

Identification of Differential Expressed Genes and Key Pathways in Monocytes and CD3+ T-Cells of Familial Hypercholesterolemia (FH) Patients

Zuhier Awan^{1,2*}

¹Department of Clinical Biochemistry, Faculty of Medicine, King Abdulaziz University, Jeddah, Saudi Arabia.

²Department of Genetics, Al Borg Diagnostics, Jeddah, Saudi Arabia

*Email: bioinfo.pacer@gmail.com, zawan@kau.edu.sa

ABSTRACT

Familial hypercholesterolemia (FH) is an autosomal dominant disorder, where molecular defects in PCSK9, APOE, and LDLR genes cause the elevation of serum LDL cholesterol levels. This condition results in an excess deposition of cholesterol in vascular tissues that hastens the process of atherosclerosis and elevates the risk of developing premature cardiovascular disease (CVD). The exact molecular mechanisms through which genetic defects contribute to alterations in lipid content and lipoprotein metabolism remain an under-researched area to date. Therefore, this study has sought to integrate the microarray gene expression data of two different cells (human circulating monocytes and CD3 positive T cells) (E-GEOD-6054 and E-GEOD-6088) of FH (with familial hypercholesterolemia, homozygous or heterozygous mutant for LDLR) in order to identify potential candidate genes and pathways associated with FH. We analyzed gene expression data in this study using the R software program using the Affy and Limma packages. In the human circulating monocytes expression dataset, a total of 957 and 618 genes were identified, which were expressed as LDLR heterozygous and homozygous, respectively, in FH patients, compared to the experimental controls. In the CD3+ T-cells, 431 and 410 genes were expressed as LDLR homozygous and heterozygous in FH patients compared to the experimental controls. In summary, by using integrated bioinformatics analysis, we were able to discern potential pathways and genes that play an important role in the development of FH, and could therefore be fruitful for future research efforts in this domain.

Keywords:-Familial hypercholesterolemia; Gene expression; PPI; CVD

INTRODUCTION

Familial hypercholesterolemia (FH) is the inherited form of hypercholesterolemia and is characterized by high serum LDL cholesterol levels. This condition results in an excess deposition of cholesterol in tissues that hastens the process of atherosclerosis and heightens the risk of premature cardio vascular disease (CVD)(Smilde, van Wissen, Wollersheim, Kastelein, & Stalenhoef, 2001); a leading cause of death in many developed countries as well as in Saudi Arabia (Sharifi, Futema, Nair, & Humphries, 2017). It is widely known that a defective LDL

clearance is the underlying pathogenic cause of FH. Three different pathogenic variants are responsible for the dominant inheritance of FH – namely the LDLR gene, apolipoprotein B (ApoB) and pro-protein convertase subtilisin/kexin type 9 (PCSK9). Mutations in ApoE, STAP1, lysosomal acid lipase (LIPA), ABCG5 or ABCG8 can also cause autosomal dominant FH, though this is uncommon.

Additionally, the presence of an extremely rare autosomal recessive type of FH can be due to a genetic alteration in the LDLR adaptor protein 1 (LDLRAP1) gene (3). The estimated risk of premature CVD is 20-fold higher for patients with FH, compared to the general population. Furthermore, those with FH are 3.9 times more likely to have CVD events in their lifetimes compared to those with similar risk factors without the presence of FH (Villa et al., 2017). According to epidemiological studies, the prevalence of mutations that cause HeFH is approximately 1/250 to 1/300, although this is not necessarily an indicator that this will become symptomatic. Moreover, prevalence rates of around 1:1,000,000 have been reported for HoFH (4). FH prevalence is yet to be determined in Saudi Arabia, although the indications are that it could be higher than normal due to high rates of consanguinity. Recently, several studies have been conducted to identify other candidate genes associated with FH, which could lead to a better understanding of some of the precise molecular mechanisms behind this disorder. Such gene includes CYP7A, SREBP-2 or SCAP. However, the primary goal of the majority of these studies has been to focus purely on the genes themselves, paying little attention to the link between these genes (Ye et al., 2013). It is hoped this study will shed some light on this area, by identifying novel genes potentially involved in FH along with the significant pathways associated with this disorder. This has been done by using gene expression datasets which aid in providing novel FH biomarkers. Differentially expressed genes (DEGs) in two datasets from two different cells (human circulating monocytes and in CD3 positive T cells) have been performed, followed by Gene Ontology analysis, and building the protein-protein interaction (PPI) network (Awan et al., 2021; Banaganapalli et al., 2020; A. I. Bima et al., 2022; A. I. H. Bima et al., 2022).

