ISSN 0006-3509, Biophysics, 2006, Vol. 51, Suppl. 1, pp. S100–S109. © Pleiades Publishing, Inc., 2006.

Multidimensional Analysis and Functional Assignment of DNA-Microarray Transcription Profiles of Genes Involved in Adipogenesis

A. V. Katokhin, V. M. Efimov, M. Sh. Badratinov, O. K. Kamneva, and V. A. Mordvinov

Institute of Cytology and Genetics, Siberian Branch, Russian Academy of Sciences, pr. akademika Lavrent'eva 10, Novosibirsk, 630090 Russia
e-mail: katokhin@bionet.nsc.ru

Abstract—The results of two independent DNA-microarray experiments concerning adipogenesis in the murine preadipocyte 3T3-L1 cell line, which covered the first two days after the induction of differentiation, were analyzed using the multidimensional scaling (MDS) method. In both data arrays, the first three scaling components accounted for 73.5–73.8% of the total dispersion. This result implies that both arrays of the gene expression profiles are in fact three-dimensional and each component reflects a definite principal process involved in one of the three early stages of adipogenesis: (i) determination of the fibroblast-like stem cells, (ii) clonal expansion of adipoblasts, and (iii) preadipocyte conversion into a mature adipocyte phenotype. Each profile of the gene expression is characterized by coefficients of correlation with the first three scaling components. The functional annotation in terms of the Gene Ontology database profiles (sorted according to the correlations with each component) generally corresponds to a regular change of elementary biological processes during the three early stages of adipogenesis. Analysis of correlations with the principal scaling components for the genes previously classified as subject to differential expression in the course of adipogenesis in mice suggests a complicated role of these genes in early adipogenesis (in some cases, described in the literature). The MDS analysis of the gene expression profiles and the analysis of correlations between these profiles and the main scaling components provides a deeper insight into the fine role of these genes and makes possible the search for new biomarkers of various differentiation stages.

Key words: murine preadipocytes, 3T3-L1 cell line, adipogenesis, DNA microarray, multidimensional analysis, functional annotation of genes

DOI: 10.1134/S0006350906070207

INTRODUCTION

Adipocytes (fat cells) are highly specialized cells playing a critical role in regulation of the energy homeostasis. Their primary and most thoroughly studied functions are storing energy in the form of lipids (triacylglycerides) and releasing it in the form of free fatty acids formed in the course of lipolysis. In addition, they participate in the regulation of energy balance of the whole organism by means of secretion of the endocrine, paracrine, and autocrine signals [1]. However, detailed signal pathways and molecular processes regulating the expression and secretion of regulatory factors and the general functioning of adipocytes, as well as the consequences of various possible violations are still insufficiently studied [2, 3].

In recent years, many advanced countries have encountered the problem of obesity (adiposis), which leads to increased risk of many diseases such as type II diabetes, dyslipidemia, coronary heart diseases [4], and some types of malignancy [5]. Excess growth of the fat tissue may be caused both by the accumulation of excess lipids in the existing adipocytes and by an increase in the population of fat cells as a result of the transformation of stromal stem cells into adipocytes [2, 6]. Elucidation of a detailed pattern of these processes and development of biomarkers for distinguishing them will probably provide more effective strategies for the treatment of adiposis and related diseases.

Now it is commonly accepted that polypotent fibroblast-like stem cells of mesodermal origin are the precursors of adipocytes [7]. In most mammals, these preadipocytes begin to differentiate into adipocytes at the early embryonal stage, but this process is also continued in the early postnatal period [8]. The close morphology of fibroblasts and preadipocytes and the impossibility to distinguish between preadipocytes on various stages of differentiation make the investigation of adiposis in vivo a difficult task. For this reason, cell cultures of two types are widely used for the investigation of adipocyte differentiation: (i) preadipocyte lines (presently available only for mice) and (ii) primary cultures of stromal precursor cells (which are isolated from fat tissues of humans and other mammals [1]). Using these cells, it is possible to separate species-, sex-, age-dependent and many other features of adipogenesis.

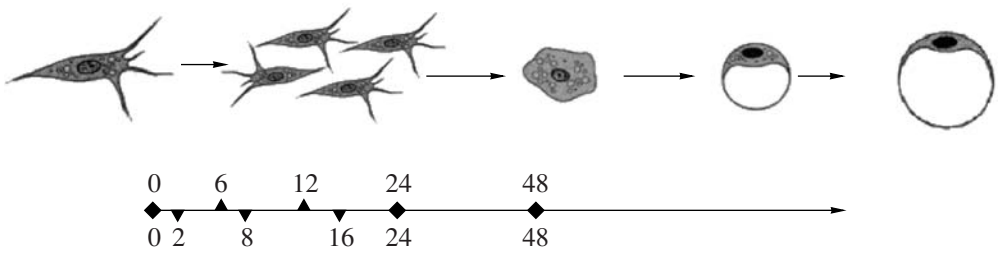

Fig. 1. Schematic diagram illustrating the transformation of cell phenotype from fibroblast-like stem cells via adipoblasts to preadipocytes (top) and the temporal series corresponding to the two sets of data taken from [17] (upper scale) and [15] (lower scale).

