

## Supplemental Methods

### Composition of adipogenic differentiation media

Induction media was prepared by adding 1 mL FBS, 500  $\mu$ l Penicillin-Streptomycin, 15  $\mu$ l human Insulin (0.5  $\mu$ M, Sigma-Aldrich, I2643-50MG), 10  $\mu$ l T3 (2 nM, Sigma-Aldrich, T6397-100MG), 50  $\mu$ l Biotin (33  $\mu$ M, Sigma-Aldrich, B4639-100MG), 100  $\mu$ l Pantothenate (17  $\mu$ M, Sigma-Aldrich, P5155-100G), 1  $\mu$ l Dexamethasone (0.1  $\mu$ M, Sigma-Aldrich, D2915-100MG), 500  $\mu$ l IBMX (500  $\mu$ M, Sigma-Aldrich, I7018-100mg), and 12.5  $\mu$ l Indomethacin (30  $\mu$ M, Sigma-Aldrich, I7378-5G) to 48.5 mL DMEM medium and sterile filter.

### Composition of NP-40 based lysis buffer

To 14.7 mL nuclease-free water (Qiagen), 150  $\mu$ l of Tris-Hydrochloride (Sigma-Aldrich, T2194), 30  $\mu$ l of Sodium Chloride (5M; Sigma-Aldrich, 59222C), 45  $\mu$ l of Magnesium Chloride (1M; Sigma-Aldrich, M1028), and 75  $\mu$ l of NP-40 (Sigma-Aldrich, 74385) was added.

### Single-cell and single-nuclei sequencing library preparation

For 10x technology, single cells or single nuclei were loaded into each channel of a Chromium single-cell 3' Chip. Single cells/nuclei were partitioned into droplets with gel beads in the Chromium Controller. After emulsions were formed, barcoded reverse transcription of RNA took place. This was followed by cDNA amplification, fragmentation and adapter and sample index attachment, all according to the manufacturer's recommendations.

For mcSCRB-seq experiment with white and brown adipocytes (day 20), 96-well plates were first preloaded with rows of 10 uniquely barcoded primers and lysis buffer according to the mcSCRB-seq protocol, with the only difference being the use of  $\mu$ CB-seq RT primers (Chen et al., 2020) instead of standard mcSCRB-seq ones. The sequence of barcodes used were: TCACAGCA, GTAGCACT, ATAGCGTC, CTAGCTGA, CTACGACA, GTACGCAT, ACATGCGT, GCATGTAC, ATACGTGC, and

GCAGTATC. CellenONE X1 instrument was used to individually deliver a single adipocyte into each well for a total of 200 cells. Following cell delivery, the mcSCRB-seq protocol was followed directly, but with the following two modifications:

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

### **scRNA-seq and snRNA-seq data analysis**

scRNA-seq white & brown preadipocytes dataset was processed using cellranger-3.0.2 with default parameters, and the human GRCh38-3.0.0 genome (November 19, 2018) as input. A custom pre-mRNA GTF file was created using the GRCh38-3.0.0 FASTA file as input to include intronic reads in UMI counts. Sample demultiplexing, doublet removal, and empty droplet removal was performed using the Seurat function HTODemux (Note S1). Cell barcodes were further filtered to have more than 200 genes. Post demultiplexing and filtering, scVI was used to infer a 20-dimensional latent space based on the expression of the top 2000 most variable genes. This latent space was then used in Seurat to generate the UMAP visualization using the RunUMAP command. Downstream clustering (resolution = 0.4) and differential expression analysis ( $\log_{2}FC > 0.5$ ) was performed using Seurat's SCTransform pipeline (Hafemeister and Satija, 2019). Clusters with > 5% mean mitochondrial content were removed from downstream analyses. In the identified high-quality clusters, cells had minimal cell-cycle effects as calculated using Seurat (Supplemental Fig. S1Q). Gene ontology analysis was performed at [geneontology.org](http://geneontology.org) (Mi et al. 2019; The Gene Ontology Consortium et al. 2000; Carbon et al. 2019) and results were further confirmed using the goana package in R with genome wide human annotation

derived from org.Hs.eg.db Bioconductor package. Transcription factor enrichment analysis was performed using the ChEA3 tool (Keenan et al., 2019). GRCh38-ref20202A (2020) reference was used for analysis involving lncRNAs, keeping everything else the same. Independent sub-clustering of cluster 0 and cluster 1 identified differences in cellular states based on cell-cycle only, suggesting the absence of any cellular subtypes (Supplemental Fig. S1N and S1O). However, sub-clustering of cluster 2 revealed a *PI16+* adipocyte progenitor population (Supplemental Fig. S1P; Merrick et al. 2019; Ruan 2020), which was also identified in snRNA-seq dataset (Supplemental Fig. S4G). In this manuscript, we focused on only the major cell-types identified within human white and brown preadipocytes (Fig. 1B). For sub-clustering, resolution was set to 0.3, the smallest value at which distinct clusters were first identified within clusters 0, 1, and 2.

