European Review for Medical and Pharmacological Sciences 2013; 17: 1909-1915

Pathway analysis detect potential mechanism for familial combined hyperlipidemia

Y.L. WANG, F. XUE[1], L.Z. LIU[1], Z.H. HE[1]

Department of Orthopedics and [1]Department of Interventional Radiology; Feng Xian Center Hospital, Shanghai, China

Abstract. – BACKGROUND: Familial combined hyperlipidemia (FCHL) is the most commonly inherited hyperlipidemia in men. It constitutes a substantial risk factor for atherosclerosis patients.

AIM: To delineating the potential mechanism of FCHL by bioinformatics tools.

MATERIALS AND METHODS: In this study, Protein-Protein Interaction (PPI) network was constructed to identify the potential functional proteins and their interactive relationships in familial combined hyperlipidemia.

RESULTS: Our results showed that androgen receptor (AR) might play an important role in familial combined hyperlipidemia by interaction with TGIF1, NR3C1, KLK2, etc. Some pathways were also identified, such as Hedgehog signaling pathway, Phosphatidylinositol signaling system, and Long-term depression, which were all demonstrated participating in lipid metabolism in previous experiments.

CONCLUSIONS: Although lack of direct evidence, by PPI network construction it proved AR is a key factor in FCHL, and also demonstrated that PPI network construction is an alternative avenue for FCHL analysis.

Key Words:
Familial combined hyperlipidemia, Androgen receptor, Signaling pathway.

Introduction

Familial combined hyperlipidemia (FCHL) is a common genetic lipid disorder in men characterized by elevated levels of plasma total cholesterol, triglycerides, apolipoprotein (apo) B, and apoC3 resulting from hepatic VLDL secretion[1]. In addition, insulin resistance of adipose tissue and muscle has been documented in FCHL as well. Adipose tissue is one of the major contributors of free fatty acids (FFA) in the circulation. High levels of FFA in the circulation may lead to both a decrease in insulin-stimulated glucose uptake in skeletal muscle and an increase in hepatic VLDL lipoprotein synthesis[2]. Consequently, FCHL showed clinical manifestations of premature coronary heart disease i.e., atherosclerosis.

In recent years, many candidate genes have been identified to be related with FCHL phenotype. Such as the upstream transcription factor 1 (USF1), is a transcription factor known to regulate the expression of a number of genes participating in glucose and lipid metabolism, therefore, it provides an excellent candidate for FCHL. USF1 further regulates the expression of several genes that participate in glucose and lipid metabolism, such as apolipoprotein C3 (APOC3), apolipoprotein A2 (APOA2), hormone sensitive lipase (LIPE), hepatic lipase (LIPC), glucokinase(GCK), insulin(INS), glucagon receptor (GCGR), fatty acid synthase (FAS), acetyl-CoA carboxylase alpha (ACACA) and plasminogen activator inhibitor-1 (PAI1)[3]. The hepatic nuclear factor 4 alpha (HNF4A) is also associated with high serum lipid levels and the metabolic syndrome in FCHL families, which also could regulate multiple genes expression. Besides, the tumor necrosis factor receptor superfamily member 1B (TNFRSF1B) gene and cholesteryl ester transfer protein (CETP) have been shown to influence the FCHL phenotype in recent researches[4-5]. In brief, these candidate gene studies could lay certain theoretical foundation for FCHL treatment.

PPI network analysis as a shortcut approach is applied to investigate physiological mechanisms in health and disease[6]. In this study, we used a similar strategy to identify functional genes and signaling pathways in FCHL, with hoping to establish a theoretical foundation for the future research.

Materials and Methods

Data Source

Affymetrix Microarray Data

The transcription profile of GSE1010 was obtained from NCBI GEO database (http://www.ncbi.nlm.nih.gov/geo/) which is based

Corresponding Author: Yongli Wang, MD; e-mail: xuefengdr01@gmail.com

on the Affymetrix Human Genome U133A Array. Here, 12 FCHL patients and 12 normal samples as control were used to study the related genes.