Investigations using murine preadipocyte 3T3-L1 cell line revealed several stages characteristic of the process of adipocyte differentiation (Fig. 1). The first several hours are necessary for the determination of totipotent embryonal stem cells and the direction of their development into fibroblast-like stem cells, adipoblasts, and preadipocytes. The adipoblasts exhibit proliferation (*exponential growth stage*) until the formation of a continuous layer (by the middle of the first day), which is terminated at the stage of the G1/S cell cycle (*stage of growth arrest* [1]). The second day involves the *phase of mitotic clonal expansion* of preadipocytes, initiated by differentiation agents and consisting of one or two cell divisions. After this phase, the main mechanisms ensuring realization of the adipogenic program set in. By the fourth day, an apparatus for the accumulation, synthesis, and decomposition of lipids is formed, the level of expression of the corresponding enzymes is increased. Then, the cytoskeleton is rearranged and the specific morphology of adipocytes is established [1, 2].

The growth in the number of adipocytes is a result of differentiation and maturation of the precursor cells, which is controlled by the endocrine, paracrine, and autocrine signals. However, detailed molecular mechanisms governing these regulatory factors of adipogenesis are yet unclear [2, 3].

In recent years, investigations of gene expression on the genome scale by means of cDNA arrays and high-density oligonucleotide arrays, supplemented by proteomic data, filled many gaps in our knowledge of adipocyte differentiation. A comparative analysis of data obtained using various microarray technologies (interplatform verification of conclusions) should significantly decrease erroneous predictions concerning the differential expression of genes and their belonging to co-expression groups and gene expression clusters. However, this approach has certain limitations, since it is applicable only to data for arrays with comparable temporal points [9].

At present, a large volume of information is accumulated concerning the pathways of signals regulating the process of adipogenesis, but the available data are fragmentary and many of the molecular mechanisms are still unclear. Several attempts have been undertaken to study adipogenesis of murine preadipocyte 3T3-L1 cells using microarray techniques, which revealed many new genes with significant differential expression and complex expression patterns, but the attention of researchers was mostly devoted to the late stages of adipogenesis [10–13]. For this reason, the early events of preadipocyte differentiation are still insufficiently studied, including the process of totipotent stem cell commitment, transcription regulation mechanisms, and the functioning of principal transcription factors [3, 9, 15]. We believe that the integration of microarray data concerning the differential and consistent expression of key adipogenesis-related genes with information about the structure and functions of the corresponding proteins will provide more adequate reconstruction of the molecular processes of determination and proliferation of adipoblasts and will reveal the key regulatory elements in these gene networks.

The main aim of this investigation was analysis of gene expression profiles at the early stages of adipogenesis in mice, obtained with the aid of two microarray platforms for the murine preadipocyte 3T3-L1 cell line. The analysis was performed using an approach based on the multidimensional scaling (MDS) of the data array. Then, functional annotation was carried out in terms of the Gene Ontology (GO) database lists of genes sorted according to the results of correlation analysis. Finally, the efficacy of this approach was qualitatively evaluated by comparison with the results for well-known gene biomarkers of early adipogenesis.

DATA AND METHODS

Data on the gene expression refer to the murine preadipocyte 3T3-L1 cell line originating from a clone obtained using 17- to 19-day embryos of the Swiss 3T3 mice line. The cells of this line are partly committed (i.e., have certain limitations of the development potential) and eventually yield cells of cartilage, fatty, or connective tissue, depending on the conditions [1].

We have used two data sets. The first set was taken from the data reported by Burton et al. [15] (below, Burton array), which were obtained using Affymetrix MG-U74Av2 oligonucleotide chips. This microarray platform contained 12 488 probe sets. The Burton array is available in the format of primary *.CEL files [16].

The set of data used in the present analysis included the results of 12 hybridizations representing six temporal points 0, 2, 8, 16, 24, and 48 h (each in two repeats) after inducing differentiation with an IDX (insulin, dexamethasone, 3-isobutyl-1-methylxanthine) hormonal mixture.

The second set was taken from the data reported by Hackl et al. [17] (below, Hackl array), which were obtained using TIGR-Chips cDNA microarrays comprising 27 648 elements (including control). The set of data used in the present analysis included the results of 30 two-color hybridizations, representing five temporal points 0, 6, 12, 24, and 48 h after induction of differentiation with a mixture of insulin, dexamethasone, 3-isobutyl-1-methylxanthine, pantotenoic acid, and biotin, and three biological replicas. Each replica and temporal point was studied in two hybridizations with alternation of fluorochrome and a common control (RNA from initial cells). The Hackle array is available in E-MARS-2 notation from the ArrayExpress database [18].

The common elements in both platforms (oligonucleotide and cDNA microarrays) were revealed using UniGene identifiers. Elements of one array not having analogs in the other were exempt from the further analysis. As a result of this preparation, the Burton array was represented in our analysis by 3395 expression profiles and 12 samples (0, 2, 8, 16, 24, and 48 h points, each in two repeats), while the Hackl array was represented by 5437 expression profiles and 30 samples (six samples for five points: 0, 6, 12, 24, and 48 h).

The data arrays were converted into logarithmic scale, centered, and normalized (first, over samples and then over expression profiles). Each line of the matrix of gene expression can be considered as a point in a multidimensional space. Moreover, since all distance from each point to the origin after normalization is unity, all points (profiles) occur on the surface of a multidimensional sphere with a unit radius ad a dimension of $N - 1$, where N is the number of samples. The coefficient of correlation between two profiles is equal to the cosine of the angle between the radii drawn from the origin to the corresponding points. The correlation with a profile occurring on the opposite side of the sphere is –1, which implies that the higher is the expression of one gene, the lower is that of the other in the same samples (and vice versa).

