



Identification and functional analysis of differentially expressed genes related to obesity using DNA microarray

J.Y. Du, H. Yang, D.R. Tian, Q.M. Wang and L. He

Department of Anatomy, Tianjin Medical University, Tianjin, China

Corresponding author: D.R. Tian

E-mail: jianyingderunddt@hotmail.com

Genet. Mol. Res. 13 (1): 64-72 (2014)

Received November 27, 2012

Accepted May 16, 2013

Published January 8, 2014

DOI <http://dx.doi.org/10.4238/2014.January.8.5>

ABSTRACT. We looked for differentially expressed genes at different stages of preadipocyte differentiation and examined their functions, based on DNA microarrays of preadipocytes obtained from healthy subjects undergoing cosmetic liposuction. We downloaded gene expression profile GSE25910 from the Gene Expression Omnibus database and identified the differentially expressed genes with packages in R language. The selected differentially expressed genes were further analyzed using bioinformatics methods. Comparing gene expression at different stages of preadipocytes differentiation, we found that expression of 668 and 1007 genes were altered in middle and late stages compared with the early stage, respectively. Function analysis revealed that the differentially expressed genes were mainly related to fatty acid metabolic processes in the former two stages.

Key words: Preadipocytes; Microarray; Differentially expressed gene; Function enrichment analysis

INTRODUCTION

Obesity is an abnormal accumulation of body fat, usually 20% or more over an individual's ideal body weight (He et al., 1995). With the increased intake of high-fat food today, the incidence of obesity is increasing, and a variety of diseases accompanying obesity, such as coronary heart disease, type-2 diabetes, coronary hypertension, sleep apnea syndrome, fatty liver, respiratory syndrome and cardiovascular diseases, also threaten human health (Zhang et al., 1994). Besides smoking, obesity has been regarded as the second preventable death risk factor and one of four major health problems with AIDS, drug abuse and alcoholism (Hoffmann et al., 2001).

Before the mesoderm pluripotent stem cells differentiate into mature adipocytes, the cells in different stages are called preadipocytes. The quantity of adipocytes has a great influence on obesity, and the differentiation of preadipocytes has a direct impact on the quantity of mature adipocytes (Hsu and Yen, 2007). Therefore, preadipocytes are closely related to obesity. The disorder of preadipocyte differentiation will lead to changes in the quantity of mature adipocytes, and it will be harmful to human health regardless of the increase or decrease in adipocytes (Kume et al., 2005).

In this study, we screened differentially expressed genes in the different differentiation stages on the basis of the gene expression profile of preadipocyte samples. We tried to identify the marker genes that can control the quantity of adipocytes and analyzed the enrichment functions of the differentially expressed genes in the differentiation of preadipocytes. Current studies indicate that obesity is a polygenic and multifactorial metabolic disorder (Maffei et al., 1996). Also, the most important reason causing obesity is abnormal fat metabolism. Therefore, besides weight management, diet improvement and increased exercise, the development of therapy targeted at the specific genes and metabolic pathways is expected to become an effective auxiliary means in the future.

MATERIAL AND METHODS

Affymetrix microarray data

The dataset GSE25910 (Arner et al., 2012) was downloaded from Gene Expression Omnibus database (<http://www.ncbi.nlm.nih.gov/geo/>), which was deposited by Arner and colleagues. Subcutaneous adipose tissue was obtained from healthy subjects undergoing cosmetic liposuction ($n = 12$). Preadipocytes were isolated and differentiated them into adipocytes *in vitro*. Then microarray analyses were performed at day 4/5 (early), 8 (middle) and 12 (late) of differentiation.

The platform information was GPL6244 [HuGene-1_0-st] Affymetrix Human Gene 1.0 ST Array [transcript (gene) version]. The microarray probe annotation information was from Affymetrix containing all information of Affymetrix ATH1 (25K) microarray. The original file and the annotation information file of the platform were downloaded as well.

