Genomics Data 2 (2014) 246–248

Contents lists available at ScienceDirect

Genomics Data

journal homepage: http://www.journals.elsevier.com/genomics-data/

Data in Brief

Transcriptomics profiling of human SGBS adipogenesis

Mafalda Galhardo [a], Lasse Sinkkonen [a], Philipp Berninger [b], Jake Lin [c], Thomas Sauter [a,*], Merja Heinäniemi [d,*]

[a] Life Sciences Research Unit, University of Luxembourg, 162a Avenue de la Faïencerie, L-1511 Luxembourg, Luxembourg
[b] Biozentrum, Universität Basel and Swiss Institute of Bioinformatics, Klingelbergstrasse 50-70, 4056 Basel, Switzerland
[c] Luxembourg Centre for Systems Biomedicine, University of Luxembourg, House of Biomedicine, 7 Avenue des Hauts-Fourneaux, L-4362 Esch/Alzette, Luxembourg
[d] Institute of Biomedicine, School of Medicine, University of Eastern Finland, FI-70120 Kuopio, Finland

ARTICLE INFO

Article history:
Received 29 June 2014
Accepted 13 July 2014
Available online 7 August 2014

Keywords:
Adipocyte
Differentiation
Metabolism

ABSTRACT

Obesity is an ever-growing epidemic where tissue homeostasis is influenced by the differentiation of adipocytes that function in lipid metabolism, endocrine and inflammatory processes. While this differentiation process has been well-characterized in mice, limited data is available from human cells. Applying microarray expression profiling in the human SGBS pre-adipocyte cell line, we identified genes with differential expression during differentiation in combination with constraint-based modeling of metabolic pathway activity. Here we describe the experimental design and quality controls in detail for the gene expression and related results published by Galhardo et al. in Nucleic Acids Research 2014 associated with the data uploaded to NCBI Gene Expression Omnibus (GSE41352).

Specifications [standardized info for the reader]; where applicable, please follow the Ontology for Biomedical Investigations: http://obi-ontology.org/page/Main_Page	
Organism/cell line/tissue	Human/SGBS pre-adipocyte/adipose tissue
Sex	Male
Sequencer or array type	Illumina HumanHT-12V3.0 expression beadchip
Data format	Raw and analyzed data
Experimental factors	Time point of differentiation to adipocytes. Cells were cultured 2 days in serum-free OF medium prior to differentiation
Experimental features	Time series of differentiation (20 samples, 7 time points in duplicate or triplicate). SGBS pre-adipocyte cells originate from patient with SGB syndrome. See Wabitsch M. et al. Int J Obes Relat Metab Disord. 2001 for more details on differentiation protocol and origin of cells.
Consent	See Wabitsch M. et al. Int J Obes Relat Metab Disord. 2001 [2]
Sample source location	See Wabitsch M. et al. Int J Obes Relat Metab Disord. 2001 [2]

Direct link to deposited data

http://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE41352.

* Corresponding authors.
E-mail addresses: thomas.sauter@uni.lu (T. Sauter), merja.heinaniemi@uef.fi (M. Heinäniemi).

Experimental design, materials and methods

Cell differentiation and experimental design

Gene expression levels during adipocyte differentiation were obtained by stimulating the SGBS pre-adipocyte cell line with a mix of differentiation inducing compounds and collecting RNA samples at 0, 4, 8 and 12 h and on days 1, 3 and 12 of adipogenesis for hybridization on Illumina HT-12 microarrays. Triplicate samples were prepared following the differentiation protocol modified from [2] (exception is 12 h time point that has only duplicate samples) as shown in Table 1. SGBS cells differentiate within 10–12 days as determined by microscopic analysis (Oil red O staining). At this time point the cells are filled with small sized lipid droplets and are most responsive, whereas at later time points (20 days) the lipid droplets fuse and cells are less active (personal communication, Dr. Martin Wabitsch).

Specifically, SGBS cells were cultured in Dulbecco's modified Eagle's medium (DMEM)/Nutrient Mix F12 (Gibco) containing 8 mg/L biotin, 4 mg/L pantothenate, 0.1 mg/mg streptomycin and 100 U/mL penicillin (OF medium) supplemented with 10% FBS in a humidified 95% air/5% CO_2 incubator. The cells were seeded into 10 cm plates, which were coated with a solution of 10 μL/mL fibronectin and 0.05% gelatine in phosphate-buffered saline. Confluent cells were cultured in serum-free OF medium for 2 days followed by stimulation to differentiate with OF media supplemented with 0.01 mg/mL human transferrin, 200 nM T3, 100 nM cortisol, 20 nM insulin, 500 μM IBMX and 100 nM rosiglitazone (Cayman Chemicals). After day 4, the differentiating cells

http://dx.doi.org/10.1016/j.gdata.2014.07.004

Table 1
Microarray sample description from the SGBS pre-adipocyte differentiation experiment (GSE41578). GEO sample identifiers are presented for the 20 samples prepared, as well as their differentiation time point and replicate number.

