THLBB2021_PUB001: RNAseq data set

Cohort: Twin Study/TWINFAT
Data set accession number in THL Biobank: THLBB2021_PUB001

Publication information

Molecular pathways behind acquired obesity: adipose tissue and skeletal muscle multiomics in monozygotic twin pairs discordant for BMI
Birgitta W. van der Kolk[1#*], Sina Saari[1#], Alen Lovric[2,3], Muhammad Arif [2], Marcus Alvarez[4], Arthur Ko[5], Zong Miao[4,6], Navid Sahebekhtiari[1], Maheswary Muniandy[1], Sini Heinonen[1], Ali Oghabian[1], Riikka Jokinen[1], Sakari Jukarainen[7], Antti Hakkarainen[8], Jesper Lundbom[8,9], Juho Kuula[8,10], Per-Henrik Groop[11-14], Taru Tukiainen[7], Nina Lundbom[8], Aila Rissanen[1], Jaakko Kaprio[7], Evan G. Williams[15], Nicola Zamboni[15], Adil Mardinoglu[2,16], Päivi Pajukanta[4,6,17], Kirsi H. Pietiläinen[1,18§]

Published article's journal and reference information

Cell Reports Medicine 2, 100226, April 20, 2021
DOI: https://doi.org/10.1016/j.xcrm.2021.100226

Participant and transcriptomics description

Twin participants
The twin pairs were recruited from population-based longitudinal Twin-studies of the University of Helsinki, based on their responses to questions regarding weight and height.

Here, the study included 49 monozygotic twin pairs discordant for BMI (within-pair difference, ΔBMI ≥ 2.5 kg/m^2), from two age groups (27–42 years old and 57–69 years old) and for whom adipose tissue and skeletal muscle multiomics data were available. Twenty-seven pairs were female. Eight pairs were discordant and four pairs were concordant for T2DM, while other pairs reported no T2DM.

Adipose tissue and muscle biopsy collection
All biopsy collections took place during the fasting (12 h) state. The adipose tissue (49 twin pairs) and skeletal muscle biopsies (44 twin pairs) were 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. A needle muscle biopsy was taken from the vastus lateralis muscle. An incision was made through the skin, after which the sample was taken using a 5-mm Bergström needle. Both tissue specimens were immediately snap-frozen in liquid nitrogen and stored in liquid nitrogen until further analysis.

Adipose tissue and muscle transcriptomics
For total RNA extraction, ~250 mg of frozen adipose tissue and skeletal muscle biopsies was used. RNA was extracted using the AllPrep RNA, DNA, miRNA Universal Kit (Qiagen, Nordic, Solletuna, Sweden) with a DNase I (Qiagen) digestion according to the manufacturer's instructions. The resulting DNA-free RNA samples were analyzed for quality on a 2100 Bioanalyzer according to the manufacturer's protocol (Agilent Technologies, Santa Clara, CA, USA). The RNA integrity numbers (RINs) were calculated automatically using the 2100 expert software prior to RNA sequencing.

For the RNA sequencing, the libraries were prepared using Illumina Stranded mRNA preparation and sequenced the samples on the Illumina HiSeq2000 platform to an average sequence depth of 40 to 50 M paired-ends. Adipose tissue RNA reads were sequenced to a length of 75 bp and skeletal muscle RNA reads

to a length of 69 bp. The reads were aligned from the samples against the human reference genome hg38 using STAR v2.5.2b and its two-pass protocol with Gencode v26 annotations *(Dobin A, Davis CA, Schlesinger F, et al. (2012) STAR: ultrafast universal RNA-seq aligner. Bioinformatics 29:15–21).* An RNAseq sample was required to include at least 20 M uniquely mapped reads and the correct Library strandedness. The sample quality was assessed using Picard (Broad Institute Picard Tools). To avoid mixing up samples, the genotype array and RNAseq data were matched using exonic SNPs with VerifyBamID *(Jun G, Flickinger M, Hetrick KN, et al. (2012) Detecting and estimating contamination of human DNA samples in sequencing and array-based genotype data. Am. J. Hum. Genet. 91:839–848).* Read counts were calculated using HTSeq v0.6.1p (*Anders S, Pyl PT, Huber W (2015) HTSeq--a Python framework to work with high-throughput sequencing data. Bioinformatics 31:166–169).*

Differential expression analysis

Differential expression analyses between co-twins were performed using the R package Limma (Voom) [76]. Prefiltering of genes was applied by retaining genes that have at least ten counts in 70% of samples and only selecting protein coding genes. We identified the altered genes, proteins and metabolites that associated with the heavier compared with the leaner co-twin within each sample. We adjusted the regression model for the sex, age group and diabetes status of the individuals. To ensure pairwise comparisons between twins, we used the family ID as the identifier. We corrected p values for multiple testing (using the Benjamini and Hochberg method) and, for the adipose tissue transcriptomics, we considered FDR $p < 0.05$ statistically significant. For adipose tissue metabolomics and for all skeletal muscle omics analyses, applying multiple test corrections proved statistically too conservative and hampered the biological interpretation. Therefore, we considered nominal $p < 0.05$ significant for these datasets.

Description of RNA sequencing datasets

Twin Study (TWINFAT) adipose tissue and muscle transcriptomics

*Type: RNA sequencing (RNAseq)
*Project: Twin Study
*Owner: THL Biobank
*Original date received at THL Biobank: 10.5.2021
*Original file names:

adipose_tissue_count.tsv
RNA seq data for adipose tissue

adipose_tissue_sample_annotation.tsv
This file contains all the relevant phenotype information for the RNAseq data for adipose tissue. It has the discordance specified ('leaner_or_heavier' variable) as well as the variable 'batch' which contains all the information about the covariate adjustments which were applied in the paper.

muscle_tissue_count.tsv
RNA seq data for muscle tissue

muscle_tissue_sample_annotation.tsv
This file contains all the relevant phenotype information for the RNAseq data for skeletal muscle. It has the discordance specified ('leaner_or_heavier' variable) as well as the variable 'batch' which contains all the information about the covariate adjustments which were applied in the paper.

*Original study name: TWINFAT
*Sequencing year: 2018
*Measured at: University of California, Los Angeles (UCLA)
*Coverage: average sequence depth of 40 to 50 M paired-ends
*Reads: adipose tissue RNA reads length of 75 bp and skeletal muscle RNA reads length of 69 bp

*Reference Genome: GRCh38
*Sequencing: libraries prepared by using Illumina Stranded mRNA preparation and sequencing on the Illumina HiSeq2000 platform (Illumina Inc., San Diego, CA, USA)
*Possible comments regarding the use of RNAseq or any other additional information:

Tools used in data processing
-Reads from the samples against the human reference genome hg38 using STAR v2.5.2b and its two-pass protocol with Gencode v26 annotations
-Read counts were calculated using HTSeq v0.6.1p
-The sample quality was assessed using Picard