Data preprocessing and analysis

The gene expression data of the three differentiation stages were divided into two groups: early versus middle and early versus late. The missing values were imputed using the method of KNN in R language (Howrylak et al., 2009) and the original data were standardized with mid-value clustering (Troyanskaya et al., 2001). The LIMMA package was then applied to analyze the gene expression profile of two groups (Fujita et al., 2006), and the Bayesian ap-

proach was used for multiple test correction (Smyth, 2005). FDR (false discovery rate) < 0.05 and $|\logFC| > 1$ were selected as the threshold for screening differentially expressed genes.

Differentially expressed genes in functional categories

The sequences of differentially expressed genes from two groups were compared with the COG (cluster of orthologous groups of proteins) database using BLASTX (Altschul et al., 1990; Tatusov et al., 2001) (E-value $< 1e-05$) to obtain the functional annotation and COG functional classification in categories of cellular component, molecular function and biological process. We then calculated the relative increment rate (R) of the differentially expressed genes in each enriched term using equation (1) to find the category with the largest relative increment rate during preadipocyte differentiation from middle to late stage.

$$R (\%) = \frac{m-n}{n} \quad (\text{Equation 1})$$

where m represents the number of differentially expressed genes in the early versus late group of a functional term, and n represents the number of differentially expressed genes in the early versus middle group.

Comparison of the differentially expressed genes in two groups

The differentially expressed genes screened from the two groups were further compared to obtain the common gene set and specific gene set. For the common gene set, the Pearson correlation coefficient of their fold-change value in each group was calculated to investigate the consistency of upregulation and downregulation.

Functional analysis of the differentially expressed genes

EASE (Expressing Analysis Systematic Explorer) is a customizable software application for rapid biological interpretation of gene lists that result from the analysis of microarray, proteomics, SAGE and other high-throughput genomic data (Hosack et al., 2003). We analyzed the functions of the differentially expressed genes in each set using the Fisher algorithm in EASE and observed significant enrichment of these genes in multiple Gene Ontology (GO) categories with FDR < 0.05 . Besides, we compared the expression values of the genes that were enriched in the most significant function between different differentiation stages.

RESULTS

Identification of the differentially expressed genes

After data preprocessing, we identified the differentially expressed genes in the middle and late stages compared with early stage, respectively. FDR < 0.05 and $|\logFC| > 1$ were used as the thresholds to select differentially expressed genes. Finally, we identified 668 and 1007 differentially expressed genes in the early versus middle group and early versus late group, respectively.

Differentially expressed genes in functional categories

According to the level of similarity between the sequences of the differentially expressed genes and the sequences of each GO node in the Gene Ontology, the differentially expressed genes in two groups were classified into 13 categories (Figure 1). In the early versus middle group, the category which enriched the most genes was metabolic process (330 genes), followed by biological regulation (283 genes) and multicellular organismal process (188 genes). In the early versus late group, the enriched categories were the same as in the early versus middle group, but with a different number of enriched genes in each category. We calculated the relative increment rate of the differentially expressed genes in different functional categories in preadipocyte differentiation from middle to late stages (Table 1). We could find that the largest relative increment rate was for the category metabolic process (relative increment rate = 48.79%).

