Genetics and Molecular Research

Prediction of genetic risk factors of atherosclerosis using various bioinformatic tools

H.X. Wang and Y.X. Zhao

Department of Neurology, Department of Dermatology,
Jinan Central Hospital Affiliated to Shandong University, Jinan, China

Corresponding author: Y.X. Zhao
E-mail: zhyyxin@163.com

Genet. Mol. Res. 15 (2): gmr.15027347
Received September 21, 2015
Accepted November 19, 2015
Published April 4, 2016
DOI http://dx.doi.org/10.4238/gmr.15027347

ABSTRACT. The aim of this study was to identify potential markers of atherosclerosis development in familial hypercholesterolemia (FH) patients. GSE13985 microarray data, generated using blood samples from 5 FH patients and 5 matched controls, was downloaded from the Gene Expression Omnibus. Differentially expressed genes (DEGs) between FH and controls were identified and a protein-protein interaction (PPI) network was constructed. Module and hub proteins were screened in this network. The module genes were subjected to a gene ontology (GO) analysis, and a Kyoto Encyclopedia of Genes and Genomes enrichment analysis was also performed. A total of 394 genes, including 125 up- and 269 down-regulated genes, were differentially expressed. Ribosomal proteins L9 (RPL9), L35 (RPL35), and S7 (RPS7) were designated as hub nodes in the PPI network. The DEGs were found to be significantly enriched in ribosomal and oxidative phosphorylation pathways. Ribosomal protein genes were found to be involved in the ribosomal pathway. The cytochrome-c oxidase (*COX*) genes COX subunit VIIa polypeptide 2 (*COX7A2*), COX subunit VIIb (*COX7B*), COX subunit VIIc (*COX7C*), and COX subunit VIc (*COX6C*) were enriched in the oxidative phosphorylation pathway. Module analysis

and GO enrichment analysis identified ribosomal proteins as important regulators of FH. Ribosomal and oxidative phosphorylation pathways may be closely associated with atherosclerosis development. Ribosomal protein genes and cytochrome-coxidase genes may be potential therapeutic targets for atherosclerosis.

Key words: Atherosclerosis; Familial hypercholesterolemia; Differentially-expressed genes; Bioinformatic analysis; Mechanism

INTRODUCTION

Atherosclerosis is a chronic disease caused by the deposition of lipid (mainly cholesterol and cholesterol esters) to the intima of the arterial wall, increasing its thickness (Moghadasian, 2002). Atherosclerosis contributes to high disease mortality in coronary artery disease and cerebrovascular disease worldwide. Patients in the initial stages of atherosclerosis are often asymptomatic, and can be so for decades. Therefore, very few cases of atherosclerosis are diagnosed early, increasing the need for methods to predict patients at a greater atherogenic risk. Familial hypercholesterolemia (FH) is a proven significant risk factor for the development of atherosclerosis (Cheung and Lam, 2010). FH is an autosomal dominant disorder characterized by high cholesterol levels, specifically low-density lipoproteins (LDL). FH atherosclerosis is a leading cause of death worldwide.

Many molecular biological studies have attempted to clarify the pathogenesis of atherosclerosis in FH patients. High LDL-cholesterol is a risk factor for atherosclerosis, and deletion of the LDL receptor gene (*LDLR*) has been proven to be associated with FH (Hobbs et al., 1987). Smilde et al. (2001) reported that aggressive reduction in LDL-cholesterol levels was associated with a corresponding reduction in atherosclerosis progression in FH patients. Previous research at the genomic level has indicated that the over-expression of cholesteryl ester transfer protein reduces cholesterol levels in the body (Ishibashi et al., 1993). Moreover, a decrease in the production of interleukin-12, which links the 12/15-lipoxygenase pathway, was associated with reduced atherosclerosis in a mouse model of FH (Zhao et al., 2002). However, the molecular mechanism of atherosclerosis is not fully understood, and gene therapy available for atherosclerosis remains insufficient.

This study utilized GSE13985 microarray data obtained from blood samples of FH patients and controls. The differentially expressed genes (DEGs) between FH and normal blood samples were analyzed using various bioinformatic methods. Significant DEGs and DEG-related functions and pathways were analyzed. The purpose of this study was to explore the potential markers of atherosclerosis in FH patients, and uncover specific targets that could prevent the development and progression of atherosclerosis.