snRNA-seq white and brown preadipocyte dataset was also processed using cellranger-3.0.2. For white preadipocyte, barcodes with < 200 genes were removed from downstream analyses. CellBender (Fleming et al., 2019) was used to remove empty droplets. For downstream analyses, only barcodes called as cells by both cellranger and CellBender were used and barcodes with UMI count > 49000 were filtered out as possible doublets. For brown preadipocyte, barcodes with < 200 genes were removed and scVI was used to infer a 20-dimensional latent space. First round of clustering was performed in Seurat with the resolution set to 0.06. We identified 3 clusters, with cluster 1 having most of the barcodes called as empty by CellBender. Therefore, cluster 1 was removed from downstream analysis as well other barcodes that were called as "cell-containing" by cellranger but not by CellBender. Cluster 2 was marked with high mitochondrial content (>20%) and hence was also removed from downstream analyses. After filtering out low-quality barcodes and clusters, Scrublet (Wolock et al., 2019) was used to remove any potential doublets. After individual QC of white and brown preadipocyte libraries, the two datasets were integrated together using scVI with no batch effect correction. The output from scVI analysis was a 20-dimensional latent space representation with cell embeddings for

both white and brown nuclei. This latent space was then used in Seurat to generate the UMAP visualization using the RunUMAP command. Downstream clustering (resolution = 0.24) and differential expression analysis ( $\log_{2}FC > 0.5$ ) was performed using Seurat's SCTransform pipeline (see Supplemental Fig. S4). Cells in each cluster had no significant cell-cycle effects (Supplemental Fig. S4C). For gene ontology, and differential expression analyses, the same tools as mentioned in the above paragraph were used. GRCh38-ref20202A (2020) reference was used for analysis involving lncRNAs, keeping everything else the same.

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

snRNA-seq white adipocyte dataset was processed using cellranger-3.1.0. Barcodes with < 200 genes were removed from downstream analyses and scVI was used to infer a 20-dimensional latent space. For clustering using Seurat, the resolution parameter was set to 0.45. We identified 7 clusters, with cluster 3 having most of the barcodes called as empty by CellBender. Therefore, cluster 3 was removed from downstream analysis as well other barcodes that were called as "cell-containing" by cellranger but not by CellBender. Cluster 5 was marked with high mitochondrial content and hence was

also removed from downstream analyses. Cluster 2 had the greatest number of doublets identified by the doubletDetection (Gayoso et al., 2019) tools and was filtered out, as well as cluster 4 which was enriched for ribosomal proteins suggesting cellular debris contamination.

### **Transcriptional signature analysis using primary white and brown preadipocytes**

Primary white and brown preadipocytes were isolated from the neck region of 6 individuals and subjected to microarray gene expression profiling (Tews et al., 2014). Data was accessed using GEO Accession # GSE54280 and analyzed using GEO2R. Differentially expressed genes were identified using a white vs brown test. List of genes enriched in white or brown primary preadipocytes were defined as signatures for respective cell-types and used as input in Vision to assign score to in vitro preadipocytes analyzed in our study (Fig. 1).

### **Identifying number of lncRNAs detected as a function of sequencing depth**

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

### **Silhouette coefficient analysis**

Both scRNA-seq and snRNA-seq datasets for brown preadipocytes were subsampled as described above. snRNA-seq dataset was further randomly subset to have the same number of total barcodes as scRNA-seq. At each sequencing depth, top 20 principal components were calculated using Seurat's standard pipeline. Three resolution coefficients based on the Silhouette index, Calinski Harabasz index,

and Davies Bouldin index were then calculated based on Euclidean distance between cells in the PCA space using the clusterCrit package in R. For analyzing cluster separation resolution between brown cluster 1 and 2 as a function of UMI count, the same analysis was performed except that downsampling was performed to have the same number of UMI rather than reads between scRNA-seq and snRNA-seq dataset using the downsampleMatrix command in the DropletUtils package in R (Lun et al., 2019).

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

For integrating scRNA-seq white preadipocyte (day-0) & white-adipocyte (day-20) and snRNA-seq white preadipocyte (day-0) & white-adipocyte (day-20) datasets (a total of 4 datasets), we first created a single anndata object with UMI count-matrices from each dataset as input. Each of the four UMI matrices were generated by processing the originals FASTQ files (no downsampling of reads), and subset to only have high-quality barcodes as outlined in Methods section # 5. During concatenation, each of the four datasets was assigned a "batch" key. The concatenated anndata object was then used as input to scvi-tools for integration using the commands outlined in the tutorial here: [https://docs.scvi-tools.org/en/stable/user\\_guide/notebooks/harmonization.html](https://docs.scvi-tools.org/en/stable/user_guide/notebooks/harmonization.html). The output of following these steps was a 10-dimensional latent space with batch-corrected embedding for cells from each of the four datasets. UMAP visualization was then generated using the RunUMAP command in Seurat with the 10-dimensional latent space as input. The dendrogram was generated using the BuildClusterTree command in Seurat, which constructs a phylogenetic tree relating the 'average' cell from each identity class. Tree is estimated based on the eigenvalue-weighted euclidean distance matrix constructed in latent-space.