METHODOLOGY

Gene datasets:

Two different datasets were obtained from the array express (E-GEOD-6054 and E-GEOD-6088) available at <https://www.ebi.ac.uk/arrayexpress>. (GEOD-6054) comprised the transcription profile of monocytes which were derived from FH patients, including 25 samples (5 homozygous FH, 7 heterozygous FH and 13 control), while (E-GEOD-6088) included the transcription profile of CD3 positive T cells derived from FH patients with 23 T cell samples (3 homozygous-Homo FH, 6 heterozygous-Hetero FH, 14 control).

R Normalization:

To ensure that different arrays were compared and measured accurately within the study, background correction was conducted, and involved removing local artifacts that can affect neighbouring measurements and the normalizing of data.

Data Preprocessing and Differentially Expressed Genes (DEGs) Analysis:

DEGs analysis in this study was conducted with the R software program using Affy and Limma packages. The individual analysis of each dataset was carried out using the Benjamini–Hochberg's False Discovery Rate (FDR) < 0.01 with cut-off p-values of < 0.5. A heatmap of DEGs was created using the visual inspection tools from the Network Analyst (<http://www.heatmapper.ca>) and were clustered using the single linkage method.

Gene ontology and signaling pathway enrichment analysis of DEGs:

Gene ontology and enrichment of the DEGs including the molecular function, biological process and cellular component were conducted in Panther (<http://www.pantherdb.org/>). In addition, the KEGG pathway databases were used to perform the signaling pathway.

Construction of the PPI network and module analysis:

The PPI network was performed on the string database (<http://www.string-db.org/>) (Irizarry et al., 2003). Cytoscape was also used to visualize and build up the PPI network (Shannon et al., 2003) and CFinder (<http://www.cfinder.org>) was used for cluster analysis (Adamcsek, Palla, Farkas, Derenyi, & Vicsek, 2006; Shannon et al., 2003) to obtain significant genes, which were used for enrichment analysis.

RESULTS

Identification of DEGs:

Two datasets obtained from the array express (GEOD-6054) and (E-GEOD-6088) were screened for DEGs in homozygous and heterozygous conditions. For the monocyte dataset, 618 DEGs were identified in a homozygous condition and 957 were heterozygous. In the CD3+ T Cell dataset, 431 DEGs were identified as homozygous, while 410 DEGs were heterozygous. The Volcano plots along with scatter plots of DEGs are shown in figure 1A.

Identification of biological pathways and gene ontology:

Gene ontology analysis including molecular function, biological process and cellular component for the DEGs was performed using panther and string for KEGG pathways. Among these, GO annotations were transferase activity, transferring phosphorus-containing groups (GO:0016772), kinase binding (GO:0019900), protein domain specific binding (GO:0019904), pathways in cancer and protein serine/threonine kinase activity (GO:0004674). These were recorded as the top five involved in the FH monocyte profile. In contrast, complement activation classical pathway (GO:0006958), antigen binding (GO:0003823), B cell receptor signaling pathway (GO:0050853),

MHC class II protein complex (GO:0042613) and the Tuberculosis pathway were the top five involved in FH CD3 positive T Cell samples (Table 1).

Analysis of PPI network and modules:

The DEGs PPI network complex for the CD3 positive homo condition was constructed and contained 323 nodes and 953 edges (see figure 6). The top genes' centrality is shown in table 3. Using CFinder with a k-cliques value of > 10 , one module was extracted from the PPI network. Pathway enrichment analysis illustrated that the aforementioned module consisted of 14 nodes and 62 edges which are mainly linked to a necroptotic signaling pathway and apoptosis. While in the heterozygous condition, 341 nodes and 2091 edges were identified. Using CFinder again with a k-cliques value of > 10 , two modules (1 and 2) were extracted from the constructed PPI network. Pathway enrichment analysis showed that Module 1 consisted of 18 nodes and 136 edges which are mainly associated with a positive regulation of the hippocampal neuron apoptotic process and a positive regulation of the microglial cell mediated cytotoxicity. Module 2 consisted of 10 nodes and 45 edges which are mainly associated with the sphingosine-1-phosphate receptor signaling pathway, the sphingolipid mediated signaling pathway and the neuroactive ligand-receptor interaction (Figure 1B).

