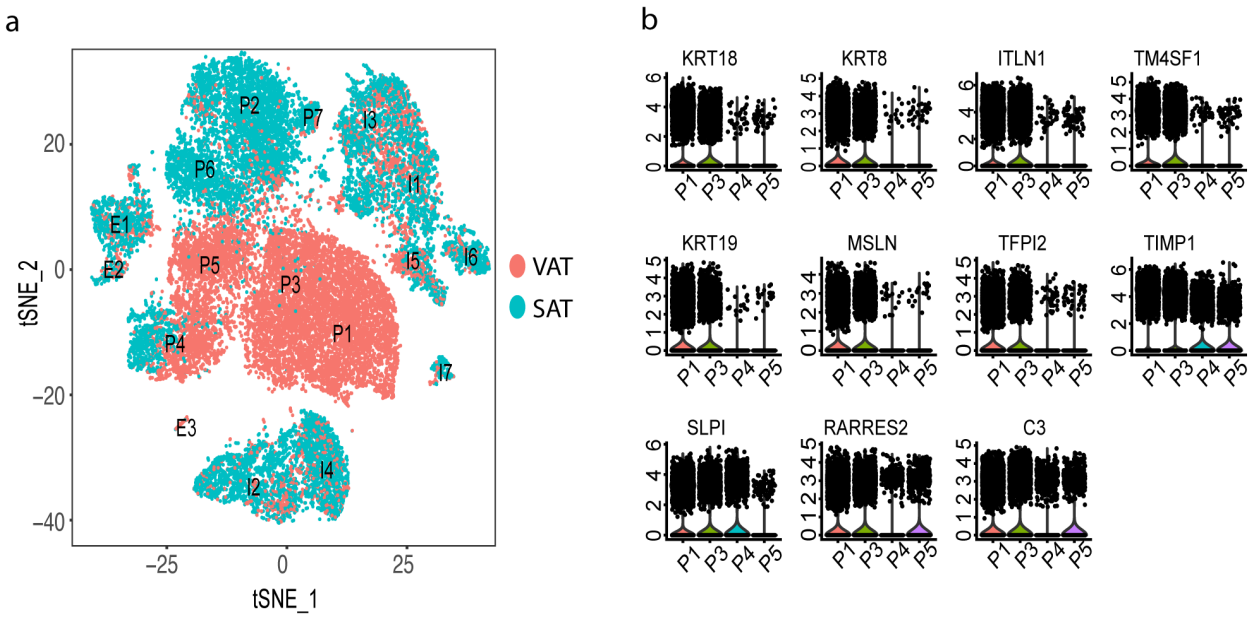



Figure 4: Main cell clusters in SVF based on depot. (a) Clustering results of all SVF samples that underwent single-cell RNA sequencing which identified 17 clusters from 26,350 cells. Clusters were classified as Progenitors (P), Immune cells (I) and Endothelial cells (E) and labelled accordingly. Cells are labelled as VAT (red)- or SAT (blue)-derived, respectively. (b)Violin plots of expression density of 11 genes that are specific to VAT progenitor clusters. The y axis indicate log transformed expression values and the width indicate number of cells expressing the particular gene.

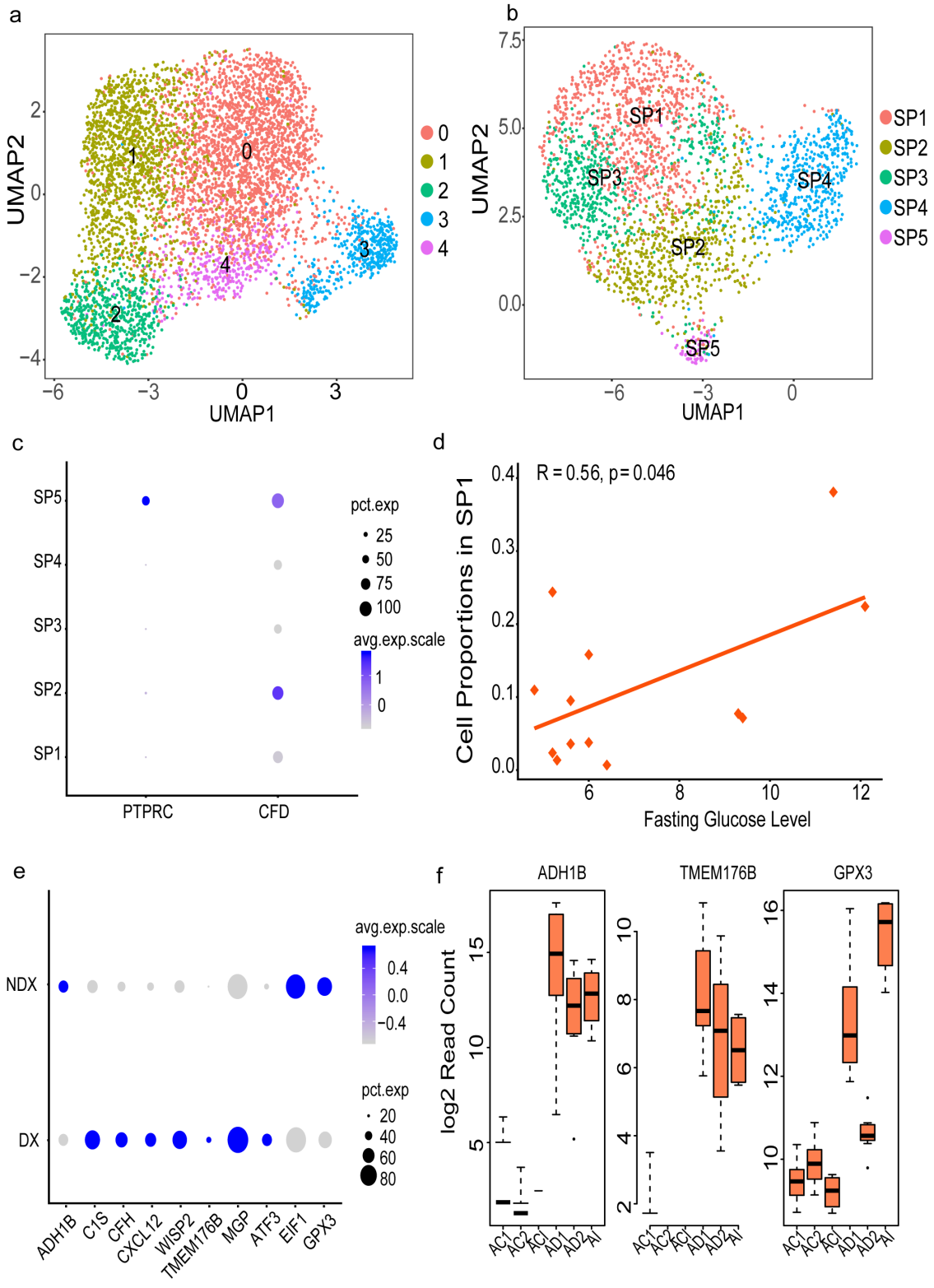



Figure 5: Progenitor clusters specific to SAT. (a) Re-clustering of SAT-specific progenitors from the complete sample set of 26,350 cells (b) Re-clustering of SAT-specific progenitors (2,705 cells)

by subsampling the high coverage library to 1500 cells (c) Dot plot of the expression of PTPRC (CD45) and CFD across SAT progenitors (SP1-SP5). (d) Pearson correlation of fasting glucose levels (mmol/L) and SP1 proportion across all samples obtained from 13 individuals (e) Dot plot of differentially expressed SP1 genes in cells derived from T2D (DX) versus non-T2D (NDX) samples previously validated in the MuTHER study (f) Box plot of T2D-associated SP1 gene expression at different time points of adipocyte differentiation from mesenchymal stem cells (MSCs, n= 8 individuals). AI corresponds to the first time point three days after culturing MSCs in the induction media. AD1 and AD2 are 1 and 2 weeks of differentiation in adipogenic media after the AI timepoint. ACI, AC1 and AC2 are corresponding control sets for AI, AD1 and AD2 without any adipocyte differentiation treatments. Each timepoint with corresponding control includes three independent MSC cultures and shown by average log 2 read count and error bars corresponding to standard deviation. The black line inside the boxplot represents and median value and the size of the box is determined by the $25^{th}$ and $75^{th}$ percentile of the data. The wiskers of box plot represents the maximum and minimum values of the data shown. The size of the dot in (c) and (e) corresponds to the percentage of cells expressing the genes in each cluster and the color represents the average expression level.

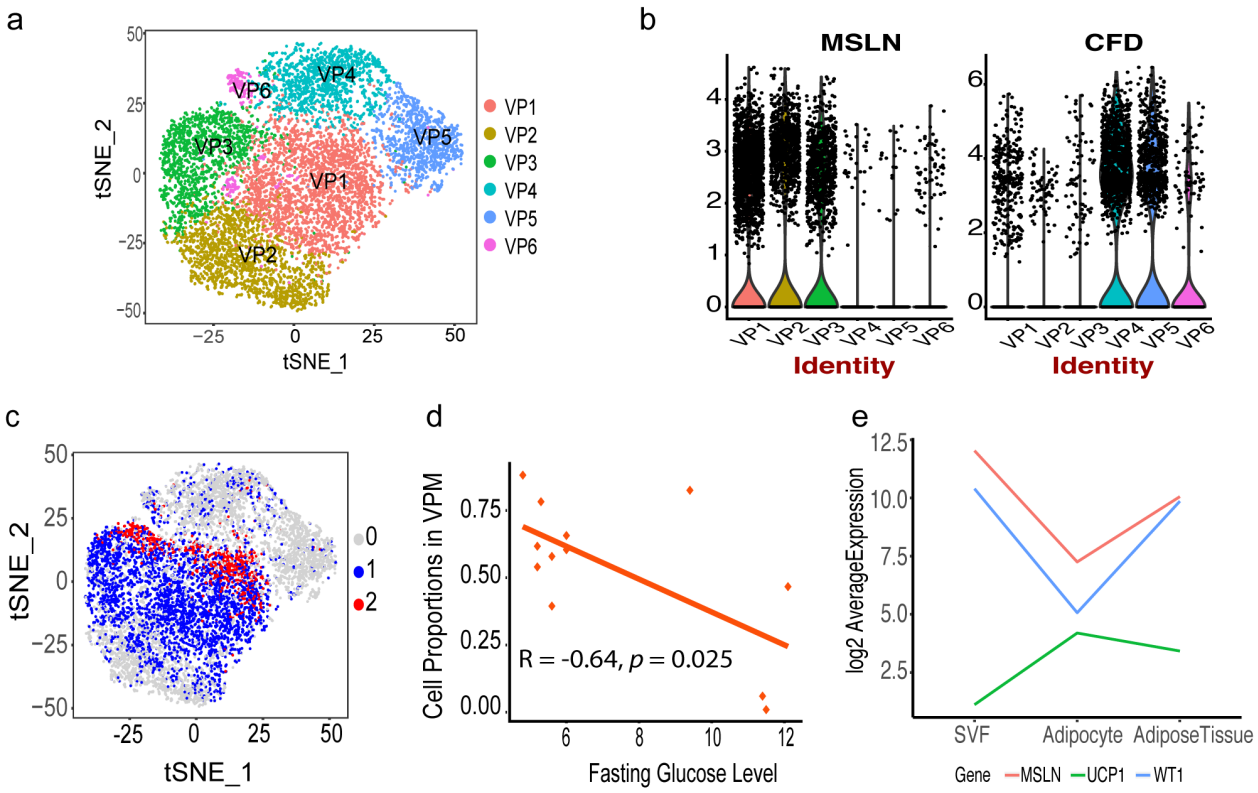



Figure 6: Progenitor clusters specific to VAT derived from individuals with obesity. (a) Re-clustering of VAT-specific progenitors (9,847 cells) from the complete sample set (b) Violin plots showing expression density of *MSLN* and *CFD* in 6 VP clusters. The y axis indicate log transformed expression values and the width indicate number of cells expressing the particular gene (c) VAT progenitor cells labelled based on mitochondrial gene distribution: group 0 (grey, 5,018 cells) shows cells with mitochondrial gene expression <=5%, group 1 (blue, 4,100 cells) represents 6% to 14% and group 2 (red, 729 cells) represents 15% to 24% expression. See also supplementary figure 5 (d) Pearson correlation of fasting glucose levels (mmol/L) and VPM cell proportion across all samples from 12 individuals (e) Expression pattern of *UCP1, MSLN* and *WT1* using bulk RNA-Seq showing change in expression in mature adipocytes.

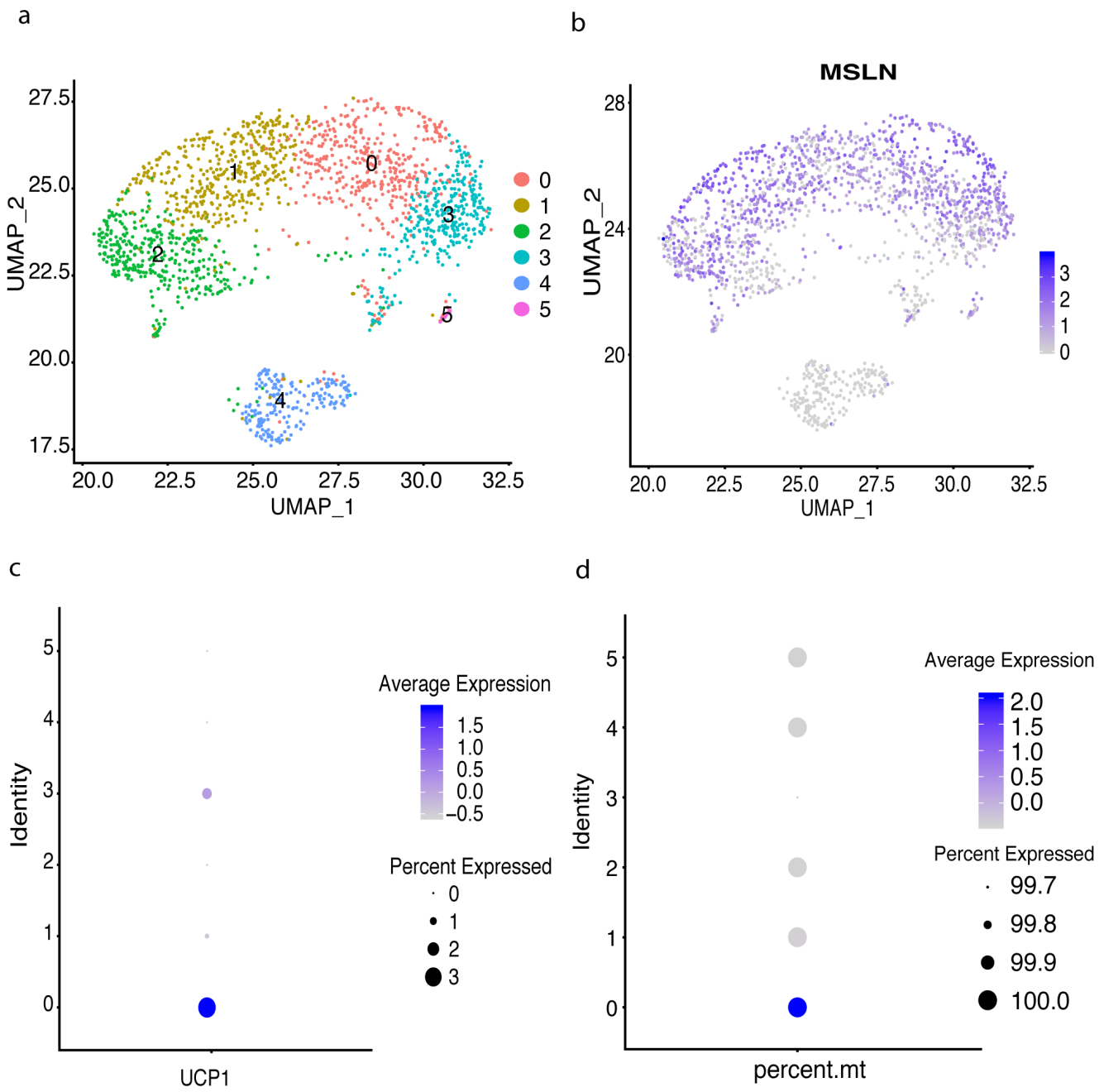
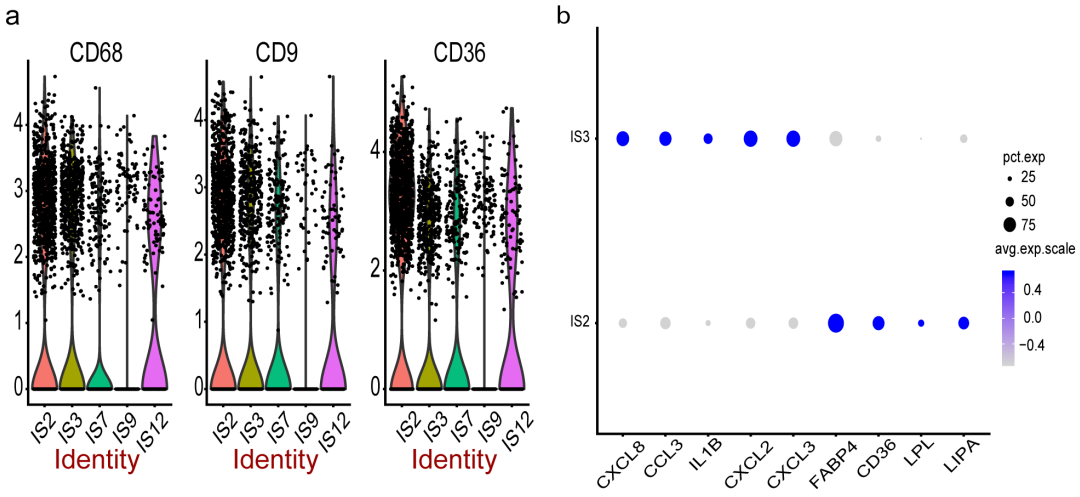
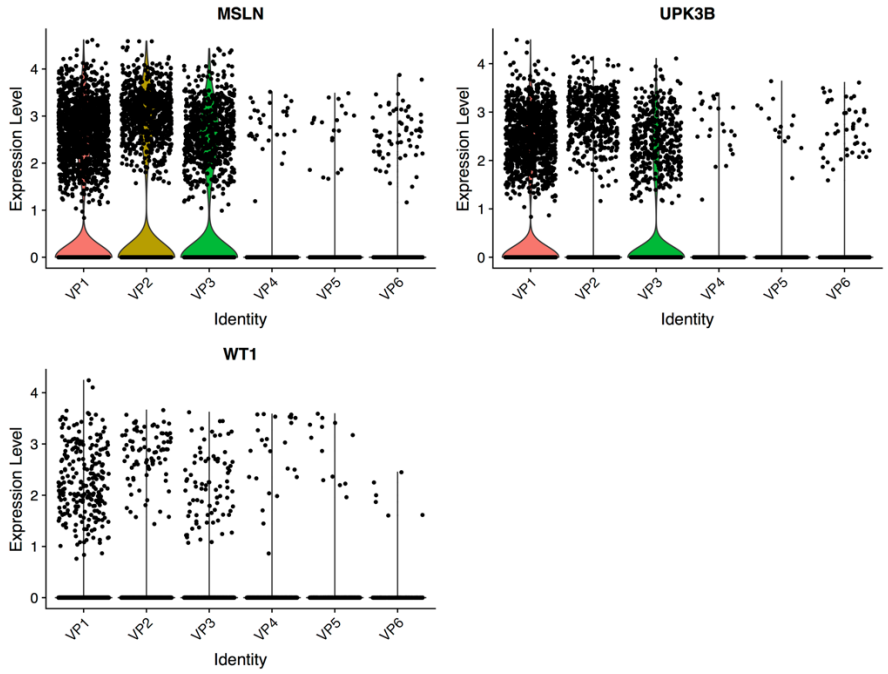



Figure 7: Progenitor clusters specific to VAT derived from a healthy individual. (a) Re-clustering of progenitors identified 6 clusters comprising 1,781 cells (b) Umap highlighting expression of *MSLN* in all progenitor clusters (c) *UCP1* expression was detected in cluster 0 which had high mitochondrial gene expression (d)

Extended Data




Extended Data 1: Multiple macrophage clusters were identified in SVF from both SAT and VAT depots (a) 4 distinct macrophage clusters showing varying expression of *CD68* (19 - 52% of cells), *CD9* (10 – 51% of cells) and *CD36* (25 – 72% of cells). The y axis of the violin plot indicate log transformed expression values and the width indicate number of cells expressing the particular gene. (b) Genes involves in lipid metabolism is found expressed in macrophage cluster – IS2, whereas IS3 is rich in inflammatory markers.




Extended Data 2: Gene expression of marker genes in 6 visceral specific progenitor clusters. The y axis of the violin plot indicate log transformed expression values and the width indicate number of cells expressing the particular gene.

# CHAPTER3: Adipose Tissue Transcriptome and Epigenome Characterization During Development and in Health and Disease

## 3.1 Bridging Statement between Chapter 2 and 3

As described in Chapter 2, we were able to successfully catalogue the cellular landscape in human AT providing the first of its kind single-cell atlas across two metabolic disease-relevant AT depots (SAT and VAT). We also provided significant insight into obesity-induced metabolic complications by contrasting cellular and transcriptomic profiles in obese individuals with or without T2D. However, the limitations of this study were the inclusion of only obese individuals which restricted the comparative analysis between healthy (lean) and obese phenotypes. In addition, the study also only included adult individuals leaving the pediatric population unstudied. To this end, in Chapter 3, we extended these efforts by characterizing adipose tissue derived from both healthy (lean) and extremely obese children. In addition to single-cell RNA sequencing we also studied the epigenome by open chromatin profiling using snATAC-Seq to identify epigenetic regions that regulates gene expression pattern identified across disease relevant cell population.

