Single-cell transcriptional networks in differentiating preadipocytes suggest drivers associated w...

This work was made openly accessible by BU Faculty. Please share how this access benefits you. Your story matters.

| Version                                      | Published version                                                                                                                                 |
|----------------------------------------------|---------------------------------------------------------------------------------------------------------------------------------------------------|
| Citation (published version):                | Alfred K Ramirez, Simon N Dankel, Bashir Rastegarpanah, Weikang Cai, Ruidan Xue, Mark Crovella, Yu-Hua Tseng, C Ronald Kahn, Simon Kasif. 2020. "Single-cell transcriptional networks in differentiating preadipocytes suggest drivers associated with tissue heterogeneity."
  Nat Commun, Volume 11, Issue 1, pp. 2117 - ?.
  https://doi.org/10.1038/s41467-020-16019-9 |

https://hdl.handle.net/2144/41945
Boston University
Single-cell transcriptional networks in differentiating preadipocytes suggest drivers associated with tissue heterogeneity

Alfred K. Ramirez1,2, Simon N. Dankel3,4, Bashir Rastegarpanah5, Weikang Cai1, Ruidan Xue1,7, Mark Crovella5,6, Yu-Hua Tseng1, C. Ronald Kahn1✉ & Simon Kasif2,6✉

White adipose tissue plays an important role in physiological homeostasis and metabolic disease. Different fat depots have distinct metabolic and inflammatory profiles and are differentially associated with disease risk. It is unclear whether these differences are intrinsic to the pre-differentiated stage. Using single-cell RNA sequencing, a unique network methodology and a data integration technique, we predict metabolic phenotypes in differentiating cells. Single-cell RNA-seq profiles of human preadipocytes during adipogenesis in vitro identifies at least two distinct classes of subcutaneous white adipocytes. These differences in gene expression are separate from the process of browning and beiging. Using a systems biology approach, we identify a new network of zinc-finger proteins that are expressed in one class of preadipocytes and is potentially involved in regulating adipogenesis. Our findings gain a deeper understanding of both the heterogeneity of white adipocytes and their link to normal metabolism and disease.
Adipose tissue is a heterogeneous organ and composed of several cell types, including mature adipocytes, preadipocytes, stem cells, endothelial cells, and various blood cells. At least three distinct developmental types of adipocytes have been identified: white, brown, and beige adipocytes. Different adipose depots have distinct physiological functions associated with their anatomical location and cell composition. Accumulation of visceral (intra-abdominal) white adipose tissue is associated with insulin resistance and metabolic syndrome, whereas accumulation of subcutaneous adipose tissue is metabolically benign and may be even associated with increased insulin sensitivity. Determining the mechanisms for these phenotypic differences could lead to development of therapies for diabetes, obesity, and their associated morbidities.

The formation of new adipocytes, adipogenesis, is a well-characterized cellular model of differentiation. The process depends on several key transcriptional regulators including PPARγ, CEBPβ, and various Krüppel-like Factors (KLFs). Advances in technologies such as high-throughput sequencing have revealed additional transcriptional regulators ranging from zinc-finger proteins (ZFPs/ZNFs) to long non-coding RNAs. Each of the different types of adipocytes also have distinct transcription factors, such as ZFP423 for white adipocytes and EBF2 for brown/beige adipocytes, or developmental markers including TBX15 and various homeobox genes.

A central challenging question in research of metabolic disease is the contributions of genetic, epigenetic and environmental factors. A variant of this question is whether disease risk is intrinsic to a subset of fat cells that interact with environmental stresses in disease pathogenesis. This is especially relevant since recent studies in mice have suggested developmental heterogeneity even among white adipocytes. We attempted to address this question for human white fat using a synergistic application of several methodologies: single-cell transcriptional profiling coupled with clonal expansions in relevant tissues, network analysis and data integration using gene signatures.

Single-cell RNA sequencing is an ideal technique to profile gene expression of heterogeneous cell populations obtained from a single tissue such as blood or brain. However, the mechanisms driving cellular heterogeneity are not well understood. We aim to develop methods that could lead to a better understanding of the potential drivers of cellular heterogeneity. We describe a network algorithm and apply it to single-cell data in reference to what we integrated with new RNA seq data in clonally expanded preadipocyte cell lines obtained to generate metabolic phenotypes. Such phenotyping is challenging in single-cell gene expression profiling. The algorithm and additional analysis reveals a gene network that is predicted to be associated and potentially drive adipocyte differentiation and heterogeneity.

Results

scRNA-seq reveals at least two distinct cell populations in differentiating preadipocytes. Raw data from single-cell RNA sequencing of white adipocytes undergoing differentiation was obtained from Soumillon et al. For these experiments, primary human abdominal subcutaneous preadipocytes derived from a single donor were seeded at 100% or 80% confluence (day 0), by day 3 (pink), the differentiating preadipocytes separated into two distinct clusters, and this separation remained evident at day 7 (red). The experiment beginning at 80% confluence and followed over 14 days of differentiation showed a similar two-cluster separation, although in this experiment there was greater separation at the day 0, i.e., preconfluent, cells, which then became somewhat mixed at days 1 and 2 before separating again at days 3–14 (Supplementary Fig. 1A).