MATERIAL AND METHODS

Affymetrix microarray data

The public Gene Expression Omnibus was searched, and microarray data (GenBank accession No.: GSE13985) related to atherosclerosis were obtained. The dataset was deposited

by Režen et al. (2008) in the National Center for Biotechnology Information (NCBI, http://www.ncbi.nlm.nih.gov/geo/) database on December 18, 2008. Gene expression data was developed from the white blood cells of 5 FH patients and 5 matched controls. Raw data and probe annotation information were downloaded using the platform GPL570 (Affymetrix Human Genome U133 Plus 2.0 Array; Affymetrix Inc., Santa Clara, CA, USA).

Data preprocessing and differential expression analysis

Probe-level data in CEL files were converted to expression measures. The data were preprocessed using the Affy package (http://www.bioconductor.org/packages/release/bioc/html/affy.html) (Team RC, 2012) in the R language. If multiple probes corresponded to the same gene, the mean expression value was calculated to represent the gene expression level of that gene. The missing values were input using the K-nearest neighbors method (Troyanskaya et al., 2001). The obtained data were subjected to quantile normalization using the Affy package in R.

The DEGs in blood cells of FH patients were compared against those of matched controls with Limma (http://master.bioconductor.org/packages/release/bioc/html/limma.html) (Smyth, 2004). P value < 0.05 and |fold change (FC)| > 1.5 ($|\log_2 FC| > 0.585$) were defined as threshold values.

Protein-protein interaction (PPI) network

Search Tool for the Retrieval of Interacting Genes (STRING, http://www.string-db.org/) (Szklarczyk et al., 2011) is a public database for the prediction of gene interactions with a confidence score. The interactions among DEGs were screened using STRING and the protein pairs with a confidence score > 0.9, in order to construct PPI networks. Connectivity degrees are the number of edges linked to a given node. Important nodes with high degrees in the network were analyzed as hub proteins. Interactions with confidence scores >0.9 were identified as candidates for further analysis. The PPI network was visualized using Cytoscape (http://cytoscape.org/plugins.html) (Smoot et al., 2011).

Module analysis for PPI network

Gene products with similar functions are clustered in the same module to regulate the various biological processes (BPs). Molecular Complex Detection (MCODE) (Bader and Hogue, 2003) is a method used to detect densely connected regions in PPI networks. MCODE was used to perform a module analysis of the PPI network. Modules with degree cutoffs ³2 (each node in module with a degree of at least 2) and K-core values ³2 (each node with at least 2 neighboring nodes) were identified.

Gene Ontology (GO) database (http://geneontology.org/) (Ashburner et al., 2000) is a large collection of annotated genes. The functions of the modules were determined by gene ontology (GO) analysis of the module genes using the Bingo plugin of Cytoscape (Maere et al., 2005). In this study, GO terms with fold discovery rate <0.05 were considered to be significant.

Pathway analysis

The Kyoto Encyclopedia of Genes and Genomes (KEGG) knowledge database (http://

www.kegg.jp/) (Altermann and Klaenhammer, 2005), a major database assisting in pathway analysis, is used to identify significantly enriched metabolic or signal pathways for target genes. The Database for Annotation, Visualization, and integrated discovery (DAVID, http://david.abcc.ncifcrf.gov/) (Huang et al., 2009) is an online tool that provides a comprehensive set of functional annotations for a number of genes. DEGs in the PPI network were subjected to a KEGG pathway enrichment analysis using the DAVID online tool. P values < 0.05 were the cutoff criterion for functional enrichment analysis.

RESULTS

Identification of DEGs

As shown in Figure 1, the raw expression data was normalized after preprocessing. A total of 394 DEGs, including 125 up-regulated and 269 down-regulated genes, were obtained.