### 3.2 Title, Authors and Affiliations

High-resolution characterization of cellular imbalance and epigenome dysregulation of adipose tissue during human development and disease

Jinchu Vijay[1], Warren A Cheung[2], Guillaume Bourque[1], Jason D Fraser[3], Tomi Pastinen[1,2], Elin Grundberg[1,2*]

[1]Department of Human Genetics, McGill University, Montreal, Québec, Canada
[2]Genomic Medicine Center, Children's Mercy Research Institute, Children's Mercy Kansas City, MO, USA
[3]Department of Surgery, Children's Mercy Kansas City, MO, USA

[*]Correspondence should be addressed to E.G. (egrundberg@cmh.edu)

### 3.3 Abstract

Adipose tissue found in various locations of the human body is known to be implicated in development of obesity and associated metabolic disorders. We recently profiled adipose tissue derived from adults with obesity using single cell RNA sequencing (scRNA-Seq) and identified novel dysfunctional T cells and adipocyte progenitors associated with metabolic complications in a depot-specific manner. Here, we expand these efforts to the pediatric population as well as characterization across healthy and obese tissue including paired single nuclei open chromatin analysis (snATAC-Seq) combined with scRNA-Seq. We provide higher resolution characterization of endothelial and smooth muscle cells and show striking tissue and disease association of adipose resident immune cells. We observed a Natural Killer (NK) cell subset with anticancer gene expression signatures enriched in lean individuals, suggesting role of similar protective mechanism against obesity. We confirm the distinct relationships of dysfunctional and protective adipocyte progenitors with obesity: dysfunctional *CFD*-expressing progenitors are enriched in children with obesity (36%) compared to lean children (15.8%). On the other hand, *MSLN*-expressing cells are found more abundantly in adipose tissue from lean children (43%) compared to obese children (9.4%) and possess the signature of a brown-like 'protective' progenitor population. Further analysis using snATAC-Seq showed open chromatin in *MSLN* clusters being enriched for binding sites for TEAD-family of transcription factors – *TEAD2, TEAD4, TEAD3* and *TEAD1*. On the other hand, open chromatin defining *CFD* progenitors were enriched for key regulators of adipogenesis: *CEBPA* and *CEBPD*. Our expanded adipose cell atlas provides insight into cellular patterns and biological pathways linked to obesity across the lifespan.

### 3.4 Introduction

Adipose tissue (AT) is a specific connective tissue that are embryonically derived from the mesoderm and consist of extracellular matrix and several cell types of the adipocyte and non-adipocyte cellular fraction. The latter, known as the stromal vascular fraction (SVF), include mesenchymal stem cells, adipocyte progenitor cells, fibroblasts, endothelial cells and immune cells, each cell type with multiple sub-populations and distinct role in disease risk and progression. In mammals, two specific types of AT exist - white (WAT) and brown (BAT) based on microscopic characteristics of adipocytes where white adipocytes consist of a single large lipid droplet and possess only a few mitochondria, whereas brown adipocytes contain multiple lipid droplets per cell and are packed with mitochondria and have thermogenic properties[130]. It was initially believed that classical BAT was only relevant during infancy and restricted to interscapular depot but recent studies have suggested that brown adipocytes, and/or adipocytes possessing characteristics of both brown and white adipocytes (known as 'beige' or 'brite' adipocytes), may be more common in adults than had been previously appreciated[158,159]. We recently applied single-cell RNA sequencing (scRNA-Seq) in a cohort of AT derived from adult individuals with extreme obesity and identified a novel depot-specific preadipocyte population within adult SVF marked by *WT1* and *MSLN* (indicative of a mesothelial origin) with high mitochondria content and possessing browning features as well as with a protective function against obesity-induced metabolic complications such as Type 2 Diabetes (T2D)[101]. This progenitor subpopulation was shown to be exclusive to visceral depots and distinct from the most common progenitor in subcutaneous AT which was marked by high *CFD* expression. This unbiased and high-resolution analysis of SVF at single-cell level also allowed us to map the immune cellular landscape in adult AT identifying a multitude of common and latent immune cell

types including but not limited to $CD9^+$ macrophages and $CD8^+$ T cells associating with metabolic disease status. In all, this study contributed novel insight into AT architecture underlying obesity in the adulthood providing a resource at higher resolution than previous population-based genome-wide catalogues of adult AT established by us[43] or other large consortia[45,160]. However, there continue to be a sparse representation of AT derived from children leaving a large gap of developmental aspect of tissue architecture and childhood disease association unexplored. Indeed, childhood obesity continues to be a growing health problem globally with an estimation of almost 400 million children being obese worldwide[161]. Childhood obesity is also found to be strongly associated with prolonged obesity and metabolic complications in adulthood[162,163] and believed to be the cause of the increased incidence seen for T2D in children and adolescence[164]. In fact, obesity-induced T2D in children and adolescents has a worse prognosis and progresses more rapidly than adult-onset T2D or early-onset type 1 diabetes including greater mortality[165]. Although there have been an increasing number of genome-wide association studies focusing on childhood obesity[166] and T2D[167] to disentangle age-specific genetic factors underlying these conditions, the cellular landscape of human AT during development and across disease groups (i.e. lean vs obese children) has not been characterized at high-resolution to date.

To this end, we apply single-cell and single-nuclei approaches to assess the complex cellular landscape of intra-abdominal visceral adipose tissue (VAT) in both lean and extremely obese children during development contrasting to our recently released single-cell atlas of adult AT. We show that differential signatures of gene expression and open chromatin regions in response to obesity is distributed among distinct cell population of VAT.

## 3.5 Results

### 3.5.1 Cellular Characterizing of Adipose Tissue in Children

We collected 24 VAT samples in our discovery cohort derived from children (0 - 19 years) undergoing elective surgery including six with a healthy weight and 11 with extreme childhood obesity (average BMI of 49.4kg/$m^2$). A subset (N=8) of the childhood obesity sample group had biospecimens collected from two different visceral depots (see Methods) referred to as "up" and "down".

Focusing on the SVF, we applied a donor-multiplexed capture approach (see Methods) and performed scRNA-Seq on the 24 VAT samples obtaining 28,529 high-quality cells (average 1,141 cells/sample) after initial QC and filtering including only genetically unique singlets removing doublets and ambiguous calls. We identified 24 cell clusters across the main cell categories: stem cells/progenitors/preadipocytes (N=9), immune cells (N=10), endothelial cells (N=3) and smooth muscle cells (N=2) based on the expression of the marker genes – *CD34, PECAM1* and *PTPRC* (Figure 1A,1C). The overall distribution of these broad cell groups varied significantly based on disease state (i.e., healthy weight vs. extreme obese) and also based on specific VAT depot (Figure 1B). We noted increased resolution of the vascular endothelium (i.e., endothelial cells and smooth muscle cells) compared to our earlier scRNA-Seq efforts in AT derived from adults[101] as described in detail below.

### 3.5.2 Vascular Endothelium Dysfunction in Childhood Obesity

We identified four different clusters from the vascular endothelium which included two endothelial cell clusters (EK1 and EK2) and two smooth muscle cell clusters (SK1 and SK2). We also

identified a cluster of lymphatic endothelial cells (EK3), validating our observation of different endothelial cells in AT from adult samples. In children, endothelial clusters EK1 and EK2 were identified to be adipose resident microvascular endothelial cells and vascular endothelial cells respectively. Differential expression analysis between vascular endothelial clusters and smooth muscle cell clusters showed EK1 and EK2 have high expression of markers of angiogenesis including *TIE1*[168], *ADGRL4*[169], *KDR*[170], *ENG*[171] and *EGFL7*[172] (Adjusted P < 5.48E-302, Table 1). Smooth muscle cells were found to be marked by the expression of *NEXN*[173], *CNN1*[174], and *TAGLN*[175] . Interestingly, the SK1 and SK2 clusters also expressed *ADCY3*, a gene which loss of function is known to increase the risk of obesity and T2D[176]. Comparing the gene expression of SK1 and SK2 showed SK1 had high expression of collagen genes and insulin like growth factor binding proteins 2-7 (Table 2) which plays a key role in smooth muscle differentiation, migration, and DNA synthesis[177,178]. SK1 also showed an upregulation for genes involved in fatty acid handling such as *LPL, CD36* and *FABP4*[179]. Top genes also showed characteristic beneficial signatures which included *STEAP4* that acts against adipogenesis and insulin resistance in response to pro inflammatory stimuli[180], *MARCKS* involved in smooth muscle and endothelial cell proliferation[181], *TIMP3* regulates of homeostasis of endothelium[182]. Interestingly, SK2 showed expression of *MYH11* and *TRPV1* (Figure 2A) which are recently identified to be markers of smooth muscle cells potent to develop into beige adipocytes[96]. Further, pathway analysis using Metascape[183] identified enrichment for oxidative phosphorylation (q-value = 5.09E-14) and mitochondrial electron transport, ubiquinol to cytochrome c (q-value = 1.88E-12) among the down regulated genes (Adjusted P. <= 1E-4) in SK1 in comparison with SK2.

Among the vascular endothelial clusters, EK1 cells were found to be more prevalent in obese individuals compared to the lean counterparts which corresponds to 12.8% and 5.9% of their total population. Differential gene expression analysis between obese and lean of EK1 markers identified 394 significant genes (Adjusted P <= 1E-4). To validate the obesity-association of these EK1 marker genes we used the population-based MuTHER and METSIM AT resource each with ~700 AT samples profiled with corresponding individual-based phenotyping: 323 (82%) EK1 genes were associated (P <= 0.05) with obesity traits in either MuTHER (BMI, Trunk fat%) or METSIM (BMI, WHR) of which 152 genes (47%) were validated in both studies. Top differential expressed genes between obese and lean patients in EK1 showed higher expression of stress response genes - *NNMT*[184], *MT2A, MT1M, MT1X, MT1A* and *MT1E*[185] (Adjusted P <= 1.34E-23, Table 3) in obese individuals. Further, pathway analysis using Metascape on genes with positive fold change and adjusted P <= 1E-04 identified GO:0009636 – Response to toxic substances as the top enriched pathway with a q-value of 9.57E-08. Interestingly, genes involved in fatty acid handling *FABP4, RBP7, LPL, CD36* and *PPARG*[179] (Adjusted P <= 2.63E-5) were also upregulated in obese. Among the down regulated EK1 genes in obese individuals, we found an enrichment towards the key genes involved in notch signalling pathway (Adjusted P <= 8.61E-5, Figure 2B) including *DLL4, SMAD1, SMAD5, ADAM10, KDR, MAML2* and *TM4SF18*[186]. Pathway analysis on genes with negative fold change and adjusted P <= 1E-04 identified GO:0001568 - blood vessel development as the top pathway with q-value of 2.58E-18.

Comparing the proportion of cells from each cluster across the samples, we found that both EK1 and EK2 were positively correlated with smooth muscle cell clusters SK1 and SK2 (Pearson p <= 2.5E-05) in obese individuals. However, the pattern was not observed in individuals with a healthy

weight. Further, EK1 showed strong negative correlation with CD4+ T cell clusters IKS1 and monocytes IKS4 (Pearson p <= 1.76E-04, Cluster details provided in the following section) in obese patients but the data from healthy weight individuals showed opposite trend with IKS4 and no correlation with IKS1. EK2 also showed a similar pattern of correlation in obese individuals with a weaker association. We also found significant correlation between EK1 with CD8+ T cell cluster (IKS5) following a similar pattern as that of monocytes (Pearson p <= 0.014). In addition, cell to cell communication analysis using ligand receptor interaction showed link between EK1 and EK2 as well as between EK1 and CD8+ T cell cluster (IKS5).

### 3.5.3 Differential Accumulation of Immune Cells Based on Obesity Status

We assessed the distribution of AT-resident immune cells across our study cohorts and noted that adults have proportionally more immune cells (T2D -33%, nonT2D – 35%) compared to children (Obese – 27%, Lean – 13%) although the largest differences are attributed by metabolic health status as shown by low immune cell infiltration in lean individuals.

To increase the resolution of cell type assignment, we created a subset of 6,772 cells from clusters IK1 – IK10 which were identified as immune cells by unsupervised, reference-based clustering algorithm implemented in SingleR. We identified 17 new immune clusters (IKS1 – IKS17, Figure 3A) using variable genes from the subset. As expected, we found two main groups of clusters forming T/NK cells (IKS1, IKS2, IKS3, IKS5, IKS9, IKS10, IKS11, IKS12) and Monocytes/ Macrophages (IKS4, IKS7, IKS13, IKS16). We noted that obese samples showed an enrichment of T/NK cells comprising 69% of its total immune cell population compared to the healthy weight (lean) samples which only had 52% of the immune cells being T/NK cells. The reverse pattern was

observed for Monocyte / Macrophage clusters – 31% of immune cells in health weight and 18% in obese samples, respectively, with IKS4 and IKS7 being the main contributor of the variation. We identified IKS4 as monocytes and IKS7 as lipid metabolizing M2 macrophages confirming our findings in adults (Figure 3B).

Among the immune subclusters, the variation between lean and obese were more pronounced in T and NK cell clusters. Cells from obese individuals formed the majority population in clusters IKS1, IKS2, IKS9 and IKS11. On the other hand, IKS5 and IKS12 were from lean samples. IKS1 showed high expression for genes coding ribosomal proteins. IKS1 and IKS2 were annotated to be CD4+ T cells with the expression of *MAL* in IKS1 - a marker of activated T cell[187]. IKS10 was found to be rich in metallothionein genes – *MT1X, MT1G, MT1F, MT1E* and *MT2A* (Figure 3C) and it constituted of only cells from obese samples. This further validates our observation of the development of metallothionein rich T cell clusters during obesity in our adult study. IKS9 and IKS11 were identified to be NK cells with expression of cytotoxic NK cell markers on IKS9[188]. Among the lean specific clusters, IKS5 showed expression of *CCL5* and *CD8B*, the markers of CD8+ T cells. IKS12 expressed NK cell markers *KLRC1 KLRD1, NKG7* and chemokines *XCL1*, *XCL2* (Figure 3D) which are shown to have beneficial effects in cancer immunotherapy[189].

3.5.4 Characterization of Diverse Adipose Progenitors

We identified nine distinct clusters of progenitors expressing either *CFD* (CK1-CK5) or *MSLN* (MK1- MK4). In line with our previous study using adult participants, we found that MK clusters are found in healthy individuals (43%) compared to obese (9.4%) and possess the signature of a brown-like progenitor population. On the other hand, CK clusters were found to be comprise of

15.8% and 36% of total cell population in healthy and obese children, respectively. Interestingly, we also noted depot specific variation of VAT in cell proportion of MK clusters. MK cells were sparse in “up” samples (greater omentum but proximal to stomach) with a total cell proportion of only 3%, whereas in “down” samples (greater omentum but distal to stomach) it was around 18%. We expanded these analyses by including snATAC-Seq which confirmed the cell distribution. Specifically, snATAC-Seq identified the MK cluster to be corresponding to 43% of cells from healthy children, 4% from ‘up” and 17% from “down". To further disentangle the regulatory machinery underlying the specific CK and MK cell lineages we identified the open chromatin regions specific to the snATAC-Seq clusters (Figure 4A). First, we performed differential peak calling between corresponding chromatin regions in CK vs. MK clusters and identified 1,369 regions (log2FC >= 0.05 and Adjusted P <= 0.05) that were specific to CK clusters. Transcription factor binding motif (TFBM) enrichment analysis on these CK specific open chromatin regions showed enrichment for transcription factors ZNF384, TWIST1, ZBTB18, CEBPA, EBF3, CEBPD, ATF2, MEF2A, TAL1::TCF3, ZBTB26, EBF1. On the other hand, 786 regions/peaks were identified to be specific to the *MSLN* expressing clusters, the MK cells. Using GREAT[190] resource, we identified 129 proximal genes that are situated +/- 5kb from these peaks. We found *KRT18, KRT19, KRT8, MSLN* and *TM4SF1* among the genes identified by GREAT which further validates our observation of specificity of these genes in adult *MSLN* specific clusters. We also observed the presence of the mesothelial markers[1] *LRRN4* and *UPK3B* among them. TFBM analysis on the MK-specific peaks identified transcription factors that are enriched in *MSLN* clusters which include MAZ, ZNF148, TEAD2, ZNF263, KLF5, TEAD4, GATA3, TEAD1, TEAD3, NFIB. Interestingly, TEAD1 and TEAD2 are negatively associated with BMI (FDR <= 1.41E-03) in METSIM. We also observed a significant peak region upstream of *WT1* gene specific

to the MK cluster (chr11-32466080-32466828, Adjusted P = 2.57E-201) compared to CK clusters (Figure 4B). Further analysis showed that the peak region indeed has the binding site for TEAD1 suggesting key role of the TEAD transcription factors in regulating the VAT-specific MK cells in children and adults.

We then subclustered the 13,889 progenitor cells from scRNA-Seq to obtain higher resolution and to overcome inter sample variations, we refined the clusters using Conos[71], which resulted in seven new clusters (Figure 5A)- 2 MSLN expressing clusters (MKS1, MKS2) and 5 *CFD* rich clusters (CKS1 – CKS5; Figure 5B). The distribution of MKS1 cells were similar as the main clustering with higher proportion in lean, followed by obese "down" sample and almost absent in obese "up" samples. However, this correlation of cell proportion with metabolic health status was not observed for the other MSLN sub-cluster, MKS2, but only for depot-specificity. Specifically, MKS2 cells were comprised of 15% and 16% of total progenitor population from lean, and obese "down" samples but only 4% of total progenitor population from obese "up" samples indicating a depot-specific cellular subpopulation. Differential gene expression analysis between MKS1 and MKS2 showed increased expression of *COL6A3* and its transcriptional regulator *PRRX1* (Adjusted. P <= 9.92E-173) which are known to be overexpressed during obesity[191].In addition, MKS2 is also marked by genes involved in adipogenesis such as *COL14A1*[192], *DCN*[193], *DPT*[194] and *SERPINF1*[195] (Adjusted. P <= 1.19E-250).

Focusing on the sub-clusters for the *CFD* cells, we noted that CKS1 was marked by genes involved in adipogenesis including *RASD1*[196], *PLIN2, CEBPB, CEBPD* and *MYC*[197]. CKS1 also showed cluster specific expression for *AKRIC1* and *AKR1C2*, which were identified to be a marker of

abdominal obesity[198,199]. CKS2 showed expression of *FOS, JUNB* and *FOSB* indicating they are induced pre-adipocytes[200]. Further, selective expression of *PPARG* in CKS1, CKS2 and CKS4 confirm they are committed preadipocytes at different stages of differentiation. CKS3 is identified to be multipotent progenitors expressing *DPP4, WNT2, CD34* and *CD24*[97,201] (Figure 5C). CKS5 showed expression of secreted frizzled related proteins (SFRP) – *SFRP4*, *SFRP2* and *SFRP1* (Adjusted P <= 3.71E-133). Expression of SFRPs and *MFAP5* suggests CKS5 might have a regulatory role in adipogenesis and inflammation[127,202,203]. Interestingly, CKS5 is enriched in obese patients, particularly in "Up" samples with a proportion of 13.4% of its total progenitor population. Whereas it comprised of 6.6% and 2.7% in "Down" and "lean" progenitors. We then did differential gene expression analysis between CSK5 and rest of the CSK clusters (CSK1 – CSK4). The analysis identified 144 genes with an adjusted P <= 1E-4 and positive log2FC. These 144 genes were compared with differential gene expression analysis results between obese and lean from bulk RNA-Seq of isolated adipocytes. Interestingly, we could not find similar pattern of disease specificity matured adipocytes suggesting the pattern is observed during early adipocyte differentiation stages. Our theory was further confirmed using bulk RNA-Seq data from adult patients from Chapter 2. Among the 144 genes, 69 genes were upregulated in SVF in comparison with matured adipocytes (Adjusted P <= 1E-4, positive log2FC) and 16 were found to have increased expression in adipocytes.

### 3.6 Discussion

We have performed a comparative study between cellular composition of VAT derived from individuals with healthy weight and obesity, leveraging gene expression and open chromatin profiling (as an indication of active regulatory regions) at single-cell resolution. We observed

interesting dynamics in vascular endothelium of VAT. Healthy adipose tissue expansion is accompanied by the expansion of endothelial cells to facilitate microcirculation and thereby maintaining tissue homeostasis[204]. Endothelial cells also contribute to the metabolism of excess free fatty acids during adipose tissue expansion [205]. Studies have shown that fatty acid handling by endothelial cells is facilitated by the cross talk with preadipocytes through *PPARG* activation[110,179]. However, obesity induced hypoxia can lead to endothelial dysfunction and tissue inflammation[206,207]. Our analysis using lean and obese individuals show that the obesity related endothelial variation is pronounced in adipose specific vascular endothelium - EK1 with increased expression of oxidative stress response genes in obese patients. On the other hand, genes involved in notch signalling pathway that negatively regulates endothelial differentiation are down regulated signalling active endothelial expansion in response to adipose tissue expansion. *DLL4* is one of the notable down regulated genes among them which is shown to negatively inhibit angiogenesis by modulating the expression of vascular endothelial growth factor[208]. Inhibition of *DLL4* in mouse models resulted in improved fat accumulation, insulin resistance and atherosclerosis with detectable attenuation of inflammation through polarization of macrophages to proinflammatory M1 phenotype[209]. Endothelial cells are known to remain quiescent until they are mediated by angiogenic signatures from adipocytes. Down regulation of Notch signalling and upregulation of fatty acid handling genes including *FABP4* (Adjusted. P = 2.54E-55) suggests active proliferation of EK1 in obese compared to lean individuals in response to obesity[210,211]. The increase in expression of fatty acid handling genes is also seen in a cluster of smooth muscle cells (SK1) further indicates the expansion of endothelium[212]. Endothelial signals for expansion together with low abundance of macrophage population[213] in children suggests healthy endothelium and immune function in obese individuals.