PAGODA identified a set of gene signatures associated with the transcriptional heterogeneity during adipogenesis in the 7-day (Fig. 1b) and the 14-day experiment (Supplementary Fig. 1B). The genes within the highest-ranking gene sets reflected the stage of differentiation (Supplementary Fig. 1A–E). The t-SNE plots, however, showed cell clustering within each stage of differentiation. In order to determine the genes associated with clusters observed on the t-SNE plots, we performed differential gene expression between the left and right cluster for each day of analysis. First, the 7-day experiment beginning at 100% confluency in the preadipocytes (day 0), differential expression analysis between the two cell clusters produced genes primarily related to protein synthesis (Supplementary Fig. 3A). At day 3, the separation was primarily driven by genes involved in remodeling the extracellular matrix (Supplementary Fig. 3B). Finally, at day 7, in addition to the previous differences in gene expression, differential expression of genes related to metabolism appeared. These include gene sets related to oxygen consumption and protein metabolism (Fig. 1c). Genes in each of these metabolic gene sets were higher in the right cluster compared to the left, suggesting that the cells in the cluster were more metabolically active.

Whereas some differences between the 7-day and 14-day experiments were evident due to the differences in experimental design (Supplementary Fig. 4A–C), both showed a similar pattern with early differences in the two clusters being mainly in growth-related genes and late differences mainly in metabolic genes. Independently, markers of adipocytes or preadipocyte differentiation, such as FABP4 and PDGFRα (Fig. 1d), were not differentially expressed between the clusters. The left cluster, as compared to the right cluster, also showed higher expression of TBX15, a mesodermal developmental gene, which has previously been shown to be a marker of glycolytic adipocytes in mice.

There was no differential expression or no detection of markers of beige or brown adipocytes, including TMEM26 or UCP1, between clusters, indicating that this heterogeneity was not due to a mixture of white and beige preadipocytes.

Integrative analysis predicts differences in glucose uptake and extracellular acidification. To corroborate these gene expression differences and their predictive association with metabolic phenotypes, we sought to determine whether white preadipocytes from a single adipose depot clustered by metabolic phenotypes. A total of 35 clonally expanded white preadipocytes from human neck white subcutaneous tissue were generated as previously described. Gene expression profiles were obtained by RNA-seq and analyzed with PAGODA. Principal component analysis revealed at least two clusters with the 35 clonally expanded cell lines (Fig. 2a). The gene sets capturing significant aspects of heterogeneity reflected differences in cell adhesion (GO:0034330), response to type I interferon (GO:0017357), and RNA processing (GO:0006395) (Fig. 2b). Some genes within these aspects linked to heterogeneity of white adipocytes in mature fat, such as...
Cadherin-6 (CDH6) and Nucleophosmin (NPM1), showed different levels of expression in cells within the same clusters, suggesting subclusters within both clusters A and B (Fig. 2b).

Phenotypic profiling was performed on each clonal cell lines before or after differentiation to obtain profiles in extracellular acidification (ECAR), oxygen consumption (OCR), glucose uptake, and adipogenic capacity (PPARG expression after differentiation) (Fig. 2b). The correlation analysis was performed with phenotype profiles and the five significant genes sets identified between Cluster A and B, and showed that GO:00071357 was correlated with glucose uptake, whereas the other gene sets were not correlated to any of the measured phenotypes (Supplementary Fig. 9). When compared to the differentially expressed genes from the day 0 clusters (Fig. 1a), we observed approximately 100 genes shared (Fig. 2c). Comparing the seven most extreme clones in the gene expression of cluster A against the seven most extreme clones of cluster B showed a difference in glucose uptake and PPARG expression (i.e., the capacity of differentiation into adipocytes) (Fig. 2d).

We then sought to associate heterogeneous populations identified in single-cell RNA seq with quantitative measurements in clonally expanded cell lines. Using the clonal gene expression data, phenotype-gene correlation vectors were created between every gene and each measured phenotype (Fig. 2e). These gene-
phenotype correlations were projected to the single-cell data to determine whether any gene signatures would preferentially mark the heterogeneous populations of single cells. The results indicated that the clusters showed differences in the expression of genes correlated to extracellular acidification and glucose uptake across multiple days (Fig. 2f, Supplementary Fig. 5A–H).

We then proceeded to develop a network approach to identify the potential drivers of heterogeneity in single-cell populations that revealed a network of zinc-finger proteins associated with adipogenesis-resistant cells. Combining clustering gene expression with network analysis is routinely applied to identify biological processes associated with disease-related phenotypes.
in different cell types. We developed an algorithm to detect a network of highly-activated genes that potentially code for protein complexes (protein–protein interaction (PPI) networks) in single-cell data (Fig. 3a). Our method identified at least three distinct protein networks: (1) the ubiquitin-proteasome complex (residual error = 2583), (2) the ribosome (residual error = 2587), and (3) an uncharacterized protein network (residual error = 2593). As the two higher-ranking networks captured known biology in adipogenesis, we focused on the module consisting of 30 connected genes in the PPI (Fig. 3b). These genes were enriched for proteins containing KRAB domains and C2H2 zinc-finger regions (Fig. 3c).