However, the DEGs PPI network complex for the monocyte homozygous condition revealed 497 nodes and 1606 edges. The top genes' centrality is shown in table 4. Using CFinder with a k-cliques value of > 10 , one module was constructed for the monocyte homozygous condition containing 10 nodes and 45 edges which are mainly associated with CCR5 chemokine receptor binding, CXCR chemokine receptor binding and the toll-like receptor signaling pathway.

Meanwhile, the monocyte heterozygous condition was constructed, with 795 nodes and 3661 edges, using CFinder with a k-cliques value of > 10 for two modules. Module 1 contained 22 nodes and 230 edges and was linked to protein heterodimerization activity, protein dimerization activity and DNA binding. Module 2, which contained 28 nodes and 316 edges, was associated with nucleosomal DNA binding, nucleosome binding and DNA replication-dependent nucleosome assembly (Figure 1C).

Table 1: Gene ontology and KEGG pathway analysis for module 1 in cluster finder for CD3+T cell (homozygous)

Description	No. genes	(fold Enrichment)	(raw P-value)	(FDR)
Molecular function				
death effector domain binding (GO:0035877)	3	> 100	1.27E-08	2.94E-05
cysteine-type endopeptidase activity involved in apoptotic signaling pathway (GO:0097199)	2	> 100	2.44E-05	1.62E-02
death receptor binding (GO:0005123)	4	> 100	1.63E-09	7.57E-06
cysteine-type endopeptidase activity involved in apoptotic process (GO:0097153)	3	> 100	2.94E-07	2.73E-04
tumor necrosis factor receptor binding (GO:0005164)	3	> 100	2.34E-06	1.81E-03
tumor necrosis factor receptor superfamily binding (GO:0032813)	4	> 100	5.48E-08	6.38E-05
Biological Process				
necroptotic signaling pathway (GO:0097527)	3	> 100	3.03E-08	4.00E-05
TRAIL-activated apoptotic signaling pathway (GO:0036462)	2	> 100	1.14E-05	2.37E-03
death-inducing signaling complex assembly (GO:0071550)	2	> 100	2.44E-05	4.54E-03
toll-like receptor 3 signaling pathway (GO:0034138)	2	> 100	3.57E-05	6.07E-03
positive regulation of cysteine-type endopeptidase activity involved in apoptotic signaling pathway (GO:2001269)	2	> 100	4.22E-05	6.95E-03
regulation of necroptotic process (GO:0060544)	3	> 100	4.10E-07	2.09E-04
Cellular component				
CD95 death-inducing signaling complex (GO:0031265)	5	> 100	5.90E-14	1.20E-10
riposome (GO:0097342)	4	> 100	4.69E-11	3.19E-08
death-inducing signaling complex (GO:0031264)	5	> 100	2.55E-13	2.60E-10
membrane raft (GO:0045121)	5	20.76	3.07E-06	1.25E-03
membrane microdomain (GO:0098857)	5	20.7	3.12E-06	1.06E-03
cytosolic part (GO:0044445)	4	20.42	3.76E-05	6.97E-03
KEGG				
Description	No. genes	FDR	Matching genes	
Apoptosis	8	1.31E-	CASP10,CASP8,CFLAR,FADD,FAS,FASLG,GZMB,PR	

		12	F1
Autoimmune thyroid disease	6	3.72E-11	CTLA4,FAS,FASLG,GZMB,IL4,PRF1
Chagas disease (American trypanosomiasis)	6	1.23E-09	CASP8,CFLAR,CXCL8,FADD,FAS,FASLG
Allograft rejection	5	1.23E-09	FAS,FASLG,GZMB,IL4,PRF1