The role of endothelial cells in adipose tissue inflammation is studied extensively[206]. We were able to identify the cross talk between sub clusters of endothelial cells and CD8+ T cells. Recent single cell studies on mice model shows CD8+ T cells can modulate vascular regeneration[214] . On the other hand, CD4+ T cells can induce apoptosis of endothelial cells[215] and they are found to negatively correlated in obese patients. In all, we observe the pattern favoring expansion of endothelium. Among the other immune cells, we observed *XCL1* and *XCL2* expressing NK cells which are studied in cancer immunotherapy leveraging its anticancer role by recruiting dendritic cells[189]. Presence of such NK cells (IKS12), predominantly in lean samples suggests the chances of similar immune response in obesity. Though, the dendritic cell population in our data set (IKS16) also shows concentration of cells from lean samples, the cluster does not have enough power (N=65) to explore further.

We confirmed the presence of two different progenitor population and the association with obesity. By characterising open chromatin regions specific to each cell cluster, we expanded our adipose tissue catalogue with potent regulatory regions. *MSLN* expressing cells are found to be enriched for TEAD family of transcription factors. TEAD transcription factors are shown to be involved in adipogenesis of murine preadipocytes through hippo signalling pathway[216]. Further, *TEAD4* is found to play critical role in development of adipocytes from multipotent muscle derived stem cells[217]. Taken together, TEAD transcription factors could be the key regulators of alternate adipogenic pathway of development of *MSLN* expressing adipocytes from mesothelium.

In conclusion, we have performed a comprehensive comparative study of adipose resident cell population between lean and obese children identifying dynamics in gene expression and open chromatin regions specific to multiple cell population.

### 3.7 Methods

Study Subjects:

Participants were selected through an institutionally approved study with ongoing recruitment at Children's Mercy Kansas City for pediatric patients undergoing hernia repair or bariatric sleeve surgery. Consent was obtained through the management of the study protocol approved by the Institutional Review Board at Children's Mercy Kansas City (Protocol # 17110653)

For consented patients undergoing bariatric surgery, adipose tissue was collected from two different fat compartments of the greater omentum with one proximal and one distal to the stomach referred as "up" and "down" in this article. For consented patients undergoing hernia repair, an easily accessible portion of the greater omentum at the lower end of the omental drape was identified and a portion was pulled out and a 2cm segment was amputed to allow for areas of cautery changes to be discarded. All specimens were passed off the field in a jar containing saline.

Adipose Tissue Preparation

The adipose tissue samples were weighed before being transferred to a petri dish and rinsed with PBS. Each sample was then minced with a scalpel into small pieces, cutting until there was no discernible fragments and transferred to a 50 mL conical tube and digestion buffer was added at 1mL/0.25g of tissue. The digestion buffer consisted of Krebs-Ringers-HEPES+ 2.5mM glucose+

2% FBS +200uM adenosine+ 1mg/mL collagenase type 1A. Krebs-Ringers-HEPES contained 120mM NaCl, 4.7mM KCl, 2.2mM CaCl2, 10mM HEPES, 1.2mM KH2PO4, and 1.2mM MgSO4. The tissue in digestion buffer was then incubated at 37ºC for one hour at 100 rpm on a heated orbital shaker. After digestion, the samples were filtered through a 150um mesh Celltrics filter (Fisher, Cat. No. NC9021438) into a 15mL tube, followed by an equal volume of HBSS+ 2% FBS +200uM adenosine. The suspensions were then centrifuged at 200g for 8 minutes. The adipocytes were transferred into a new 15 mL conical tube using a 1mL pipette tip with 1-2mm snipped off the end of the tip to allow gentler transfer of the adipocytes. The cell pellet was resuspended in 1mL of Hank's Balanced Salt Solution (Fisher, Cat. No. MT21022CV) + 2% FBS +200uM adenosine. Both the adipocytes and the SVF were washed in HBSS+ 2% FBS +200uM adenosine and centrifuged at 200g for 8 minutes. The adipocytes were then transferred to another tube and saved for further processing. The media was removed from the SVF pellet, and the pellet was resuspended in ACK Lysis Buffer (Fisher, Cat. No. A1049201). The SVF suspension was incubated for 5 minutes at room temperature and then centrifuged at 300g for 8 minutes. After centrifugation, the buffer was removed from the pellet, the SVF was washed again in HBSS+ 2% FBS +200uM adenosine and centrifuged again at 300g for 8 minutes. The media was then removed and the SVF pellet was resuspended in 0.5mL of HBSS+ 2% FBS +200uM adenosine and counted on a Countess II automated cell counter to determine the total cell count, viability, and live/dead size. After counting, the SVF was centrifuged at 300g for 8 minutes, resuspended in Recovery Cell Culture Freezing Media (Fisher, Cat. No. 12-648-010) and transferred into a 2mL cryovial. The cryovial was placed in a cell cooler and placed at -80ºC overnight.

Cell pooling

Two pools totaling 25 SVF samples were made. Thawing Medium for the SVF samples consisted of DMEM/F-12 (Thermo Fisher Cat No. 11320033) supplemented with 10% heat-inactivated FBS (GE Healthcare Cat No. SH30088.03HI), 100 units/mL of penicillin, and 100 µg/mL of streptomycin (Thermo Fisher Cat No. 15140122). For each sample to be thawed, 10 mL of Thawing Medium was prewarmed in a 37°C bead bath. Cells were thawed in groups of up to five samples at a time. The cryovials were placed in a 37°C bead bath. Once thawed, the cryovials and prewarmed 15-mL conical tubes of Thawing Medium were aseptically transferred into a biosafety cabinet. For each sample, 1 mL of Thawing Medium was added, dropwise, to the cell suspension. The cell suspension was mixed by gently pipetting and then diluted in the remaining 9 mL of Thawing Medium. After thawing and dilution, the cells were left at room temperature while the remaining cells required for pooling were thawed in a similar fashion. Once all samples were thawed, the samples were centrifuged at 300 x g for 6 min for SVF and 300 x g for 8 min for PBMCs. The supernatant was carefully aspirated, and the cell pellets were resuspended in 0.5 mL of room-temperature Thawing Medium. All samples were then placed on ice to be pooled. Up to 7.5 × 105 cells from each sample were pooled together. The pool was then passed through a 40-µm nylon mesh cell strainer to remove cell aggregates. The pool was centrifuged at 300 x g for 6 min for SVF and 300 x g for 8 min for PBMCs at 4°C, and the supernatant was carefully aspirated. The cell pellet was resuspended in 1 mL of cold Thawing Medium and the cell count and viability were assessed using a Countess II automated cell counter. No fewer than three aliquots per pool were cryopreserved by centrifuging at 300 x g for 6 min for SVF and 300 x g for 8 min for PBMCs at 4°C and resuspending the cell pellets in Recovery Cell Culture Freezing Medium. The cell

suspensions were transferred to cryogenic storage vials and were slow-frozen overnight to a temperature of -80°C in a Corning CoolCell FTS30.

Single-cell sequencing

Aliquots of the two SVF pools were thawed and serially diluted in DMEM/F12 supplemented with 10% fetal bovine serum and 1% penicillin-streptomycin in 15ml conical tube to a volume of 14ml. They were centrifuged at 300 rcf for 10 minutes and the supernatant was removed and discarded. The pelleted cells were suspended in an adequate amount of DMEM/F12 supplemented with 10% fetal bovine serum and 1% penicillin-streptomycin and transferred to a 1.5ml microfuge tube for counting. 10 µl of the suspension was used for a cell count and viability assay on a Countess II Automated Cell Counter (Invitrogen) to determine the volume of suspension to be used in the 10x Chromium Single Cell Library protocol. For scRNA-Seq, 10x Genomics Gene Expression v3 was used according to the manufacturer's protocol to target approximately 12,500 SVF cells per capture with a total of two captures per pool. For scATAC-Seq, 10x Genomics Single Cell ATAC Solution v1.0 was used according to the manufacturer's protocol to target approximately 10,000 nuclei per capture with a total of four captures per pool. Nuclei isolation was carried out according to the 10x Genomics Demonstrated Protocol: Nuclei Isolation for Single Cell ATAC Sequencing, using the optional DNase Treatment in the appendix for the SVF pools. The final libraries were analyzed using a TapeStation to determine library size and a Qubit Broad Range dsDNA assay to determine library concentration. Libraries were sequenced using an Illumina NovaSeq6000 sequencer using 28x8x94 and 52x8x16x52 cycle paired-end for scRNA-Seq and snATAC-Seq libraries, respectively.

QC and clustering of single-cell data of scRNA-Seq data

The primary processing of single cell libraries from scRNA-Seq were done using Cell Ranger[58] from 10X Genomics using GRCh38 as reference genome. The sequenced samples were aggregated using cellranger aggr pipeline where read depth difference between libraries was normalized by subsampling higher depth libraries. The samples were then separated using Demuxlet[65] using genetic deconvolution. Seurat[218] was used for further analysis of clustering, dimensionality reduction and differential gene expression analysis. For filtering, we used only cells with a minimum gene count of 200 and a maximum of 8,000 genes/cell based on their distribution in the sample. We also removed cells with unique molecular identifiers (UMIs) greater than 75,000 per cell. Cells with percentage of mitochondrial gene expression greater than 50 was also removed. After filtering we obtained 28,529 cells. We regressed out the variability due to UMI distribution, mitochondrial gene expression and the difference between G2M and S phase scores based on the gene expression of cell cycle gene. Top 75 dimensions were used to generate final clusters using principal component analysis (PCA) and graph-based clustering. Top 100 genes in each cluster were identified in comparison with all other cells using the function – 'FindAllMarkers' in Seurat. Clusters were annotated using SingleR[98] and literature mining. Differential gene expression analysis between cell population was done using Wilcoxon rank sum test in 'FindMarkers' function. The subset for the specific cell types were created from the clustering results using 'SubsetData' function of Seurat. Conos[71] was used to remove inter sample variations among progenitor subclusters.

QC and clustering of snATAC-Seq

The primary analysis of snATC-Seq was done using Cell Ranger ATAC[85] and then deconvoluted using Demuxlet. The secondary analysis was done using Signac[86]. For filtering low quality cells, we used the following criteria: Total number of fragments in peak region greater than 2,000 and less than 20,000, percentage of reads in peaks greater than 15, ratio of reads in blacklisted regions identified by ENCODE less than 0.05, nucleosome signal less than 4 and TSS enrichment greater than 2. After QC and filtering 15,601 cells remained. MACS2[219] was then used to call peaks from snATAC-Seq data. The peaks are then normalised using frequency-inverse document frequency (TF-IDF) normalization. The top variable features were then selected using FindTopFeatures function of Signac and using them dimension reduction was achieved through singular value decomposition. UMAP was then used for nonlinear dimension reduction and followed by graph-based clustering. The peaks within clusters were then assigned to coding genes using GeneActivity function in Signac which in turn was used for annotating the clusters using scRNA-Seq cluster identities through Canonical Correlation Analysis (CCA) implemented in Seurat[220]. The overrepresented peaks in cluster of interest were identified using FindMarkers function. Motif representation among the peaks were then calculated using chromVAR[221].

Bulk RNA-Seq of adipocytes:

In total of 0.5 to 3 million cells were re-suspended in 500 uL TRIzol Reagent and total RNA was extracted using the miRNeasy Mini Kit (Qiagen) according to the manufacturer's protocol. RNA library preparations were carried out on 500 ng of RNA with RNA integrity number (RIN)>7 using the Illumina TruSeq Stranded Total RNA Sample preparation kit, according to manufacturer's protocol. Final libraries were analyzed on a Bioanalyzer and sequenced on the NovaSeq6000

RNA-Seq Data Analysis:

Raw reads were trimmed for quality (phred33 ≥30) and length (n≥32), and Illumina adapters were clipped off using Trimmomatic v.0.35[154]. Filtered reads were then aligned to the GRCh38 human reference using STAR v.2.5.3a[155]. Raw read counts of genes were obtained using htseq-count v.0.6.0[156]. Differential gene expression analysis was done using DeSeq2 v.1.18.1[157].

### 3.8 Acknowledgements

This work was supported by a Canadian Institutes of Health Research (CIHR) team grant awarded to E.G (EGM141898) and the CIHR funded Epigenome Mapping Centre at McGill University (EP1-120608) awarded to T.P. E.G. holds the Roberta D. Harding & William F. Bradley, Jr. Endowed Chair in Genomic Research and T.P. holds the Dee Lyons/Missouri Endowed Chair in Pediatric Genomic Medicine. We thank Rebecca Biswell, Daniel Louiselle and Margaret Gibson at the Genomic Medicine Center at Children's Mercy Research Institute for technical assistance.

### 3.9 Additional Information

Author Contributions

E.G conceived the study. E.G and T.P designed experiments. J.D.F provided the clinical samples. W.A.C, B. Y and G.B provided bioinformatics support. J.V. analyzed the data, interpreted the results, and drafted the manuscript with major inputs from E.G.

Competing interests

The authors declare no competing interests.

### 3.10 Main Tables and Figures

Tables

Table 1: Marker genes identified in differential expression analysis between vascular endothelial cells (EK1, EK2) and smooth muscle cells (SK1, SK2). The genes were identified using Wilcoxon rank sum test implemented in Seurat.

| | | | | | |
|---|---|---|---|---|---|
| **Gene Name** | **P value** | **avg_log2FC** | **Percentage of cells in cluster expressing the gene** | **Percentage of cells in all other clusters expressing the gene** | **Adj. P value** |
| *TIE1* | 0 | 1.5165445 | 0.555 | 0.008 | 0 |
| *PIK3R3* | 0 | 1.40791587 | 0.474 | 0.035 | 0 |
| *NEXN* | 0 | -1.2966169 | 0.035 | 0.455 | 0 |
| *ADGRL4* | 0 | 2.62373681 | 0.783 | 0.025 | 0 |
| *ADCY3* | 0 | -1.3233146 | 0.053 | 0.495 | 0 |
| *KDR* | 0 | 1.40585025 | 0.459 | 0.008 | 0 |
| *ENG* | 0 | 1.81255118 | 0.669 | 0.136 | 0 |
| *EGFL7* | 0 | 2.70219042 | 0.759 | 0.033 | 0 |
| *TAGLN* | 0 | -4.9109743 | 0.251 | 0.941 | 0 |
| *CNN1* | 0 | -1.2867982 | 0.006 | 0.363 | 0 |

Table 2: Marker genes identified in differential expression analysis between smooth muscle cell clusters SK1 and SK2. The genes were identified using Wilcoxon rank sum test implemented in Seurat.

| Gene Name | P value | avg_log2FC | Percentage of cells in cluster expressing the gene | Percentage of cells in all other clusters expressing the gene | Adj. P value |
|---|---|---|---|---|---|
| *COL6A3* | 5.141E-204 | 1.92582687 | 0.64 | 0.086 | 1.291E-199 |
| *STEAP4* | 8.606E-204 | 1.97812276 | 0.824 | 0.275 | 2.161E-199 |
| *MARCKS* | 1.053E-180 | 1.66176045 | 0.74 | 0.256 | 2.644E-176 |
| *CD36* | 1.941E-164 | 2.16740349 | 0.73 | 0.25 | 4.875E-160 |
| *IGFBP2* | 2.458E-131 | 1.96970589 | 0.561 | 0.141 | 6.173E-127 |
| *COL3A1* | 3.676E-103 | 1.89029266 | 0.729 | 0.481 | 9.231E-99 |
| *IGFBP3* | 6.9324E-95 | 1.42641775 | 0.399 | 0.073 | 1.7408E-90 |
| *COL1A2* | 1.1267E-88 | 1.33921631 | 0.783 | 0.623 | 2.8293E-84 |
| *IGFBP7* | 4.7156E-72 | 0.53312316 | 0.988 | 0.996 | 1.1841E-67 |
| *IGFBP4* | 2.3538E-70 | 0.86892075 | 0.798 | 0.648 | 5.9106E-66 |
| *COL25A1* | 1.0199E-58 | 0.88824006 | 0.364 | 0.115 | 2.5611E-54 |
| *IGFBP5* | 4.7251E-56 | 0.96289185 | 0.895 | 0.97 | 1.1865E-51 |
| *IGFBP6* | 9.805E-52 | 1.03674476 | 0.539 | 0.334 | 2.4621E-47 |
| *COL4A1* | 6.698E-48 | 1.06703237 | 0.621 | 0.457 | 1.6819E-43 |
| *FABP4* | 1.877E-44 | 1.56698366 | 0.878 | 0.838 | 4.7133E-40 |
| *COL5A3* | 4.6845E-44 | 0.87194862 | 0.377 | 0.165 | 1.1763E-39 |
| *COL1A1* | 6.8833E-43 | 1.72635132 | 0.648 | 0.573 | 1.7285E-38 |
| *COL6A2* | 9.7033E-42 | 0.70397306 | 0.847 | 0.857 | 2.4366E-37 |
| *COL6A1* | 1.044E-37 | 0.76713222 | 0.735 | 0.712 | 2.6216E-33 |
| *COLEC11* | 2.7555E-35 | 0.73688326 | 0.194 | 0.043 | 6.9193E-31 |
| *LPL* | 9.4169E-24 | 0.60281939 | 0.15 | 0.039 | 2.3647E-19 |
| *COL5A1* | 1.7158E-21 | 0.56700767 | 0.264 | 0.134 | 4.3085E-17 |
| *COL4A2* | 2.2603E-19 | 0.64410821 | 0.597 | 0.552 | 5.6758E-15 |
| *COL12A1* | 1.1254E-13 | 0.55828489 | 0.283 | 0.187 | 2.8259E-09 |
| *COL21A1* | 2.7442E-13 | 0.41867115 | 0.193 | 0.103 | 6.891E-09 |
| *COL14A1* | 2.1839E-08 | 0.4867296 | 0.594 | 0.651 | 0.0005484 |
| *COL5A2* | 3.3963E-06 | 0.42483834 | 0.321 | 0.276 | 0.08528499 |

Table 3: Marker genes identified in differential expression analysis between cells from healthy weight and obese in EK1. The genes were identified using Wilcoxon rank sum test implemented in Seurat.

| Gene Name | p_val | avg_log2FC | Percentage of cells in cluster expressing the gene | Percentage of cells in all other clusters expressing the gene | Adj. P value |
|---|---|---|---|---|---|
| *FABP4* | 1.0097E-59 | 1.18480781 | 0.96 | 0.805 | 2.5354E-55 |
| *NNMT* | 1.2719E-51 | 1.62173999 | 0.65 | 0.288 | 3.1938E-47 |
| *MT2A* | 2.1711E-48 | 1.81472053 | 0.915 | 0.863 | 5.4518E-44 |
| *MT1M* | 6.4812E-45 | 2.56694069 | 0.603 | 0.298 | 1.6275E-40 |
| *MT1X* | 8.2603E-42 | 2.4899569 | 0.683 | 0.469 | 2.0743E-37 |
| *MT1A* | 3.4485E-35 | 2.86753148 | 0.512 | 0.245 | 8.6594E-31 |
| *RBP7* | 5.0403E-28 | 0.7685901 | 0.83 | 0.695 | 1.2657E-23 |
| *MT1E* | 5.3207E-28 | 1.845696 | 0.606 | 0.404 | 1.3361E-23 |
| *LPL* | 3.3849E-27 | 1.61420539 | 0.419 | 0.156 | 8.4999E-23 |
| *SMAD1* | 6.4518E-22 | -0.5904665 | 0.202 | 0.404 | 1.6201E-17 |
| *CD36* | 7.1059E-20 | 0.5379689 | 0.898 | 0.784 | 1.7844E-15 |
| *KDR* | 6.6562E-18 | -0.6605249 | 0.418 | 0.606 | 1.6714E-13 |
| *TM4SF18* | 2.5935E-15 | -0.5462767 | 0.395 | 0.565 | 6.5126E-11 |
| *MAML2* | 1.7136E-14 | -0.3460279 | 0.134 | 0.279 | 4.3029E-10 |
| *DLL4* | 2.5548E-13 | -0.4616037 | 0.237 | 0.406 | 6.4154E-09 |
| *ADAM10* | 2.4519E-10 | -0.3019825 | 0.24 | 0.385 | 6.1569E-06 |
| *PPARG* | 1.0465E-09 | 0.65391322 | 0.423 | 0.3 | 2.6279E-05 |
| *SMAD5* | 3.4303E-09 | -0.3252063 | 0.271 | 0.411 | 8.6139E-05 |