This module had a small tightly-connected subnetwork consisting of largely-uncharacterized zinc-finger proteins (ZNFs). Although the target genes of these ZNFs are not known, one of the highest-scoring motifs in the promoters of these genes was Interferon Regulator Factor 1 (IRF1) (Supplementary Fig. 7), a transcription factor which has been previously identified as a negative regulator of adipogenesis. Analysis of the top 100 cells with the detected module over the time-course of adipogenesis revealed that expression of several of these ZNFs (ZNF264, ZNF490, ZNF587, and ZNF714) decreased as the level of expression of FABP4, a marker of adipocyte differentiation, increased (Fig. 3d). Interestingly, the ZNF cluster did not significantly vary in the gene expression profiles of preadipocytes cultures across the time course of preadipocyte differentiation into adipocytes (Fig. 3e). This ZNF module is only detected at the single-cell level and not at the multi-cellular population level, suggesting the module is found in a subpopulation of preadipocytes and not in progenitor cells of other mesodermal fates (Supplementary Fig. 8A–C).

**Discussion**

Understanding the heterogeneity of white adipocytes is an important question in adipocyte biology since it is known that different patterns of fat distribution are linked to insulin resistance, different metabolic states, and the risk of diseases like type 2 diabetes and metabolic syndrome. However, this type of analysis has been hindered by technological challenges (e.g., sorting of large and fragile adipocytes) and reliance on predefined markers, and is especially difficult to apply to human fat tissue. Here, we have addressed the question by unbiased analysis of single-cell RNA-seq of human subcutaneous preadipocytes undergoing differentiation into white adipocytes. The results show the presence of at least two distinct clusters of cells in both the preadipocyte and adipocyte stages. Differential gene expression reveals that the early differentiation differences are due to genes related to cell cycle, protein synthesis, and growth-associated pathways. Late-stage differences are primarily due to genes related to glycolytic metabolism, which was also confirmed by phenotypic profiling in clonally-derived white preadipocyte cell lines. Lastly, a subset of cells within one of the clusters is marked by the activation of a set of zinc-finger proteins, alluding to additional potential subsets of preadipocytes including a pool of adipocyte precursors resistant to adipogenesis.

Previous research in mouse models have identified subpopulations of preadipocytes marked by the expression of certain genes such as *Scal* and *Mfy5* (refs 13,27). It was also shown that different fat depots had different proportions of these *Scal*+ or *Mfy5*+ subpopulations. For example, preadipocytes derived from a *Mfy5*+ lineage comprised 9% of perigonadal white adipose tissue and as much as 50% of anterior subcutaneous white adipose tissue. Additional studies found that knocking out PTEN in Myf5+ cells resulted in lipodystrophy, suggesting that these different preadipocyte populations have a functional importance and could serve as important therapeutic targets. Historically, research into human white adipose heterogeneity was limited to stromal vascular cells sorted by pre-defined markers.

Recently, there have been several efforts to utilize single-cell sequencing to determine the cell types in human white adipose tissue. For example, Schwalie et al. identified a group of anti-adipogenic stromal cells, marked by *CD142* expression, which they termed Aregs. Consistent with this, our data suggest *CD142* (gene F3) may be differentially expressed between the day 0 clusters in the two separate differentiation experiments (FDR ~0.44 and ~0.20). It is worth noting that these studies, as well as ours, have thus far been limited to preadipocytes and adipose tissue derived from healthy humans. The single-cell heterogeneity of adipose in diseased states, such as diabetes and metabolic syndrome, could reveal new mechanistic insights.

In adipose tissue, one of the primary challenges is to determine the transcriptional programs that give rise to different cells, potentially controlled by uncharacterized transcription factors such as zinc-finger proteins. Zinc-finger proteins are one of the largest families of DNA-binding proteins. The zinc-finger proteins ZNF264 and ZNF490 have been shown to be DNA-binding, and ZFP423 has been shown to be a marker of preadipocytes in the mouse. In addition, the expression of ZNF714 is higher in visceral adipose tissue from insulin-resistant obese populations compared to insulin-sensitive obese populations. In the context of developmental biology, transcriptional and cellular heterogeneity have been shown to increase tissue robustness as a response mechanism to different perturbations. In adipose tissue, the robustness is needed to maintain...
energy homeostasis in response to perturbations such as diet, frequency of eating, and many other factors. Earlier studies have made observations consistent with at least two types of preadipocytes defined by several differences including adipogenic capacity, replicative capacity, and apoptosis susceptibility. Thus, these ZNFs may be involved in the transcriptional regulation of white adipocyte differentiation contributing to tissue heterogeneity and robustness.