Sample name	GSM identifier	Title	Time	Replicate
Sample 1	GSM1015366	SGBS_day0_1	0 h	1
Sample 2	GSM1015367	SGBS_day0_2	0 h	2
Sample 3	GSM1015368	SGBS_day0_3	0 h	3
Sample 4	GSM1015369	SGBS_4h_1	4 h	1
Sample 5	GSM1015370	SGBS_4h_2	4 h	2
Sample 6	GSM1015371	SGBS_4h_3	4 h	3
Sample 7	GSM1015372	SGBS_8h_1	8 h	1
Sample 8	GSM1015373	SGBS_8h_2	8 h	2
Sample 9	GSM1015374	SGBS_8h_3	8 h	3
Sample 10	GSM1015375	SGBS_12h_1	12 h	1
Sample 11	GSM1015376	SGBS_12h_2	12 h	2
Sample 12	GSM1015377	SGBS_day1_1	Day 1	1
Sample 13	GSM1015378	SGBS_day1_2	Day 1	2
Sample 14	GSM1015379	SGBS_day1_3	Day 1	3
Sample 15	GSM1015380	SGBS_day3_1	Day 3	1
Sample 16	GSM1015381	SGBS_day3_2	Day 3	2
Sample 17	GSM1015382	SGBS_day3_3	Day 3	3
Sample 18	GSM1015383	SGBS_day12_1	Day 12	1
Sample 19	GSM1015384	SGBS_day12_2	Day 12	2
Sample 20	GSM1015385	SGBS_day12_3	Day 12	3

were kept in OF media supplemented with 0.01 mg/mL human transferrin, 100 nM cortisol and 20 nM insulin.

Gene expression analysis

Total RNA was extracted using TriSure (Bioline). 1 mL of TriSure was added per a confluent 10 cm dish to lyse the cells. RNA was extracted with 200 μL chloroform and precipitated from the aqueous phase with 400 μL isopropanol by incubating at −20 °C overnight. The longer isopropanol incubation allowed the precipitation of microRNAs and other small RNAs from the same samples. The total RNA samples were processed according to the manufacturer instructions to prepare cDNA that was hybridized on microarrays (Turku Centre for Biotechnology, Microarray and Sequencing Facility, Turku, Finland). Total RNA integrity was confirmed using an Agilent 2100 Bioanalyzer (Agilent Technologies, Santa Clara, CA, USA).

Data processing and normalization

The raw data files were processed and quality controlled using the R/Bioconductor lumi package. Raw and normalized expression values are available via GEO (GSE41352). Control probe data was included and used to background correct the signal values with the lumiB "bgAdjust" method. We provide this data and sample data in a format that is directly compatible with the lumi analysis package through our web resource at http://systemsbiology.uni.lu/idare.html. The data was then transformed with the "vst" method and normalized with robust spline normalization (rsn) method. The probe intensity value distribution and sample relation are plotted in Figs. 1 and 2, with sample naming described in Table 1. No outliers were detected based on data value range at this step and the samples clustered according to the biological sample group. The code that can be used to download processed data from GEO or to process them from the files that we provide through our website is available (see Additional Data File 1).

Statistical analysis

The negative probe signals were used to filter non-expressed genes. Only genes that had a detection p-value < 0.05 within all samples of at least one time point were selected for statistical analysis, resulting in a total of 12756 detected probes. The statistical analysis was performed using the R/Bioconductor limma package. The F-test was used to assess

