THLBB2020_PUB002: Single nuclei RNA sequencing dataset

Cohort: Twin Study/TWINFAT
Cohort PI: Kirsi Pietiläinen
Data set accession number in THL Biobank: THLBB2020_PUB002

Publication information

Enhancing droplet-based single-nucleus RnA-seq resolution using the semi-supervised machine learning classifier DIEM
Marcus Alvarez, Elior Rahmani, Brandon Jew, Kristina M. Garske, Zong Miao, Jihane N. Benhammou, Chun Jimmie Ye, Joseph R. Pisegna, Kirsi H. Pietiläinen, Eran Halperin & Päivi Pajukanta

Corresponding author: Päivi Pajukanta

Published article's journal and reference information

Scientific Reports (2020) 10:11019
DOI: https://doi.org/10.1038/s41598-020-67513-5

Data set availability information

Data set availability statement in paper: "The human single nucleus RNA-seq datasets generated and analyzed during the current study are available upon request."
Data available through THL Biobank: human single nucleus RNA-seq datasets. The single nucleus RNA-seq data cannot be linked with any other twin cohort data available at THL Biobank for purpose of biobank research.

Data set accession number at THL Biobank: THLBB2020_PUB002
Other datasets used in the publication are not made available through THL Biobank. If you are interested in any other data used in the publication, please contact the corresponding author.

Participant and transcriptomics description

Participants
Adipose tissue samples were obtained from 6 participants of the Twin Study.

Adipose tissue biopsy collection
All biopsy collections took place during the fasting (12 h) state. The adipose tissue was taken in sterile conditions under local anesthesia (lidocaine). The subcutaneous adipose tissue biopsies were taken from superficial abdominal adipose tissue near the umbilicus using a surgical technique or through a needle biopsy. Tissue specimens were immediately snap-frozen in liquid nitrogen and stored in liquid nitrogen until further analysis.

Single-nucleus RNA-seq of human subcutaneous adipose tissue
Frozen subcutaneous adipose tissue was processed separately for each of the 6 samples. Tissue was minced over dry ice and transferred into ice-cold lysis buffer consisting of 0.1% IGEPAL, 10 mM Tris–Hcl, 10 mM NaCl, and 3 mM MgCl2. After a 10 min incubation period, the lysate was gently homogenized using a dounce homogenizer and filtered through a 70 µm MACS smart strainer (Miltenyi Biotec #130-098-462) to remove debris. Nuclei were centrifuged at 500×g for 5 min at 4 °C and washed in 1 ml of resuspension

buffer (RSB) consisting of 1X PBS, 1.0% BSA, and 0.2 U/µl RNase inhibitor. We further filtered nuclei using a 40 µm Flowmi cell strainer (Sigma Aldrich # BAH136800040) and centrifuged at 500×g for 5 min at 4 °C. Pelleted nuclei were re-suspended in wash buffer and immediately processed with the 10X Chromium platform following the Single Cell 3′ v2 protocol. After library generation with the 10X platform, libraries were sequenced on an Illumina NovaSeq S2 at a sequencing depth of 50,000 reads per cell. Reads were aligned to the GRCh38 human genome reference with Gencode v26 gene annotations23 using the 10X CellRanger 2.1.1 pipeline. A custom pre-mRNA reference was generated to account for unspliced mRNA by merging all introns and exons of a gene into a single meta-exon.

Description of single nuclei RNA sequencing datasets

Twin Study (TWINFAT) adipose tissue single nuclei transcriptomics datasets
*Type: single nuclei RNA sequencing (snRNAseq)
*Project: Twin Study
*Owner: THL Biobank
*Original date recieved at THL Biobank: 11.5.2021
*Original file names:

barcodes.tsv
This file contains the oligonucleotide barcode tags that link mRNA reads originating from the same droplet. This lists all (unfiltered) oligonucleotide barcodes generated by 10X. The barcodes correspond to the columns in the matrix.mtx expression file.

diem_fltr_ids.txt
This file contains the barcodes that were filtered using DIEM. These barcodes tag the droplets that contain nuclei, with empty droplet barcodes removed.

features.tsv
This file contains the Gencode v26 Ensembl IDs and affiliated HGNC symbols for which the reads were mapped to. The genes correspond to the rows in the matrix.mtx expression file.

matrix.mtx
This file contains the sparse matrix with the count data in the Matrix Market Exchange format (https://math.nist.gov/MatrixMarket/formats.html). All 10X barcodes are included. Non-zero UMI counts of the matrix are associated with a feature (row) and a barcode (column). After the header, the first column contains the feature (row) index, the second column contains the barcode (column) index, and the third index contains the UMI count.

*Original study name: TWINFAT
*Sequencing year: 2018
*Measured at: University of Helsinki
*Coverage: sequencing depth of 50,000 reads per cell
*Reads: adipose tissue RNA reads length of 75 bp and skeletal muscle RNA reads length of 69 bp
*Reference Genome: GRCh38 human genome reference with Gencode v26 gene annotations
*Sequencing: libraries prepared by using 10X Chromium platform following the Single Cell 3′ v2 protocol. After library generation with the 10X platform, libraries were sequenced on an Illumina NovaSeq S2.
*Possible comments regarding the use of snRNAseq or any other additional information:

Tools used in data processing
- 10X CellRanger 2.1.1 pipeline