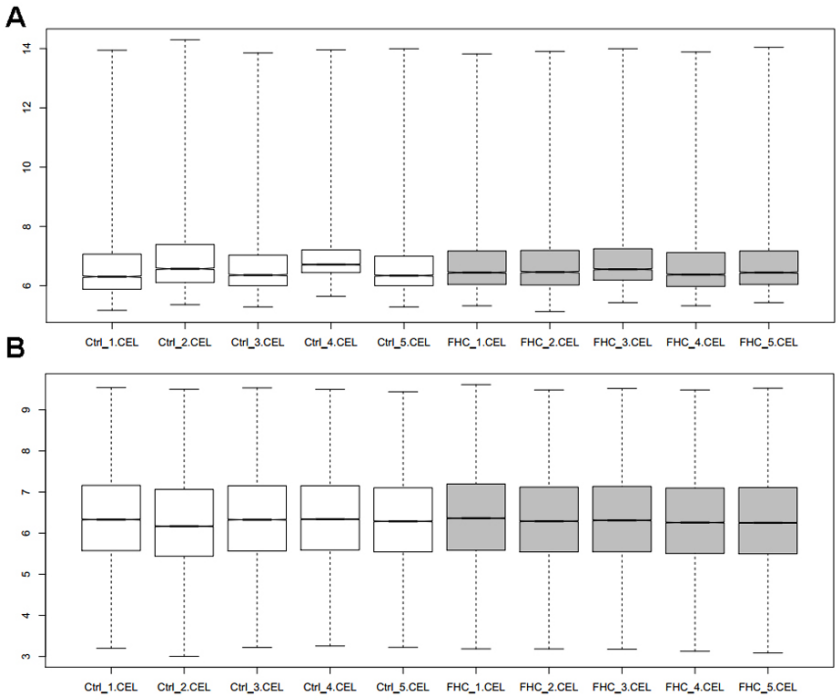

Figure 1. Box plots of data normalization. The x-coordinate represents samples; y-coordinate represents gene expression values. White boxes represent normal samples and gray boxes represent familial hypercholesterolemia samples.

PPI network analysis

The PPI network was constructed with 94 nodes (containing 25 up- and 69 down-regulated genes) and 220 edges (Figure 2). Genes with a high connectivity degree, such as ribosomal protein

L9 (*RPL9*, degree = 22), ribosomal protein L35 (*RPL35*, degree = 20), ribosomal protein S7 (*RPS7*, degree = 19), and ribosomal protein L23 (*RPL23*, degree = 18), were selected as hub nodes in this network. All of the hub node genes were down-regulated.

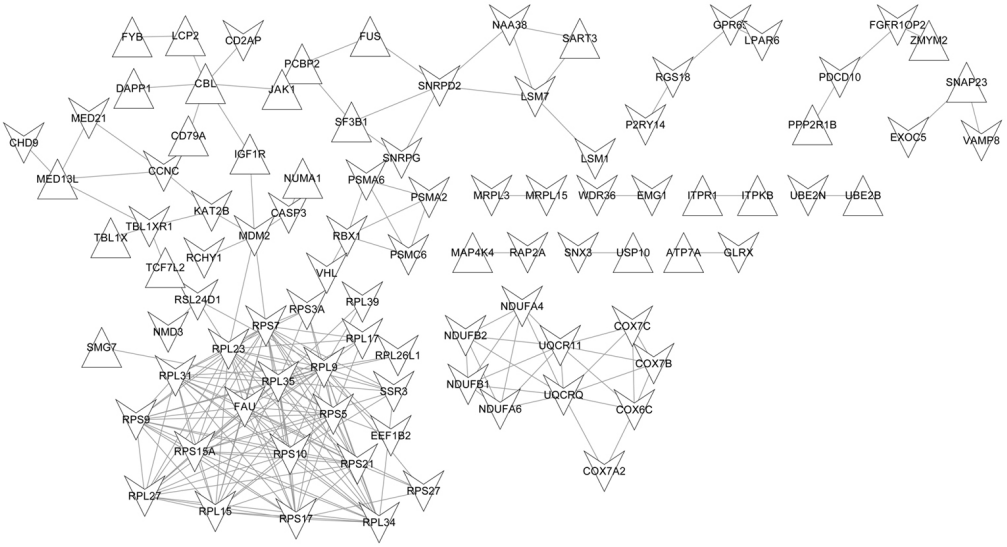

Figure 2. Protein-protein interaction network of differentially expressed genes. Triangles represent up-regulated genes and inverted triangles represent down-regulated genes.