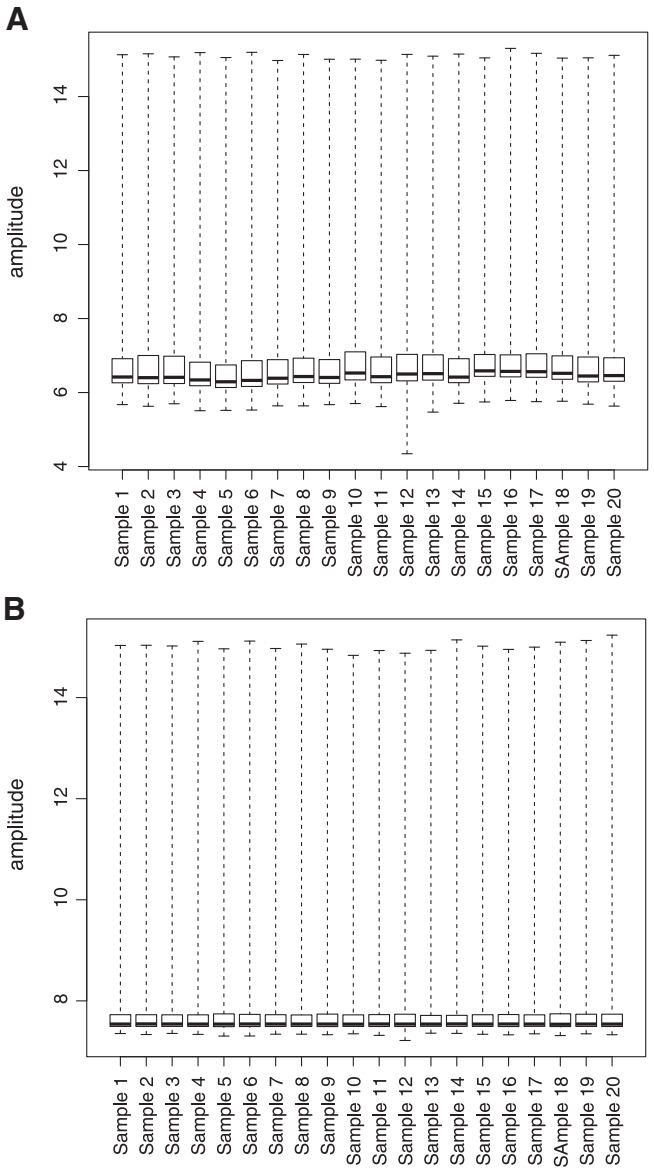

Fig. 1. Probe intensity plots for the 20 SGBS differentiation samples in GSE41578. A) Box plots of raw probe intensities. B) Box plots of normalized probe intensities indicate the absence of outliers and comparable data mean intensities.

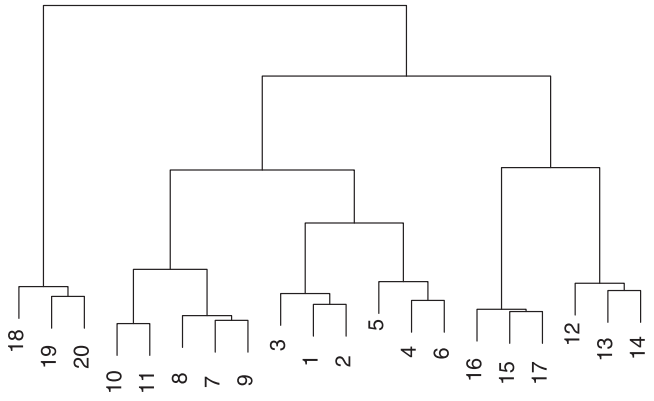

Fig. 2. Hierarchical clustering of the SGBS differentiation microarray samples. The dendrogram shows high similarity between replicates and grouping based on differentiation time progression.

significance of overall dynamic response over the differentiation while a two-tailed t-test was performed to compare specific time points to day 0 undifferentiated cells. In both analyses Benjamini–Hochberg adjusted p-value < 0.01 was considered statistically significant. In total, 1936 Refseq transcripts changed their expression more than 2-fold up or down during the differentiation time series. The code that can be used to filter non-expressed genes and to perform the statistical analysis is available (see Additional Data File 1).

Several of these genes were metabolic genes, represented by 2-fold more differentially expressed genes compared to other gene categories with similar numbers of genes (extracted from the GO Online SQL Environment, as of 12th of August 2013: cell projection, envelope, locomotion and receptor activity).

Analysis of metabolic genes in Recon1

The annotation data from Recon1 was obtained and checked against the current EntrezGene and Refseq annotations (hg19 Refseq; Feb 02 2012). The reaction to gene mappings were updated with current gene IDs (see Table S1). Withdrawn IDs and pseudogenes present a difficulty in the Recon1 annotation. As there were only few such genes (see Table S1), they were left out from visualizations and assigned expression level 0 in modeling. *LPIN1* was missing and due to its central role in adipocytes, it was added to the triacylglycerol pathway reaction catalyzed by *Phosphatidic Acid Phosphatase* (PPAP).

The expression profiles of metabolic genes (from Recon 1 [3]) or TFs (from [4]) were clustered for visualization using self-organizing maps (GEDI software [5]) and AutoSOME [6] as instructed in the tool documentation. The settings to reproduce the results presented in [1] were the following: GEDI grid size was adjusted based on input gene number and settings were tuned in order to minimize data missing grid points (gene density map) (see Table S2). AutoSOME GUI was used following the description in the manual without data filtering. Clustering was done for columns (samples) on "precision" mode, with the "Fuzzy Cluster Network" option and network visualization with Cytoscape [7]. Enriched pathways of the human metabolic reconstruction [3] were determined using a hypergeometric test.