Figures

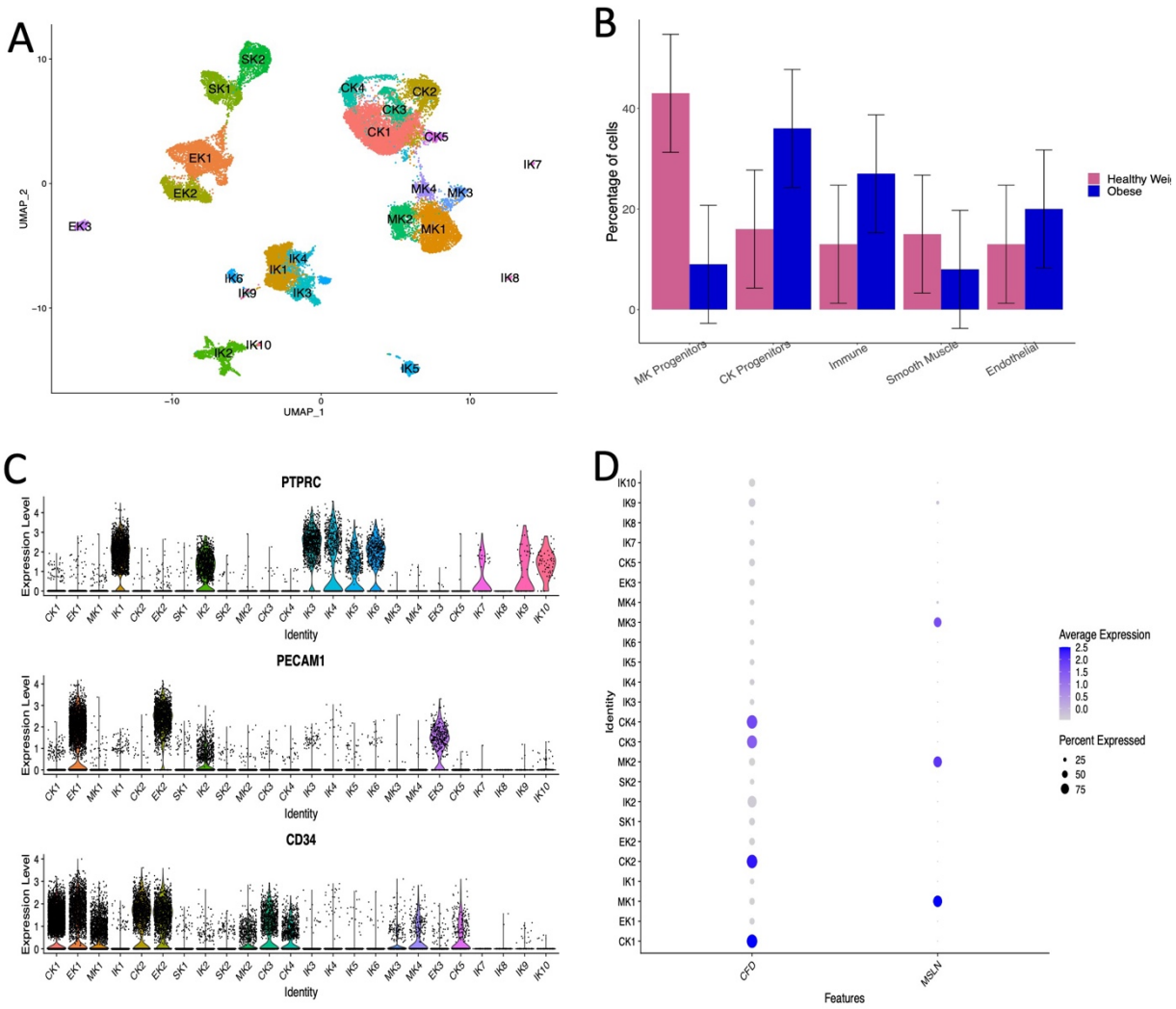



Figure 1: Classification of cell population in SVF. Clustering of scRNA-Seq resulted in 24 clusters (A) which are broadly classified as MSLN (MK) and CFD (CK) expressing progenitors, Immune cells (IK), Smooth muscle cells (SK) and Endothelial cells (EK) based on the marker genes (C and D). The cell population differed in tissue concentration depending on disease status (B)

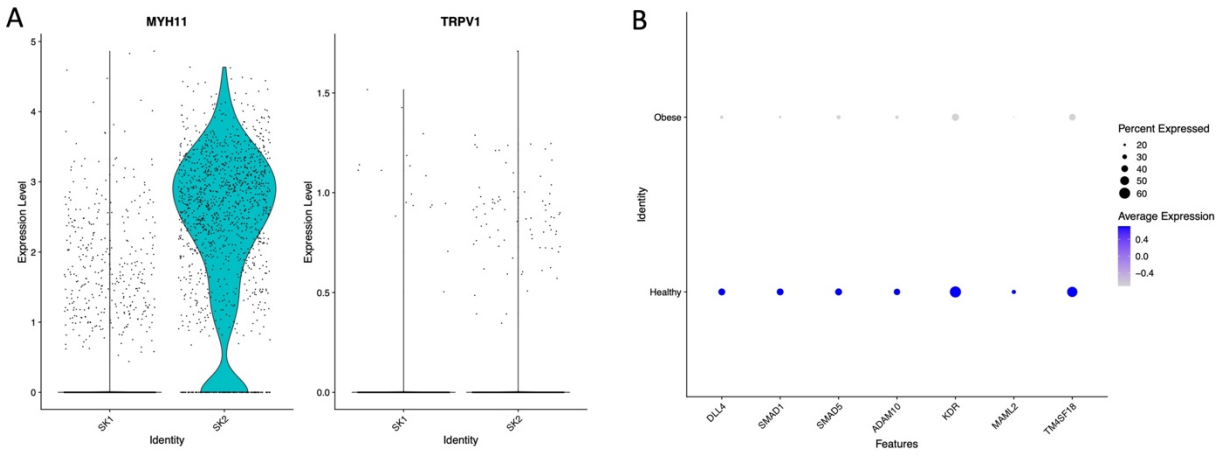



Figure 2: Cluster specific markers of vascular endothelium. (A)Smooth muscle cell cluster SK2 shows expression of markers of brown adipogenesis. (B) Endothelial Cell cluster EK1 shows down regulation of genes involved in notch signaling pathway in obese.

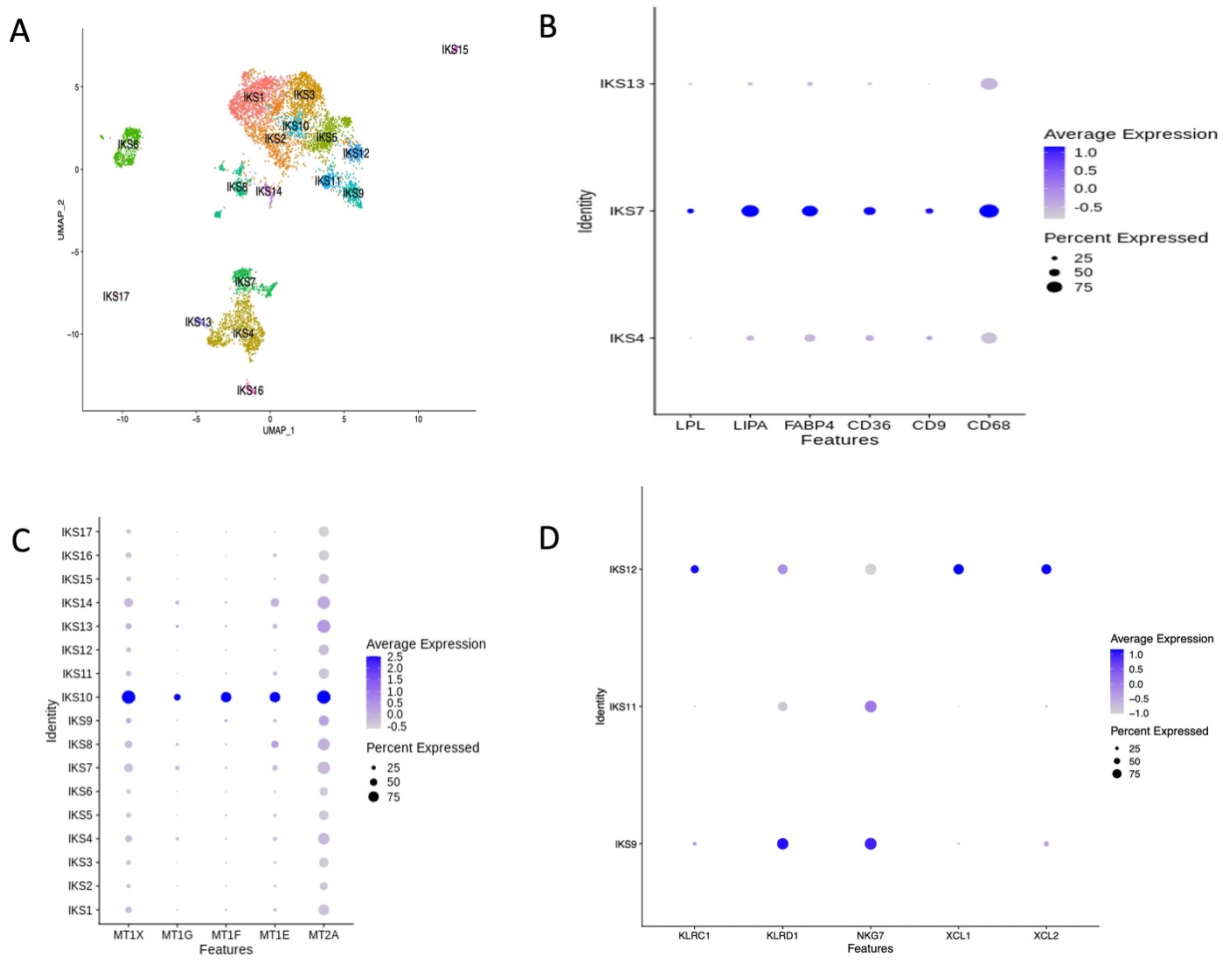



Figure 3: Adipose resident immune cells. The sub-clustering of immune cells identified 17 distinct clusters (A) which included lipid handling macrophages -IKS7 (B), Metallothionein rich T cells – IKS10 (C) and NK cells with anticancer properties (D).

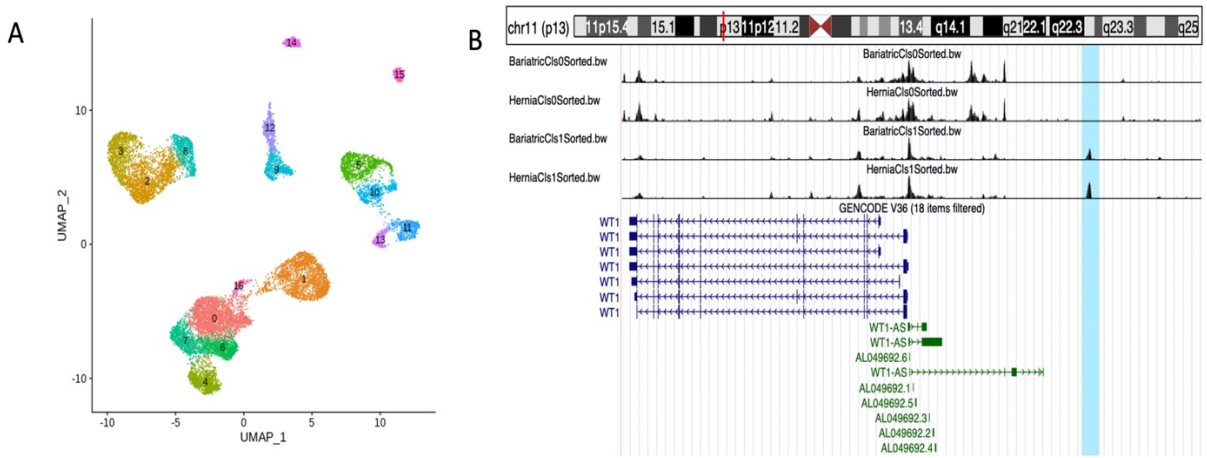



Figure 4: Clustering using snATAC-Seq. (A) Cell clustering using data from snATAC-Seq identified 17 clusters with cluster 1 annotated as *MSLN* expressing progenitors and clusters 0,4,6 and 7 as *CFD* expressing progenitors. Differential peak calling between *MSLN* and *CFD* expressing clusters identified a regulatory region specific to cluster 1 which is found to be in the upstream of promoter of *WT1* gene (B).

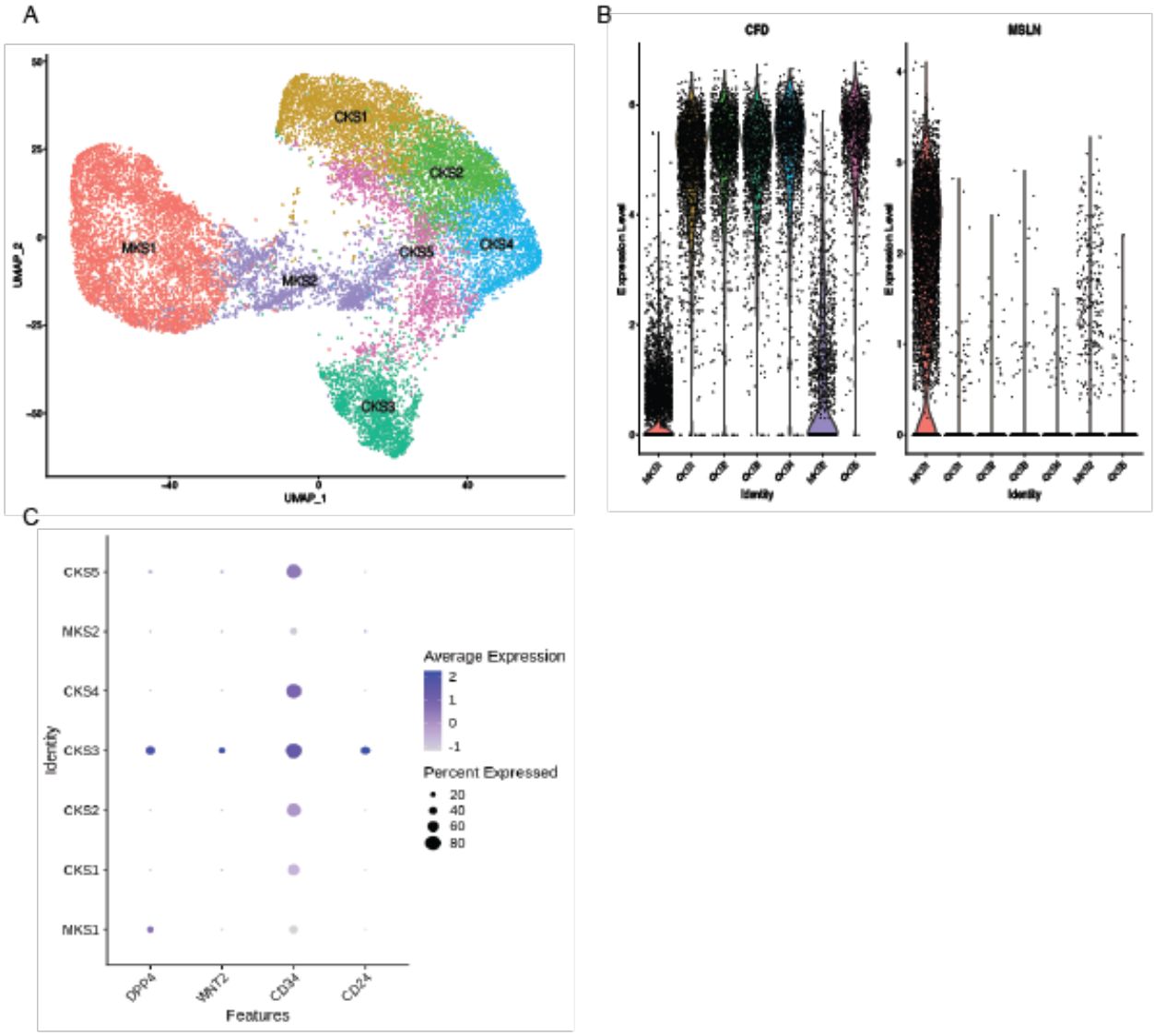



Figure 5: Subpopulation among progenitors using scRNA-Seq. 5 distinct clusters of progenitors were identified (A) with 2 *MSLN* expressing clusters and 5 *CFD* expressing clusters. The CK clusters were found to be in different stages of differentiation.

# CHAPTER 4: General Discussion

Development of single cell technologies revolutionized genomic understanding of biological samples. It helped to rediscover significant transcriptomic and epigenetic signatures which are otherwise diluted by the signals from the whole tissue. Through the two research chapters included in this thesis we attempted to catalogue and characterize different cell populations in adipose tissue across two different adipose tissue depots of the human body with a specific focus on the metabolic disease relevant intra-abdominal visceral compartment, VAT. We also attempted to understand variation among these cell population in response to disease phenotypes such as obesity and T2D. Our observations in these chapters that are linked to obesity can be broadly classified as alteration in three processes – adipocyte expansion, fatty acid handling and immune cell dysfunction. Our results suggest coordinated functional changes of multiple cell types that are involved in the development of obesity and associated metabolic disorders.

## 4.1. Adipocyte Expansion in Obesity

Adipocyte expansion is known to be play a significant role in obesity and hence numerous studies have focused on profiling adipocytes. Expansion of adipocyte volume is achieved either by increasing the adipocyte size by lipid accumulation or by recruiting new adipocytes[16,222]. In either case, excess adipocyte expansion leads to hypoxia and adipocyte dysfunction[223]. Efforts to identify genetic, epigenetic, transcriptomic and metabolomic signatures contributing to adipocyte dysfunction using bulk sequencing approaches found numerous markers associated with the disease status. However, imputing them to the cell type or stages of the cell differentiation is crucial to understand the biological relevance and to assess the therapeutic potential. Conventional

functional studies require considerable time and resources to attain this goal. Alternatively, single cell sequencing approach helped to solve this issue to a certain extent. For instance, studies have shown the presence of two distinct kinds of adipocytes in VAT[224] and a subset of VAT adipocyte is developed from mesothelium – a different origin than the remaining SAT adipocytes. However apart from a few well studied genes such as *WT1, MSLN* and *UPK3B*, the progenitor subsets were not characterized in detail[23]. In addition, focus on VAT association in development of metabolic disorders also make characterization of these subsets crucial.

Using scRNA-Seq we characterized the two subpopulations of VAT adipocytes in adults and children. The progenitor subsets are marked by either *CFD* or *MSLN* corresponding to their embryonic origin. snATAC-Seq identified distinct open chromatin regions and transcription factor motifs for corresponding *CFD* and *MSLN* expressing clusters further confirming the two subsets have distinct regulatory machineries. Our data suggests *MSLN* expressing adipocyte progenitors show protective function and the population of these progenitors tend to decrease with development of obesity early in development as shown by the trend across children and adults. We also found that SAT contains only *CFD* expressing progenitors and their population expands in VAT during obesity. Further, one of the *CFD* expressing cluster showed strong association to T2D status in adults, supporting our hypothesis that these progenitors contribute to pathogenic abdominal obesity. Accumulation of fat in SAT is considered as a metabolically healthy obesity[225]. However, further studies would be required to understand how the similar progenitors in two different depots have opposite pathological fate.

Acquiring intra-abdominal VAT samples requires invasive surgical procedures which makes it often not feasible to obtain across health and disease as well as across the lifespan. One of the main advantages of this study is the inclusion of human adipose tissue samples from multiple depots, SAT and VAT. We were able to acquire intra-abdominal VAT from both adults and children through collaboration with bariatric surgical teams performing procedures where tissue can be accessed without any harm to the patients and delivered to the laboratory in timely manner. In addition, we were also able to successfully access and profile tissue from healthy individuals through the collaborations with general surgeons. However, the availability of biospecimen depends on the specific surgical procedure. For instance, samples from lean and obese children for our study was obtained from patients undergoing hernia repair or bariatric surgery. Differences in technical procedure for sample procurement led to challenges in joint data analysis showing clear batch effects. Recent developments in computational methods to mitigate those variations are carefully considered and applied successfully to our dataset[71,88,226]. However, we were unable to identify regulatory adipocyte signatures identified by Schwalie et. al[94]. We suspect it might be due to difference in adipose tissue biology between human and rodents. Adipose tissue in rodents is distributed to distinct anatomical regions different from that of human adipose tissue depots which makes knowledge translation challenging[227]. Additionally, in Chapter 3, we tried to study VAT from two different regions of intra-abdominal area from obese patients. We found differences in concentration of cell population, gene expression and open chromatin regions of progenitor cell clusters between the two regions of same fat depot. Our results emphasize the importance of appropriate disease related tissue for fine resolution analysis using single cells.

### 4.2. Coordinated fatty acid handling

Adipose tissue is well known for fatty acid storage and handling. VAT in particular gained interest because of high lipid turn over, direct access to liver through the portal vein and its association with metabolic syndromes[228-231]. Our data identified population of a macrophage and endothelial cells are involved in fatty acid handling. These cell types also exhibited preferential accumulation and function depending on the obesity phenotype with increased concentration in healthy samples. Involvement of cell types other than adipocytes indicated a coordinated network in maintaining energy homeostasis in the tissue. In-depth sequencing enabled us to characterize three different endothelial population in adipose tissue. Endothelial dysfunction in adipose tissue is known to cause metabolic disorders and are investigated for its therapeutic potential[204]. We show that the adipose resident microvascular endothelial cells are the subset implicated by obesity and metabolic disorders. Fatty acid handling by endothelial cells is managed by the cross talk with adipocytes and impairing angiogenesis by obesity induced hypoxia leads to dysfunction of endothelial cells[232]. We and others[233] also identified presence of a group of lipid-associated macrophage which has a M2 phenotype. Inclusion of samples from healthy children helped us to successfully identify gene expression and open chromatin regions that differs in these cell clusters in response to obesity. We expect this rich catalogue of genomic signatures will help to understand the mechanisms leading to development of dysfunctional cell types.

### 4.3. Dysfunctional Immune cells

Immune cells showed an interesting pattern in our study. One of the CD8+ T cell cluster was identified to be dysfunctional with notable cell concentration in obese patients than healthy individuals. The cells are marked by expression of metallothionein genes *MT1F, MT1X, MT2A,*

*MT1E* and *MT1G*. T cell exhaustion or loss of function is seen during chronic infection or invasive cancer[234]. Singer et. al. profiled the gene expression markers in these dysfunctional T cells and identified metallothionein gene[137]. Factors affecting T cell dysfunction during obesity is reviewed in detail by Aguilar et. al.[235] According to them, T cell exhaustion is promoted by a variety of factors which include increased levels of adipokines, altered levels of metabolites, continuous T cell activation and macrophage polarization signals. As a result, T cells show decreased proliferation, cytokine expression and reduced proportion of naïve and regulatory T cells[235]. In addition, T cell mediated autoimmune activity linked to T2D have been documented previously[236]. In all dysfunctional T cells appears to be worth exploring to ascertain the therapeutic potential against systemic inflammation in obesity. Further, the importance of metallothionein genes in obesity go beyond the scope of T cell function. Metallothionein genes are involved in stress response and studies have shown that they are expressed on adipocytes during obesity[237]. Our results from chapter 3 also show that metallothionein gene expression is elevated in adipose resident micro vascular endothelial cells from obese patients. It is clear metallothionein genes are induced in multiple cell types during obesity. However, further studies are required to understand their adipose specific function in obesity and direction of effect in pathogenesis.