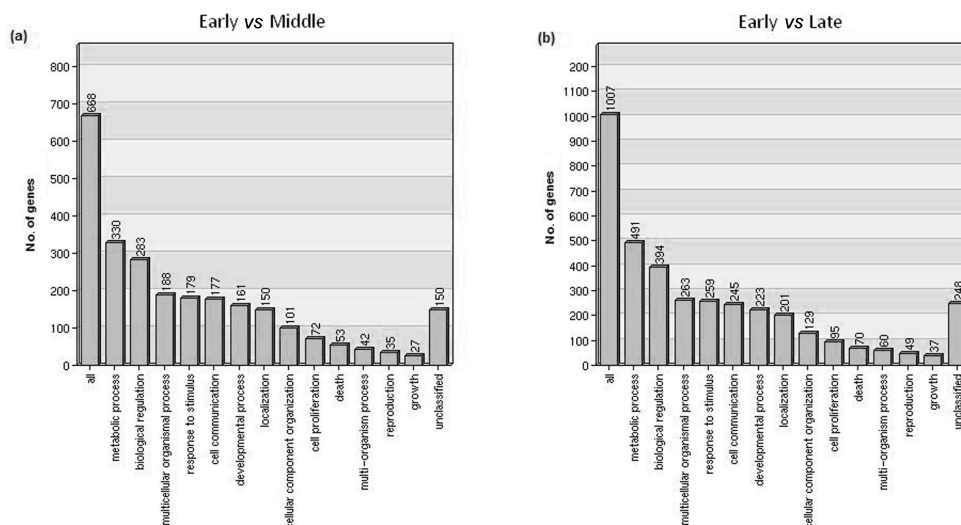


Figure 1. Cluster of orthologous groups of proteins (COG) functional classification of differentially expressed genes in early-middle group and early-late group. The abscissa shows the different functions and the ordinate shows the number of genes.

Table 1. The relative increment rate of the differentially expressed genes in different cluster of orthologous groups of proteins (COG) functional classification.

Function	Gene count in early-middle	Gene count early-late	Relative increment rate (%)
Metabolic process	330	491	48.79
Response to stimulus	179	259	44.69
Multi-organism process	42	60	42.86
Reproduction	35	49	40
Multicellular organismal process	188	263	39.89
Biological regulation	283	394	39.22
Developmental process	161	223	38.51
Cell communication	177	245	38.42
Growth	27	37	37.04
Localization	150	201	34
Death	53	70	32.08
Cell proliferation	72	95	31.94
Cellular component organization	101	129	27.72

Comparison of differentially expressed genes in two groups

We further compared the differentially expressed genes in the two groups and identified the common and specific gene sets in the two groups. As shown in Figure 2, there were 584 common differentially expressed genes in the early versus middle and early versus late groups and 84 specific differentially expressed genes in the early versus middle group and 423 specific differentially expressed genes in the early versus late group.

We then calculated the Pearson correlation coefficient of the fold change of the overlapping genes. As shown in Figure 3, the Pearson correlation coefficient was higher than 0.97. This suggested that the expression pattern of the common genes was almost coincident in the two differentiation groups.

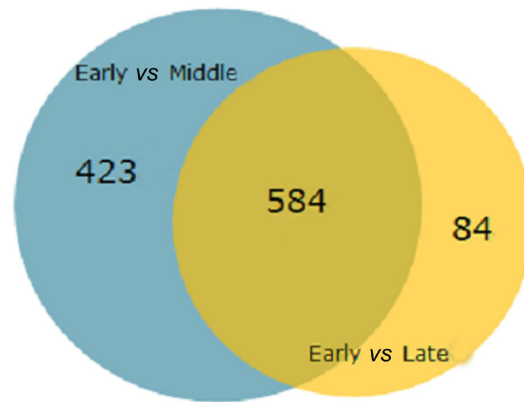


Figure 2. Venn graph displays the number of genes in each set. There are 584 overlapping differentially expressed genes.

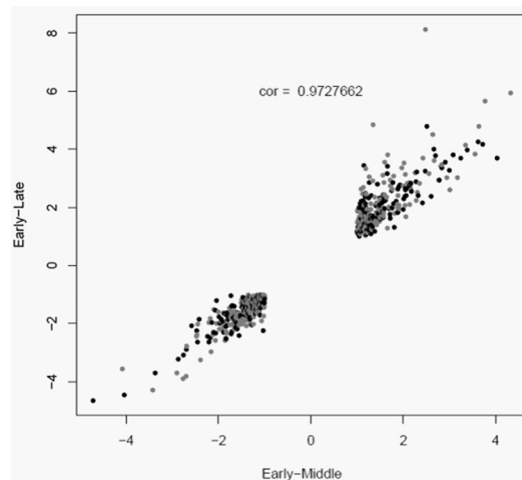


Figure 3. Pearson correlation coefficient of the fold-change value of the overlapping genes in group early-middle and early-late. The abscissa and black point shows logFC of group early-middle. The ordinate and gray point shows logFC of group early-late.