Previous work in systems biology developed numerous methods to model a biological system as a linear system. In the context of gene expression, this framework intuitively implies that each gene in a system is a gate whose output can be modeled as a linear combination of outputs of other gates with some error. While solving such systems is relatively tractable, the noise, complexity and lack of linearity in biological systems makes these efforts challenging. Here, we introduce a useful technical insight to integrate biological intuitions with computational modeling. Instead of assuming the entire system is linear, we assume the system can be decomposed into connected network modules where the system can be approximated by a low dimensional linear factorization of the union of these network modules. The possibility of such factorization without the constraint of connected PPI modules has been previously suggested in other biological systems. The network constraint adds significant technical challenge to obtaining such a model. The advent of modern omics, single-cell technologies and machine learning opens the door to methods utilizing integrative algebraic network decompositions with numerous applications in systems biology.

Our work provides preliminary support for the hypothesis that heterogeneity might be coded for early in adipogenesis. Since heterogeneity of mature adipocytes has been associated with metabolic disease risk, we can speculatively hypothesize that metabolic disease might be intricately tied to early and specific differentiation trajectories. These developmental trajectories might be affected by genetics. Alternatively, pre-metabolic disease states via yet to be fully understood signaling or communication mechanisms may modify transcriptional control in preadipocytes, driving increase in cell populations with higher disease risk. Thus, this research direction may lead to therapeutic directions repurposing specific differentiation paths.
In summary, we have demonstrated a synergistic combination of three systems biology frameworks (single-cell transcriptional profiling, integration and joint analysis with clonal gene expression data and network modeling). We have provided evidence of the presence of at least two types of human white subcutaneous adipocytes. These adipocytes show large differences in metabolic gene sets and are not accounted for by differences in brown or beige fat. We have also shown that a subset of preadipocytes are marked by the expression of a ZNF network, which may be involved in regulating white adipocyte differentiation. Changes in the balance of these types of white adipocytes may have important physiological consequences in adipose tissue and whole-body metabolism. Dissecting intrinsic cellular heterogeneity in white pre-adipocytes and adipose tissue may significantly impact our understanding of the role of differentiation programs, genetics and environment in metabolic disease risk.

**Methods**

**Base-level analysis of single-cell RNA-seq.** Single-cell RNA-seq profiles of abdominal subcutaneous human preadipocytes undergoing adipogenesis were obtained as previously described. In these data, the primary human preadipocytes were taken from commercially-available liposaparies (single donor; Life Technologies) sorted by fluorescence-activated cell sorting (FACS) based on the positive selection (i.e., express the protein) for CD29, CD44, CD73, CD90, CD105, CD16, and negative selection (i.e., do not express) against CD4, CD14, CD31, CD54, Lin1. The cells were maintained in 2% reduced serum growth media for up to 3 passages (MesnPro, Life Technologies). For differentiation, cells were grown to either 80% or 100% confluence and then incubated with differentiation media (StemPro adipogenesis differentiation media, Life Technologies) for 7 or 14 days in two separate experiments as described. For single-cell sorting, culture plates were treated with trypsin, and the released cells pelleted by centrifugation at 1000 rpm for 5 min. Pellets were resuspended with Dulbecco’s phosphate-buffered saline (DPBS), stained with Hoescht 33342, and individual cells FACS sorted using the above markers into prepared 384-well plates. Library generation and RNA-seq analysis was performed as described.

For data analysis, we filtered out cells containing less than 1000 detected genes. On average, each cell contained ~1600 detected genes. We used PAGODA to analyze the expression profiles in experiment ‘Di’ (differentiation beginning at 80% confluency, 2092 cells after QC) and experiment ‘B’ (differentiation beginning at 100% confluency, 4319 cells after QC) (GSE53638). PAGODA finds non-redundant genes sets which first principal components (i.e., aspects) are oversampled. Significant aspects were chosen by FDR < 0.05. Single-cell differential gene expression was performed with the package SCDE. As the t-SNE plots revealed two distinct clusters, we performed differential gene expression between the left and right cluster in each day and in each separate experiment.

**Cell culture and adipogenesis.** No specific human studies approval was needed for the current study as the human cell lines used in this study were existing biobank specimens that had been collected under an approved human studies protocol for the research previously published in . Human preadipocyte cell lines were obtained from superficial fat depots in the neck of patients undergoing cervical spine surgery and immortalized with stable transfection of human telomerase (hTERT) as previously described. Cells were passaged at 80% confluency and maintained in growth media (DMEM-H, 10% FBS, 1% penicillin-streptomycin). For differentiation, cells were grown to 100% confluence and incubated with pre-induction media (DMEM-H, 2% FBS, 1% Pen-Strep, 0.2% norepinephrine, 500 nM insulin, 2 nM T3, 1 μM rosiglitazone) for 6 days, followed by the induction media (DMEM-H, 2% FBS, 1% Pen-Strep, 0.2% norepinephrine, 500 nM insulin, 500 μM IBMX, 2 nM T3, 1 μM rosiglitazone, 33 μM biotin, 17 μM pantothenate, 100 nM dexamethasone, and 30 μM indomethacin) for 7–14 days.