### 4.4. Knowledge and Data Sharing

In Chapter 2 we created the first of its kind adipose tissue cell atlas with comprehensive characterization of different cell population in SAT and VAT of adult obese individuals. We aimed to provide a single cell reference data set for the broader research community. These results were published in Nature Metabolism in December of 2019 and has already been cited by 28 other articles with more than 10,000 article access. In addition, we have submitted all the raw and

processed data in Gene Expression Omnibus (GEO) for unrestricted access. Visualization is a crucial step to understand underlying patterns in the data. UCSC (University of California, Santa Cruz) recently launched a web browser for single cell data visualization similar to the well adapted UCSC genome browser[238]. The data from Chapter 2 is in the process of integration to UCSC single cell browser (https://cells.ucsc.edu/). The dataset is also used as a reference resource for other in-house and collaborative projects. One of such projects was aimed to characterize *AKR1C2* which is involved in androgen metabolism. Expression of *AKR1C2* is found to be associated with body fat distribution. Using single cell data reference, we identified that the *AKR1C2* expression was contributed by *CFD* expressing progenitors.[198]

### 4.5. Challenges in single cell sequencing of adipocytes

Though single cell technologies offer an abode of genomic information about the samples studied, it has its inherent challenges. One of the main issues we encounter when we study the adipose tissue is the cell size limitation from 10X genomics. Current compatibility of 10x droplet based system is less than 30 μm (https://kb.10xgenomics.com/hc/en-us/articles/218170543-What-is-the-range-of-compatible-cell-sizes-). Reports indicate adipocytes may range from 20 to 300 μm in diameter[239] which makes profiling of adipocytes challenging. Consequently, we focused on SVF at single cell resolution and adipocytes are characterized using bulk sequencing approaches. Further, recent developments in high throughput single nuclei RNA-Seq offers an alternative method for adipocyte profiling but limiting the RNA profile to nuclear transcripts[240,241].

# CHAPTER 5: Conclusions and Future Directions

We have performed a comprehensive epigenetic and transcriptomic analysis of adipose tissue - dissecting its features to different cellular levels and tissue localizations. Our single cell dataset forms the first attempt to catalog depot-and obesity-specific adipose cellular population in the first-of-its kind "adipose tissue cell atlas across development and disease". Integrating transcriptomic and epigenetic data have allowed us to uncover gene expression signatures and molecular mechanisms regulating adipose tissue and obesity development.

Our results show involvement of multiple cell types that are either induced or repressed functionally during obesity. Correlation in cell proportions and cell to cell communications show presence of cross talk between different cells. A comprehensive network analysis would be required to understand further about the interactions between cell types. Our current data comprise of snATAC-Seq and scRNA-Seq captured independently from each other. snATAC-Seq offers high granularity in cluster specification and provides insight into regulatory regions in the genome at single cell level. Even so, annotating snATAC-Seq data using predicted gene expression pattern is challenging especially when it comes to subpopulation of cell types. The future direction for this issue would be applying a "multiome" approach with joint analysis of RNA and open chromatin region from the same cells. Access to the regulatory regions will help to ascertain gene regulatory network that coordinates development of different obesity associated cell types. However, functional validation would be necessary to fully elucidate the role of key genomic signatures.

The research projects described in two chapters were focused on different stages of life either adult samples or from children using a cross-sectional study design. Longitudinal assessment over different time periods would be required to understand the impact of inherited or acquired cellular populations and transcriptomic signatures' role in the development of obesity later in life. The continuation of enrollment of lean patients undergoing hernia repair early in life as described in Chapter 3 would allow for such assessment and to identify genomic signatures that are present in early childhood which facilitate obesity in later stages and to identify acquired obesity related changes during development.

## References


1 WHO. Obesity and overweight. Geneva: World Health Organization (2016).

2 WHO. Levels and trends in child malnutrition. Geneva: World Health Organization (2020).

3 Swinburn, B. A. et al. The global obesity pandemic: shaped by global drivers and local environments. Lancet 378, 804-814, doi:10.1016/S0140-6736(11)60813-1 (2011).

4 Conway, B. & Rene, A. Obesity as a disease: no lightweight matter. Obes Rev 5, 145-151, doi:10.1111/j.1467-789X.2004.00144.x (2004).

5 Alberti, K. G. & Zimmet, P. Z. Definition, diagnosis and classification of diabetes mellitus and its complications. Part 1: diagnosis and classification of diabetes mellitus provisional report of a WHO consultation. Diabet Med 15, 539-553, doi:10.1002/(SICI)1096-9136(199807)15:7<539::AID-DIA668>3.0.CO;2-S (1998).

6 Eckel, R. H., Grundy, S. M. & Zimmet, P. Z. The metabolic syndrome. Lancet 365, 1415-1428, doi:10.1016/S0140-6736(05)66378-7 (2005).

7 Vague, J. Sexual differentiation; Factor determining forms of obesity. Presse Med 55, 339 (1947).

8 Lapidus, L. et al. Distribution of adipose tissue and risk of cardiovascular disease and death: a 12 year follow up of participants in the population study of women in Gothenburg, Sweden. Br Med J (Clin Res Ed) 289, 1257-1261, doi:10.1136/bmj.289.6454.1257 (1984).

9 Ibrahim, M. M. Subcutaneous and visceral adipose tissue: structural and functional differences. Obes Rev 11, 11-18, doi:10.1111/j.1467-789X.2009.00623.x (2010).

10 Matsuzawa, Y. et al. Pathophysiology and pathogenesis of visceral fat obesity. Obes Res 3 Suppl 2, 187S-194S, doi:10.1002/j.1550-8528.1995.tb00462.x (1995).

11 Tchernof, A. & Despres, J. P. Pathophysiology of human visceral obesity: an update. Physiol Rev 93, 359-404, doi:10.1152/physrev.00033.2011 (2013).

12 Arner, P., Lithell, H., Wahrenberg, H. & Bronnegard, M. Expression of lipoprotein lipase in different human subcutaneous adipose tissue regions. J Lipid Res 32, 423-429 (1991).

13 Ramirez, M. E. et al. Evidence for sex steroid inhibition of lipoprotein lipase in men: comparison of abdominal and femoral adipose tissue. Metabolism 46, 179-185, doi:10.1016/s0026-0495(97)90299-7 (1997).

14 Palmer, B. F. & Clegg, D. J. The sexual dimorphism of obesity. Mol Cell Endocrinol 402, 113-119, doi:10.1016/j.mce.2014.11.029 (2015).

15 Mauvais-Jarvis, F. Sex differences in metabolic homeostasis, diabetes, and obesity. Biol Sex Differ 6, 14, doi:10.1186/s13293-015-0033-y (2015).

16 Spalding, K. L. et al. Dynamics of fat cell turnover in humans. Nature 453, 783-787, doi:10.1038/nature06902 (2008).

17 Hirsch, J. & Han, P. W. Cellularity of rat adipose tissue: effects of growth, starvation, and obesity. J Lipid Res 10, 77-82 (1969).

18 Laforest, S., Labrecque, J., Michaud, A., Cianflone, K. & Tchernof, A. Adipocyte size as a determinant of metabolic disease and adipose tissue dysfunction. Crit Rev Clin Lab Sci 52, 301-313, doi:10.3109/10408363.2015.1041582 (2015).

19 Lefterova, M. I. & Lazar, M. A. New developments in adipogenesis. Trends Endocrinol Metab 20, 107-114, doi:10.1016/j.tem.2008.11.005 (2009).

20 Brey, C. W., Nelder, M. P., Hailemariam, T., Gaugler, R. & Hashmi, S. Kruppel-like family of transcription factors: an emerging new frontier in fat biology. Int J Biol Sci 5, 622-636, doi:10.7150/ijbs.5.622 (2009).

21 Kirstein, A. S. et al. PTEN regulates adipose progenitor cell growth, differentiation, and replicative aging. J Biol Chem, 100968, doi:10.1016/j.jbc.2021.100968 (2021).

22 Chen, K. et al. HMGB2 orchestrates mitotic clonal expansion by binding to the promoter of C/EBPbeta to facilitate adipogenesis. Cell Death Dis 12, 666, doi:10.1038/s41419-021-03959-3 (2021).

23 Chau, Y. Y. et al. Visceral and subcutaneous fat have different origins and evidence supports a mesothelial source. Nat Cell Biol 16, 367-375, doi:10.1038/ncb2922 (2014).

24 Chau, Y. Y. & Hastie, N. Wt1, the mesothelium and the origins and heterogeneity of visceral fat progenitors. Adipocyte 4, 217-221, doi:10.4161/21623945.2014.985009 (2015).

25 Chondronikola, M. et al. Brown adipose tissue improves whole-body glucose homeostasis and insulin sensitivity in humans. Diabetes 63, 4089-4099, doi:10.2337/db14-0746 (2014).

26 Nicholls, D. G. & Locke, R. M. Thermogenic mechanisms in brown fat. Physiol Rev 64, 1-64, doi:10.1152/physrev.1984.64.1.1 (1984).

27 Nedergaard, J. et al. UCP1: the only protein able to mediate adaptive non-shivering thermogenesis and metabolic inefficiency. Biochim Biophys Acta 1504, 82-106, doi:10.1016/s0005-2728(00)00247-4 (2001).

28 van Marken Lichtenbelt, W. D. et al. Cold-activated brown adipose tissue in healthy men. N Engl J Med 360, 1500-1508, doi:10.1056/NEJMoa0808718 (2009).

29 Vitali, A. et al. The adipose organ of obesity-prone C57BL/6J mice is composed of mixed white and brown adipocytes. J Lipid Res 53, 619-629, doi:10.1194/jlr.M018846 (2012).

30 Harms, M. & Seale, P. Brown and beige fat: development, function and therapeutic potential. Nat Med 19, 1252-1263, doi:10.1038/nm.3361 (2013).

31 Xu, H. et al. Chronic inflammation in fat plays a crucial role in the development of obesity-related insulin resistance. J Clin Invest 112, 1821-1830, doi:10.1172/JCI19451 (2003).

32 Grant, R. W. & Dixit, V. D. Adipose tissue as an immunological organ. Obesity (Silver Spring) 23, 512-518, doi:10.1002/oby.21003 (2015).

33 Lumeng, C. N., Bodzin, J. L. & Saltiel, A. R. Obesity induces a phenotypic switch in adipose tissue macrophage polarization. J Clin Invest 117, 175-184, doi:10.1172/JCI29881 (2007).

34 Cipolletta, D. et al. PPAR-gamma is a major driver of the accumulation and phenotype of adipose tissue Treg cells. Nature 486, 549-553, doi:10.1038/nature11132 (2012).

35 Yang, J. et al. Genetic variance estimation with imputed variants finds negligible missing heritability for human height and body mass index. Nat Genet 47, 1114-1120, doi:10.1038/ng.3390 (2015).

36 Locke, A. E. et al. Genetic studies of body mass index yield new insights for obesity biology. Nature 518, 197-206, doi:10.1038/nature14177 (2015).

37 Shungin, D. et al. New genetic loci link adipose and insulin biology to body fat distribution. Nature 518, 187-196, doi:10.1038/nature14132 (2015).

38 Goodarzi, M. O. Genetics of obesity: what genetic association studies have taught us about the biology of obesity and its complications. Lancet Diabetes Endocrinol 6, 223-236, doi:10.1016/S2213-8587(17)30200-0 (2018).

39 Yeo, G. S. H. Genetics of obesity: can an old dog teach us new tricks? Diabetologia 60, 778-783, doi:10.1007/s00125-016-4187-x (2017).

40 Pigeyre, M., Yazdi, F. T., Kaur, Y. & Meyre, D. Recent progress in genetics, epigenetics and metagenomics unveils the pathophysiology of human obesity. Clin Sci (Lond) 130, 943-986, doi:10.1042/CS20160136 (2016).

41 Mendelson, M. M. et al. Association of Body Mass Index with DNA Methylation and Gene Expression in Blood Cells and Relations to Cardiometabolic Disease: A Mendelian Randomization Approach. PLoS Med 14, e1002215, doi:10.1371/journal.pmed.1002215 (2017).

42 Allum, F. et al. Characterization of functional methylomes by next-generation capture sequencing identifies novel disease-associated variants. Nat Commun 6, 7211, doi:10.1038/ncomms8211 (2015).

43 Grundberg, E. et al. Mapping cis- and trans-regulatory effects across multiple tissues in twins. Nat Genet 44, 1084-1089, doi:10.1038/ng.2394 (2012).

44 Grundberg, E. et al. Global analysis of DNA methylation variation in adipose tissue from twins reveals links to disease-associated variants in distal regulatory elements. Am J Hum Genet 93, 876-890, doi:10.1016/j.ajhg.2013.10.004 (2013).

45 Laakso, M. et al. The Metabolic Syndrome in Men study: a resource for studies of metabolic and cardiovascular diseases. J Lipid Res 58, 481-493, doi:10.1194/jlr.O072629 (2017).

46 Forsberg, E. C., Serwold, T., Kogan, S., Weissman, I. L. & Passegue, E. New evidence supporting megakaryocyte-erythrocyte potential of flk2/flt3+ multipotent hematopoietic progenitors. Cell 126, 415-426, doi:10.1016/j.cell.2006.06.037 (2006).

47 Kvist, T., Ahring, B. K., Lasken, R. S. & Westermann, P. Specific single-cell isolation and genomic amplification of uncultured microorganisms. Appl Microbiol Biotechnol 74, 926-935, doi:10.1007/s00253-006-0725-7 (2007).

48 Tang, F. et al. 220-plex microRNA expression profile of a single cell. Nat Protoc 1, 1154-1159, doi:10.1038/nprot.2006.161 (2006).

49 Tang, F. et al. mRNA-Seq whole-transcriptome analysis of a single cell. Nat Methods 6, 377-382, doi:10.1038/nmeth.1315 (2009).

50 Sitnicka, E. et al. Key role of flt3 ligand in regulation of the common lymphoid progenitor but not in maintenance of the hematopoietic stem cell pool. Immunity 17, 463-472, doi:10.1016/s1074-7613(02)00419-3 (2002).

51 Wen, C. Y. et al. Quick-response magnetic nanospheres for rapid, efficient capture and sensitive detection of circulating tumor cells. ACS Nano 8, 941-949, doi:10.1021/nn405744f (2014).

52 Enam, S. et al. Association of human polyomavirus JCV with colon cancer: evidence for interaction of viral T-antigen and beta-catenin. Cancer Res 62, 7093-7101 (2002).

53 Yasen, A. et al. Progress and applications of single-cell sequencing techniques. Infect Genet Evol 80, 104198, doi:10.1016/j.meegid.2020.104198 (2020).

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

55 Matula, K., Rivello, F. & Huck, W. T. S. Single-Cell Analysis Using Droplet Microfluidics. Adv Biosyst 4, e1900188, doi:10.1002/adbi.201900188 (2020).

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

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

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

59 D, C. R. Chromium Next GEM Single Cell 3′ Reagent Kits v3.1. 10x Genomics (November 2019).

60 Keren-Shaul, H. et al. MARS-seq2.0: an experimental and analytical pipeline for indexed sorting combined with single-cell RNA sequencing. Nat Protoc 14, 1841-1862, doi:10.1038/s41596-019-0164-4 (2019).

61 Natarajan, K. N. et al. Comparative analysis of sequencing technologies for single-cell transcriptomics. Genome Biol 20, 70, doi:10.1186/s13059-019-1676-5 (2019).

62 Ilicic, T. et al. Classification of low quality cells from single-cell RNA-seq data. Genome Biol 17, 29, doi:10.1186/s13059-016-0888-1 (2016).

63 Xi, N. M. & Li, J. J. Benchmarking Computational Doublet-Detection Methods for Single-Cell RNA Sequencing Data. Cell Syst 12, 176-194 e176, doi:10.1016/j.cels.2020.11.008 (2021).

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

65 Kang, H. M. et al. Multiplexed droplet single-cell RNA-sequencing using natural genetic variation. Nat Biotechnol 36, 89-94, doi:10.1038/nbt.4042 (2018).

66 Lahnemann, D. et al. Eleven grand challenges in single-cell data science. Genome Biol 21, 31, doi:10.1186/s13059-020-1926-6 (2020).

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

68 Bacher, R. Normalization for Single-Cell RNA-Seq Data Analysis. Methods Mol Biol 1935, 11-23, doi:10.1007/978-1-4939-9057-3_2 (2019).

69 Lun, A. T., Bach, K. & Marioni, J. C. Pooling across cells to normalize single-cell RNA sequencing data with many zero counts. Genome Biol 17, 75, doi:10.1186/s13059-016-0947-7 (2016).

70 Butler, A., Hoffman, P., Smibert, P., Papalexi, E. & Satija, R. Integrating single-cell transcriptomic data across different conditions, technologies, and species. Nat Biotechnol 36, 411-420, doi:10.1038/nbt.4096 (2018).

71 Barkas, N. et al. Joint analysis of heterogeneous single-cell RNA-seq dataset collections. Nat Methods 16, 695-698, doi:10.1038/s41592-019-0466-z (2019).

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

73 McInnes, L., Healy, J. & Melville, J. in arXiv e-prints arXiv:1802.03426 (2018).

74 Slovin, S. et al. Single-Cell RNA Sequencing Analysis: A Step-by-Step Overview. Methods Mol Biol 2284, 343-365, doi:10.1007/978-1-0716-1307-8_19 (2021).

75 Andrews, T. S., Kiselev, V. Y., McCarthy, D. & Hemberg, M. Tutorial: guidelines for the computational analysis of single-cell RNA sequencing data. Nat Protoc 16, 1-9, doi:10.1038/s41596-020-00409-w (2021).

76 Luecken, M. D. & Theis, F. J. Current best practices in single-cell RNA-seq analysis: a tutorial. Mol Syst Biol 15, e8746, doi:10.15252/msb.20188746 (2019).

77 Davey, C. A., Sargent, D. F., Luger, K., Maeder, A. W. & Richmond, T. J. Solvent mediated interactions in the structure of the nucleosome core particle at 1.9 a resolution. J Mol Biol 319, 1097-1113, doi:10.1016/S0022-2836(02)00386-8 (2002).

78 Luger, K., Mader, A. W., Richmond, R. K., Sargent, D. F. & Richmond, T. J. Crystal structure of the nucleosome core particle at 2.8 A resolution. Nature 389, 251-260, doi:10.1038/38444 (1997).

79 Dorigo, B. et al. Nucleosome arrays reveal the two-start organization of the chromatin fiber. Science 306, 1571-1573, doi:10.1126/science.1103124 (2004).

80 Marino-Ramirez, L., Kann, M. G., Shoemaker, B. A. & Landsman, D. Histone structure and nucleosome stability. Expert Rev Proteomics 2, 719-729, doi:10.1586/14789450.2.5.719 (2005).

81 Buenrostro, J. D., Giresi, P. G., Zaba, L. C., Chang, H. Y. & Greenleaf, W. J. Transposition of native chromatin for fast and sensitive epigenomic profiling of open chromatin, DNA-binding proteins and nucleosome position. Nat Methods 10, 1213-1218, doi:10.1038/nmeth.2688 (2013).

82 Cusanovich, D. A. et al. Multiplex single cell profiling of chromatin accessibility by combinatorial cellular indexing. Science 348, 910-914, doi:10.1126/science.aab1601 (2015).

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

84 Rondini, E. A., Ramseyer, V. D., Burl, R. B., Pique-Regi, R. & Granneman, J. G. Single cell functional genomics reveals plasticity of subcutaneous white adipose tissue (WAT) during early postnatal development. Mol Metab, 101307, doi:10.1016/j.molmet.2021.101307 (2021).

85 Satpathy, A. T. et al. Massively parallel single-cell chromatin landscapes of human immune cell development and intratumoral T cell exhaustion. Nat Biotechnol 37, 925-936, doi:10.1038/s41587-019-0206-z (2019).

86 Tim Stuart, A. S., Caleb Lareau, Rahul Satija. Multimodal single-cell chromatin analysis with Signac. bioRxiv, doi:10.1101/2020.11.09.373613 (2020).

87 Amemiya, H. M., Kundaje, A. & Boyle, A. P. The ENCODE Blacklist: Identification of Problematic Regions of the Genome. Sci Rep 9, 9354, doi:10.1038/s41598-019-45839-z (2019).

88 Korsunsky, I. et al. Fast, sensitive and accurate integration of single-cell data with Harmony. Nat Methods 16, 1289-1296, doi:10.1038/s41592-019-0619-0 (2019).

89 Chen, H. et al. Assessment of computational methods for the analysis of single-cell ATAC-seq data. Genome Biol 20, 241, doi:10.1186/s13059-019-1854-5 (2019).

90 Mathelier, A. et al. JASPAR 2016: a major expansion and update of the open-access database of transcription factor binding profiles. Nucleic Acids Res 44, D110-115, doi:10.1093/nar/gkv1176 (2016).

91 Heinz, S. et al. Simple combinations of lineage-determining transcription factors prime cis-regulatory elements required for macrophage and B cell identities. Mol Cell 38, 576-589, doi:10.1016/j.molcel.2010.05.004 (2010).

92 Pliner, H. A. et al. Cicero Predicts cis-Regulatory DNA Interactions from Single-Cell Chromatin Accessibility Data. Mol Cell 71, 858-871 e858, doi:10.1016/j.molcel.2018.06.044 (2018).

93 Hildreth, A. D. et al. Single-cell sequencing of human white adipose tissue identifies new cell states in health and obesity. Nat Immunol 22, 639-653, doi:10.1038/s41590-021-00922-4 (2021).

94 Schwalie, P. C. et al. A stromal cell population that inhibits adipogenesis in mammalian fat depots. Nature 559, 103-108, doi:10.1038/s41586-018-0226-8 (2018).

95 Ragni, E. et al. Identification of miRNA Reference Genes in Extracellular Vesicles from Adipose Derived Mesenchymal Stem Cells for Studying Osteoarthritis. Int J Mol Sci 20, doi:10.3390/ijms20051108 (2019).

96 Shamsi, F. et al. Vascular smooth muscle-derived Trpv1(+) progenitors are a source of cold-induced thermogenic adipocytes. Nat Metab 3, 485-495, doi:10.1038/s42255-021-00373-z (2021).

97 Merrick, D. et al. Identification of a mesenchymal progenitor cell hierarchy in adipose tissue. Science 364, doi:10.1126/science.aav2501 (2019).