Pathway Data

KEGG (Kyoto Encyclopedia of Genes and Genomes) is a collection of online databases dealing with genomes, enzymatic pathways, and biological chemicals[7]. The pathway database records networks of molecular interactions in the cells, and variants of them specific to particular organisms (http://www.genome.jp/kegg/). Total 130 pathways, involving 2287 genes, were collected from KEGG.

Protein-Protein Interaction (PPI) Data

The Human Protein Reference Database (HPRD)[8] is a protein database accessible through the internet. The Biological General Repository for Interaction Datasets (BIOGRID[9] is a curated biological database of protein-protein and genetic interactions.

Collect the protein-protein interaction (PPI) data from the HPRD and BIOGRID database. Total 326119 unique PPI pairs were collected in which 39240 pairs are from HPRD and 379426 pairs are from BIOGRID.

Methods

Differentially Expressed Genes (DEGs) Analysis

For the GSE1010 dataset, the limma method[10] was used to identify differentially expressed genes (DEGs). The original expression datasets from all conditions were extracted into expression estimates, and then constructed the linear model. The DEGs only with the fold change value larger than 2 and *p*-value less than 0.05 were selected.

Co-expression Analysis

For demonstrating the potential PPI relationship, the Pearson Correlation Coefficient (PCC) was calculated for all pair-wise comparisons of gene-expression values between TFs and the DEGs. The PPI relationships whose absolute PCC are larger than 0.6 were considered as significant.

Gene Ontology Analysis

DAVID[11], a high-throughput and integrated data-mining environment, analyzes gene lists derived from high-throughput genomic experiments. Use the DAVID to identify over-represented gene ontology (GO) categories in biological process.

PPI Network Construction

Using the PPI data that have been collected from HPRD and BIOGRID database, we matched the interactions between two DEGs. Then selected the interaction that PCC of the two DEGs is large than 0.6 as the candidate relationships.

Significance Analysis of Pathway Under PPI Data

To determine the co-expressed significance of a gene pair in disease cases, we used the PCC test to calculate the *p*-value.

Map those *p*-values to the nodes and edges in the PPI network. The following formula is used to define a function as the combination of statistical significance of an interaction by a scoring scheme. The detail description could be seen in Liu et al[12].

$$S(\mathrm{e}) = f\,[diff(x), corr(x, y), diff(y)]$$

$$= -2 \sum_{i=1}^{k} \log_e (pi)$$

The diff(x) and diff(y) are differential expression assessments of gene x and gene y, respectively. corr(x,y) represents their correlation between gene x and gene y. f is a general data integration method that can handle multiple data sources differing in statistical power. Where k = 3, p1 and p2 are the p-values of differential expression of two nodes, p3 is the p-value of their co-expression.

To define the significance of a pathway P, Sp, we summarize all the scores of edges S(e) of every pathway,

$$\mathrm{Sp} = \sum_{e \in p} S(e)$$

To estimate a *p*-value for significance of this pathway, we iteratively compute similar scores 10^5 times on randomly generated pathways of the same size as that of pathway P. The frequency of scores that are larger than Sp is used as the significance *p*-value of pathway P.

Significance Analysis of Pathway

We adopted an impact analysis that includes the statistical significance of the set of pathway genes but also considers other crucial factors

such as the magnitude of each gene's expression change, the topology of the signaling pathway, their interactions, etc[13]. In this model, the Impact Factor (IF) of a pathway Pi is calculated as the sum of two terms:

$$IF(Pi) = \log\left(\frac{1}{pi}\right) + \frac{\sum_{g \in Pi} [PF(\mathrm{g})]}{[\mathrm{VE}] \cdot \mathrm{N}_{\mathrm{de}}(Pi)}$$

The first term is a probabilistic term that captures the significance of the given pathway Pi from the perspective of the set of genes contained in it.

It is obtained by using the hyper geometric model in which Pi is the probability of obtaining at least the observed number of differentially expressed gene, Nde, just by chance[14-15].