Each data array was characterized by a matrix of Euclidean distances between samples, and this matrix was processed using the MDS method with a maximum dimension of 9 accessible in the Statistica program package. Several first MDS axes (components) correspond (by analogy with the method of principal components) to the directions of maximum variability between samples. Note that straightforward calculation of the principal components via the correlation matrix of profiles is impossible in view of a vast number of profiles. Then, each profile was characterized by the coefficients of correlation with the MDS components. Components related to the temporal points reflect the main processes involved in adipogenesis, while their correlations with expression profiles characterize the participation of genes in these processes. Positive coefficients correspond to an increase in the expression during the process, while negative coefficients are indicative of suppression of the gene expression.

The data were processed using Microsoft Excel and Statistica program packages, and a Transcriptomics routine included in a software complex Systems Biology [19] developed at the Theoretical Genetics Laboratory, Institute of Cytology and Genetics (Novosibirsk).

The search for identifiers in published databases associated with genes presented in microarrays and the functional annotation of genes were performed using a NETAFIX server supported by Affymetrix [20].

RESULTS

The literature and online databases present several data arrays obtained using various microarray platforms for the investigation of processes involved in the adipogenesis induced in fibroblast-like 3T3-L1 murine cells [10–13, 15, 17]. However, only the data reported in [15, 17] contain a sufficiently detailed series of temporal points for the early stages of adipogenesis developing within the first two days. These investigations were performed using different microarrays: oligonucleotide Affymetrix chips [15] and cDNA microarrays of the Institute of Genome Research (USA). As is known, the results of interpretation of the data of microarray experiments strongly depend on the platform, protocol, and the features of algorithm used in the analysis [21]. This situation stimulates the search for additional means of increasing the reliability of conclusions. In our study, this was achieved through a comparison of the results of MDS analysis f the profiles of gene expression obtained using different microarray platforms.

The coordinates of samples on the axes of scaling related to the time past the induction of differentiation can be considered (by analogy with the method of principal components) as corresponding to the principal components of the gene expression profiles. In the Burton array, the first three scaling components accounted for 73.5% and in the Hackl array, for 73.8% of the total dispersion. This result implies that both arrays of the gene expression profiles are essentially three-dimensional (despite the vast number of individual profiles) and each of these components reflects a definite principal process involved in adipogenesis.

It was found that, judging from the character of temporal variations, the first scaling component in the Burton array corresponded to the third component in the Hackl array, the third component in the Burton array corresponded to the first component in the Hackl array, and the second components of both arrays corresponded to each other. Accordingly, Figs. 2–4 present the results of analysis so that the components corre-

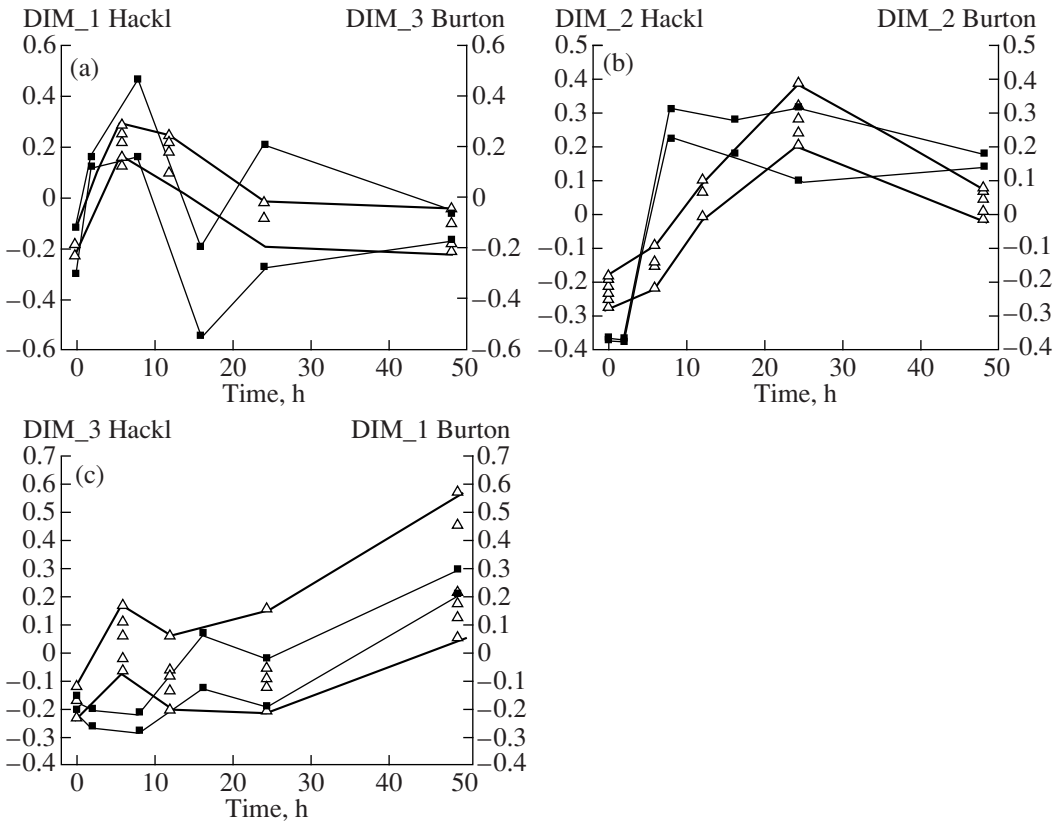

Fig. 2. Temporal variation of the scaling components corresponding to each other in the Burton (squares) and Hackl (triangles) data arrays: (a) third component in the Burton array and first component in the Hackl array; (b) second components in both arrays; (c) first component in the Burton array and third component in the Hackl array.