**Total RNA isolation and RNA-seq of clonal preadipocytes.** Clonal human neck-derived subcutaneous preadipocyte cell lines were maintained in culture and grown to 100% confluence before RNA harvest. Total RNA was harvested and purified with Trizol reagent. RNA quality was measured on the Agilent 2100 Bioanalyzer to ensure RNA quality was greater than 8.0. To construct libraries for sequencing, we used the NEBNext Ultra Directional Library Prep Kit and Poly-A selection (Kit (New England Biolabs). Library enrichment and multiplexing was performed using the NEBNext High-Fidelity PCR Master Mix (14 cycles of PCR) and NEBNext Multiplex Oligos. The cDNA libraries were multiplexed on the Illumina HiSeq 2000 and 2500 to generate raw reads and deposited on GEO (GSE128253). Raw 50 bp paired-end reads were aligned with STAR (v2.3.10a) to the human genome build hg19. After alignment, the genome v5.4 was used for HTseq-sequencing analysis. Libraries were prepared in technical duplicates from each donor and analyzed using HTseq-seq (v0.5.4a). Analyses were carried out in R and various packages including DESeq2 (ref. 51) and SCDE. Subject-specific gene expression profiles were normalized with SCDE. To estimate the number of clusters in the subject-corrected expression profiles, we performed k-medoids clustering and selected the k that maximized the average silhouette width (k = 2–10). This resulted in phenotypically-distinct subclusters within these two large clusters defined by gene expression, we performed hierarchical clustering and cut the tree to yield five clusters with average cluster size of 7 cell lines.

**Extracellular flux profiling.** Oxygen consumption rate (OCR) and extracellular acidification rate (ECAR) were measured using a Seahorse XF24 (Seahorse Bioscience). Cells were counted with a Cellometer® Vision (Nexcelom Bioscience) and seeded in 24-well Seahorse Cell Culture Microplates (6000 cells/well). Each clonal cell line was seeded in duplicate or triplicate (one 24-well plate per subject). 1.5 h after the seeding in 100 μl medium (4.5 g/L glucose DMEM, 10% FBS, 1% penicillin/streptomycin), an additional 150 μl medium were added to ensure optimal distribution of cells. The metabolic profiles of the cells were analyzed by measurement of oxygen consumption and ECAR every 24–31 day different from day 0. 48 h after the measurements above. 2 h prior to measurements, 150 μl of culture medium was removed from each well. The cells were washed once with 900 μl of XF Assay Medium (1 g/L glucose, 2 mM L-glutamine, 2 mM Na pyruvate), followed by the addition of XF Assay Medium to a final volume of 500 μl and incubation at 37 °C with no CO2. Basal oxygen consumption and extracellular acidification were measured for 30 min. DNA was extracted by washing cells with PBS, incubating in 100 μl 50 mM NaOH for 30 min, heating at 96 °C for 5 min, and adding 25 μl 1 M Tris pH 6.5. DNA concentrations were measured by NanoDrop. The OCR and ECAR values of preadipocytes were normalized to the DNA contents in the corresponding well. The OCR and ECAR values of adipocytes were measured in 12-well plates (50,000 cells/well, duplicates) and normalized per cell. For measurement, 48 h after the measurements described above. 400 μl of each sample were transferred to scintillation tubes and 4 ml Cytoscent scintillation fluid were added before rigorous shaking and measurement in a Beckman LS6000 scintillation counter. Disintegration in 3 min performed by the following program: 95 °C for 10 s and 45 °C for 45 s. Melting curve analysis (increments of 0.5 °C from 60 °C to 95 °C) confirmed specificity of the primers. Mean target mRNA levels based on technical triplicates were calculated by the ΔACT method and normalized to the mRNA level of HKPR. Primers for PPARγ (Forward: 5′-GAA ACC GAT TCC TTC ACT GAT-3′; Reverse: 5′-TCA AAG GAC TGG GAG TGG IC-3′) and HKPR (Forward 5′-TGA AAA GGC CCC CAC GAA G-3′; reverse 5′-TGA CAG ATG GCC AGA CAA GTA C-3′) were ordered from IDT.

**Subnetwork detection by integrating the protein-protein interactome and module detection.** In this paper, we deploy a network algorithm to analyze single-cell RNA-seq data in differentiating preadipocytes. We believe that the integration of network analysis and single-cell RNA-seq is a very promising direction for understanding both modularity and heterogeneity of gene expression in complex biological processes. Our hypothesis is that the gene expression state of the cell is, in part, influenced by an underlying biological network defined by the interaction between proteins. This hypothesis is consistent with a number of publications in system biology. This framework is different from using a regulatory network that captures protein-DNA interactions. Our underlying network is a wiring diagram where the nodes are proteins and the edges are literature-reported experimentally-observed interactions. It should be noted that we assume these literature-reported interactions would be present in all cells expressing the proteins, i.e., they are tissue independent. Given our hypothesis, each gene expression profile of a single-cell defines a network snapshot. Previous publications using a regulatory network framework for modeling gene expression attempted to recover the causal regulators of gene expression.
expression. In contrast, we aim to produce a relatively simple representation of the activity of protein complexes, transcriptional modules or signaling pathways in different cells. In the simplest case, the goal is to explain the system behavior by predicting up-regulated and down-regulated modules and their activity in heterogeneous cell populations during adipogenesis.