A consistent version of the generic human metabolic model Recon1 [3] was used as modeling platform for prediction of network activity distributions. The Recon1 model was downloaded from the BiGG database [8] (04.11.11) and the consistent version was derived using the function "reduceModel" from the COBRA toolbox 2.0 [9], which resulted in the exclusion of 1273 reactions (34%) of the initial model (Table S3). To include the microarray data as soft-constraints for reaction activity prediction, the probes were mapped to Entrez Gene IDs. First, continuous log2 normalized expression values for the probes were discretized into three categories: lowly expressed (−1), moderately expressed (0) and highly expressed (1) based on the mean expression ± 0.5 * standard deviation cutoffs across all arrays. Then, one unique discretized value per gene was selected taking the rounded discretized mean of all probes for a gene. Each gene was then assigned to the Recon1 reaction based on gene–protein-reaction associations.

Discussion

Here we describe a time series dataset of human SGBS pre-adipocyte differentiation. This dataset is comprised of whole transcriptome gene expression profiling data derived using the Illumina BeadArrays. We demonstrated differential expression that was particularly prevalent among metabolic genes. Moreover, discretization of the metabolic gene expression levels allowed using them as soft-constrains for metabolic activity modeling. Further, this dataset is part of a GEO SuperSeries (GSE41578) and we have used it in combination with next-generation sequencing data and microRNA expression profiles to associate putative regulators to the metabolic genes in [1]. To further analyze the data in an integrative manner, we introduced gene metanodes and the web portal IDARE (Integrated Data Nodes or Regulation) in [1] for interactive data exploration of various data types within the metabolic network context, available at http://systemsbiology.uni.lu/idare.html, including a detailed user guide. The data could be similarly analyzed to interrogate the regulation of other pathways. Results from the data have increased our understanding of human adipogenesis.

Supplementary data to this article can be found online at http://dx.doi.org/10.1016/j.gdata.2014.07.004.

Acknowledgments

We would like to acknowledge the following funding sources: Fonds National de la Recherche Luxembourg [AFR, AM2c and C08/BM/01]; funding from the University of Eastern Finland and the Fondation du Pélican de Mie et Pierre Hippert-Faber under the aegis of Fondation de Luxembourg.

References

[1] M. Galhardo, L. Sinkkonen, P. Berninger, J. Lin, T. Sauter, M. Heinäniemi, Integrated analysis of transcript-level regulation of metabolism reveals disease-relevant nodes of the human metabolic network. Nucleic Acids Res. 42 (2014) 1474–1496.
[2] M. Wabitsch, R.E. Brenner, I. Melzner, M. Braun, P. Möller, E. Heinze, K.M. Debatin, H. Hauner, Characterization of a human preadipocyte cell strain with high capacity for adipose differentiation. Int. J. Obes. Relat. Metab. Disord. 25 (2001) 8–15.
[3] N.C. Duarte, S.A. Becker, N. Jamshidi, I. Thiele, M.L. Mo, T.D. Vo, R. Srivas, B.Ø. Palsson, Global reconstruction of the human metabolic network based on genomic and bibliomic data. Proc. Natl. Acad. Sci. U. S. A. 104 (2007) 1777–1782.
[4] M. Heinäniemi, M. Nykter, R. Kramer, A. Wienecke-Baldacchino, L. Sinkkonen, J.X. Zhou, R. Kreisberg, S.A. Kauffman, S. Huang, I. Shmulevich, Gene-pair expression signatures reveal lineage control. Nat. Methods 10 (2013) 577–583.
[5] G.S. Eichler, S. Huang, D.E. Ingber, Gene Expression Dynamics Inspector (GEDI): for integrative analysis of expression profiles. Bioinformatics 19 (2003) 2321–2322.
[6] A.M. Newman, J.B. Cooper, AutoSOME: a clustering method for identifying gene expression modules without prior knowledge of cluster number. BMC Bioinforma. 11 (2010) 117.
[7] M. Smoot, K. Ono, J. Ruscheinski, P.-L. Wang, T. Ideker, Cytoscape 2.8: new features for data integration and network visualization. Bioinformatics 27 (2011) 431–432.
[8] J. Schellenberger, J.O. Park, T.C. Conrad, B.Ø. Palsson, BiGG: a Biochemical Genetic and Genomic knowledgebase of large scale metabolic reconstructions. BMC Bioinformatics 11 (2010) 213.
[9] J. Schellenberger, R. Que, R.M.T. Fleming, I. Thiele, J.D. Orth, A.M. Feist, D.C. Zielinski, A. Bordbar, N.E. Lewis, S. Rahmanian, J. Kang, D.R. Hyduke, B.Ø. Palsson, Quantitative prediction of cellular metabolism with constraint-based models: the COBRA Toolbox v2.0. Nat. Protoc. 6 (2011) 1290–1307.