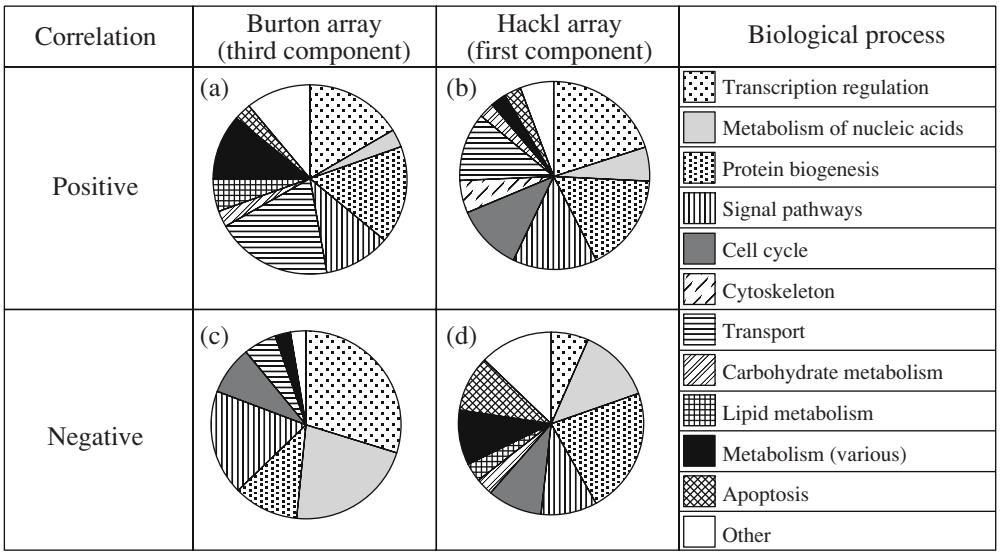

Fig. 3. Diagrams of distribution of 12 sets of biological processes for 50 gene expression profiles with maximum values of the (a, b) positive and (c, d) negative coefficients of correlation between gene expression and the values of (a, c) third component of the Burton array and (b, d) first component of the Hackl array. Annotation of biological processes is indicated in the right column.

sponding to each other are plotted on the same panel. Lines connecting points in Fig. 2 reflect the boundaries of variation of the components calculated for various samples.

Despite the presence of noncoincident points in the different sets, one can readily see that on the whole the plots reflect similar dynamics in the course of gene expression. Indeed, the behavior of the first component

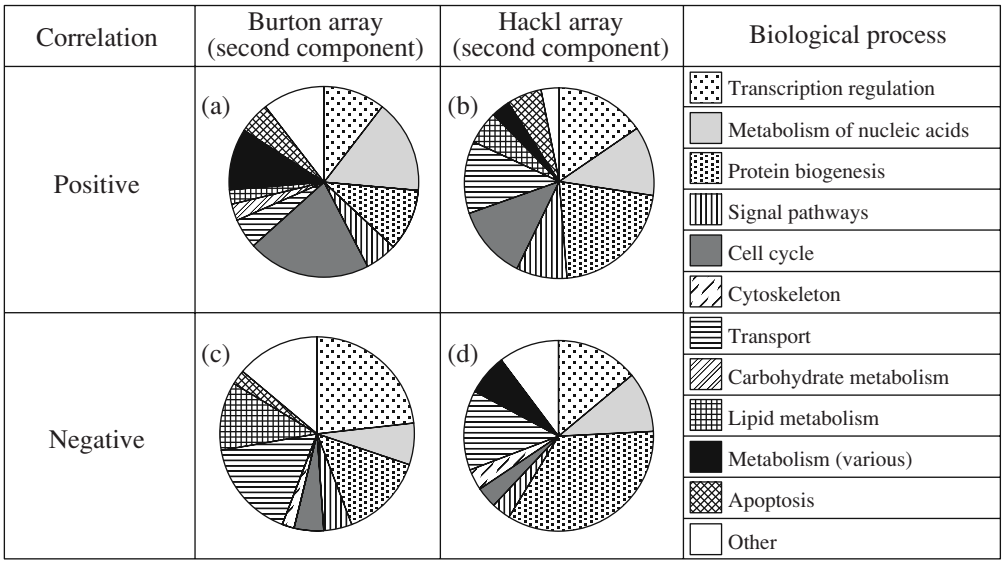

Fig. 4. Diagrams of distribution of 12 sets of biological processes for 50 gene expression profiles with maximum values of the (a, b) positive and (c, d) negative coefficients of correlation between gene expression and the values of (a, c) second component of the Burton array and (b, d) second component of the Hackl array. Annotation of biological processes is indicated in the right column.

in the Burton array and the third component in the Hackl array (Fig. 2a) with a maximum corresponding to the first half of the first day upon adipogenesis stimulation can be interpreted as reflecting changes in the gene expression at the stage of determination of the adipocyte-specific development of stem cells during the first hours of adipogenesis.

The behavior of the second components (Fig. 2b) in both arrays, which show the growth of components by the end of the first day of adipogenesis and the retention of high values during the second day, can be interpreted as reflecting changes in the gene expression in the stage of exponential growth of adipoblasts and mitotic clonal expansion of preadipocytes.

The third component of the Burton array and the first component of the Hackl array (Fig. 2c) exhibit growth by the end of the third day, which apparently reveals changes in gene expression in the course of proliferation suppression and the transition of preadipocytes to the formation of adipocyte-specific morphological and biochemical phenotypes.