DISCUSSION

FH is characterized by an elevated serum level of LDL-C which enables atherosclerosis development. Atherosclerosis is regarded as one of the primary risk factors for coronary artery diseases, heart attacks and strokes (Collaboration et al., 2016). An accumulation of LDL particles results in the deposition of fibrous plaques in the subendothelial space and a narrowing in the diameter of the arteries, which can lead to heart ischemia and myocardial infarction (Heusch et al., 2014). In this study, we applied integrated bioinformatics approach and tools to identify potential pathways and genes, which might play an important role in the development of FH. Using R Affy and Limma packages, we found 618 DEGs in a homozygous condition and 957 in a heterozygous condition in the monocyte dataset. In the CD3+ T cell dataset, 431 DEGs were identified as homozygous, while 410 DEGs were identified in the heterozygous form. Gene ontology and KEGG pathway were also identified for each dataset. Module analysis revealed that more significant enrichments of DEGs in the CD3 positive T Cells was linked to biological processes such as the necroptotic signaling pathway, apoptotic process, sphingolipid mediated signaling pathway and cell mediated cytotoxicity. The KEGG pathway enrichment identified apoptosis as an important pathway with those matched genes being involved (FAS, FASLG, GZMB, and PRF1) (see tables 5 and 6). Down regulation of this gene due to high cholesterol levels can trigger the infiltration of inflammatory cells into the vessel wall and heighten the risk of CVD (Adamcsek et al., 2006). The GZMB gene encodes a member of the granzyme subfamily of proteins which induces target cell apoptosis. Increased expression of this protein product, granzyme B, has been associated with atherosclerosis. Furthermore, it was noted that changes in sphingolipid plasma concentrations can contribute to the development of cardiac events. Many studies have reported that high levels of sphingomyelin are associated with increased atherosclerotic lesions, which can lead to premature cardiovascular problems – one of the major complications of FH (Iqbal, Walsh, Hammad, & Hussain, 2017).

Furthermore, based on the gene ontology analysis and module analysis, we observed that significant enrichments of DEGs in monocyte dataset were associated with molecular function, specifically CCR5, CXCR chemokine receptor binding and protein dimerization activity. The KEGG pathway was enriched in the toll-like receptor signaling pathway and systemic lupus erythematosus with six genes involved in the first pathway

(*CCL3, CXCL10, CXCL8, STAT1, TLR7, TNF*) and seven genes involved in the second KEGG pathway (*H3F3A, HIST1H2AD, HIST1H2BE, HIST1H2BH, HIST1H2BJ, HIST1H2BO, HIST1H3B*) (see tables 7 and 8). According to different studies which focus on the contribution of chemokines and their receptors in the pathology of atherosclerosis and related cardiovascular disease, it was noted that *CCR5* in particular plays an important role in the development of atherosclerosis along with its ligands. *CCR5* deletion polymorphism and *CCR5delta32* have been associated with a reduced risk of CVD (Hyde et al., 2010) and both the *CCR5* antagonism and gene deletion were shown to reduce atherosclerosis in mouse models of the disease. Up regulation of candidate genes in the *CCR5* chemokine receptor binding pathway in FH patients might accelerate the appearance of the disease's complication (Jones, Maguire, & Davenport, 2011). Furthermore, recent studies have established a link between innate immunity and the development of atherosclerosis (Falck-Hansen, Kassiteridi, & Monaco, 2013). The innate immune system launches the recruitment of monocytes from the blood and enhances their differentiation to macrophages in the vessel wall (Perek et al., 2018). The adaptive immune response promotes and regulates inflammation in mature lesions (Li & Ley, 2015). The activation of toll-like receptors are regarded as being the orchestrators of the progression of inflammation, which is a characteristic of atherosclerotic lesions. The *TNF* gene encodes multifunctional proinflammatory cytokine which plays a role in the regulation of important biological processes like apoptosis and lipid metabolism. Untreated FH patients exhibit elevated levels of pro-inflammatory cytokines when compared to healthy controls (El Messal et al., 2006). Based on the findings of our study, we suggest that the up regulation of the genes (*CCL3, CXCL10, CXCL8, STAT1, TLR7, TNF*) and over expression of their receptors might have implications for adverse plasma lipid concentrations, hardening of the arteries and development of CVD. The KEGG analysis in this study has highlighted the link between Systemic Lupus Erythematosus (SLE) and hypercholesterolemia, which supports previous findings that untreated patients with SLE are at a higher risk of developing CVD when compared to the general population due to dyslipidemia (Aranow & Ginzler, 2000; Yuan, Li, Wang, Song, & Zhang, 2016).

