Bioinformatic Analysis of Expression Data of ApoE Deficient Mice

Olga Papadodima[1], Allan Sirsjo[2], and Aristotelis Chatziioannou[1]

[1] Institute of Biological Research and Biotechnology, National Hellenic Research Foundation, 48 Vas. Constantinou Ave., 11635 Athens, Greece
[2] School of Health and Medical Sciences, Division of Clinical Medicine, University of Örebro, Sweden

Abstract. Atherosclerosis is a multifactorial disease involving a lot of genes and proteins recruited throughout its manifestation. The present study aims to exploit bioinformatic tools in order to analyze microarray data of atherosclerotic aortic lesions of ApoE knockout mice, a model widely used in atherosclerosis research. In particular, a dynamic analysis was performed among young and aged animals, resulting in a list of 852 significantly altered genes. Pathway analysis indicated alterations in critical cellular processes related to cell communication and signal transduction, immune response, lipid transport and metabolism. Cluster analysis partitioned the significantly differentiated genes in three major clusters of similar expression profile. Promoter analysis applied to functional related groups of the same cluster, revealed shared putative *cis*-elements potentially contributing to a common regulatory mechanism. Finally, by reverse engineering the functional relevance of differentially expressed genes with specific cellular pathways, putative genes acting as hubs were identified, linking functionally disparate cellular processes in the context of traditional molecular description.

Keywords: Microarray analysis, Atherosclerosis, transcriptomic analysis, promoter analysis.

1 Introduction

Atherosclerosis is the leading pathological contributor to cardiovascular morbidity and mortality worldwide, characterized by the progressive accumulation of lipid and fibrous depositions in the vessel wall of medium-sized and large arteries. Although it has traditionally been viewed as simple deposition of lipids within the vessel wall, it is now assumed that atherosclerosis is a multifactorial disease that involves several genes and proteins, activated during its genesis, progress and phenotypic manifestation. During atherogenesis, a complex endothelial activation and dysfunction induced by elevated and modified low-density lipoproteins and many other factors leads to a compensatory inflammatory response (1). Current evidence supports a central role for inflammation, in all phases of the atherosclerotic process. Substantial biological data implicate inflammatory pathways in early atherogenesis, in the progression of lesions, and finally in the thrombotic complications of this disease (2).

I. Maglogiannis, V. Plagianakos, and I. Vlahavas (Eds.): SETN 2012, LNAI 7297, pp. 254–261, 2012.

Clinical investigations, population studies, and cell culture experiments have provided important clues to the pathogenesis of atherosclerosis. However, the use of animal models has had a crucial contribution in the research of the atherosclerotic course. Atherosclerosis will not be developed in laboratory mice under normal conditions. However, targeted deletion of the gene for Apolipoprotein E (ApoE knockout mice) leads to severe hypercholesterolemia and spontaneous atherosclerosis (3). For this reason, ApoE deficient mice are widely used to study atherosclerosis (4). ApoE is a ligand for receptors that clear chylomicrons and very low-density lipoprotein remnants. Furthermore, a number of population studies suggest that ApoE genotype predicts the risk of developing atherosclerosis and related diseases (5).

In this study, we present a bioinformatic analysis based on microarray data derived from ApoE knockout mice. Gene expression data of wild type and ApoE knockout 6, 32 and 78-weeks old mice were used. The dataset that we used was presented in a detailed work studying atherosclerosis and inflammatory pathways during aging (6). The workflow of our analysis consists of six basic steps: normalization, statistical selection, pathway analysis, clustering and promoter analysis. It has been successfully applied in a dynamic analysis of mastic oil treatment of cancer cells, providing novel evidence on the molecular basis of its inhibitory action on tumor growth (7). As a further step, in order to expand our knowledge regarding the functional implication of genes in various cellular processes, prioritizing them according to their centrality, we exploited the hierarchical structure of the Gene Ontology (GO) tree, and candidate hub-genes or interacting proteins were identified.

2 Materials and Methods

2.1 Microarray Data

The mouse dataset used is the GSE 10000, available at Gene Expression Omnibus (GEO) database. Microarrays were prepared following MIAME guidelines, as described in (6). Briefly, RNA from aortic tissue of apoE knockout and wild type animals was hybridized on Affymetrix 430 2.0 Arrays. Three different ages were studied: 6, 32 and 78 weeks.

2.2 Microarray Data Analysis and Statistical Analysis

Microarray data analysis was performed in Gene ARMADA (8). Briefly, background correction was performed employing its gcRMA method followed by Quantile Normalization. Data were $\log_2$ transformed to comply with the normality assumption. Differentially expressed genes in at least one among all the experimental conditions were identified using Gene ARMADA, by performing 1-way ANOVA on $\log_2$ transformed fold changes. The resulting gene list was obtained by setting the p-value threshold to 0.01, the False Discovery Rate (FDR) threshold to 0.05 and by removing genes that presented a fold change below |1|, in log2 scale, in all conditions.