Functional analysis of the differentially expressed genes

We performed functional analysis of the three sets of differentially expressed genes using EASE (Tables 2 and 3). As shown in Table 2, the 584 common differentially expressed genes were enriched in 11 GO categories, and the most significant category was fatty acid metabolic process (GO: 0006631), including 29 differentially expressed genes. As shown in Table 3, the specific 423 genes in the early versus late group were enriched in 3 significant GO categories, and the most significant function was fatty acid metabolic process (GO: 0006631) as well. Also, there were 19 differentially expressed genes enriched in the fatty acid metabolic process, including 16 upregulated and 3 downregulated genes.

Table 2. Significant enriched categories of 584 differentially expressed genes.

Term	Description	P value	FDR
GO:0006631	Fatty acid metabolic process	6.03E-11	1.07E-07
GO:0009725	Response to hormone stimulus	3.08E-10	5.44E-07
GO:0055114	Oxidation reduction	1.09E-09	1.92E-06
GO:0010033	Response to organic substance	1.32E-09	2.33E-06
GO:0009611	Response to wounding	1.25E-08	2.22E-05
GO:0008610	Lipid biosynthetic process	2.38E-08	4.21E-05
GO:0035295	Tube development	3.30E-06	0.005831
GO:0016042	Lipid catabolic process	3.93E-06	0.006951
GO:0048545	Response to steroid hormone stimulus	4.97E-06	0.008779
GO:0007155	Cell adhesion	5.14E-06	0.009078
GO:0022610	Biological adhesion	5.16E-06	0.009118

Table 3. Significant enriched categories of differentially expressed genes in group early-late.

Term	Description	P value	FDR
GO:0006631	Fatty acid metabolic process	5.19E-07	8.89E-04
GO:0055114	Oxidation reduction	9.86E-07	0.001689
GO:0008610	Lipid biosynthetic process	3.92E-06	0.006707

We further compared the expression profile of the 29 common differentially expressed genes which were enriched in the category fatty acid metabolic process in the progression from early stage to late stage (Figure 4 and Table 4). As shown in Figure 4, the expression pattern of these 29 genes was coincident from early stage to late stage; that is, the 23 upregulated genes were always upregulated and the 6 downregulated genes were always downregulated. It proved that in preadipocyte differentiation, the genes that controlled the fatty acid metabolic process were specifically expressed regardless of stage.

DISCUSSION

In this study, we analyzed the gene expression profile in different stages of preadipocyte differentiation to select the differently expressed genes in the middle and late stages. A total of 668 and 1007 differentially expressed genes were identified in the early versus middle group and early versus late group. Functional analysis revealed that the common differentially expressed genes in different differentiation stages and the specific genes in early versus late group were both significantly enriched in fatty acid metabolism. In addition, the quantity of

differentially expressed genes enriched in metabolic function increased most during differentiation. In line with previous studies, our results also suggested that obesity is apparently some kind of metabolic syndrome (Han et al., 2002; von Eyben et al., 2003).

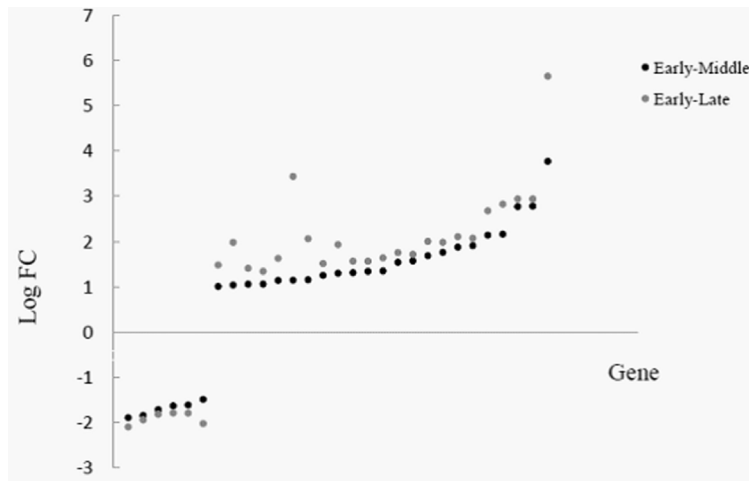


Figure 4. Log FC of the common genes in group early-middle and early-late enriched in the fatty acid metabolic process. Black and gray dots represent logFC of differentially expressed genes in group early-middle and early-late, respectively. The abscissa shows 29 genes and the ordinate shows logFC of the 29 genes.