The second term is a functional term that depends on the identity of the specific genes that are differentially expressed as well as on the interactions described by the pathway (i.e., its topology).

The second term sums up the absolute values of the perturbation factors (PFs) for all genes g on the given pathway Pi.

The PF of a gene g is calculated as follows:

$$PF(\mathrm{g}) = \mathrm{VE}(\mathrm{g}) + \sum_{u \in USg} \beta_{\mathrm{ug}} \cdot \frac{PF(\mathrm{u})}{N_{\mathrm{ds}}(\mathrm{u})}$$

In this equation, the first term $\Delta E\ (g)$ captures the quantitative information measured in the gene expression experiment. The factor $\Delta E\ (g)$ represents the normalized measured expression change of the gene g. The first term $\Delta E\ (g)$ in the above equation is a sum of all PFs of the genes u directly upstream of the target gene g, normalized by the number of downstream genes of each such gene $N_{ds}\ (u)$, and weighted by a factor β_{ug}, which reflects the type of interaction: $\beta_{ug} = 1$ for induction, $\beta_{ug} = -1$ for repression (KEGG supply this information about the type of interaction of two genes in the description of the pathway topology). US_g is the set of all such genes upstream of g. We need to normalize with respect to the size of the pathway by dividing the total perturbation by the number of differentially expressed genes on the given pathway, $N_{de}\ (P_i)$. In order to make the *IF*s as independent as possible from the technology, and also comparable between problems, we also divide the second term in equation 1 by the mean absolute fold change ΔE, calculated across all differentially expressed genes.

Results

PPI Network in FCHL

To get DEGs of FCHL, we obtained publicly available microarray data sets GSE1010 from GEO. After microarray analysis, the differentially expressed genes (DEGs) with the fold change > 2 and p-value < 0.05 were selected. 232 genes were selected as DEGs from GSE1010. Then we select 56 DEGs randomly to construct PPI network, and 56 edges were obtained. From the result, it showed that AR protein has the high degree in the network with the p-value 0.045 (Figure 1).

GO Analysis of the PPI Network in FCHL

We used 56 DEGs in the PPI network (Figure 1) to do the GO enrichment analysis. In order to get more credible result, the p-value < 0.01 and FDR <0.05 one were selected. Several Gene Ontology (GO) categories were enriched among these genes in the PPI network, including regulation of cell proliferation, enzyme linked receptor protein signaling pathway, protein amino acid phosphorylation and so on (Table I).

Significant Pathway Under Protein-Protein Interactions

Through weighting the network edges by integrating gene expression and co-expression data in FCHL (see Methods), we can evaluate if the interactions among those pathways are significant or not, and each pathway got a Sp score. After sorting all related pathways base on the score of Sp with the p-value < 0.05, we lists the top 10 significant pathways in Table II.

Significant Pathway in FCHL Base on the Hypergeometric Distribution

To identify the relevant pathways changed in FCHL, we used a statistical approach on pathway level. Significance analysis at single gene level may suffer from the limited number of samples and experimental noise that can severely limit the power of the chosen statistical test. Pathway can provide an alternative way to relax the significance threshold applied to single genes and may lead to a better biological interpretation. So, we adopted a pathway based impact analysis method that contained many factors including the statistical significance of the set of differentially expressed genes in the pathway, the magnitude of each gene's expression change, the topology of the signaling path-