In order to determine gene profiles contributing to each particular process, we calculated the coefficients of correlation with the first three scaling components and sorted the profiles with the maximum absolute value of these coefficients, that is, with the increasing or decreasing expression. Each of the three principal components was represented by 50 gene expression profiles with maximum values of the positive and negative correlation coefficients and brought into correspondence with the functional annotation of their participation in biological processes in terms of the GO database [22]. Complete results of the functional annotation of genes of these gene profiles are available on the site of the Functional Genomics Sector, ICG [20]. The diagrams of distribution of the sets of biological processes including terms of the GO database with close meaning are presented in Figs. 3–5.

The diagrams in Figs. 3a–3d show that, in the course of determination of the adipocyte-specific development of the stem cells during the first hours after the induction of adipogenesis (with the values of scaling components depicted in Fig. 2a), the trend most pronounced in both data arrays is the predominance of genes related to positive and negative regulation of transcription and transmission of signals from receptors on the cell surface and in the cytosol via kinase channels to the nucleus. Note that the genes are partly activated and partly suppressed, which is natural of rapid regulatory processes. One can see, on the one hand, enhanced expression of genes encoding proteins that provide the transport of molecules, ions, and electrons and, on the other hand, low expression of the genes responsible for the metabolism of nucleic acids (DNA replication and repair, RNA metabolism) and the cell cycle process.

The diagrams in Figs. 4a–4d correspond to the stage of adipoblast and preadipocyte proliferation (with the values of scaling components depicted in Fig. 2b) and show that, in comparison to Fig. 3, both data arrays are characterized, on the one hand, by an increase in the expression of genes involved in the cell cycle, metabolism of nucleic acids, and apoptosis and, on the other hand, a decrease in the expression of genes encoding the cytoskeleton proteins and involved in the transport.

The diagrams in Figs. 5a–5d reflect most significant changes in the expression of genes in the third phase involving the suppression of proliferation and the conversion of preadipocytes into mature adipocytes (with the values of scaling components depicted in Fig. 2c). In comparison with the preceding diagrams, these data

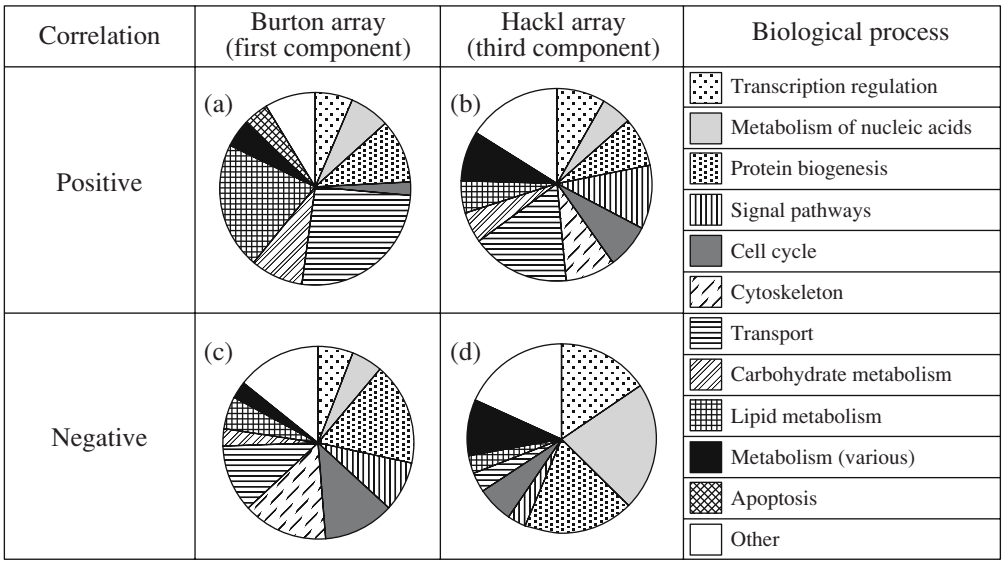

Fig. 5. Diagrams of distribution of 12 sets of biological processes for 50 gene expression profiles with maximum values of the (a, b) positive and (c, d) negative coefficients of correlation between gene expression and the values of (a, c) first component of the Burton array and (b, d) third component of the Hackl array. Annotation of biological processes is indicated in the right column.

show, on the one hand, increased expression of genes responsible for the transport and the metabolism of carbohydrates and lipids and, on the other hand, decreased expression of genes encoding cytoskeletal proteins and genes involved in the cell cycle.

The above diagrams reveal certain differences between the two data arrays. In particular, data of the Hackl array appear as a mixture of adjacent temporal points and, hence, it can be suggested that the cell culture in these samples was more desynchronized than that in the samples used in the Burton array. This distinction can also be due to some differences in the experimental procedures. Nevertheless, the general trends are well manifested in both data arrays.

Let us consider the correlations with the principal MDS components, which were calculated in both data arrays for some profiles of gene biomarkers from the GO database list, showing more than twofold difference (increase or decrease) in expression within the first 24 h after the induction of adipogenesis as compared with that presented in review [9]. This list is supplemented with the well-known markers of early adipogenesis [3]. A comparison of the MDS results was possible not for all genes included in the list, since not all of them remained in the analyzed data arrays after filtration with respect to zero values and the presence of common UniGene identifiers. The results of our comparison are presented in Table 1.