2.3 Prioritized Pathway/Functional Analysis

Statistical enrichment analysis was performed using StRAnGER (9), in order to highlight biological processes including statistically significant numbers of the ANOVA derived genes. In order to expand our knowledge regarding the functional implication of genes in various cellular processes, prioritizing them according to their centrality, we used the online tool GORevenge (10) with the following settings: Aspect: BP (Biological Process), Distance: Graph, Algorithm: BubbleGene and Relaxation: 2.

2.4 Cluster and Promoter Analysis

The list derived from ANOVA was subjected to hierarchical clustering (linkage method: Average, distance: Cosine) in Gene ARMADA. Promoter analysis was performed as previously described in (5), with the difference that only mouse promoters were considered.

3 Results

3.1 Statistically Significant Differentiated Genes

To obtain the aortic gene expression profile of ApoE deficient mice in 6, 32 and 78 week old mice we analyzed the GSE 10000 dataset, containing expression data of aortic tissue from wild type and ApoE knockout mice. Specifically, in order to identify significant alterations among all three tested ages, 1-way ANOVA was applied to expression fold changes between expression in ApoE knockout and wild type animals (p value<0.01 and FDR<0.05) coupled with further filtering on fold change (>|1| in at least one condition in log2 scale). A list of 1033 significantly differentiated probsets was obtained, depicted per time point using a volcano plot representation (Figure 1). These 1033 probsets correspond to 852 annotated genes. It is characteristic that in 6 weeks old mice the number of significantly altered genes is very limited, in 32 weeks old mice the majority of differentiated genes are upregulated while in 78 weeks old mice we have the greater number of differentiated genes.

3.2 Pathway Analysis

For the scope of gaining further insight concerning the biological functionalities of gene expression alterations in a more systematic way, the list of 852 significantly differentiated genes yielded from ANOVA was subjected to statistical enrichment analysis using StRAnGER, exploiting GO terms and Kegg pathways for the task of the functional annotation of the interrogated genes. GO-based analysis focused on the categories of “Biological Process” with a hypergeometric p-value<0.001 suggested several processes as possibly differentiated which are presented in Table 1. A lot of central molecular mechanisms emerge as altered, as indicated by the GO categories listed in Table 1, like differentiation, proliferation (inferred by cell cycle and cell

division GO terms) apoptosis, cell adhesion, signal transduction, and immune response. Kegg–based analysis also indicates alterations in cell adhesion and signal transduction. It is important to note that in conformity to the well established relationship of atherosclerosis and inflammation, the majority (29 out of 32) of the genes under the category "immune response" are upregulated suggesting a stimulation of the immunological mechanisms.

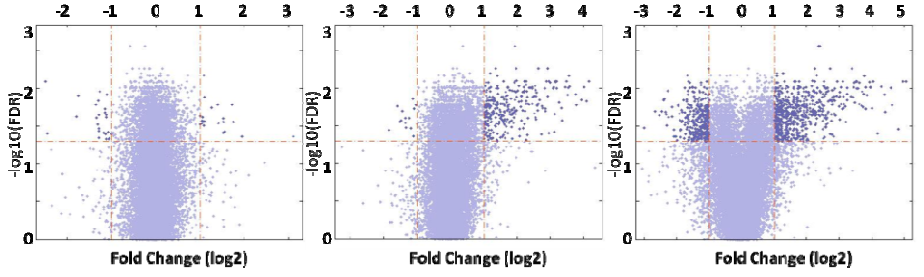

Fig. 1. Volcano plots of the gene list as yielded by ANOVA**.** Each panel represents filtered and normalized data from each experimental condition (3, 6 and 78 weeks old mice). The horizontal axes depict the fold change ratio between ApoE deficient and wild type mice, for each age in $\log_2$ scale while the vertical axes represent statistical significance by depicting the $-\log_{10}$(FDR).

3.3 Cluster Analysis

In order to identify groups of genes presenting similar expression and possibly comprising regulated "waves" of transcription, the list of 1033 significantly differentiated probesets was subjected to hierarchical clustering (Figure 2). Three major clusters can be distinguished: the first one contains transcripts downregulated in 78 week old mice, while their expression remains close to the control (wild type) level at 6 and 32 weeks. The second cluster groups genes which are upregulated at 32 weeks and their expression at ApoE knockout mice remains at high levels, as compared to wild type, also at 78 weeks. The third cluster groups genes whose expression is late upregulated at 78 week old ApoE knockout, as compared to age-matched wild type mice.