Module analysis

Only one module was selected in the PPI network (Figure 3). The module network was constructed with 15 nodes. All DEGs in this module were ribosomal protein genes with high connectivity degrees.

The results of GO functional annotation of the DEGs in this module are summarized in Table 1. The most significantly enriched GO BP term was translational elongation (P value = 8.03E-34). Other significant GO BP terms included translation and cellular macromolecule biosynthetic process.

KEGG pathway analysis

The KEGG pathway analysis yielded 2 pathways (Table 2). The DEGs were significantly enriched in the ribosomal (P value = 4.89E-21) and oxidative phosphorylation (P value = 2.96E-05) pathways. Twenty-one DEGs, including *RPL9*, *RPL35*, *RPS7*, and *RPL23*, were enriched in the ribosomal pathway, all of which were ribosomal protein-related genes. Some of the genes included in oxidative phosphorylation included cytochrome-c oxidase genes, such as cytochrome c oxidase (COX) subunit VIIa polypeptide 2 (*COX7A2*), COX subunit VIIb (*COX7B*), COX subunit VIIc (*COX7C*), and COX subunit VIc (*COX6C*).

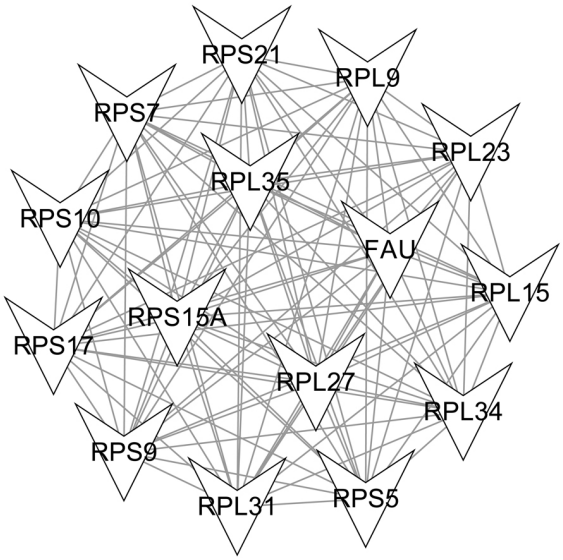

Figure 3. Module selected from protein-protein interaction network. Inverted triangles represent down-regulated genes.

Table 1. Gene Ontology (GO) analysis of differentially expressed genes involved in the module.

Category	Description	P value	FDR
BP	Translational elongation	8.03E-34	6.82E-32
BP	Translation	2.66E-26	1.13E-24
BP	Cellular macromolecule biosynthetic process	6.70E-18	1.90E-16
BP	Macromolecule biosynthetic process	9.08E-18	1.93E-16
BP	Gene expression	1.38E-16	2.35E-15
BP	Cellular biosynthetic process	1.10E-14	1.56E-13
BP	Biosynthetic process	2.86E-14	3.48E-13
BP	Cellular protein metabolic process	4.36E-13	4.63E-12
BP	Protein metabolic process	7.96E-12	7.52E-11
BP	Cellular macromolecule metabolic process	6.77E-10	5.75E-09
BP	Macromolecule metabolic process	5.19E-09	4.01E-08
BP	Cellular metabolic process	1.35E-07	9.58E-07
BP	Primary metabolic process	3.23E-07	2.11E-06
BP	Metabolic process	1.95E-06	1.18E-05

BP = biological process; FDR = false discovery rate.

Table 2. Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway enrichment analysis of differentially expressed genes (DEGs).

Term	Count	P value	FDR
hsa03010:Ribosome	21	4.89E-21	4.90E-18
hsa00190:Oxidative phosphorylation	10	2.96E-05	0.029692

Count = enriched gene number in the KEGG term; FDR = false discovery rate.