The well-known adipogenesis biomarker *Cebpb* in the Burton array exhibits a high positive correlation with the third component and moderate correlation with the second component and shows an insignificant positive correlation with the second component in the Hackl array. This behavior of the gene profile can be explained by the fact that the expression of this gene begins immediately after the induction of adipogenesis and its product is necessary to start the program of mitotic clonal expansion of adipoblasts. However, this product is also necessary for activation of *Cebpa* and *Pparg* genes, which initiate the adipocyte-specific program after the arrest of proliferation [2, 3, 25]. Therefore, it is possible that the expression of *Cebpb* is also not suppressed in the course of proliferation. Thus, calculation of correlations allowed us to reveal the expression of this gene during two phases of the early adipogenesis.

The profile of another biomarker of the early adipogenesis, *Klf5*, which is a target for *Cebpb* [3], exhibits maximum correlations with the second MSD component in the Burton array and the third component in the Hackl array, which implies that the activation of this gene expression follows one step behind the expression of its regulator.

The well-known circadian rhythm regulator, *Arntl* (*Bmall*), is necessary for activation of the process of accumulation and hydrolysis of lipids in adipocytes (although the mechanism of its functioning is still unknown [26]). The profile of this gene exhibits maximum correlations with the second MSD component in the Burton array and the first component in the Hackl array.

The expression of *Gas6* shows a high positive correlation with the second MSD component in the Burton array and a high positive correlation with the first component in the Hackl array. It should be noted that experimental data confirm the active expression of this gene in the phase of clonal expansion of adipoblasts [27]. Therefore, the result observed for the Hackl array is probably indicative of a lack of synchrony of the cell processes in this experiment. It was shown [2] that

Table 1. Correlation with the principal MDS components of the Burton and Hackl arrays for the genes induced within the first two days of adipogenesis in 3T3-L1 cells according to list [9] and review [3] (data in parentheses refer to obsolete gene notations encountered in literature or obsolete identifiers used by manufacturers in biochip descriptions)

Gene name	Notation	Identifier of Affymetrix probe chip	Identifier of reference transcript	Identifier of UniGene	Correlation of Burton array with component			Correlation of Hackl array with component		
					3	2	1	1	2	3
A disintegrin and metallopeptidase domain 8	*Adam8*	103024_at	NM_007403	Mm. 15969	−0.52	0.66	−0.20	0.94	−0.03	0.01
Aryl hydrocarbon receptor nuclear translocator-like	*Arntl*	102382_at; 164698_at	NM_007489	Mm.440371 (Mm.12177)	−0.04	0.42	−0.46	0.50	0.14	0.02
CCAAT/enhancer binding protein (C/EBP), beta	*Cebpb*	92925_at	NM_009883	Mm.439656 (Mm.4863)	0.071	0.38	−0.57	−0.03	0.20	−0.19
Cystein and glycinerich protein 2 (Double LIM protein-1)	*Csrp2*	168753_at; 93550_at	NM_007792	Mm.2020	−0.38	0.38	−0.79	−0.79	0.33	−0.07
Growth arrest specific 6	*Gas6*	99067_at	NM_019521	Mm.3982	−0.36	0.87	−0.22	0.89	−0.14	0.15
Kruppel-like factor 5	*Klf5*	97937_at	NM_009769	Mm.30262	−0.06	0.60	−0.34	−0.47	−0.45	0.06
Latexin	*Lxn*	96065_at	NM_016753	Mm.212039 (Mm.2632)	−0.22	0.79	−0.46	0.54	−0.53	−0.06
signal transducer and activator of transcription 5A	*Stat5a*	100422_i_at	NM_011488	Mm.277403 (Mm.4697)	−0.27	−0.17	0.59	0.30	−0.60	−0.63

exact synchronism and correct chronology of molecular events during adipogenesis is very difficult to maintain because, depending on various stimuli, adipoblasts can make various numbers of divisions and even pass to the state of preadipocytes without division.

An analogous situation was observed for *Adam8* and *Lnx* genes, but no experimental data on their role in the adipogenesis is available and they have been mentioned only as candidate genes for marking the entire process of adipogenesis [9].

The expression profile of the induced *Csrp2* gene shown in Table 1 exhibits a positive correlation with the second components in both Burton and Hackl arrays, so that it probably involved in the proliferation of adipoblasts. Previously, it was believed that the product of this gene, representing a small protein associated with the actin cytoskeleton, participated in the development of vessels; however, more recently it was shown that this gene is expressed in various developing tissues of mesodermal origin [24]. The role of *Csrp2* in the differentiation of cells—and, in particular, adipocytes—is still unclear.

Among the genes repressed during the first day of adipogenesis (Table 2), it is necessary to mention *Ddit3* (*Chop10*), which plays the role of the apoptosis activator and antiadipogenic factor . The expression of this gene is repressed in the course of adipogenesis [3, 9], which is reflected by negative correlations with all components in both data arrays.

A more complicated behavior is observed for the anti-adipogenic factors such as *Gata2*, *Gata3*, and *Klf2*, the profiles of which (judging from the correlation with the first two MDS components) show an increase in the expression at some moments during the first day of adipogenesis but are repressed by the end of the second day. The roles of many genes revealed by microarray experiments, the expression of which is inhibited (so that the profiles exhibit a significant negative correlation with some of the principal components), are still unclear and have to be studied. It is the absence of their expression that is very important for the normal course of cell processes, which has led them out of the focus of interest of researchers employing classical molecular biology approaches.