98 Aran, D. et al. Reference-based analysis of lung single-cell sequencing reveals a transitional profibrotic macrophage. Nat Immunol 20, 163-172, doi:10.1038/s41590-018-0276-y (2019).

99 Hu, B. C. The human body at cellular resolution: the NIH Human Biomolecular Atlas Program. Nature 574, 187-192, doi:10.1038/s41586-019-1629-x (2019).

100 Lindeboom, R. G. H., Regev, A. & Teichmann, S. A. Towards a Human Cell Atlas: Taking Notes from the Past. Trends Genet 37, 625-630, doi:10.1016/j.tig.2021.03.007 (2021).

101 Vijay, J. et al. Single-cell analysis of human adipose tissue identifies depot and disease specific cell types. Nat Metab 2, 97-109, doi:10.1038/s42255-019-0152-6 (2020).

102 Porter, S. A. et al. Abdominal subcutaneous adipose tissue: a protective fat depot? Diabetes Care 32, 1068-1075, doi:10.2337/dc08-2280 (2009).

103 Denis, G. V. & Obin, M. S. 'Metabolically healthy obesity': origins and implications. Mol Aspects Med 34, 59-70, doi:10.1016/j.mam.2012.10.004 (2013).

104 Michaud, A. et al. Relevance of omental pericellular adipose tissue collagen in the pathophysiology of human abdominal obesity and related cardiometabolic risk. Int J Obes (Lond) 40, 1823-1831, doi:10.1038/ijo.2016.173 (2016).

105 Schipper, H. S., Prakken, B., Kalkhoven, E. & Boes, M. Adipose tissue-resident immune cells: key players in immunometabolism. Trends Endocrinol Metab 23, 407-415, doi:10.1016/j.tem.2012.05.011 (2012).

106 Nawaz, A. et al. CD206(+) M2-like macrophages regulate systemic glucose metabolism by inhibiting proliferation of adipocyte progenitors. Nat Commun 8, 286, doi:10.1038/s41467-017-00231-1 (2017).

107 Olsen, T. K. & Baryawno, N. Introduction to Single-Cell RNA Sequencing. Curr Protoc Mol Biol 122, e57, doi:10.1002/cpmb.57 (2018).

108 Hepler, C. et al. Identification of functionally distinct fibro-inflammatory and adipogenic stromal subpopulations in visceral adipose tissue of adult mice. Elife 7, doi:10.7554/eLife.39636 (2018).

109 Ehrlund, A. et al. The cell-type specific transcriptome in human adipose tissue and influence of obesity on adipocyte progenitors. Sci Data 4, 170164, doi:10.1038/sdata.2017.164 (2017).

110 Briot, A. et al. Senescence Alters PPARgamma (Peroxisome Proliferator-Activated Receptor Gamma)-Dependent Fatty Acid Handling in Human Adipose Tissue Microvascular Endothelial Cells and Favors Inflammation. Arterioscler Thromb Vasc Biol 38, 1134-1146, doi:10.1161/ATVBAHA.118.310797 (2018).

111 Banerji, S. et al. LYVE-1, a new homologue of the CD44 glycoprotein, is a lymph-specific receptor for hyaluronan. J Cell Biol 144, 789-801 (1999).

112 Schluns, K. S., Kieper, W. C., Jameson, S. C. & Lefrancois, L. Interleukin-7 mediates the homeostasis of naive and memory CD8 T cells in vivo. Nat Immunol 1, 426-432, doi:10.1038/80868 (2000).

113 Michelet, X. et al. Metabolic reprogramming of natural killer cells in obesity limits antitumor responses. Nat Immunol 19, 1330-1340, doi:10.1038/s41590-018-0251-7 (2018).

114 Schall, T. J. et al. A human T cell-specific molecule is a member of a new gene family. J Immunol 141, 1018-1025 (1988).

115 Small, K. S. et al. Identification of an imprinted master trans regulator at the KLF14 locus related to multiple metabolic phenotypes. Nat Genet 43, 561-564, doi:10.1038/ng.833 (2011).

116 Civelek, M. et al. Genetic Regulation of Adipose Gene Expression and Cardio-Metabolic Traits. Am J Hum Genet 100, 428-443, doi:10.1016/j.ajhg.2017.01.027 (2017).

117 Kratz, M. et al. Metabolic dysfunction drives a mechanistically distinct proinflammatory phenotype in adipose tissue macrophages. Cell Metab 20, 614-625, doi:10.1016/j.cmet.2014.08.010 (2014).

118 Jager, N. A. et al. Folate receptor-beta imaging using 99mTc-folate to explore distribution of polarized macrophage populations in human atherosclerotic plaque. J Nucl Med 55, 1945-1951, doi:10.2967/jnumed.114.143180 (2014).

119 Liao, X. et al. Kruppel-like factor 4 regulates macrophage polarization. J Clin Invest 121, 2736-2749, doi:10.1172/JCI45444 (2011).

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

121 Acosta, J. R. et al. Increased fat cell size: a major phenotype of subcutaneous white adipose tissue in non-obese individuals with type 2 diabetes. Diabetologia 59, 560-570, doi:10.1007/s00125-015-3810-6 (2016).

122 Ribeiro, R. et al. Human periprostatic white adipose tissue is rich in stromal progenitor cells and a potential source of prostate tumor stroma. Exp Biol Med (Maywood) 237, 1155-1162, doi:10.1258/ebm.2012.012131 (2012).

123 Yang, R. Z. et al. Identification of omentin as a novel depot-specific adipokine in human adipose tissue: possible role in modulating insulin action. Am J Physiol Endocrinol Metab 290, E1253-1261, doi:10.1152/ajpendo.00572.2004 (2006).

124 de Souza Batista, C. M. et al. Omentin plasma levels and gene expression are decreased in obesity. Diabetes 56, 1655-1661, doi:10.2337/db06-1506 (2007).

125 Watanabe, T., Watanabe-Kominato, K., Takahashi, Y., Kojima, M. & Watanabe, R. Adipose Tissue-Derived Omentin-1 Function and Regulation. Compr Physiol 7, 765-781, doi:10.1002/cphy.c160043 (2017).

126 Winnier, D. A. et al. Transcriptomic identification of ADH1B as a novel candidate gene for obesity and insulin resistance in human adipose tissue in Mexican Americans from the Veterans Administration Genetic Epidemiology Study (VAGES). PLoS One 10, e0119941, doi:10.1371/journal.pone.0119941 (2015).

127 Vaittinen, M. et al. MFAP5 is related to obesity-associated adipose tissue and extracellular matrix remodeling and inflammation. Obesity (Silver Spring) 23, 1371-1378, doi:10.1002/oby.21103 (2015).

128 Hou, S. et al. S100A4 protects mice from high-fat diet-induced obesity and inflammation. Lab Invest 98, 1025-1038, doi:10.1038/s41374-018-0067-y (2018).

129 Kuefner, M. S. et al. Secretory phospholipase A2 group IIA modulates insulin sensitivity and metabolism. J Lipid Res 58, 1822-1833, doi:10.1194/jlr.M076141 (2017).

130 Perdikari, A. et al. BATLAS: Deconvoluting Brown Adipose Tissue. Cell Rep 25, 784-797 e784, doi:10.1016/j.celrep.2018.09.044 (2018).

131 Wankhade, U. D. et al. TGF-beta receptor 1 regulates progenitors that promote browning of white fat. Mol Metab 16, 160-171, doi:10.1016/j.molmet.2018.07.008 (2018).

132 Zou, Y. et al. IRX3 Promotes the Browning of White Adipocytes and Its Rare Variants are Associated with Human Obesity Risk. EBioMedicine 24, 64-75, doi:10.1016/j.ebiom.2017.09.010 (2017).

133 Roberts, A. C. & Porter, K. E. Cellular and molecular mechanisms of endothelial dysfunction in diabetes. Diab Vasc Dis Res 10, 472-482, doi:10.1177/1479164113500680 (2013).

134 Escobedo, N. & Oliver, G. The Lymphatic Vasculature: Its Role in Adipose Metabolism and Obesity. Cell Metab 26, 598-609, doi:10.1016/j.cmet.2017.07.020 (2017).

135 Nishimura, S. et al. CD8+ effector T cells contribute to macrophage recruitment and adipose tissue inflammation in obesity. Nat Med 15, 914-920, doi:10.1038/nm.1964 (2009).

136 Wu, H. et al. T-cell accumulation and regulated on activation, normal T cell expressed and secreted upregulation in adipose tissue in obesity. Circulation 115, 1029-1038, doi:10.1161/CIRCULATIONAHA.106.638379 (2007).

137 Singer, M. et al. A Distinct Gene Module for Dysfunction Uncoupled from Activation in Tumor-Infiltrating T Cells. Cell 166, 1500-1511 e1509, doi:10.1016/j.cell.2016.08.052 (2016).

138 Russo, L. & Lumeng, C. N. Properties and functions of adipose tissue macrophages in obesity. Immunology 155, 407-417, doi:10.1111/imm.13002 (2018).

139 Coats, B. R. et al. Metabolically Activated Adipose Tissue Macrophages Perform Detrimental and Beneficial Functions during Diet-Induced Obesity. Cell Rep 20, 3149-3161, doi:10.1016/j.celrep.2017.08.096 (2017).

140 Song, N. J. et al. Small Molecule-Induced Complement Factor D (Adipsin) Promotes Lipid Accumulation and Adipocyte Differentiation. PLoS One 11, e0162228, doi:10.1371/journal.pone.0162228 (2016).

141 Li, C. Y. et al. Comparative analysis of human mesenchymal stem cells from bone marrow and adipose tissue under xeno-free conditions for cell therapy. Stem Cell Res Ther 6, 55, doi:10.1186/s13287-015-0066-5 (2015).

142 Acosta, J. R. et al. Single cell transcriptomics suggest that human adipocyte progenitor cells constitute a homogeneous cell population. Stem Cell Res Ther 8, 250, doi:10.1186/s13287-017-0701-4 (2017).

143 Chung, S. S. et al. Glutathione peroxidase 3 mediates the antioxidant effect of peroxisome proliferator-activated receptor gamma in human skeletal muscle cells. Mol Cell Biol 29, 20-30, doi:10.1128/MCB.00544-08 (2009).

144 Hammarstedt, A. et al. WISP2 regulates preadipocyte commitment and PPARgamma activation by BMP4. Proc Natl Acad Sci U S A 110, 2563-2568, doi:10.1073/pnas.1211255110 (2013).

145 Jang, M. K. & Jung, M. H. ATF3 inhibits PPARgamma-stimulated transactivation in adipocyte cells. Biochem Biophys Res Commun 456, 80-85, doi:10.1016/j.bbrc.2014.11.037 (2015).

146 Kim, J. Y. et al. Activating transcription factor 3 is a target molecule linking hepatic steatosis to impaired glucose homeostasis. J Hepatol 67, 349-359, doi:10.1016/j.jhep.2017.03.023 (2017).

147 Wu, J. et al. Beige adipocytes are a distinct type of thermogenic fat cell in mouse and human. Cell 150, 366-376, doi:10.1016/j.cell.2012.05.016 (2012).

148 Tchernof, A. et al. Regional differences in adipose tissue metabolism in women: minor effect of obesity and body fat distribution. Diabetes 55, 1353-1360, doi:10.2337/db05-1439 (2006).

149 Laitinen, A. et al. A robust and reproducible animal serum-free culture method for clinical-grade bone marrow-derived mesenchymal stromal cells. Cytotechnology 68, 891-906, doi:10.1007/s10616-014-9841-x (2016).

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

151 Aran, D. et al. Reference-based annotation of single-cell transcriptomes identifies a profibrotic macrophage niche after tissue injury. bioRxiv 284604, doi:https://doi.org/10.1101/284604 (2018).

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

153 Huang da, W., Sherman, B. T. & Lempicki, R. A. Bioinformatics enrichment tools: paths toward the comprehensive functional analysis of large gene lists. Nucleic Acids Res 37, 1-13, doi:10.1093/nar/gkn923 (2009).

154 Bolger, A. M., Lohse, M. & Usadel, B. Trimmomatic: a flexible trimmer for Illumina sequence data. Bioinformatics 30, 2114-2120, doi:10.1093/bioinformatics/btu170 (2014).

155 Dobin, A. et al. STAR: ultrafast universal RNA-seq aligner. Bioinformatics 29, 15-21, doi:10.1093/bioinformatics/bts635 (2013).

156 Anders, S., Pyl, P. T. & Huber, W. HTSeq--a Python framework to work with high-throughput sequencing data. Bioinformatics 31, 166-169, doi:10.1093/bioinformatics/btu638 (2015).

157 Love, M. I., Huber, W. & Anders, S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome Biol 15, 550, doi:10.1186/s13059-014-0550-8 (2014).

158 Lidell, M. E. et al. Evidence for two types of brown adipose tissue in humans. Nat Med 19, 631-634, doi:10.1038/nm.3017 (2013).

159 Wang, W. & Seale, P. Control of brown and beige fat development. Nat Rev Mol Cell Biol 17, 691-702, doi:10.1038/nrm.2016.96 (2016).

160 Consortium, G. T. The GTEx Consortium atlas of genetic regulatory effects across human tissues. Science 369, 1318-1330, doi:10.1126/science.aaz1776 (2020).

161 Ogden, C. L. et al. Trends in Obesity Prevalence Among Children and Adolescents in the United States, 1988-1994 Through 2013-2014. JAMA 315, 2292-2299, doi:10.1001/jama.2016.6361 (2016).

162 Parsons, T. J., Power, C., Logan, S. & Summerbell, C. D. Childhood predictors of adult obesity: a systematic review. Int J Obes Relat Metab Disord 23 Suppl 8, S1-107 (1999).

163 Whitaker, R. C., Wright, J. A., Pepe, M. S., Seidel, K. D. & Dietz, W. H. Predicting obesity in young adulthood from childhood and parental obesity. N Engl J Med 337, 869-873, doi:10.1056/NEJM199709253371301 (1997).

164 Mayer-Davis, E. J., Dabelea, D. & Lawrence, J. M. Incidence Trends of Type 1 and Type 2 Diabetes among Youths, 2002-2012. N Engl J Med 377, 301, doi:10.1056/NEJMc1706291 (2017).

165 Constantino, M. I. et al. Long-term complications and mortality in young-onset diabetes: type 2 diabetes is more hazardous and lethal than type 1 diabetes. Diabetes Care 36, 3863-3869, doi:10.2337/dc12-2455 (2013).

166 Vogelezang, S. et al. Novel loci for childhood body mass index and shared heritability with adult cardiometabolic traits. PLoS Genet 16, e1008718, doi:10.1371/journal.pgen.1008718 (2020).

167 Srinivasan, S. et al. The First Genome-Wide Association Study for Type 2 Diabetes in Youth: The Progress in Diabetes Genetics in Youth (ProDiGY) Consortium. Diabetes 70, 996-1005, doi:10.2337/db20-0443 (2021).

168 Patan, S. TIE1 and TIE2 receptor tyrosine kinases inversely regulate embryonic angiogenesis by the mechanism of intussusceptive microvascular growth. Microvasc Res 56, 1-21, doi:10.1006/mvre.1998.2081 (1998).

169 Masiero, M. et al. A core human primary tumor angiogenesis signature identifies the endothelial orphan receptor ELTD1 as a key regulator of angiogenesis. Cancer Cell 24, 229-241, doi:10.1016/j.ccr.2013.06.004 (2013).

170 Millauer, B. et al. High affinity VEGF binding and developmental expression suggest Flk-1 as a major regulator of vasculogenesis and angiogenesis. Cell 72, 835-846, doi:10.1016/0092-8674(93)90573-9 (1993).

171 Li, C. et al. CD105 antagonizes the inhibitory signaling of transforming growth factor beta1 on human vascular endothelial cells. FASEB J 14, 55-64, doi:10.1096/fasebj.14.1.55 (2000).

172 Richter, A. et al. EGFL7 Mediates BMP9-Induced Sprouting Angiogenesis of Endothelial Cells Derived from Human Embryonic Stem Cells. Stem Cell Reports 12, 1250-1259, doi:10.1016/j.stemcr.2019.04.022 (2019).

173 Zhu, B. et al. Nexilin/NEXN controls actin polymerization in smooth muscle and is regulated by myocardin family coactivators and YAP. Sci Rep 8, 13025, doi:10.1038/s41598-018-31328-2 (2018).

174 Takahashi, K., Hiwada, K. & Kokubu, T. Vascular smooth muscle calponin. A novel troponin T-like protein. Hypertension 11, 620-626, doi:10.1161/01.hyp.11.6.620 (1988).

175 Lees-Miller, J. P., Heeley, D. H. & Smillie, L. B. An abundant and novel protein of 22 kDa (SM22) is widely distributed in smooth muscles. Purification from bovine aorta. Biochem J 244, 705-709, doi:10.1042/bj2440705 (1987).

176 Grarup, N. et al. Loss-of-function variants in ADCY3 increase risk of obesity and type 2 diabetes. Nat Genet 50, 172-174, doi:10.1038/s41588-017-0022-7 (2018).

177 Hsieh, T., Gordon, R. E., Clemmons, D. R., Busby, W. H., Jr. & Duan, C. Regulation of vascular smooth muscle cell responses to insulin-like growth factor (IGF)-I by local IGF-binding proteins. J Biol Chem 278, 42886-42892, doi:10.1074/jbc.M303835200 (2003).

178 Koyama, H., Raines, E. W., Bornfeldt, K. E., Roberts, J. M. & Ross, R. Fibrillar collagen inhibits arterial smooth muscle proliferation through regulation of Cdk2 inhibitors. Cell 87, 1069-1078, doi:10.1016/s0092-8674(00)81801-2 (1996).

179 Gogg, S., Nerstedt, A., Boren, J. & Smith, U. Human adipose tissue microvascular endothelial cells secrete PPARgamma ligands and regulate adipose tissue lipid uptake. JCI Insight 4, doi:10.1172/jci.insight.125914 (2019).

180 Ozmen, F. et al. STEAP4 and HIF-1alpha gene expressions in visceral and subcutaneous adipose tissue of the morbidly obese patients. Mol Immunol 73, 53-59, doi:10.1016/j.molimm.2016.03.008 (2016).

181 Yu, D. et al. MARCKS Signaling Differentially Regulates Vascular Smooth Muscle and Endothelial Cell Proliferation through a KIS-, p27kip1- Dependent Mechanism. PLoS One 10, e0141397, doi:10.1371/journal.pone.0141397 (2015).

182 Farrokhi, E., Ghatreh-Samani, K., Salehi-Vanani, N. & Mahmoodi, A. The effect of resveratrol on expression of matrix metalloproteinase 9 and its tissue inhibitors in vascular smooth muscle cells. ARYA Atheroscler 14, 157-162, doi:10.22122/arya.v14i4.1484 (2018).

183 Zhou, Y. et al. Metascape provides a biologist-oriented resource for the analysis of systems-level datasets. Nat Commun 10, 1523, doi:10.1038/s41467-019-09234-6 (2019).

184 Campagna, R. et al. Nicotinamide N-methyltransferase in endothelium protects against oxidant stress-induced endothelial injury. Biochim Biophys Acta Mol Cell Res 1868, 119082, doi:10.1016/j.bbamcr.2021.119082 (2021).

185 Schulkens, I. A. et al. Expression, regulation and function of human metallothioneins in endothelial cells. J Vasc Res 51, 231-238, doi:10.1159/000365550 (2014).

186 Mack, J. J. & Iruela-Arispe, M. L. NOTCH regulation of the endothelial cell phenotype. Curr Opin Hematol 25, 212-218, doi:10.1097/MOH.0000000000000425 (2018).

187 Alonso, M. A. & Weissman, S. M. cDNA cloning and sequence of MAL, a hydrophobic protein associated with human T-cell differentiation. Proc Natl Acad Sci U S A 84, 1997-2001, doi:10.1073/pnas.84.7.1997 (1987).

188 Fu, B., Tian, Z. & Wei, H. Subsets of human natural killer cells and their regulatory effects. Immunology 141, 483-489, doi:10.1111/imm.12224 (2014).

189 Bottcher, J. P. et al. NK Cells Stimulate Recruitment of cDC1 into the Tumor Microenvironment Promoting Cancer Immune Control. Cell 172, 1022-1037 e1014, doi:10.1016/j.cell.2018.01.004 (2018).

190 McLean, C. Y. et al. GREAT improves functional interpretation of cis-regulatory regions. Nat Biotechnol 28, 495-501, doi:10.1038/nbt.1630 (2010).

191 Dankel, S. N. et al. COL6A3 expression in adipose tissue cells is associated with levels of the homeobox transcription factor PRRX1. Sci Rep 10, 20164, doi:10.1038/s41598-020-77406-2 (2020).

192 Ruehl, M. et al. The elongated first fibronectin type III domain of collagen XIV is an inducer of quiescence and differentiation in fibroblasts and preadipocytes. J Biol Chem 280, 38537-38543, doi:10.1074/jbc.M502210200 (2005).

193 Meissburger, B. et al. Regulation of adipogenesis by paracrine factors from adipose stromal-vascular fraction - a link to fat depot-specific differences. Biochim Biophys Acta 1861, 1121-1131, doi:10.1016/j.bbalip.2016.06.010 (2016).

194 Unamuno, X. et al. Dermatopontin, A Novel Adipokine Promoting Adipose Tissue Extracellular Matrix Remodelling and Inflammation in Obesity. J Clin Med 9, doi:10.3390/jcm9041069 (2020).

195 Bohm, A. et al. Common genetic variation in the SERPINF1 locus determines overall adiposity, obesity-related insulin resistance, and circulating leptin levels. PLoS One 7, e34035, doi:10.1371/journal.pone.0034035 (2012).

196 Cha, J. Y. et al. Dexras1 mediates glucocorticoid-associated adipogenesis and diet-induced obesity. Proc Natl Acad Sci U S A 110, 20575-20580, doi:10.1073/pnas.1320454110 (2013).

197 Zhang, K. et al. Molecular Mechanism of Stem Cell Differentiation into Adipocytes and Adipocyte Differentiation of Malignant Tumor. Stem Cells Int 2020, 8892300, doi:10.1155/2020/8892300 (2020).