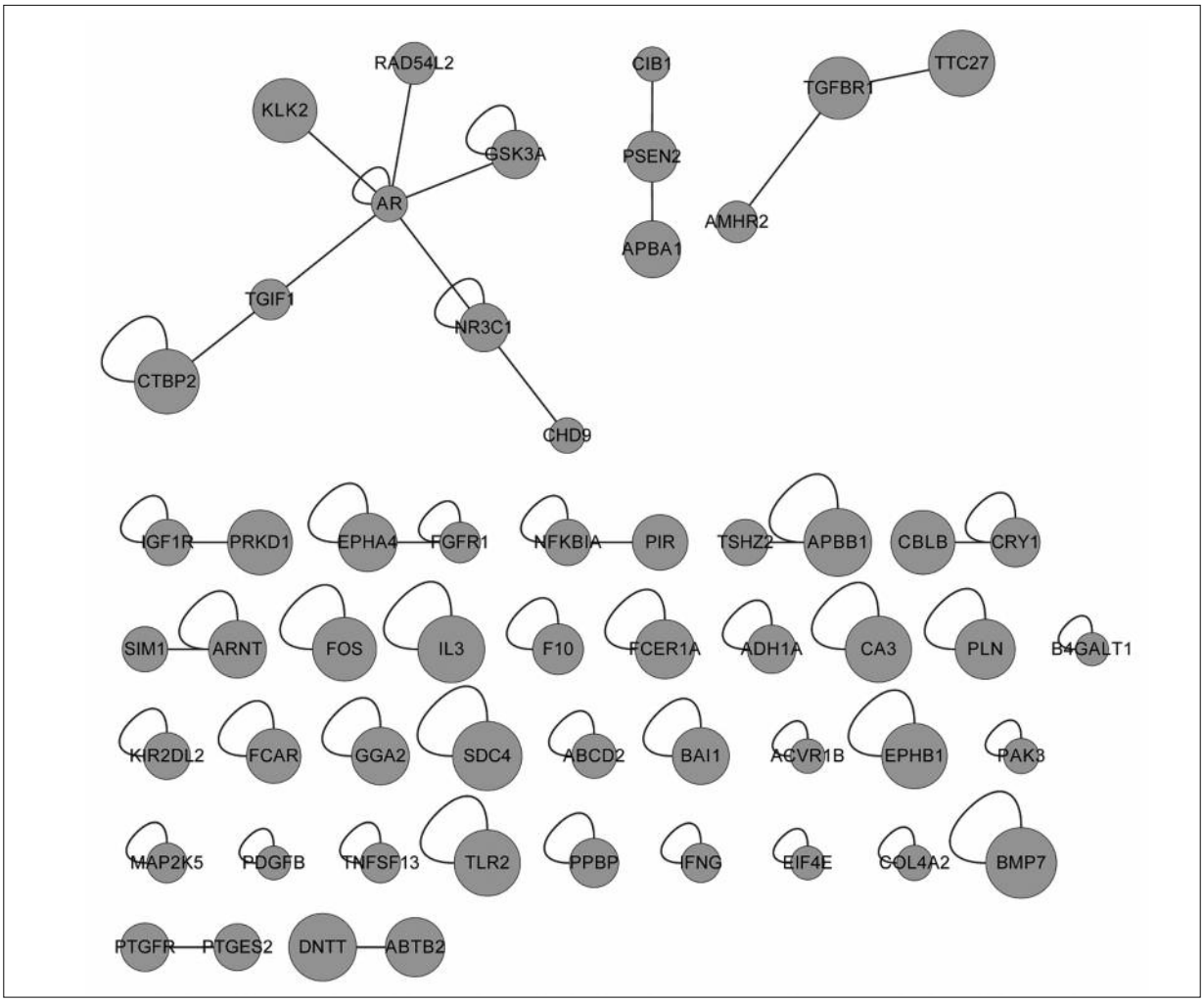

Figure 1. PPI network of FCHL. The nodes strand for the DEGs with the *p*-value range from 0 to 0.05, less *p*-value large node size. The edges stand for the connections between the two nodes.

Table I. GO enrichment analysis of biological process.

Category	Term	Count	Percent	*p*-value	FDR
BP	GO:0042127~regulation of cell proliferation	16	28.57143	2.03E-07	3.26E-04
BP	GO:0007167~enzyme linked receptor protein signaling pathway	11	19.64286	6.94E-07	0.001115
BP	GO:0006468~protein amino acid phosphorylation	14	25	1.23E-06	0.001971
BP	GO:0008284~positive regulation of cell proliferation	11	19.64286	3.88E-06	0.006242
BP	GO:0016310~phosphorylation	14	25	9.08E-06	0.014604
BP	GO:0010604~positive regulation of macromolecule metabolic process	14	25	1.90E-05	0.03053
BP	GO:0045935~positive regulation of nucleobase, nucleoside, nucleotide and nucleic acid metabolic process	12	21.42857	2.39E-05	0.038434

Table II. Significant pathway analysis base on the PPI.