Thus the MDS analysis of gene expression profiles and the analysis of correlations between these profiles and the principal scaling components show evidence for more complicated roles of some well-known biomarkers, the activity of which is distributed between three principal phases of the early adipogenesis. These data are necessary in order to refine the ability of using these factors as biomarkers for particular phases of cell differentiation.

Table 2. Correlation with the principal MDS components of the Burton and Hackl arrays for the genes repressed within the first two days of adipogenesis in 3T3-L1 cells according to list [9] and review [3] (data in parentheses refer to obsolete gene notations encountered in literature or obsolete identifiers used by manufacturers in biochip descriptions)

Gene name	Notation	Identifier of Affymetrix probe chip	Identifier of reference transcript	Identifier of UniGene	Correlation of Burton array with component			Correlation of Hackl array with component		
					3	2	1	1	2	3
Activated leukocyte cell adhesion molecule (Transmembrane glycoprotein)	*Alcam*	104407_at; 107969_at; 166517_f_at	NM_009655	Mm.288282 (Mm.2877)	0.13	−0.52	−0.29	−0.17	0.23	−0.04
DNA-damage inducible transcript 3	*Ddit3* (*Chop10*)	101429_at	NM_007837	Mm.110220 (Mm.7549)	−0.10	−0.87	−0.39	−0.17	−0.62	−0.09
Endothelial differentiation, lysophosphatidic acid G-protein-coupled receptor, 2 (G-protein coupled receptor 26)	*Edg2*	100435_at; 170802_i_at	NM_010336	Mm.4772	−0.54	−0.27	−0.47	−0.57	−0.71	0.18
Growth arrest specific 1	*Gas1*	94813_at	NM_008086	Mm.22701 (Mm.288282)	−0.56	−0.35	−0.43	0.01	−0.61	−0.10
GATA binding protein 2	*Gata2*	102789_at; 161820_f_at	NM_008090	Mm.272747 (Mm.1391)	0.13	−0.26	−0.76	−0.01	0.45	−0.36
GATA binding protein 3	*Gata3*	100924_at; 111479_at	NM_008091	Mm.313866 (Mm.606)	0.10	−0.50	−0.52	0.36	0.11	−0.14
Interferon regulatory factor 1	*Irf1*	102401_at	NM_008390	Mm.105218 (Mm.1246)	−0.16	−0.40	0.03	0.47	0.76	−0.04
Kruppel-like factor 2	*Klf2*	96109_at	NM_008452	Mm.26938	0.46	−0.27	−0.66	−0.59	−0.72	−0.003
Myristoylated alanine rich protein kinase C substrate	*Marcks*	96865_at	NM_008538	Mm.30059	−0.54	−0.59	−0.01	0.51	0.24	−0.34
Nuclear factor I/X	*Nfix*	101930_at	NM_001081981; NM_001081982; NM_010906	Mm.9394	−0.31	−0.58	−0.65	0.85	0.19	0.13
Polycystic kidney disease 2	*Pkd2*	100951_at	NM_008861	Mm.6442	−0.41	−0.18	−0.67	−0.79	0.06	−0.07
Short stature homeobox 2	*Shox2*	99043_s_at; 99042_s_at	NM_013665	Mm.39093	−0.12	−0.11	−0.87	−0.23	−0.08	0.18
Zinc finger protein 62 (ZT3 zinc finger factor)	*Zfp62*	103063_at	NM_009562	Mm.16650	−0.75	−0.02	−0.25	−0.36	−0.11	−0.51
Zinc finger protein 68 (KRAB-containing zinc finger protein KRAZ2)	*Zfp68*	95521_s_at; 111488_at; 167978_i_at; 95522_i_at	NM_013844; NM_001044747	Mm.27575	−0.32	−0.71	−0.01	0.64	0.18	0.24

DISCUSSION

Several classifications were proposed for the patterns of gene expression variations in the course of adipogenesis, which were based on an analysis using different algorithms for the clustering of data obtained using various microarray platforms. Five types of variation of the expression of 1889 genes during two days after the induction of adipogenesis were revealed by means of hierarchical clustering [13]. For these patterns of expression, the following biologically descriptive notation of the adipogenesis phases was proposed: (i) loss of preadipocyte phenotype (cluster 1, gene expression is irreversibly repressed by 16 h); (ii) mitotic clonal expression (cluster 2, a peak of gene expression by 16 h); (iii) phenotype transformation (cluster 3, gene expression is sharply activated by 16 h and keeps increasing further); (iv) early adipocyte phenotype (cluster 4, gene expression gradually increases up to 48 h); (v) late adipocyte phenotype (cluster 5, gene expression grows after 48 h). This description is similar to ours, which separates three main processes: (i) determination and readjustment of transcription programs; (ii) clonal expansion; and (iii) development of transcription programs for the adipocyte phenotype.

The microarray data obtained by Burton et al. [15] and Hackl et al [17] were also processed by the means of cluster analysis, in both cases using controlled hierarchical clustering schemes with the distribution of profiles between 12 clusters. In the first case, 1016 profiles of gene expression were processed using the algorithm of self-organized maps, while the latter analysis employed *k*-averaging algorithm. Such a big number of preset clusters (subclasses), probably, allow various subtypes of the behavior of gene expression to be revealed, but it does not allow the main biological processes involved in the adipogenesis to be identified.

Our approach, which is based on MDS analysis of gene expression profiles, evaluation of the minimum dimension of a data array (without substantial loss of information), and calculation of correlations between the individual profiles and the separated principal MDS components, makes possible the separation of gene expression profile sets (regions on the unit sphere) producing the maximum contribution to the observed dynamics of gene expression. The application of this approach to two data arrays showed that the maximum part of dispersion in both cases was related to the first three scaling components, which could be interpreted as the manifestation of three main biological processes in the profiles of gene expression.