The input to our algorithm is:

(a) A matrix containing the single-cell expression profiles in experiment ‘D1’ was first transformed to log scale (log2(UMI + 1)) and then transformed to z-scores.
(b) The input protein–protein interaction network was the union of all interactions in the network database Pathway Commons v5 (network named ‘Pathway Commons.5.4.ELIXIR_BINARY_SIF’).

Integrating the two data sources, we can interpret the N single-cell gene expression profiles as a set of N network snapshots in different time points of adipocyte differentiation. Each node (protein in the network) is associated with a score that corresponds to the differences in gene expression of the corresponding gene in this cell as compared to distribution of all values of this gene in the data. In this analysis, the edges are unweighted, although the method can be generalized to edges associated with confidences as describe in STRING and related databases.

The single-cell network snapshots can be described as a matrix of row vectors \( Y_1, Y_2, \ldots, Y_n \) where each \( Y_i \) corresponds to a gene expression profile of a single cell. This yields a matrix where the columns are nodes in the network and the rows are network snapshots (gene expression profiles of cells). The goal of the algorithm is to represent the behavior of the network during differentiation as the activity profiles of relatively small connected components of the network. Within each network component, the activity is captured by a relatively simple explanation. For example, the activity of the network could be explained by a connected component in which the expression of the genes is relatively high on average, which could define a module associated or driving a state of differentiation. Alternatively, we may be able to identify a protein complex that consists of subunits of the same protein, such as the ribosome (a protein complex) and its many subunits (the proteins RSPS2, RSPS3, etc.).

Unlike standard statistical analysis that relies on methods such as piecewise linear regression or clustering we add two relevant biological constraints:

(a) The genes in each module are connected in the PPI network (module).
(b) The activity of each genes across the cells can be mathematically described as a linear function of multiple connected networks that the gene appears in. For example, a single gene can be controlled by two or more networks: one or more of which could be activated and the other(s) repressing.

This approach extends the class of complex systems we can explain relatively succinctly with connected modules using matrix algebra. Pathway enrichment of the resulting networks was performed with DAVID. Further detail is provided in Supplementary Methods and code is available at https://github.com/mcrovella/CG-Demo.

**Analysis of ZNF expression in a single-cell RNA seq dataset of adipogenesis, chondrogenesis, and osteogenesis.** We analyzed a single-cell RNA seq data set containing profiles of the differentiation of fibroblasts into adipocytes, chondrocytes, and osteocytes (GSE37521). The normalized, pre-processed RNA-seq data were downloaded from GEO and analyzed in R via the packages GEOquery, Limma, and BioBase. For comparing the ZNF expression to FABP4, a marker for white adipocyte differentiation, each node (protein in the network) is associated with a score that corresponds to the differences in gene expression of the corresponding gene in this cell as compared to distribution of all values of this gene in the data. In this analysis, the edges are unweighted, although the method can be generalized to edges associated with confidences as describe in STRING and related databases.

**Motif analysis.** Motif analysis of the genes was performed with the R package motifmatch. The promoter region was defined as the region between 2000 bp of the transcription start site in the human genome build hg19. A significant match was set to \( p < 0.01 \). The transcription factor database was from JASPAR via the R package JASPAR2018.

**Statistical analyses.** Statistical tests were performed as indicated in the figure captions. Briefly, single-cell differential gene expression was performed as described in the R package SCDE. Statistical significance was assessed by an unpaired, two-tailed t-test or an ANOVA followed by Bonferroni correction. All statistical analyses were performed in R.

**Reporting summary.** Further information on research design is available in the Nature Research Reporting Summary linked to this article.

**Data availability**

The RNA-seq data on clonally expanded human neck-derived white preadipocytes are available at GSE12853. The single-cell RNA-seq data of white preadipocytes during in vitro differentiation are available at GSE35638. The single-cell RNA-seq data of adipocyte-derived stem cells during differentiation into three different mesenchymal lineages are available at GSE37521.