DISCUSSION

Atherosclerosis is a major cause of death worldwide; FH is an important risk factor for atherosclerosis (Cheung and Lam, 2010). Understanding the changes in FH gene expression is of critical importance towards understanding the mechanism of atherosclerosis procession. In this

study, we analyzed the gene expression profiles of blood cells in FH patients, compared to the controls; moreover, we attempted to predict the biomarkers of atherosclerosis development.

Three hundred and ninety five DEGs, including 125 up- and 269 down-regulated genes, were selected in this study. PPI network analysis identified several ribosomal proteins, such as RPL9, RPL35, and RPS7, as hub nodes. The KEGG pathway analysis revealed that DEGs were significantly enriched in ribosome and oxidative phosphorylation pathways. Ribosomal protein-related genes and cytochrome-c oxidase genes (*COX7A2*, *COX7B*, *COX7C*, and *COX6C*) were enriched in these pathways. Additionally, module analysis also identified ribosomal proteins as important proteins affecting FH. These significant DEGs and their functions were theorized to contribute to atherosclerosis development in FH patients.

Ribosome is a large and complex organelle found in all living cells. We observed that the ribosomal pathway was significantly enriched by DEG function. Ribosomal protein-related genes, such as *RPL9*, *RPL35*, and *RPS7*, were identified in this pathway. Ribosomal proteins are reported to be associated with important biological processes, including protein synthesis (Ruvinsky and Meyuhas, 2006), cell proliferation (Volarević et al., 2000), and cell apoptosis (Khanna et al., 2003). Dzau et al. (2002) reported that vascular cell proliferation contributed to the pathobiology of atherosclerosis. *RPL17*, a member of the ribosomal protein family, is an inhibitor of vascular smooth muscle cell growth, and therefore is a potential therapeutic target used to limit the thickening of the carotid intima-media (Smolock et al., 2012). Limiting the thickening of the carotid intima-media could reduce the risk of atherosclerosis (Lorenz et al., 2006). Furthermore, ribosomal protein S6 kinase (RPS6K) is the key mammalian target of the rapamycin (mTOR) effector; the mTOR pathway is a key regulator of cell growth through the regulation of protein synthesis (Jastrzebski et al., 2007). Mueller et al. (2008) reported that the mTOR inhibitor strongly inhibited the development of atherosclerosis in $LDLR^{-/-}$ mice. In this study, ribosomal protein-related genes were down-regulated in the blood cells of FH patients; however, these genes functioned as the hub nodes in the PPI network. This suggested that the abnormal expression of ribosomal protein-related genes could increase the risk of atherosclerosis development through the ribosomal pathway.

The oxidative phosphorylation pathway was found to be another important function enriched by DEGs. Oxidative phosphorylation is a metabolic pathway wherein energy is released from adenosine triphosphate (ATP) molecules in mitochondrial cells (Hatefi, 1985). The accumulation of reactive oxygen species causes an increase in oxidative stress in the cells (Scherz-Shouval and Elazar, 2007. Oxidative stress has been demonstrated to play an important role in the development of vascular diseases such as hypercholesterolemia and atherosclerosis (Magenta et al., 2013). In this study, cytochrome-c oxidase genes (*COX7A2*, *COX7B*, *COX7C*, and *COX6C*) were identified to perform this function. Cytochrome-c oxidase has been reported to regulate oxidative phosphorylation in eukaryotic enzymes (Ludwig et al., 2001). Dutta et al. (2006) discovered *COX7B* to be a mitochondrial electron transport chain gene whose expression was down-regulated in multiple sclerosis patients, leading to reduced ATP production and mitochondrial dysfunction. Mitochondrial dysfunction contributes to the development of cardiovascular diseases (including atherosclerosis) by inducing changes in the mitochondrial morphology and apoptosis (Williamson et al., 2010). Therefore, we speculated that cytochrome-c oxidases play a key role in the development of atherosclerosis. In this study, the expression of cytochrome-c oxidase genes was down-regulated in blood samples obtained from FH patients, suggesting that the aberrant expression of these genes may contribute to the development of atherosclerosis by regulating the oxidative phosphorylation pathway.

In summary, the results of our study indicate that the ribosomal and oxidative phosphorylation pathways may be closely associated with the development of atherosclerosis. Ribosomal protein genes and cytochrome-c oxidase genes may also function as potential markers for the development of atherosclerosis. However, further experiments with larger sample sizes are required to confirm our findings.