The functional annotation of genes whose expression profiles exhibited maximum or minimum correlation with the three principal MDS components showed that the maximum contributions to the first component in the Hackl array and the third component in the Burton array are due to the profiles of genes involved in the signal transmission from receptors on the cell surface and in the cytosol via kinase channels to the nucleus and in the regulation of transcription. The role of genes involved in the transcription regulation and signal transmission is also quite expected in the process of cell determination, and several gene biomarkers of this type have been determined so far [2, 3]. Genes with the expression profiles exhibiting high correlation with the second MDS component are most frequently involved in biological processes such as cell cycle, DNA metabolism, and apoptosis, which is characteristic of the stage of active cell proliferation. Such genes are well known as biomarkers for this phase of adipogenesis [2, 3]. Finally, genes with the expression profiles exhibiting high correlation with the third MDS component are more frequently involved in the transport and metabolism of carbohydrates and lipids, which are the signs of adipocyte phenotype development. However, it should be noted that the third component is characterized by a much lower dispersion as compared to the first two components. This can be related to the fact that the available data cover only the beginning of this process and/or to a decrease in the general consistency of biological processes with increasing distance from the starting point.

Thus, analysis of the distribution of correlations between gene expression profiles and the first three principal MDS components for microarray data makes it possible to refine the role of genes by distributing their profiles between several components. In this respect, the proposed approach is similar to the methods of fuzzy clustering and, in combination with the analysis of functional annotations (experimentally established or theoretically predicted), it helps studying a complicated behavior of the expression profiles of the well-known biomarkers of adipocyte differentiation and, probably, finding new markers of the particular phases of adipogenesis.

REFERENCES

1. F. M. Gregoire, C. M. Smas, and F. S. Sul, Physiol. Rev. **78** (3), 783 (1998).
2. F. M. Gregoire, Exp. Biol. Med. **226** (11), 997 (2001).
3. E. D. Rosen and O. A. MacDougald, Natl. Rev. Mol. Cell Biol **7** (12), 885 (2006).
4. V. V. Konstantinov, A. D. Deev, A. V. Kaoustina, et al., Kardiologiya **42**, (10), 45 (2002).
5. W. H. Chou, J. K. McLaughlin, J. S. Mandel, et al., Cancer Epidemiol. Biomark. Prev. **5** (1), 17 (1996).
6. R. L. Weinster, G. R. Huntger, A. F. Heini, et al., Am. J. Med. **105**, (2), 145 (1998).
7. P. Cornelius, O. A. MacDougald, and M. D. Lane, Annu. Rev. Nutr. **14**, 99 (1994).
8. T. C. Otto and M. D. Lane, Crit. Rev. Biochem. Mol. Biol. **40** (4), 229 (2005).
9. G. R. Burton and R. E. McGehee, Jr., Nutrition **20** (1), 109 (2004).
10. X. Guo and K. Liao, Gene **251** (1), 45 (2000).
11. A. Soukas, N. D. Socci, B. D. Saatkamp, et al., J. Biol. Chem. **276** (36), 34167 (2001).

12. B. A. Jessen and G. J. Stevens, Gene, **299** (1–2), 95 (2002).
13. S. E. Ross, R. L. Erickson, I. Gerin, et al., Mol. Cell. Biol. **22** (16), 5989 (2002).
14. L. Fajas, R. L. Landsberg, Y. Huss-Gercia, et al., Dev. Cell **3** (1), 39 (2002).
15. G. R. Burtton, R. Nagarajan, C. A. Peterson, and R. E. McGehee, Jr., Gene **329**, 167 (2004).
16. Arkansas Cancer Research Center (http://www.acrc.uams.edu/directory/bio.asp?id=65).
17. H. Hackl, T. R. Burkard, A. Sturn, et al., Genome Biol. **6**, (13) Epub. (2005).
18. ArrayExpress (http: www.ebi.ac.uk/arrayexpress/).
19. Systems Biology: Transcriptomics (http: samurai.bionet.nsc.ru/pages/ASNI02/content/pages/ASNI-02_transcriptomics.html).
20. Affymetrix: NETAFFX (http://www.affymetrix.com/analysis/netaffx/) [System Biology: Transcriptomics].
21. D. B. Allison, X. Gui, G. P. Page, and M. Sabripour, Nat. Rev. Gen. **7** (1), 55 (2006).
22. Gene Ontology (http://www.geneontology.org/GO.doc.shtml).
23. Results of Multidimensinal Scaling Analysis of Microarray Data for Adipogenesis (http://www.bionet.nsc.ru/gep/research/adipogenesis.html) [Results of analysis of microarray data on adipogenesis by multidimensional scaling].
24. J. R. Hendsrson, D. Brown, J. A. Richardson, et al., J. Histochem. Cytochem. **50** (1), 107 (2002).
25. Q. Q. Tang, T. C. Otto, and M. D. Lane, Proc. Natl. Acad. Sci. USA **100** (3), 850 (2003).
26. S. Shimba, N. Ishii, Y. Ohta, and T. Ohno, Proc. Natl. Acad. Sci. USA **102** (34), 12071 (2005).
27. E. C. Shugart, A. S. Levenson, C. M. Constance, and R. M. Umek, Cell Growth Differ. **6** (12), 1541 (1995).