**References**

1. Jones, M. D. & Tseng, Y. H. Deciphering adipose tissue heterogeneity. Ann. NY Acad. Sci. 1411, 5–20 (2017).
2. Ussar, S. et al. ASC-1, PA2T, and P2RX5 are cell surface markers for white, beige, and brown adipocytes. Sci. Transit. Med. 6, 247ra103 (2014).
3. Tchkonia, T. et al. Mechanisms and metabolic implications of regional differences among fat depots. Cell Metab. 17, 644–56 (2013).
4. Tran, T. T. & Kahn, C. R. Transplantation of adipose tissue and stem cells: role in metabolism and disease. Nat. Rev. Endocrinol. 6, 195–213 (2010).
5. Farmer, S. R. Transcriptional control of adipocyte formation. Cell Metab. 4, 263–273 (2006).
6. Wang, Q. A. et al. Distinct regulatory mechanisms governing embryonic versus adult adipocyte maturation. Nat. Cell Biol. 17, 1099–1111 (2015).
7. Zhao, Y.-Y., Li, S., Wang, G.-X., Yu, Q. & Liu, J. D. A long noncoding RNA transcriptional regulatory circuit drives thermogenic adipocyte differentiation. Mol. Cell 55, 372–82 (2014).
8. Longo, M. et al. Epigenetic modifications of the Zfp/ZNF423 gene control murine adipogenic commitment and are dysregulated in human hypertrophic gene set. Diabetologia 61, 369–380 (2018).
9. Shao, M. et al. Zfp423 maintains white adipocyte identity through suppression of the beige cell thermogenic gene article Zfp423 maintains white adipocyte identity through suppression of the beige cell thermogenic gene program. Cell Metab. 23, 1167–1184 (2016).
10. Rajakamari, S. et al. EB2F determines and maintains brown adipocyte identity. Cell Metab. 17, 562–574 (2013).
11. Lee, K. Y. et al. Tbx15 defines a glycolytic subpopulation and white adipocyte heterogeneity. Diabetes 66, 2822–2829 (2017).
12. Gesta, S. et al. Mesodermal developmental gene Tbx15 impairs adipocyte differentiation and mitochondrial respiration. Proc. Natl Acad. Sci. USA. 108, 2771–2776 (2011).
13. Lee, K. Y. et al. Developmental and functional heterogeneity of white adipocytes within a single fat depot. EMBO J. 38, 1–19 (2019).
14. Campbell, J. N. et al. A molecular census of arcuate hypothalamus and median eminence cell types. Nat. Neurosci. 20, 484–496 (2017).
15. Han, X. et al. Mapping the mouse cell atlas by microwell-seq. Cell 172, 1047–11107e17 (2018).
16. Soumillon, M., Cachiairelli, D., Senraru, S., van Oudenaarden, A. & Mikkelsen, T. S. Characterization of directed differentiation by high-throughput single-cell RNA-Seq. Preprint at bioRxiv https://doi.org/10.1101/003236 (2014).
17. Fan, J. et al. Characterizing transcriptional heterogeneity through pathway and gene set overdispersion analysis. Nat. Methods 13, 241–244 (2016).
18. Xué, R. et al. Clonal analyses and gene profiling identify genetic biomarkers of the thermogenic potential of human brown and white preadipocytes. Nat. Med. 21, 760–768 (2015).
19. Takeda, K. et al. Retinoic acid mediates visceral-specific adipogenic defects of human adipose-derived stem cells. Diabetes 65, 1164–1178 (2016).
20. Hepler, C. et al. Identification of functionally distinct fibro-inflammatory and adipogenic stromal subpopulations in visceral adipose tissue of adult mice. eLife 7, 1–36 (2018).
21. Liu, M. et al. Network-based analysis of affected biological processes in type 2 diabetes models. PLoS Genet. 3, e9 (2007).
22. Chuang, H.-Y., Lee, E., Liu, Y.-T., Lee, D. & Ideker, T. Network-based classification of breast cancer metastases. Nat. Syst. Biol. 3, 1–10 (2007).
23. Vanunu, O., Magger, O., Ruppin, E., Shlomi, T. & Sharan, R. Associating genes and protein complexes with disease via network propagation. PLoS Comput. Biol. 6, e1000641 (2010).
24. Cerami, E. G. et al. Pathway commons, a web resource for biological pathway data. Nucleic Acids Res. 39, D685–D690 (2011).
25. Eguchi, J. et al. Interferon regulatory factors are transcriptional regulators of adipogenesis. Cell Metab. 7, 86–94 (2008).
26. Lee, M.-J., Wu, Y. & Fried, S. K. Adipose tissue heterogeneity: implication of depot differences in adipose tissue for obesity complications. Mol. Asp. Med. 34, 1–11 (2013).
27. Shan, T. et al. Distinct populations of adipogenic and myogenic Myf5-lineage progenitors in white adipose tissues. J. Lipid Res. 54, 2214–2224 (2013).
28. Sanchez-Gurmaches, J. & Guertin, D. A. Adipocyte lineages: tracing back the origins of fat. Biochem. Biophys. Acta 1842, 340–351 (2013).
29. Sanchez-Gurmaches, J. et al. PTEN loss in the Myf5 lineage redistributes body fat and reveals subsets of white adipocytes that arise from Myf5 precursors. Cell Metab. 16, 348–362 (2012).

30. Rada, T., Reis, R. L. & Gomes, M. E. Distinct stem cell subpopulations isolated from human adipose tissue exhibit different chondrogenic and osteogenic differentiation potential. Stem Cell Rev. 7, 64–76 (2011).

31. 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 (2017).

32. Schwale, P. C. et al. A stromal cell population that inhibits adipogenesis in mammalian fat depots. Nature 559, 103–108 (2018).