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

command. Then, the identified marker genes were tested for differential expression using the FindAllMarkers command, with a logFC threshold of 0.25.

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

For integration with Seurat, scRNA-seq white preadipocyte (day-0) & white-adipocyte (day-20) and snRNA-seq white preadipocyte (day-0) & white-adipocyte (day-20) datasets were defined as individual batches (a total of 4 batches). Integration was performed following the commands outline in this tutorial: [https://satijalab.org/seurat/articles/integration\\_introduction.html](https://satijalab.org/seurat/articles/integration_introduction.html).

### **RNA spot counting analysis for smFISH**

For analysis, imaged were first cropped to only have one cell per field of view. Then, the Find Foci plugin was used using the GUI, with Max Size = 100, Peak parameter = 0.2, Max peaks = 1000, and Minimum size = 5 (Advanced settings). For image binarization, a manually selected value was used for thresholding, with visibly best performance in selecting RNA spots as foreground over background. With total RNA spots calculated for each cell, gaussian mixture model fitting was performed using the mclust package in R (Scrucca et al., 2016), and negative binomial mixture model fitting was performed using the fitNB command in SIBERG package (Tong et al., 2013) in R.

## **REFERENCES**

Carbon, S., Douglass, E., Dunn, N., Good, B., Harris, N.L., Lewis, S.E., Mungall, C.J., Basu, S., Chisholm, R.L., Dodson, R.J., et al. (2019). The Gene Ontology Resource: 20 years and still GOing strong. *Nucleic Acids Res.* 47, D330–D338.

Chen, T.N., Gupta, A., Zalavadia, M.D., and Streets, A. (2020).  $\mu$ CB-seq: microfluidic cell barcoding and sequencing for high-resolution imaging and sequencing of single cells. *Lab Chip* 20, 3899–3913.

Fleming, S., Marioni, J., and Babadi, M. (2019). CellBender remove-background: a deep generative model for unsupervised removal of background noise from scRNA-seq datasets. *BioRxiv* 791699.

Gayoso, A., Shor, J., Carr, A.J., Sharma, R., and Pe'er, D. (2019). JonathanShor/DoubletDetection: HOTFIX: Correct setup.py installation.

Hafemeister, C., and Satija, R. (2019). Normalization and variance stabilization of single-cell RNA-seq data using regularized negative binomial regression. *Genome Biol.* 20, 1–15.

Keenan, A.B., Torre, D., Lachmann, A., Leong, A.K., Wojciechowicz, M.L., Utti, V., Jagodnik, K.M., Kropiwnicki, E., Wang, Z., and Ma'ayan, A. (2019). ChEA3: transcription factor enrichment analysis by orthogonal omics integration. *Nucleic Acids Res.* 47, W212–W224.

Lun, A.T.L., Riesenfeld, S., Andrews, T., Dao, T.P., Gomes, T., and Marioni, J.C. (2019). EmptyDrops: Distinguishing cells from empty droplets in droplet-based single-cell RNA sequencing data. *Genome Biol.* 20, 63.

Merrick, D., Sakers, A., Irgebay, Z., Okada, C., Calvert, C., Morley, M.P., Percec, I., and Seale, P. (2019). Identification of a mesenchymal progenitor cell hierarchy in adipose tissue. *Science (80-. )*. 364, eaav2501.

Mi, H., Muruganujan, A., Ebert, D., Huang, X., and Thomas, P.D. (2019). PANTHER version 14: More genomes, a new PANTHER GO-slim and improvements in enrichment analysis tools. *Nucleic Acids Res.* 47, D419–D426.

Ruan, H. Bin (2020). Developmental and functional heterogeneity of thermogenic adipose tissue. *J. Mol. Cell Biol.* 12, 775–784.

Scrucca, L., Fop, M., Murphy, T.B., and Raftery, A.E. (2016). Mclust 5: Clustering, classification and density estimation using Gaussian finite mixture models. *R J.* 8, 289–317.

Tews, D., Schwar, V., Scheithauer, M., Weber, T., Fromme, T., Klingenspor, M., Barth, T.F., Möller, P., Holzmann, K., Debatin, K.M., et al. (2014). Comparative gene array analysis of progenitor cells from human paired deep neck and subcutaneous adipose tissue. *Mol. Cell. Endocrinol.* 395, 41–50.

The Gene Ontology Consortium (2000). Gene ontology: Tool for the unification of biology. *Nat. Genet.*

25, 25–29.

Tong, P., Chen, Y., Su, X., and Coombes, K.R. (2013). SIBER: Systematic identification of bimodally expressed genes using RNAseq data. *Bioinformatics* 29, 605–613.

Wolock, S.L., Lopez, R., and Klein, A.M. (2019). Scrublet: Computational Identification of Cell Doublets in Single-Cell Transcriptomic Data. *Cell Syst.* 8, 281-291.e9.