Conflicts of interest

The authors declare no conflict of interest.

REFERENCES

Altermann E and Klaenhammer TR (2005). PathwayVoyager: pathway mapping using the Kyoto Encyclopedia of Genes and Genomes (KEGG) database. *BMC Genomics* 6: 60. http://dx.doi.org/10.1186/1471-2164-6-60

Ashburner M, Ball CA, Blake JA, Botstein D, et al.; The Gene Ontology Consortium (2000). Gene ontology: tool for the unification of biology. *Nat. Genet.* 25: 25-29. http://dx.doi.org/10.1038/75556

Bader GD and Hogue CW (2003). An automated method for finding molecular complexes in large protein interaction networks. *BMC Bioinformatics* 4: 2. http://dx.doi.org/10.1186/1471-2105-4-2

Cheung BM and Lam KS (2010). Is intensive LDL-cholesterol lowering beneficial and safe? *Lancet* 376: 1622-1624. http://dx.doi.org/10.1016/S0140-6736(10)61545-0

Dutta R, McDonough J, Yin X, Peterson J, et al. (2006). Mitochondrial dysfunction as a cause of axonal degeneration in multiple sclerosis patients. *Ann. Neurol.* 59: 478-489. http://dx.doi.org/10.1002/ana.20736

Dzau VJ, Braun-Dullaeus RC and Sedding DG (2002). Vascular proliferation and atherosclerosis: new perspectives and therapeutic strategies. *Nat. Med.* 8: 1249-1256. http://dx.doi.org/10.1038/nm1102-1249

Hatefi Y (1985). The mitochondrial electron transport and oxidative phosphorylation system. *Annu. Rev. Biochem.* 54: 1015-1069. http://dx.doi.org/10.1146/annurev.bi.54.070185.005055

Hobbs HH, Brown MS, Russell DW, Davignon J, et al. (1987). Deletion in the gene for the low-density-lipoprotein receptor in a majority of French Canadians with familial hypercholesterolemia. *N. Engl. J. Med.* 317: 734-737. http://dx.doi.org/10.1056/NEJM198709173171204

Huang W, Sherman BT and Lempicki RA (2009). Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nat. Protoc.* 4: 44-57. http://dx.doi.org/10.1038/nprot.2008.211

Ishibashi S, Brown MS, Goldstein JL, Gerard RD, et al. (1993). Hypercholesterolemia in low density lipoprotein receptor knockout mice and its reversal by adenovirus-mediated gene delivery. *J. Clin. Invest.* 92: 883-893. http://dx.doi.org/10.1172/JCI116663

Jastrzebski K, Hannan KM, Tchoubrieva EB, Hannan RD, et al. (2007). Coordinate regulation of ribosome biogenesis and function by the ribosomal protein S6 kinase, a key mediator of mTOR function. *Growth Factors* 25: 209-226. http://dx.doi.org/10.1080/08977190701779101

Khanna N, Sen S, Sharma H and Singh N (2003). S29 ribosomal protein induces apoptosis in H520 cells and sensitizes them to chemotherapy. *Biochem. Biophys. Res. Commun.* 304: 26-35. http://dx.doi.org/10.1016/S0006-291X(03)00532-1

Lorenz MW, von Kegler S, Steinmetz H, Markus HS, et al. (2006). Carotid intima-media thickening indicates a higher vascular risk across a wide age range: prospective data from the Carotid Atherosclerosis Progression Study (CAPS). *Stroke* 37: 87-92. http://dx.doi.org/10.1161/01.STR.0000196964.24024.ea

Ludwig B, Bender E, Arnold S, Hüttemann M, et al. (2001). Cytochrome C oxidase and the regulation of oxidative phosphorylation. *ChemBioChem* 2: 392-403. http://dx.doi.org/10.1002/1439-7633(20010601)2:6<392::AID-CBIC392>3.0.CO;2-N

Maere S, Heymans K and Kuiper M (2005). BiNGO: a Cytoscape plugin to assess overrepresentation of gene ontology categories in biological networks. *Bioinformatics* 21: 3448-3449. http://dx.doi.org/10.1093/bioinformatics/bti551