Table 4. Twenty-nine common differentially expressed genes.

Gene symbol	Early-middle group		Early-late group	
	P value	logFC	P value	logFC
ABCD2	2.41E-05	1.686536	6.34E-06	2.002822
ACACB	1.17E-04	1.570812	1.37E-04	1.716032
ACADL	9.55E-07	2.162105	4.09E-08	2.825378
ACADM	7.03E-07	1.311383	3.93E-06	1.565403
ACADS	9.77E-07	1.350329	2.02E-06	1.639022
ACOT1	1.16E-05	1.141147	1.09E-06	1.626921
ACOT2	2.48E-06	1.253854	8.64E-06	1.514286
ACSF2	9.55E-05	1.875617	1.78E-05	2.106306
ADIPOQ	6.50E-06	2.780835	2.02E-06	2.941052
ADIPOR2	6.14E-05	1.038777	4.30E-07	1.981088
CH25H	4.79E-04	-1.88809	1.99E-04	-2.09257
CPT1A	5.22E-06	-1.4846	1.87E-10	-2.02013
CPT2	1.76E-04	1.006993	2.23E-05	1.481031
ECHS1	6.60E-06	1.060605	4.25E-06	1.343727
EHHADH	3.67E-06	1.541462	4.99E-07	1.754712
ELOVL3	1.23E-03	1.059073	8.95E-05	1.409639
ELOVL6	9.52E-04	1.145222	1.41E-10	3.436825
ETFDH	1.07E-05	1.342232	2.58E-05	1.563782
GHR	2.03E-05	1.29917	2.67E-07	1.931839
GPAM	7.85E-14	3.769796	1.10E-17	5.64796
HADH	7.85E-07	1.760459	8.85E-07	1.982493
LPL	4.05E-07	2.770246	2.02E-07	2.941445
PDPN	1.24E-04	-1.83902	1.68E-05	-1.93735
PECR	7.05E-08	2.140099	2.58E-08	2.679342
PRKAR2B	1.89E-05	1.908457	1.50E-05	2.072796
PTGES	3.24E-03	-1.62773	2.69E-04	-1.7861
PTGIS	7.45E-04	-1.70994	2.76E-05	-1.81401
PTGS1	1.78E-04	-1.60803	6.85E-06	-1.78903
TMEM195	6.59E-05	1.155526	2.95E-07	2.061054

It is suggested that fatty acid metabolism plays a decisive role in adipocyte growth and that it is possible to regulate and control the quantity of mature adipocytes through fatty acid metabolic processes. The 29 differentially expressed genes identified in this study could be regarded as target genes of the fatty acid metabolism pathway because of their continuous specific expression. Fatty acid metabolism is regulated by a variety of enzymes. For example, acetyl-CoA carboxylase plays a key role in fatty acid synthesis and degradation, and its product, malonyl-CoA, has an important effect on the hypothalamus in the regulation of food intake, as well (McGarry and Brown, 1997; Oh et al., 2005). Therefore, it will be a feasible and practical way to control fatty acid metabolism to influence the action of the target genes and alleviate the symptoms of obesity.

The factors causing obesity are various and the most important reason is abnormal fat metabolism. The results in this study showed that the differentially expressed genes were most significantly enriched in fatty acid metabolism in the progression of preadipocyte differentiation into mature adipocytes. Moreover, these genes existed in the whole differentiation process and their number continued to increase during differentiation. Therefore, there is the possibility to regulate preadipocyte differentiation and then to prevent excessive adipocytes through control of the fatty acid metabolic process. Finally, we can control obesity and a series of diseases caused by obesity, such as cardiovascular and respiratory diseases.