KEGGID	Description	Score	*p*-value
hsa04330	Notch signaling pathway	10029834	0
hsa04062	Chemokine signaling pathway	8059321	0
hsa04140	Regulation of autophagy	6550331	0
hsa04060	Cytokine-cytokine receptor interaction	5671081	1.00E-05
hsa05322	Systemic lupus erythematosus	4477764	0
hsa03420	Nucleotide excision repair	3591123	7.00E-05
hsa04621	NOD-like receptor signaling pathway	2188797	0.0152
hsa04012	ErbB signaling pathway	2183395	0
hsa04310	Wnt signaling pathway	1936697	0.00306
hsa00534	Glycosaminoglycan biosynthesis – heparan sulfate	773256.2	0.0008

way, their interactions and so on. The impact analysis method yields many significant pathways contained Phosphatidylinositol signaling system, Gap junction, Circadian rhythm and so on (Table III).

Merge the two Significant Pathway Methods

We merged the significant signaling pathways got by protein-protein interaction method and method based on hypergeometric distribution that p-value <0.01. Only the Hedgehog signaling pathway was selected.

Discussion

From the results of PPI network construction in FCHL, we found that androgen receptor (AR) might participate in FCHL by interaction with other proteins, such as TGIF1, NR3C1, KLK2, etc. Besides, Hedgehog signaling pathway, phosphatidylinositol signaling system, and long-term depression were all identified related with FCHL. We would discuss the associated relationship based on previous studies as follows.

AR (androgen receptor) gene is more than 90 kb long and codes for a protein that has 3 major functional domains: the N-terminal domain, DNA-binding domain, and androgen-binding domain. The protein functions as a steroid-hormone activated transcription factor. Upon binding the hormone ligand, the receptor dissociates from accessory proteins, translocates into the nucleus, dimerizes, and then stimulates transcription of androgen responsive genes. Subjects with the Trp64Arg genotype of the beta3-AR gene had higher rates of glucose oxidation and lower levels of free fatty acids (FFAs) in the fasting state and during the euglycemic clamp than subjects with the Trp64Trp genotype, which indicated that in FCHL families, codon 64 polymorphism of the beta3-AR gene may influence the rate of glucose oxidation via FFA levels[16].

TGIF1 protein is a homeodomain transcription repressor that was shown to bind to the retinoid

Table III. Significant pathway analysis result.

Pathway name	Impact factor	Genes in pathway#	Input genes in pathway#	% pathway genes in input	Corrected gamma *p*-value
Phosphatidylinositol signaling system	48.595	76	4	5.263	3.90E-20
Gap junction	10.771	96	3	3.125	2.47E-04
Circadian rhythm	10.049	13	1	7.692	4.78E-04
Long-term potentiation	8.762	73	3	4.11	0.001528
Calcium signaling pathway	8.405	182	8	4.396	0.002104
Long-term depression	7.4	75	4	5.333	0.005135
Adherens junction	6.534	78	4	5.128	0.010948
Chronic myeloid leukemia	5.343	75	5	6.667	0.030329
Hedgehog signaling pathway	5.244	57	4	7.018	0.032963
Pathways in cancer	5.094	330	11	3.333	0.037377

receptor (RXR) response element and inhibit RXR-mediated transcription by competitive DNA binding to an overlapping site. In addition, the protein is an active transcriptional co-repressor of SMAD2 and may participate in the transmission of nuclear signals during development. Several studies have examined networks involved in the FCHL atherosclerosis inflammatory response in macrophages. The results showed that TGIF1 was identified as a potential novel transcriptional regulator[17]. However, another study showed that TGIF1 selectively represses AR-mediated transcription from several AR-responsive promoters. The repression is mediated through binding of TGIF to the DNA binding domain of AR[18]. Therefore, we supposed that TGIF1 may interact with AR to regulate atherosclerosis induced by FCHL.