198 Ostinelli, G., Vijay, J., Vohl, M. C., Grundberg, E. & Tchernof, A. AKR1C2 and AKR1C3 expression in adipose tissue: Association with body fat distribution and

regulatory variants. Mol Cell Endocrinol 527, 111220, doi:10.1016/j.mce.2021.111220 (2021).

199 Blanchette, S. et al. Expression and activity of 20alpha-hydroxysteroid dehydrogenase (AKR1C1) in abdominal subcutaneous and omental adipose tissue in women. J Clin Endocrinol Metab 90, 264-270, doi:10.1210/jc.2004-0583 (2005).

200 Stephens, J. M., Butts, M. D. & Pekala, P. H. Regulation of transcription factor mRNA accumulation during 3T3-L1 preadipocyte differentiation by tumour necrosis factor-alpha. J Mol Endocrinol 9, 61-72, doi:10.1677/jme.0.0090061 (1992).

201 Lowe, C. E., O'Rahilly, S. & Rochford, J. J. Adipogenesis at a glance. J Cell Sci 124, 2681-2686, doi:10.1242/jcs.079699 (2011).

202 Visweswaran, M. et al. Wnt antagonist secreted frizzled-related protein 4 upregulates adipogenic differentiation in human adipose tissue-derived mesenchymal stem cells. PLoS One 10, e0118005, doi:10.1371/journal.pone.0118005 (2015).

203 Ehrlund, A. et al. Characterization of the Wnt inhibitors secreted frizzled-related proteins (SFRPs) in human adipose tissue. J Clin Endocrinol Metab 98, E503-508, doi:10.1210/jc.2012-3416 (2013).

204 Cao, Y. Adipose tissue angiogenesis as a therapeutic target for obesity and metabolic diseases. Nat Rev Drug Discov 9, 107-115, doi:10.1038/nrd3055 (2010).

205 Hafidi, M. E., Buelna-Chontal, M., Sanchez-Munoz, F. & Carbo, R. Adipogenesis: A Necessary but Harmful Strategy. Int J Mol Sci 20, doi:10.3390/ijms20153657 (2019).

206 Nijhawans, P., Behl, T. & Bhardwaj, S. Angiogenesis in obesity. Biomed Pharmacother 126, 110103, doi:10.1016/j.biopha.2020.110103 (2020).

207 Murdoch, C., Muthana, M. & Lewis, C. E. Hypoxia regulates macrophage functions in inflammation. J Immunol 175, 6257-6263, doi:10.4049/jimmunol.175.10.6257 (2005).

208 Suchting, S. et al. The Notch ligand Delta-like 4 negatively regulates endothelial tip cell formation and vessel branching. Proc Natl Acad Sci U S A 104, 3225-3230, doi:10.1073/pnas.0611177104 (2007).

209 Fukuda, D. et al. Notch ligand delta-like 4 blockade attenuates atherosclerosis and metabolic disorders. Proc Natl Acad Sci U S A 109, E1868-1877, doi:10.1073/pnas.1116889109 (2012).

210 Kalucka, J. et al. Quiescent Endothelial Cells Upregulate Fatty Acid beta-Oxidation for Vasculoprotection via Redox Homeostasis. Cell Metab 28, 881-894 e813, doi:10.1016/j.cmet.2018.07.016 (2018).

211 Wang, Y., Chen, J., Tang, W., Zhang, Y. & Li, X. Rapamycin inhibits the proliferation of endothelial cells in hemangioma by blocking the mTOR-FABP4 pathway. Biomed Pharmacother 85, 272-279, doi:10.1016/j.biopha.2016.11.021 (2017).

212 Girona, J. et al. FABP4 induces vascular smooth muscle cell proliferation and migration through a MAPK-dependent pathway. PLoS One 8, e81914, doi:10.1371/journal.pone.0081914 (2013).

213 Weisberg, S. P. et al. Obesity is associated with macrophage accumulation in adipose tissue. J Clin Invest 112, 1796-1808, doi:10.1172/JCI19246 (2003).

214 Liang, C. et al. CD8(+) T-cell plasticity regulates vascular regeneration in type-2 diabetes. Theranostics 10, 4217-4232, doi:10.7150/thno.40663 (2020).

215 Maehara, T. et al. Cytotoxic CD4+ T lymphocytes may induce endothelial cell apoptosis in systemic sclerosis. J Clin Invest 130, 2451-2464, doi:10.1172/JCI131700 (2020).

216 Zhang, W. et al. The TEA domain family transcription factor TEAD4 represses murine adipogenesis by recruiting the cofactors VGLL4 and CtBP2 into a transcriptional complex. J Biol Chem 293, 17119-17134, doi:10.1074/jbc.RA118.003608 (2018).

217 Wang, J. et al. Effect of TEAD4 on multilineage differentiation of muscle-derived stem cells. Am J Transl Res 10, 998-1011 (2018).

218 Hao, Y. et al. Integrated analysis of multimodal single-cell data. Cell 184, 3573-3587 e3529, doi:10.1016/j.cell.2021.04.048 (2021).

219 Gaspar, J. M. Improved peak-calling with MACS2. bioRxiv, 496521, doi:10.1101/496521 (2018).

220 Stuart, T. et al. Comprehensive Integration of Single-Cell Data. Cell 177, 1888-1902 e1821, doi:10.1016/j.cell.2019.05.031 (2019).

221 Schep, A. N., Wu, B., Buenrostro, J. D. & Greenleaf, W. J. chromVAR: inferring transcription-factor-associated accessibility from single-cell epigenomic data. Nat Methods 14, 975-978, doi:10.1038/nmeth.4401 (2017).

222 Hirsch, J. & Batchelor, B. Adipose tissue cellularity in human obesity. Clin Endocrinol Metab 5, 299-311, doi:10.1016/s0300-595x(76)80023-0 (1976).

223 Ye, J., Gao, Z., Yin, J. & He, Q. Hypoxia is a potential risk factor for chronic inflammation and adiponectin reduction in adipose tissue of ob/ob and dietary obese mice. Am J Physiol Endocrinol Metab 293, E1118-1128, doi:10.1152/ajpendo.00435.2007 (2007).

224 Tchkonia, T. et al. Abundance of two human preadipocyte subtypes with distinct capacities for replication, adipogenesis, and apoptosis varies among fat depots. Am J Physiol Endocrinol Metab 288, E267-277, doi:10.1152/ajpendo.00265.2004 (2005).

225 Primeau, V. et al. Characterizing the profile of obese patients who are metabolically healthy. Int J Obes (Lond) 35, 971-981, doi:10.1038/ijo.2010.216 (2011).

226 Hafemeister, C. & Satija, R. Normalization and variance stabilization of single-cell RNA-seq data using regularized negative binomial regression. Genome Biol 20, 296, doi:10.1186/s13059-019-1874-1 (2019).

227 Chusyd, D. E., Wang, D., Huffman, D. M. & Nagy, T. R. Relationships between Rodent White Adipose Fat Pads and Human White Adipose Fat Depots. Front Nutr 3, 10, doi:10.3389/fnut.2016.00010 (2016).

228 Pouliot, M. C. et al. Visceral obesity in men. Associations with glucose tolerance, plasma insulin, and lipoprotein levels. Diabetes 41, 826-834, doi:10.2337/diab.41.7.826 (1992).

229 Jensen, M. D. & Johnson, C. M. Contribution of leg and splanchnic free fatty acid (FFA) kinetics to postabsorptive FFA flux in men and women. Metabolism 45, 662-666, doi:10.1016/s0026-0495(96)90040-2 (1996).

230 Arner, P. Differences in lipolysis between human subcutaneous and omental adipose tissues. Ann Med 27, 435-438, doi:10.3109/07853899709002451 (1995).

231 Piche, M. E., Vasan, S. K., Hodson, L. & Karpe, F. Relevance of human fat distribution on lipid and lipoprotein metabolism and cardiovascular disease risk. Curr Opin Lipidol 29, 285-292, doi:10.1097/MOL.0000000000000522 (2018).

232 Sabaratnam, R. & Svenningsen, P. Adipocyte-Endothelium Crosstalk in Obesity. Front Endocrinol (Lausanne) 12, 681290, doi:10.3389/fendo.2021.681290 (2021).

233 Jaitin, D. A. et al. Lipid-Associated Macrophages Control Metabolic Homeostasis in a Trem2-Dependent Manner. Cell 178, 686-698 e614, doi:10.1016/j.cell.2019.05.054 (2019).

234 Kim, P. S. & Ahmed, R. Features of responding T cells in cancer and chronic infection. Curr Opin Immunol 22, 223-230, doi:10.1016/j.coi.2010.02.005 (2010).

235 Aguilar, E. G. & Murphy, W. J. Obesity induced T cell dysfunction and implications for cancer immunotherapy. Curr Opin Immunol 51, 181-186, doi:10.1016/j.coi.2018.03.012 (2018).

236 Rashba, E. J., Reich, E. P., Janeway, C. A. & Sherwin, R. S. Type 1 diabetes mellitus: an imbalance between effector and regulatory T cells? Acta Diabetol 30, 61-69, doi:10.1007/BF00578215 (1993).

237 Trayhurn, P., Duncan, J. S., Wood, A. M. & Beattie, J. H. Metallothionein gene expression and secretion in white adipose tissue. Am J Physiol Regul Integr Comp Physiol 279, R2329-2335, doi:10.1152/ajpregu.2000.279.6.R2329 (2000).

238 Speir, M. L. et al. UCSC Cell Browser: Visualize Your Single-Cell Data. bioRxiv, 2020.2010.2030.361162, doi:10.1101/2020.10.30.361162 (2020).

239 Stenkula, K. G. & Erlanson-Albertsson, C. Adipose cell size: importance in health and disease. Am J Physiol Regul Integr Comp Physiol 315, R284-R295, doi:10.1152/ajpregu.00257.2017 (2018).

240 Krishnaswami, S. R. et al. Using single nuclei for RNA-seq to capture the transcriptome of postmortem neurons. Nat Protoc 11, 499-524, doi:10.1038/nprot.2016.015 (2016).

241 Wu, H. et al. Comparative Analysis and Refinement of Human PSC-Derived Kidney Organoid Differentiation with Single-Cell Transcriptomics. Cell Stem Cell 23, 869-881 e868, doi:10.1016/j.stem.2018.10.010 (2018).

# Appendices

## Appendix A

Published abstracts of projects in which the thesis author contributed significantly are included below.

1. Ostinelli G, Vijay J, Vohl MC, Grundberg E, Tchernof A. *AKR1C2* and *AKR1C3* expression in adipose tissue: Association with body fat distribution and regulatory variants. Mol Cell Endocrinol. 2021 May 1;527:111220. doi: 10.1016/j.mce.2021.111220.

   **Abstract**

   Background: Changes in androgen dynamics within adipose tissue have been proposed as modulators of body fat accumulation. In this context, *AKR1C2* likely plays a significant role by inactivating 5α-dihydrotestosterone.

   Aim: To characterize *AKR1C2* expression patterns across adipose depots and cell populations and to provide insight into the link with body fat distribution and genetic regulation.

   Methods: We used RNA sequencing data from severely obese patients to assess patterns of *AKR1C2* and *AKR1C3* expression in abdominal adipose tissue depots and cell fractions. We additionally used data from 856 women to assess *AKR1C2* heritability and to link its expression in adipose tissue with body fat distribution. Further, we used public resources to study *AKR1C2* genetic regulation as well as reference epigenome data for regulatory element profiling and functional interpretation of genetic data.

   Results: We found that mature adipocytes and adipocyte-committed adipocyte progenitor cells (APCs) had enriched expression of *AKR1C2*. We found adipose tissue *AKR1C2* and

*AKR1C3* expression to be significantly and positively associated with percentage trunk fat mass in women. We identified strong genetic regulation of *AKR1C2* by rs28571848 and rs34477787 located on the binding sites of two nuclear transcription factors, namely retinoid acid-related orphan receptor alpha and the glucocorticoid receptor.

Conclusion: We confirm the link between *AKR1C2*, adipogenic differentiation and adipose tissue distribution. We provide insight into genetic regulation of *AKR1C2* by identifying regulatory variants mapping to binding sites for the glucocorticoid receptor and retinoid acid-related orphan receptor alpha which may in part mediate the effect of *AKR1C2* expression on body fat distribution.

**2.** Chechi K, Vijay J, Voisine P, Mathieu P, Bossé Y, Tchernof A, Grundberg E, Richard D. UCP1 expression-associated gene signatures of human epicardial adipose tissue. JCI Insight. 2019 Apr 18;4(8):e123618. doi: 10.1172/jci.insight.123618.

**Abstract**:

Multiple reports of uncoupling protein 1 (*UCP1*) expression have established its presence in human epicardial adipose tissue (eAT). Its functional relevance to eAT, however, remains largely unknown. In a recent study, we reported that adrenergic stimulation of eAT was associated with downregulation of secreted proteins involved in oxidative stress-related and immune-related pathways. Here, we explored the *UCP1*-associated features of human eAT using next-generation deep sequencing. Paired biopsies of eAT, mediastinal adipose tissue (mAT), and subcutaneous adipose tissue (sAT) obtained from cardiac surgery patients, with specific criteria of high and low expression of *UCP1* in eAT, were subjected to RNA sequencing. Although eAT exhibited a depot-specific upregulation in

the immune-related pathways relative to mAT and sAT, high *UCP1* expression in eAT was specifically associated with differential gene expression that functionally corresponded with downregulation in the production of reactive oxygen species and immune responses, including T cell homeostasis. Our data indicate that *UCP1* and adaptive immunity share a reciprocal relationship at the whole-transcriptome level, thereby supporting a plausible role for *UCP1* in maintaining tissue homeostasis in human eAT.

3. Allum F, Hedman ÅK, Shao X, Cheung WA, Vijay J, Guénard F, Kwan T, Simon MM, Ge B, Moura C, Boulier E, Rönnblom L, Bernatsky S, Lathrop M, McCarthy MI, Deloukas P, Tchernof A, Pastinen T, Vohl MC, Grundberg E. Dissecting features of epigenetic variants underlying cardiometabolic risk using full-resolution epigenome profiling in regulatory elements. Nat Commun. 2019 Mar 14;10(1):1209. doi: 10.1038/s41467-019-09184-z.

   **Abstract**:

   Sparse profiling of CpG methylation in blood by microarrays has identified epigenetic links to common diseases. Here we apply methylC-capture sequencing (MCC-Seq) in a clinical population of ~200 adipose tissue and matched blood samples (Ntotal~400), providing high-resolution methylation profiling (>1.3 M CpGs) at regulatory elements. We link methylation to cardiometabolic risk through associations to circulating plasma lipid levels and identify lipid-associated CpGs with unique localization patterns in regulatory elements. We show distinct features of tissue-specific versus tissue-independent lipid-linked regulatory regions by contrasting with parallel assessments in ~800 independent adipose tissue and blood samples from the general population. We follow-up on adipose-specific

regulatory regions under (1) genetic and (2) epigenetic (environmental) regulation via integrational studies. Overall, the comprehensive sequencing of regulatory element methylomes reveals a rich landscape of functional variants linked genetically as well as epigenetically to plasma lipid traits.

4. Morris JA, Kemp JP, Youlten SE, Laurent L, Logan JG, Chai RC, Vulpescu NA, Forgetta V, Kleinman A, Mohanty ST, Sergio CM, Quinn J, Nguyen-Yamamoto L, Luco AL, Vijay J, Simon MM, Pramatarova A, Medina-Gomez C, Trajanoska K, Ghirardello EJ, Butterfield NC, Curry KF, Leitch VD, Sparkes PC, Adoum AT, Mannan NS, Komla-Ebri DSK, Pollard AS, Dewhurst HF, Hassall TAD, Beltejar MG; 23andMe Research Team, Adams DJ, Vaillancourt SM, Kaptoge S, Baldock P, Cooper C, Reeve J, Ntzani EE, Evangelou E, Ohlsson C, Karasik D, Rivadeneira F, Kiel DP, Tobias JH, Gregson CL, Harvey NC, Grundberg E, Goltzman D, Adams DJ, Lelliott CJ, Hinds DA, Ackert-Bicknell CL, Hsu YH, Maurano MT, Croucher PI, Williams GR, Bassett JHD, Evans DM, Richards JB. An atlas of genetic influences on osteoporosis in humans and mice. Nat Genet. 2019 Feb;51(2):258-266. doi: 10.1038/s41588-018-0302-x.

   **Abstract**:

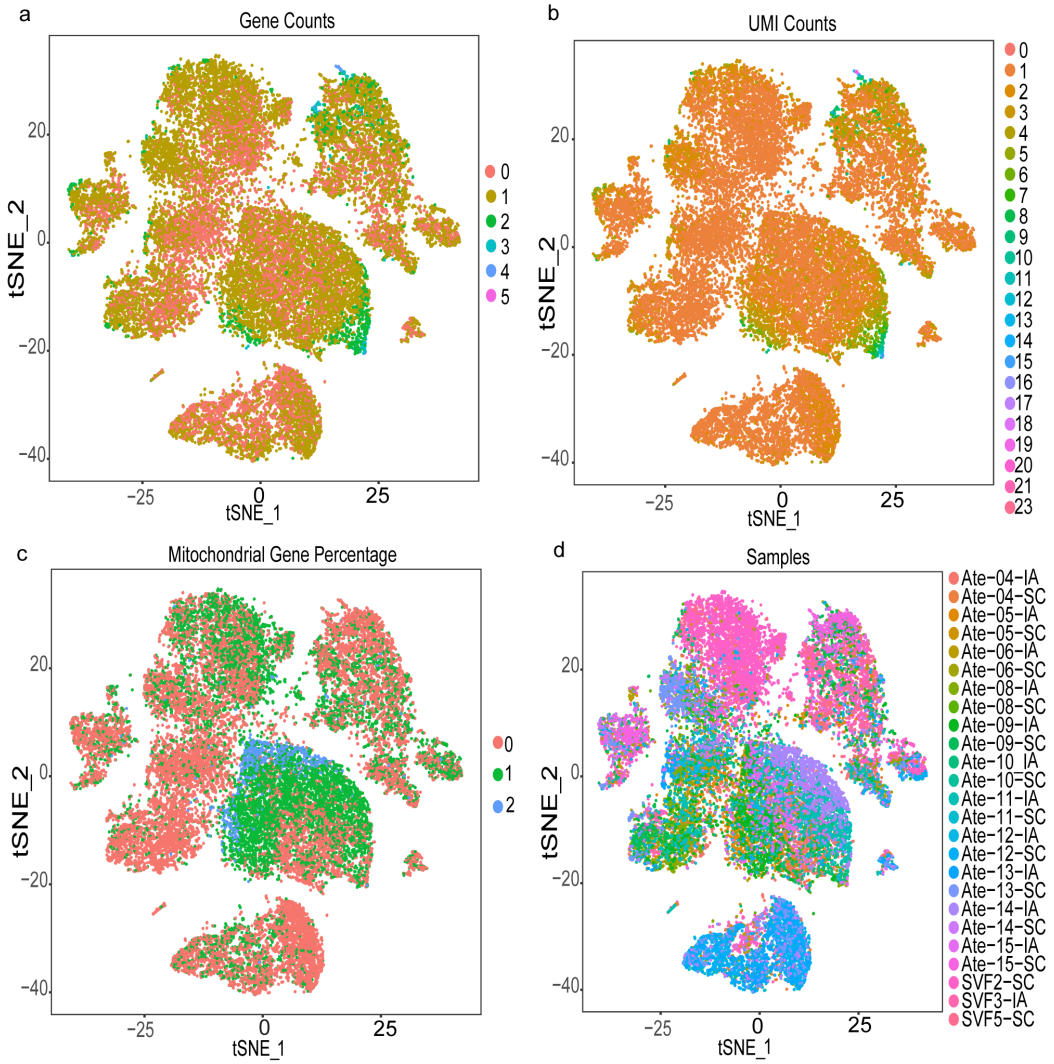
   Osteoporosis is a common aging-related disease diagnosed primarily using bone mineral density (BMD). We assessed genetic determinants of BMD as estimated by heel quantitative ultrasound in 426,824 individuals, identifying 518 genome-wide significant loci (301 novel), explaining 20% of its variance. We identified 13 bone fracture loci, all associated with estimated BMD (eBMD), in ~1.2 million individuals. We then identified target genes enriched for genes known to influence bone density and strength (maximum

odds ratio (OR) = 58, P = 1 × 10-75) from cell-specific features, including chromatin conformation and accessible chromatin sites. We next performed rapid-throughput skeletal phenotyping of 126 knockout mice with disruptions in predicted target genes and found an increased abnormal skeletal phenotype frequency compared to 526 unselected lines ($P < 0.0001$). In-depth analysis of one gene, *DAAM2*, showed a disproportionate decrease in bone strength relative to mineralization. This genetic atlas provides evidence linking associated SNPs to causal genes, offers new insight into osteoporosis pathophysiology, and highlights opportunities for drug development.