Based on these three major clusters, we performed GO-analysis to the genes of each cluster separately. Genes under cluster 1 are functionally connected to processes involved in cell differentiation, adhesion and signal transduction. Cluster 2 contains the greatest number of genes, which are related mainly to mechanisms involved in immune and inflammatory response as well as lipid metabolism. These processes emerge as significantly altered specifically in the case of cluster 2. Cluster 2 genes are also connected to key cellular processes like signal transduction, apoptosis, cell cycle and differentiation. Cluster 3 genes are mainly related to mechanisms concerning gene transcription.

Table 1. GO-analysis. The list of 852 significantly altered genes was submitted to GO analysis elucidating over-represented GO terms. GOT p-value represents the hypergeometric test p-value score for each GO term. Enrichment represents the ratio of the number of times a GO term occurs in the 852 gene list to the number of times this GO term exists in the list of the Affymetrix 430 2.0 array.

GO Annotation	GOT p-value	Enrichment
ion transport	0.00000000003	33/498
signal transduction	0.00000000004	44/803
cell differentiation	0.00000000005	36/480
immune response	0.00000000005	32/250
metabolic process	0.00000000007	38/542
cell adhesion	0.00000000011	36/387
protein amino acid phosphorylation	0.00000000059	32/497
multicellular organismal development	0.00000000128	41/770
proteolysis	0.00000000690	25/358
apoptosis	0.00000010814	24/383
lipid metabolic process	0.00000328813	15/212
protein transport	0.00003352383	22/465
G-protein coupled receptor signaling	0.00028681297	19/436
oxidation reduction	0.00034119987	21/510
cell cycle	0.00043684087	18/417
cell division	0.00050194998	12/231

3.4 Promoter Analysis

In order to investigate whether there are common regulatory transcriptional mechanisms in functional groups of genes revealed by the cluster-based GO analysis, we performed a representative promoter analysis in genes of cluster 2 belonging to the GO category “immune response” and “inflammatory response”. We selected these categories because they are functionally relevant and because they contain an adequate number of genes (the total number of genes in both categories is 36). Table 2 summarizes statistically significant transcription factor (TF) motif families common in promoter sequences of these 36 genes, sorted in descending order in terms of statistical significance.

3.5 Identification of Candidate Hub-Genes

In order to expand our knowledge regarding which genes have critical role, taking into consideration their centrality as described in the GO tree, we used the online tool GORevenge (10). The list of 852 differentiated genes was submitted to GORevenge and the derived list of genes, containing candidate hub-genes, was partitioned to contain only the genes that have been also identified as statistically significantly

differentiated. The derived list (Table 3) contains genes that were identified as significant both by ANOVA and by GORevenge analysis. Significant molecules involved in signaling and developmental mechanisms emerge as central players.

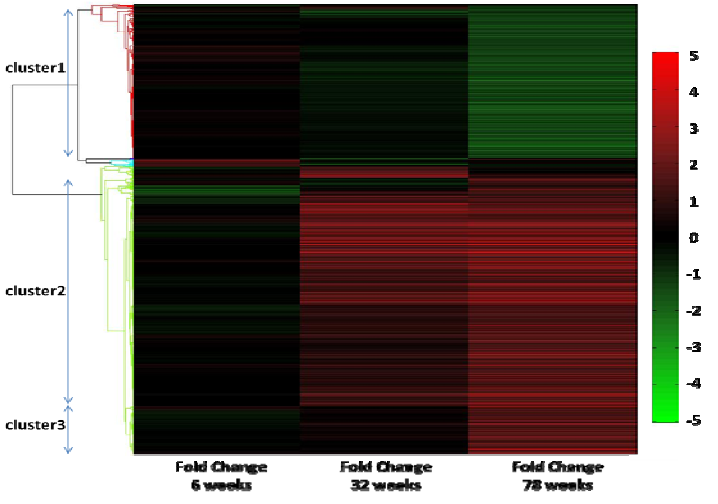

Fig. 2. Hierarchical clustering of the 1033 statistically significant differentiated probesets. Fold changes between the gene expressions in ApoE knockout as compared to age-matched wild type mice are grouped in three major clusters.

Table 2. Common TF motif families in the promoters of 36 genes belonging to cluster 2 and to the categories "immune response" and "inflammatory response". The percentage column depicts the percentage of genes whose promoters have at least one match with the respective motif family.