Hedgehog signaling was found to play a regulatory role in FCHL atherosclerosis development and progression by downstream effectors genes such as transforming growth factorβ (TGFβ) and bone morphogenic protein (BMP), and its inhibition reduced plasma cholesterol levels[19]. FCHL patients showed significantly increased levels of 7β-hydroxycholesterol compared with the control. In the Hedgehog protein processing domain, 7β-hydroxycholesterol was demonstrated to be the only oxysterol (oxidized derivatives of cholesterol) to retain full activity in the ability to substitute cholesterol in the processing domain-mediated transfer reaction. This particular property of 7β-hydrocholesterol raises the possibility that Hedgehog signaling proteins may be modified by oxidative stress driven by-products in FCHL[20].

APOA5 is suggested to reduce plasma triglycerides. Over-expression of APOA5 has been shown to lower triglycerides in mice and ApoA5 knockout mice have severe hypertriglyceridemia. However, phosphorylation of USF1 by phosphatidylinositol 3-kinase prevents it from binding to the APOA5 E-box motif promotor, thereby modulating its ability to transactivate APOA5 expression in FCHL[21].

G-substrate gene (GSBS) variation (-1323 T>C) could modify the lipoprotein phenotype of plasma cholesterol and HDL-cholesterol in FCHL patients. GSBS was originally identified as a gene coding an endogenous substrate for cyclic guanosine monophosphate (cGMP)-dependent protein kinase that exists in cerebellar Purkinje cells, and it is possibly involved in the induction of long-term depression. The expression of GSBS in the hypothalamus may indicate that GSBS effect for food intake by the hypothalamo-pituitary-adrenal axis[22].

Conclusions

Relative salient results were obtained by PPI network analysis in our work. Hedgehog signaling, phosphorylation signaling, and long-term depression were all demonstrated participating in cholesterol metabolism to influence FCHL in the end likely. However, because lack of direct evidence, the results about AR interacting with other proteins TGIF1, NR3C1, KLK2, etc are indispensable to perform further experiments to confirm this conclusion.

Conflict of Interest

None declared.

References

1) van Himbergen TM, Otokozawa S, Matthan NR, Schaefer EJ, Buchsbaum A, Ai M, van Tits LJ, de Graaf J, Stalenhoef AF. Familial combined hyperlipidemia is associated with alterations in the cholesterol synthesis pathway. Arterioscl Thromb Vasc 2010; 30: 113-120.

2) Eurlings PM, van der Kallen CJ, Geurts JM, van Greevenbroek MM, de Bruin TW. Genetic dissection of familial combined hyperlipidemia. Mol Genet Metab 2001; 74: 98-104.

3) Suviolahti E, Lilja HE, Pajukanta P. Unraveling the complex genetics of familial combined hyperlipidemia. Ann Med 2006; 38: 337-351.

4) Geurts JM, Janssen RG, van Greevenbroek MM, van der Kallen CJ, Cantor RM, Bu X, Aouizerat BE, Allayee H, Rotter JI, de Bruin TW. Identification of TNFRSF1B as a novel modifier gene in familial combined hyperlipidemia. Hum Mol Genet 2000; 9: 2067-2074.

5) Shoulders CC, Jones E, Naoumova R. Genetics of familial combined hyperlipidemia and risk of coronary heart disease. Hum Mol Genet 2004; 13: R149.

6) Beck AH, Espinosa I, Edris B, Li R, Montgomery K, Zhu S, Varma S, Marinelli RJ, van de Rijn M, West RB. The macrophage colony-stimulating factor 1 response signature in breast carcinoma. Clin Cancer Res 2009; 15: 778.