Magenta A, Greco S, Gaetano C and Martelli F (2013). Oxidative stress and microRNAs in vascular diseases. *Int. J. Mol. Sci.* 14: 17319-17346. http://dx.doi.org/10.3390/ijms140917319

Moghadasian MH (2002). Experimental atherosclerosis: a historical overview. *Life Sci.* 70: 855-865. http://dx.doi.org/10.1016/S0024-3205(01)01479-5

Mueller MA, Beutner F, Teupser D, Ceglarek U, et al. (2008). Prevention of atherosclerosis by the mTOR inhibitor everolimus in LDLR-/- mice despite severe hypercholesterolemia. *Atherosclerosis* 198: 39-48. http://dx.doi.org/10.1016/j.atherosclerosis.2007.09.019

Režen T, Cvikl A, Brecelj N, Rozman D, et al. (2008). Atherosclerotic markers in human blood - a study in patients with familial hypercholesterolemia. Available at [http://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE13985].

Ruvinsky I and Meyuhas O (2006). Ribosomal protein S6 phosphorylation: from protein synthesis to cell size. *Trends Biochem. Sci.* 31: 342-348. http://dx.doi.org/10.1016/j.tibs.2006.04.003

Scherz-Shouval R and Elazar Z (2007). ROS, mitochondria and the regulation of autophagy. *Trends Cell Biol.* 17: 422-427. http://dx.doi.org/10.1016/j.tcb.2007.07.009

Smilde TJ, van Wissen S, Wollersheim H, Trip MD, et al. (2001). Effect of aggressive versus conventional lipid lowering on atherosclerosis progression in familial hypercholesterolaemia (ASAP): a prospective, randomised, double-blind trial. *Lancet* 357: 577-581. http://dx.doi.org/10.1016/S0140-6736(00)04053-8

Smolock EM, Korshunov VA, Glazko G, Qiu X, et al. (2012). Ribosomal protein L17, RpL17, is an inhibitor of vascular smooth muscle growth and carotid intima formation. *Circulation* 126: 2418-2427. http://dx.doi.org/10.1161/CIRCULATIONAHA.112.125971

Smoot ME, Ono K, Ruscheinski J, Wang PL, et al. (2011). Cytoscape 2.8: new features for data integration and network visualization. *Bioinformatics* 27: 431-432. http://dx.doi.org/10.1093/bioinformatics/btq675

Smyth GK (2004). Linear models and empirical bayes methods for assessing differential expression in microarray experiments. *Stat. Appl. Genet. Mol.* 3.

Szklarczyk D, Franceschini A, Kuhn M, Simonovic M, et al. (2011). The STRING database in 2011: functional interaction networks of proteins, globally integrated and scored. *Nucleic Acids Res.* 39: D561-D568. http://dx.doi.org/10.1093/nar/gkq973

Team RC (2012). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria.

Troyanskaya O, Cantor M, Sherlock G, Brown P, et al. (2001). Missing value estimation methods for DNA microarrays. *Bioinformatics* 17: 520-525. http://dx.doi.org/10.1093/bioinformatics/17.6.520

Volarević S, Stewart MJ, Ledermann B, Zilberman F, et al. (2000). Proliferation, but not growth, blocked by conditional deletion of 40S ribosomal protein S6. *Science* 288: 2045-2047. http://dx.doi.org/10.1126/science.288.5473.2045

Williamson CL, Dabkowski ER, Baseler WA, Croston TL, et al. (2010). Enhanced apoptotic propensity in diabetic cardiac mitochondria: influence of subcellular spatial location. *Am. J. Physiol. Heart Circ. Physiol.* 298: H633-H642. http://dx.doi.org/10.1152/ajpheart.00668.2009

Zhao L, Cuff CA, Moss E, Wille U, et al. (2002). Selective interleukin-12 synthesis defect in 12/15-lipoxygenase-deficient macrophages associated with reduced atherosclerosis in a mouse model of familial hypercholesterolemia. *J. Biol. Chem.* 277: 35350-35356. http://dx.doi.org/10.1074/jbc.M205738200