33. Min, S. Y. et al. Diverse repertoire of human adipocyte subtypes develops from transcriptionally distinct mesenchymal progenitor cells. Proc. Natl. Acad. Sci. USA 116, 17979–17989 (2019).

34. Liu, X. et al. Data descriptor: single-cell RNA-seq of cultured human adipose-derived mesenchymal stem cells. Sci. Data 6, 1–8 (2019).

35. Merrick, D. et al. Identification of a mesenchymal progenitor cell hierarchy in adipose tissue. Science 364, eaay2501 (2019).

36. Weirauch, M. T. & Hughes, T. R. A catalogue of eukaryotic transcription factor types, their evolutionary origin, and species distribution. Subcell Biochem. 52, 25–73 (2011).

37. Schmitges, F. W. et al. Multiparameter functional diversity of human C2H2 zinc finger proteins. Genome Res. 26, 1742–1752 (2016).

38. Gupta, R. K. et al. ZIP423 expression identifies committed preadipocytes and locates adipose endothelial and perivascular cells. Cell Metab. 15, 230–239 (2012).

39. Crujeiras, A. B. et al. Genome-wide DNA methylation pattern in visceral adipose tissue differentiates insulin-resistant from insulin-sensitive obese subjects. Transl. Res. 178, 13–24.e5 (2016).

40. Paszek, P. et al. Population robustness arising from cellular heterogeneity. Proc. Natl. Acad. Sci. USA 107, 11644–11649 (2010).

41. Wilson, A., Hodgson-Garms, M., Frith, J. I. & Genever, P. Multiplicity of mesenchymal stromal cells: finding the right route to therapy. Front. Immunol. 10, 1–8 (2019).

42. Torres-Padilla, M. E. & Chambers, I. Transcription factor heterogeneity in adipocyte progenitors leads to diverse capacities for replication, adipogenesis, and apoptosis varies among fat depots. Am. J. Physiol. Endocrinol. Metab. 288, E267–E277 (2005).

43. Pruett-Marcususs, B. et al. From heterogeneity to plasticity in adipose tissues: site-specific differences. Exp. Cell Res. 312, 727–736 (2006).

44. Ortj, D. J., Thiele, I. & Palson, B. O. What is flux balance analysis? Nat. Biotechnol. 28, 245–248 (2010).

45. Jeong, H., Mason, S. P., Barabási, A. L. & Oltvai, Z. N. Lethality and centrality in protein networks. Nature 411, 41–42 (2001).

46. Segal, E., Wang, H. & Koller, D. Discovering molecular pathways from protein interaction and gene expression data. Bioinformatics 19 (Suppl 1), i264–i267 (2003).

47. Kharchenko, P. V., Silverstein, L. & Scadden, D. T. Bayesian approach to single-cell RNA-seq analysis of large gene lists using DAVID bioinformatics resources. Nat. Protoc. 4, 44–57 (2009).

48. Borchie, M. E. et al. limma powers differential expression analyses for RNA-sequencing and microarray studies. Nucleic Acids Res. 43, e47 (2015).

49. Gentleman, R. et al. Bioconductor: open software development for computational biology and bioinformatics. Genome Biol. 5, R80 (2004).

50. 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 (2017).

Acknowledgements

The authors would like to thank Evimaria Tseri for contributions to the development of the network algorithm; Drs Jonathan M. Dreyfuss and Hui Pan from Joslin Diabetes Center DRC Genomics and Bioinformatics Core for assistance and advice regarding data analysis; Mary Ellen Fitzpatrick for her assistance with the Gene Cluster Computing Environment; and Guoxiao (Grace) Wang for her assistance with CRISPR knockouts. A.K.R. was supported by NIH grant T32 DK007260-37. B.R. and M.C. were supported by NSF IIS-1421759 and NSF CNS-1618207. This work was supported in part by NIH grant R01DK082659 to C.R.K. We also received support from the Joslin NIH-funded DRC (P30DK036836).

Author contributions

A.K.R., C.R.K. and S.K. designed and provided the overall oversight of the study. A.K.R., S.N.D., B.R., M.C., W.C., C.R.K. and S.K. wrote the paper. A.K.R., R.X. and S.N.D. performed the experiments. A.K.R. analyzed gene expression data. A.K.R., B.R., M.C. and S.K. contributed to the design of the single-cell analysis methods. A.K.R., Y.H.T., C.R.K. and S.K. analyzed and discussed the data and biological results. B.R. and M.C. wrote the code for the subnetwork detection. This work was supported in part by NIH grants P30DK036836, R01DK0835659 and R37031036.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary information is available for this paper at https://doi.org/10.1038/s41467-020-16019-9.

Correspondence and requests for materials should be addressed to C.R.K. or S.K.

Peer review information Nature Communications thanks the anonymous reviewer(s) for their contribution to the peer review of this work. Peer reviewer reports are available.

Reprints and permission information is available at http://www.nature.com/reprints

Publisher’s note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article’s Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article’s Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit http://creativecommons.org/licenses/by/4.0/.

© The Author(s) 2020