**Appendix B**:

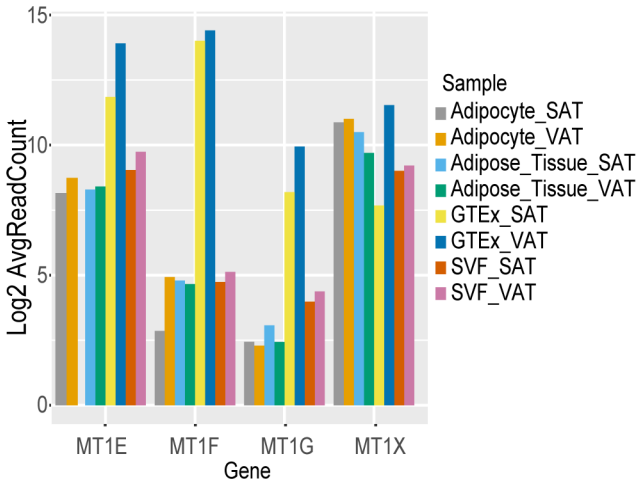
Supplementary Materials from Chapter 2 are included below

Supplementary Figures

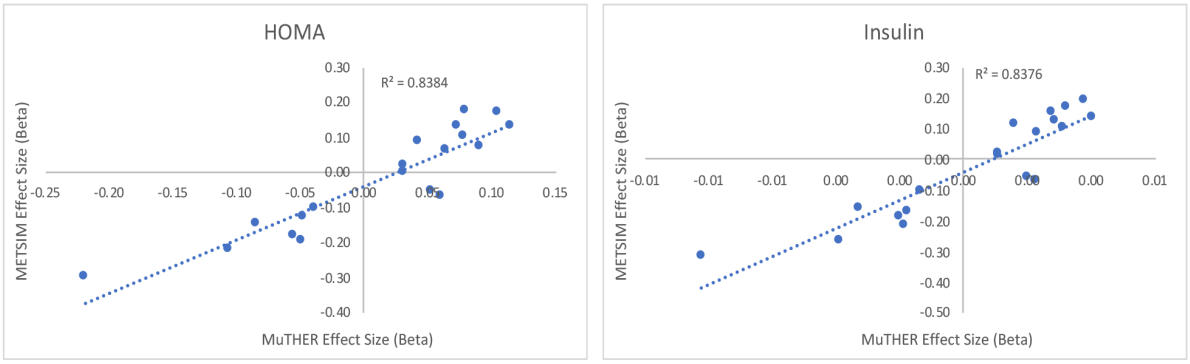



Supplementary Figure 1: UMI, nGene, mitochondrial and sample distribution of SVF clusters from all 25 samples (a) Each cell is colored based on the result obtained by dividing total number of genes expressed by 1000 and then rounding to 0 decimals. (b) Cells are colored based on the result obtained by dividing total number of UMIs expressed by 1000 and then rounding to 0

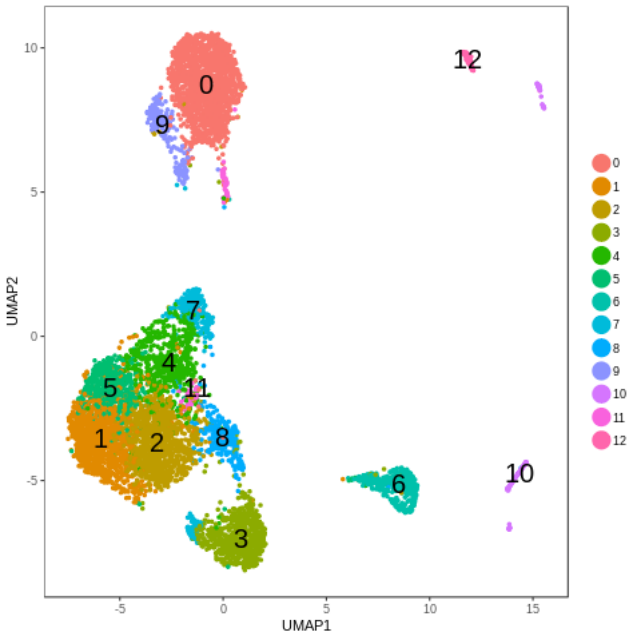
decimals. (c) cell color shows percentage of mitochondrial gene expression. Group 0 shows cells with mitochondrial gene expression <=5%, group 1 represents 6% to 14% and group 2 represents15% to 24% expression (d) Cells are colored based on the scRNAseq library



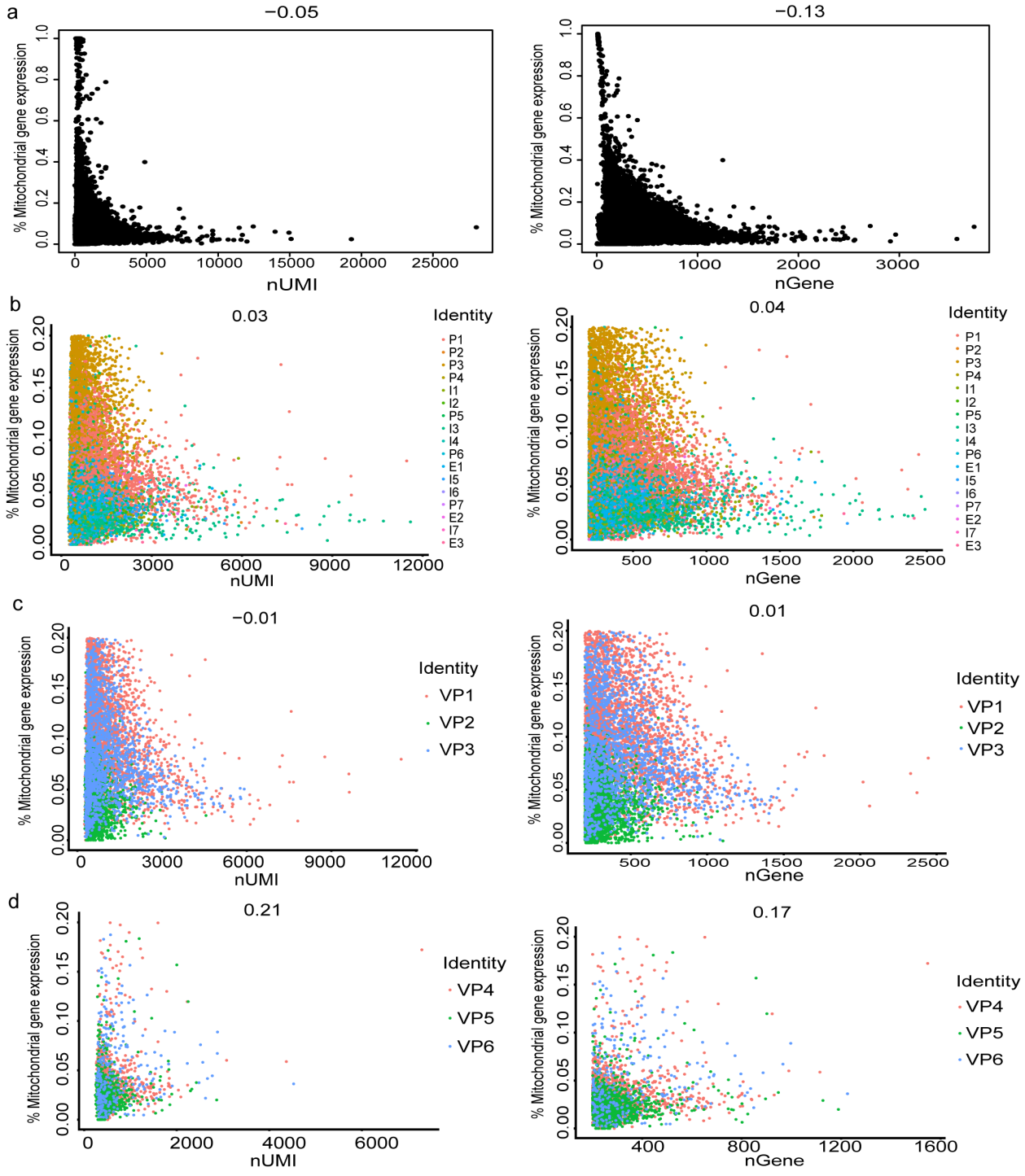

Supplementary Figures 2: Metallothionein gene shows expression in SAT and VAT depots from adipose tissue, adipocyte, SVF and GTEx



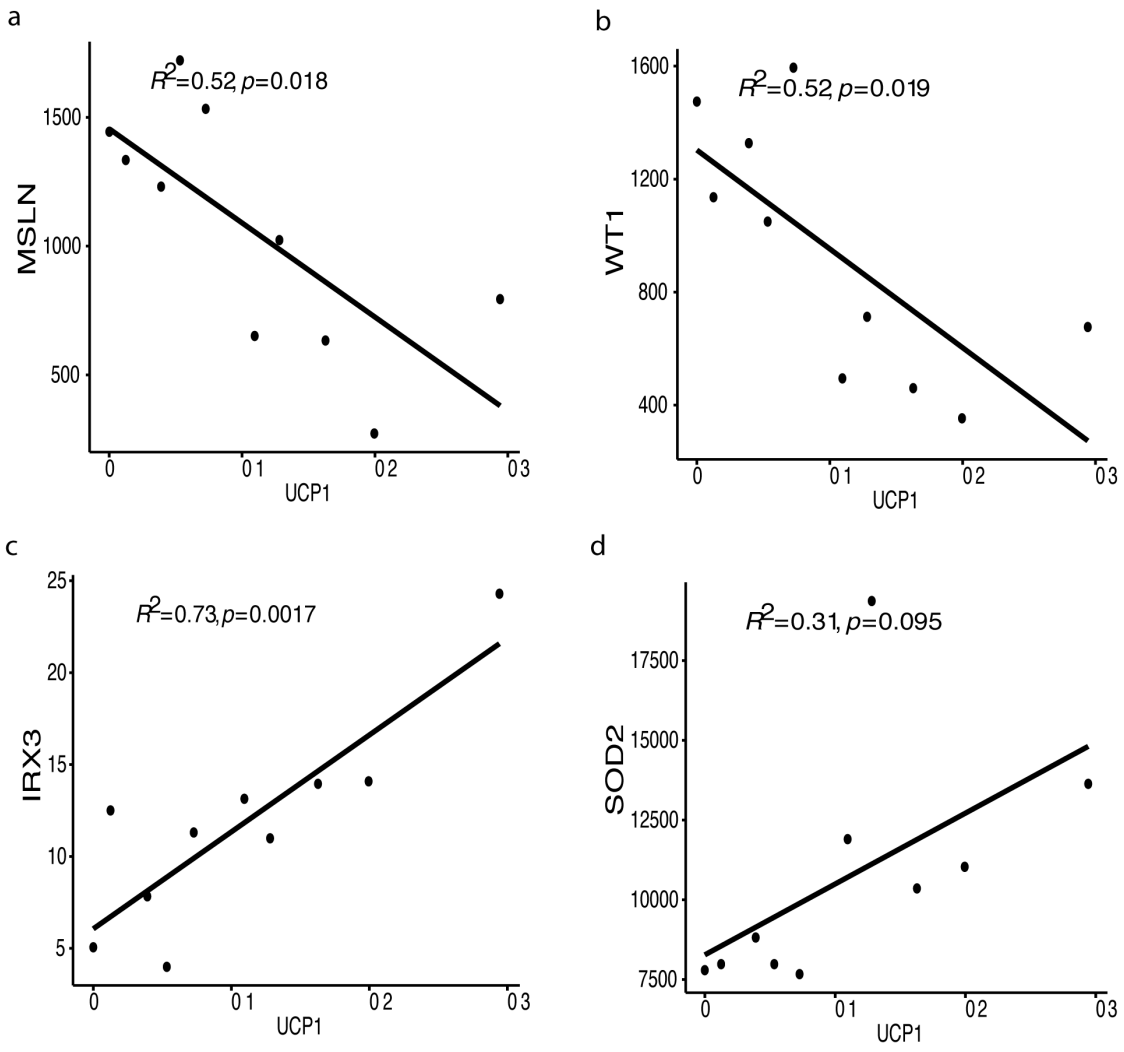

Supplementary Figures 3: T2D associated genes in MuTHER study from SP1 cluster shows strong correlation of effect sizes with METSIM Study



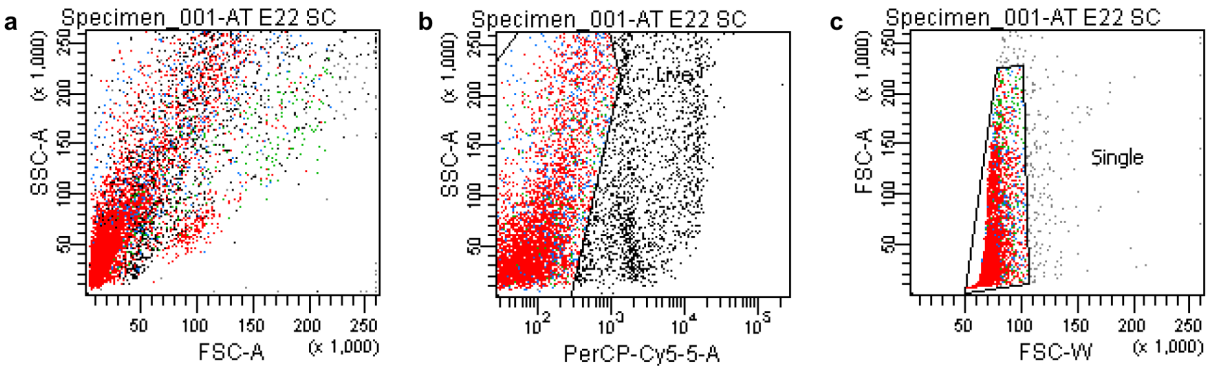

Supplementary Figures 4: Clustering results of CD34+ cell population from SVF of SAT and VAT of 2 individuals.

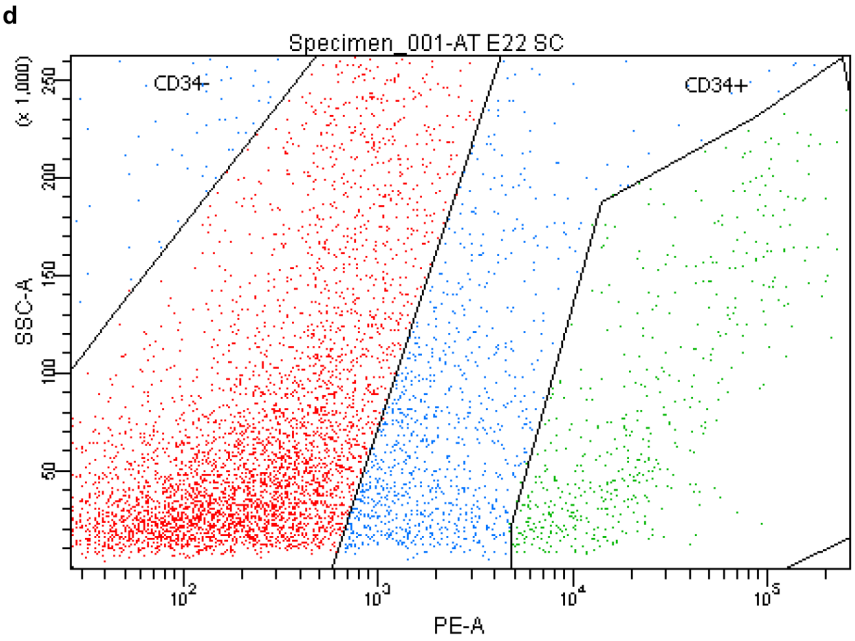



Supplementary Figures 5: Distribution of mitochondrial gene expression in relation to UMI and Gene distribution in (a) unfiltered initial data (b) All clusters (c) VPM clusters (d) VPC clusters




Supplementary Figure 6: Pearson correlation of UCP1 with MSLN (a), WT1(b), IRX3 (c) and SOD2 (d) using bulk RNA-Seq from visceral adipose tissue of 10 individuals.




Tube: AT E22 SC

| Population | #Events | %Parent | %Total |
|---|---|---|---|
| All Events | 10,000 | #### | 100.0 |
| Live | 7,364 | 73.6 | 73.6 |
| Single | 7,071 | 96.0 | 70.7 |
| CD34+ | 541 | 7.7 | 5.4 |
| CD34- | 5,345 | 75.6 | 53.4 |

Supplementary Figure 7: FACS Sorting strategies used for the representative sample included in the study. The cells are filtered for dead cells (b) , the single cells are identified (c) and sorted for CD34+ and CD34- (d)

Supplementary Tables

Supplementary tables of chapter 2 are available for public access from the published article. The link to the article is https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7025882/

Appendix C: Supplementary Materials from Chapter 3

Supplementary Tables

Supplementary Table 1: Top 20 motifs identified in *CFD* expressing clusters in comparison with *MSLN* expressing clusters

| motif | observed | background | percent.observed | percent.background | fold.enrichment | pvalue | motif.name |
|---|---|---|---|---|---|---|---|
| MA1125.1 | 569 | 10972 | 41.5631848 | 21.944 | 1.89405691 | 2.54E-61 | ZNF384 |
| MA1123.2 | 349 | 5229 | 25.4930606 | 10.458 | 2.43766118 | 8.05E-58 | TWIST1 |
| MA0698.1 | 389 | 6295 | 28.4149014 | 12.59 | 2.25694213 | 1.37E-56 | ZBTB18 |
| MA0102.4 | 308 | 4562 | 22.4981738 | 9.124 | 2.46582353 | 2.52E-51 | CEBPA |
| MA1637.1 | 329 | 5174 | 24.0321402 | 10.348 | 2.32239469 | 2.03E-49 | EBF3 |
| MA0836.2 | 277 | 3942 | 20.2337473 | 7.884 | 2.56643167 | 5.15E-49 | CEBPD |
| MA1632.1 | 306 | 4710 | 22.3520818 | 9.42 | 2.37283246 | 1.72E-47 | ATF2 |
| MA0052.4 | 430 | 8203 | 31.4097882 | 16.406 | 1.91453055 | 4.08E-44 | MEF2A |
| MA0091.1 | 266 | 3925 | 19.4302411 | 7.85 | 2.47518994 | 4.86E-44 | TAL1::TCF3 |
| MA1579.1 | 383 | 6966 | 27.9766253 | 13.932 | 2.00808393 | 3.52E-43 | ZBTB26 |
| MA0154.4 | 329 | 5568 | 24.0321402 | 11.136 | 2.15805857 | 1.30E-42 | EBF1 |
| MA0497.1 | 418 | 8032 | 30.5332359 | 16.064 | 1.90072435 | 7.33E-42 | MEF2C |
| MA1109.1 | 308 | 5166 | 22.4981738 | 10.332 | 2.1775236 | 2.39E-40 | NEUROD1 |

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| MA0152.1 | 436 | 8681 | 31.8480643 | 17.362 | 1.83435458 | 4.03E-40 | NFATC2 |
| MA1642.1 | 303 | 5090 | 22.1329438 | 10.18 | 2.1741595 | 1.69E-39 | NEUROG2(var.2) |
| MA0144.2 | 419 | 8252 | 30.606282 | 16.504 | 1.85447661 | 2.26E-39 | STAT3 |
| MA0681.2 | 320 | 5538 | 23.3747261 | 11.076 | 2.11039419 | 2.27E-39 | PHOX2B |
| MA0050.2 | 657 | 15628 | 47.9912345 | 31.256 | 1.5354247 | 2.78E-39 | IRF1 |
| MA1636.1 | 287 | 4760 | 20.9642075 | 9.52 | 2.20212263 | 3.21E-38 | CEBPG(var.2) |
| MA0488.1 | 296 | 5009 | 21.6216216 | 10.018 | 2.15827726 | 6.62E-38 | JUN |

Supplementary Table 2: Top 20 motifs identified in *MSLN* expressing clusters in comparison with *CFD* expressing clusters

| motif | observed | background | percent.observed | percent.background | fold.enrichment | pvalue | motif.name |
|---|---|---|---|---|---|---|---|
| MA1522.1 | 284 | 8583 | 36.1323155 | 17.166 | 2.10487682 | 7.41E-38 | MAZ |
| MA1653.1 | 299 | 9388 | 38.0407125 | 18.776 | 2.02602857 | 3.79E-37 | ZNF148 |
| MA1121.1 | 205 | 5115 | 26.0814249 | 10.23 | 2.5495039 | 4.26E-37 | TEAD2 |
| MA0528.2 | 308 | 9926 | 39.1857506 | 19.852 | 1.97389435 | 2.39E-36 | ZNF263 |
| MA0599.1 | 309 | 10110 | 39.3129771 | 20.22 | 1.94426197 | 3.18E-35 | KLF5 |
| MA0809.2 | 216 | 5961 | 27.480916 | 11.922 | 2.30505922 | 6.92E-33 | TEAD4 |
| MA0037.3 | 144 | 3103 | 18.3206107 | 6.206 | 2.95208036 | 7.71E-32 | GATA3 |
| MA0090.3 | 207 | 5795 | 26.3358779 | 11.59 | 2.27229317 | 1.66E-30 | TEAD1 |
| MA0808.1 | 185 | 4898 | 23.5368957 | 9.796 | 2.40270474 | 6.83E-30 | TEAD3 |
| MA1643.1 | 217 | 6309 | 27.6081425 | 12.618 | 2.18799671 | 7.34E-30 | NFIB |

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| MA0482.2 | 164 | 4075 | 20.8651399 | 8.15 | 2.56013987 | 2.66E-29 | GATA4 |
| MA0747.1 | 207 | 5968 | 26.3358779 | 11.936 | 2.20642408 | 8.47E-29 | SP8 |
| MA0516.2 | 211 | 6163 | 26.8447837 | 12.326 | 2.17789905 | 1.16E-28 | SP2 |
| MA1104.2 | 165 | 4308 | 20.9923664 | 8.616 | 2.43643993 | 4.29E-27 | GATA6 |
| MA0149.1 | 348 | 13353 | 44.2748092 | 26.706 | 1.65786 | 1.04E-26 | EWSR1-FLI1 |
| MA1107.2 | 275 | 9540 | 34.9872774 | 19.08 | 1.83371475 | 2.71E-26 | KLF9 |
| MA1564.1 | 184 | 5219 | 23.4096692 | 10.438 | 2.24273512 | 4.01E-26 | SP9 |
| MA0036.3 | 140 | 3556 | 17.8117048 | 7.112 | 2.50445794 | 6.28E-24 | GATA2 |
| MA0741.1 | 214 | 6865 | 27.2264631 | 13.73 | 1.98299076 | 9.08E-24 | KLF16 |
| MA0039.4 | 281 | 10228 | 35.7506361 | 20.456 | 1.7476846 | 1.11E-23 | KLF4 |

Appendix D: Copyright Permissions

Chapter 2 is published in Nature Metabolism doi: 10.1038/s42255-019-0152-6. Copyright of the materials described in this chapter adheres to the policies of the journal. Chapter 3 is not published and hence the reuse of its content is restricted.