Family	Description	p-value	%
V$MZF1	Myeloid zinc finger 1 factors	0.00015	91
V$PLAG	Pleomorphic adenoma gene	0.00035	91
V$EREF	Estrogen response elements	0.00038	87
V$CP2F	CP2-erythrocyte Factor, Elf related	0.00046	83
V$SRFF	Serum response element binding factor	0.00060	87
V$CTCF	CTCF and BORIS gene family	0.00062	83
V$KLFS	Krueppel like transcription factors	0.00255	100
V$STAT	Signal transducer and activator of transcription	0.00289	96
V$RXRF	RXR heterodimer binding sites	0.00329	100
V$GREF	Glucocorticoid responsive elements	0.00337	91

Table 3. GOrevenge prioritization. The second column refers to the number of GO terms remaining after Gorevenge pruning, while the third column refers to the original number of bilobical process catecory GO terms of each gene. All presented genes are also differentially expressed. Top 20 genes are shown.

gene symbol	remaining GO terms	original GO terms
Wnt5a	95	112
Fgfr2	75	92
P2rx7	62	73
Igf1	49	56
Tlr2	42	47
Thbs1	41	42
Ptgs2	36	37
Slc11a1	34	40
Psen2	34	37
Ptprc	33	40
Ccnd1	32	37
Foxf1a	32	34
Osr1	30	33
Lyn	28	33
Col1a1	26	29
Adora1	25	25
Adam17	24	29
Cxcl12	24	27
Socs3	24	27

4 Discussion

In this study we presented a detailed bioinformatic analysis of ApoE knockout mice, exploiting different approaches in order to identify critical altered molecular mechanisms and important central players. It was shown that the gene expression profile in atherosclerotic plaques containing arteries of ApoE knockout mice is profoundly different from wild type. Specifically, 852 genes were found as differentially expressed and the majority of them appear after the age of 32 weeks. The indicated altered processes, as revealed by ontology-based enrichment analysis, include adhesion and signal transduction, differentiation, apoptosis and immune response, reflecting the cellular and molecular complexity of atherosclerosis and the cross-talk of endothelial and immune cells in aortic lesions. Cluster analysis revealed three major groups of genes with similar expression profiles which were further analyzed, in order to find functional (GO-based) sub-groups in each cluster. In agreement with the notion that atherosclerosis is an inflammatory disease, immune response and inflammation were the prominent categories indicated as significantly altered in the case of cluster 2, which contains genes upregulated both in 36 and 78 weeks old mice. Promoter analysis of the genes under these categories revealed common binding elements that could

contribute to a common transcriptional regulation. In particular, all of the tested genes (100%) contain *cis*-elements of the RXR family. This family groups together motifs related to the receptors of retinoids, which are recognized by various heterodimers of retinoid X receptors (RXRs) and retinoic acid receptors. Interestingly, RXR has been reported to regulate several genes related to metabolic homeostasis and inflammation (11). Noteworthily, among the identified putative TF binding sites there are estrogen response elements (EREs) in the 87 % of the promoters as well as glucocorticoid responsive elements (GREs) in the 91 % of the tested promoters. It is well known that estrogen and glucocorticoid receptors play important roles in both physiological and pathological conditions involving immunity and inflammation (12). Finally, in the list of prioritized genes in table 3 we can distinguish several with important roles in atherosclerosis related mechanisms. It is noteworthy to mention Tlr2, a member of the Toll-like receptors family, which plays a fundamental role in activation of innate immunity (13). Furthermore, the identification of Psen2 (presenillin 2), a gene implicated in Alzheimer's disease, as candidate hub gene is interesting because genes implicated in Alzheimer have been reported to affect cholesterol or lipoprotein function and have also been implicated in atherosclerosis (14). Concluding, this bioinformatic analysis of ApoE knockout mice revealed critical altered cellular mechanisms governing atherosclerosis and indicated important molecular players.

References

1. Stoll, G., et al.: Stroke 37, 1923–1932 (2006)
2. Hansson, G.K.: N. Engl. J. Med. 352, 1685–1695 (2005)
3. Zhang, S.H., et al.: Science 258, 468–471 (1992)
4. Mahley, R.W., et al.: Annu. Rev. Genom. Hum. Genet. 1, 507–537 (2000)
5. Song, Y., et al.: Ann. Intern. Med. 141, 137–147 (2004)
6. Grabner, R., et al.: J. Exp. Med. 206, 233–248 (2009)
7. Moulos, P., Papadodima, O., et al.: BMC Med. Genomics 2, 68 (2009)
8. Chatziioannou, A., et al.: BMC Bioinformaics 10, 354 (2009)
9. Chatziioannou, A., et al.: Front. Neurosci. 5, 8 (2011)
10. Moutselos, K., et al.: IEEE Transactions on Biomedical Engineering 58, 12 (2011)
11. Nohara, A., et al.: J. Atheroscler Thromb 16, 303–318 (2009)
12. Gouni-Berthold, I., et al.: Curr. Drug Targets Cardiov. Haematol. Disord. 5, 513–523 (2005)
13. Borrello, S., et al.: Int. J. Immunopathol Pharmacol. 24(3), 549–556 (2011)
14. Carter, C.J.: Neurochem. Int. 50(1), 12–38 (2007)