7) Kanehisa M. The KEGG database. Novart Fdn Symp 2002; 247: 91-101; discussion -3, 19-28, 244-252.

8) Keshava Prasad TS, Goel R, Kandasamy K, Keerthikumar S, Kumar S, Mathivanan S, Nixon J, Van Auken K, Wang X, Shi X, Reguly T, Rust JM, Winter A, Dolinski K, Tyers M. Human Protein Reference Database–2009 update. Nucleic Acids Res 2009; 37: D767-772.

9) Stark C, Breitkreutz BJ, Chatr-Aryamontri A, Boucher L, Oughtred R, Livstone MS, Nixon J, Van Auken K, Wang X, Shi X, Reguly T, Rust JM, Winter A, Dolinski K, Tyers M. The BioGRID Interaction Database: 2011 update. Nucleic Acids Res 2011; 39: D698-704.

10) Smyth GK. Linear models and empirical bayes methods for assessing differential expression in microarray experiments. Stat Appl Genet Mol Biol 2004; 3: Article3.

11) Huang da W, Sherman BT, Lempicki RA. Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. Nat Protoc 2009; 4: 44-57.

12) Liu ZP, Wang Y, Zhang XS, Chen L. Identifying dysfunctional crosstalk of pathways in various regions of Alzheimer's disease brains. BMC Syst Biol 2010; 4: S11.

13) Draghici S, Khatri P, Tarca AL, Amin K, Done A, Voichita C, Georgescu C, Romero R. A systems biology approach for pathway level analysis. Genome Res 2007; 17: 1537-1545.

14) Tavazoie S, Hughes JD, Campbell MJ, Cho RJ, Church GM. Systematic determination of genetic network architecture. Nat Genet 1999; 22: 281-285.

15) Draghici S, Khatri P, Martins RP, Ostermeier GC, Krawetz SA. Global functional profiling of gene expression. Genomics 2003; 81: 98-104.

16) Pihlajamaki J, Rissanen J, Valve R, Heikkinen S, Karjalainen L, Laakso M. Different regulation of free fatty acid levels and glucose oxidation by the Trp64Arg polymorphism of the beta3-adrenergic receptor gene and the promoter variant (A-3826G) of the uncoupling protein 1 gene in familial combined hyperlipidemia. Metabolism 1998; 47: 1397-1402.

17) Ramsey SA, Gold ES, Aderem A. A systems biology approach to understanding atherosclerosis. EMBO Mol Med 2010; 2: 79-89.

18) Sharma M, Sun Z. 5'TG3' interacting factor interacts with Sin3A and represses AR-mediated transcription. Mol Endocrinol 2001; 15: 1918-1928.

19) Beckers L, Heeneman S, Wang L, Burkly L, Rousch M, Davidson NO, Gijbels MJ, de Winther MP, Daemen MJ, Lutgens E. Disruption of hedgehog signalling in ApoE-/- mice reduces plasma lipid levels, but increases atherosclerosis due to enhanced lipid uptake by macrophages. J Pathol 2007; 212: 420-428.

20) Micheletta F, Iuliano L. Free radical attack on cholesterol: oxysterols as markers of oxidative stress and as bioactive molecules. Immun Endocr Metab 2006; 6: 305-316.

21) Nowak M, Helleboid-Chapman A, Jakel H, Martin G, Duran-Sandoval D, Staels B, Rubin EM, Pennacchio LA, Taskinen MR, Fruchart-Najib J, Fruchart JC. Insulin-mediated down-regulation of apolipoprotein A5 gene expression through the phosphatidylinositol 3-kinase pathway: role of upstream stimulatory factor. Mol Cell Biol 2005; 25: 1537.

22) Ono S, Ezura Y, Emi M, Fujita Y, Takada D, Sato K, Ishigami T, Umemura S, Takahashi K, Kamimura K, Bujo H, Saito Y. A promoter SNP (-1323T>C) in G-substrate gene (GSBS) correlates with hypercholesterolemia. J Hum Genet 2003; 48: 447-450.