ACKNOWLEDGMENTS

Research supported by the National Natural Science Foundation of China (#81270927).

REFERENCES

- Altschul SF, Gish W, Miller W, Myers EW, et al. (1990). Basic local alignment search tool. *J. Mol. Biol.* 215: 403-410.
- Arner E, Mejhert N, Kulyté A, Balwierz PJ, et al. (2012). Adipose tissue microRNAs as regulators of CCL2 production in human obesity. *Diabetes* 61: 1986-1993.
- Fujita A, Sato JR, Rodrigues LO, Ferreira CE, et al. (2006). Evaluating different methods of microarray data normalization. *BMC Bioinformatics* 7: 469.
- Han TS, Williams K, Sattar N, Hunt KJ, et al. (2002). Analysis of obesity and hyperinsulinemia in the development of metabolic syndrome: San Antonio Heart Study. *Obes. Res.* 10: 923-931.
- He Y, Chen H, Quon MJ and Reitman M (1995). The mouse obese gene. Genomic organization, promoter activity, and activation by CCAAT/enhancer-binding protein alpha. *J. Biol. Chem.* 270: 28887-28891.
- Hoffmann IS, Jimenez E and Cubeddu LX (2001). Urinary albumin excretion in lean, overweight and obese glucose tolerant individuals: its relationship with dyslipidaemia, hyperinsulinaemia and blood pressure. *J. Hum. Hypertens.* 15: 407-412.
- Hosack DA, Dennis G Jr, Sherman BT, Lane HC, et al. (2003). Identifying biological themes within lists of genes with EASE. *Genome Biol.* 4: R70.
- Howrylak JA, Dolinay T, Lucht L, Wang Z, et al. (2009). Discovery of the gene signature for acute lung injury in patients with sepsis. *Physiol. Genomics* 37: 133-139.
- Hsu CL and Yen GC (2007). Effects of capsaicin on induction of apoptosis and inhibition of adipogenesis in 3T3-L1 cells. *J. Agric. Food Chem.* 55: 1730-1736.
- Kume S, Kato S, Yamagishi S, Inagaki Y, et al. (2005). Advanced glycation end-products attenuate human mesenchymal stem cells and prevent cognate differentiation into adipose tissue, cartilage, and bone. *J. Bone Miner. Res.* 20: 1647-1658.
- Maffei M, Stoffel M, Barone M, Moon B, et al. (1996). Absence of mutations in the human OB gene in obese/diabetic subjects. *Diabetes* 45: 679-682.
- McGarry JD and Brown NF (1997). The mitochondrial carnitine palmitoyltransferase system. From concept to molecular analysis. *Eur. J. Biochem.* 244: 1-14.

- Oh W, Abu-Elheiga L, Kordari P, Gu Z, et al. (2005). Glucose and fat metabolism in adipose tissue of acetyl-CoA carboxylase 2 knockout mice. *Proc. Natl. Acad. Sci. U. S. A.* 102: 1384-1389.
- Smyth G (2005). Limma: Linear Models for Microarray Data. In *Bioinformatics and Computational Biology Solutions using R and Bioconductor*. 397-420.
- Tatusov RL, Natale DA, Garkavtsev IV, Tatusova TA, et al. (2001). The COG database: new developments in phylogenetic classification of proteins from complete genomes. *Nucleic Acids Res.* 29: 22-28.
- Troyanskaya O, Cantor M, Sherlock G, Brown P, et al. (2001). Missing value estimation methods for DNA microarrays. *Bioinformatics* 17: 520-525.
- von Eyben FE, Mouritsen E, Holm J, Montvilas P, et al. (2003). Intra-abdominal obesity and metabolic risk factors: a study of young adults. *Int. J. Obes. Relat. Metab. Disord.* 27: 941-949.
- Zhang Y, Proenca R, Maffei M, Barone M, et al. (1994). Positional cloning of the mouse obese gene and its human homologue. *Nature* 372: 425-432.