CONCLUSION

This study has illustrated that FH patients showed differences in lipoprotein and cholesterol metabolism in circulating monocytes and lymphocytes compared to healthy controls. Based on gene enrichment analysis and PPI networks, DEGs of FH in both datasets were found to be closely linked with immune responses, hormone responses, apoptosis and chemokines receptors. In conclusion, these DEGs may be used as specific therapeutic molecular targets in the treatment of FH. The findings from this study could prove useful toward increasing the understanding in the pathogenesis of FH in future studies. However, for this and similar studies to gain greater credibility in this field, further work and research must be carried out.

REFERENCES

1. Adamcsek, B., Palla, G., Farkas, I. J., Derenyi, I., & Vicsek, T. (2006). CFinder: locating cliques and overlapping modules in biological networks. *Bioinformatics*, 22(8), 1021-1023. doi:10.1093/bioinformatics/btl039
2. Aranow, C., & Ginzler, E. M. (2000). Epidemiology of cardiovascular disease in systemic lupus erythematosus. *Lupus*, 9(3), 166-169. doi:10.1191/096120300678828208
3. Awan, Z. A., Rashidi, O. M., Al-Shehri, B. A., Jamil, K., Elango, R., Al-Aama, J. Y., . . . Shaik, N. A. (2021). Saudi Familial Hypercholesterolemia Patients With Rare LDLR Stop Gain Variant Showed Variable Clinical Phenotype and Resistance to Multiple Drug Regimen. *Front Med (Lausanne)*, 8, 694668. doi:10.3389/fmed.2021.694668
4. Banaganapalli, B., Mansour, H., Mohammed, A., Alharthi, A. M., Aljuaid, N. M., Nasser, K. K., . . . Shaik, N. A. (2020). Exploring celiac disease candidate pathways by global gene expression profiling and gene network cluster analysis. *Scientific Reports*, 10(1), 16290. doi:10.1038/s41598-020-73288-6
5. Bima, A. I., Elsamanoudy, A. Z., Alamri, A. S., Felimban, R., Felemban, M., Alghamdi, K. S., . . . Banaganapalli, B. (2022). Integrative global co-expression analysis identifies Key MicroRNA-target gene networks as key blood biomarkers for obesity. *Minerva Medica*. doi:10.23736/S0026-4806.21.07478-4
6. Bima, A. I. H., Elsamanoudy, A. Z., Albaqami, W. F., Khan, Z., Parambath, S. V., Al-Rayes, N., . . . Shaik, N. A. (2022). Integrative system biology and mathematical modeling of genetic networks identifies shared biomarkers for obesity and diabetes. *Math Biosci Eng*, 19(3), 2310-2329. doi:10.3934/mbe.2022107
7. Collaboration, E. A. S. F. H. S., Vallejo-Vaz, A. J., Akram, A., Kondapally Seshasai, S. R., Cole, D., Watts, G. F., . . . Ray, K. K. (2016). Pooling and expanding registries of familial hypercholesterolaemia to assess gaps in care and improve disease management and outcomes: Rationale and design of the global EAS Familial Hypercholesterolaemia Studies Collaboration. *Atheroscler Suppl*, 22, 1-32. doi:10.1016/j.atherosclerosissup.2016.10.001
8. El Messal, M., Beaudoux, J. L., Drissi, A., Giral, P., Chater, R., Bruckert, E., . . . Chapman, M. J. (2006). Elevated serum levels of proinflammatory cytokines and biomarkers of matrix remodeling in never-treated patients with familial hypercholesterolemia. *Clin Chim Acta*, 366(1-2), 185-189. doi:10.1016/j.cca.2005.09.027
9. Falck-Hansen, M., Kassiteridi, C., & Monaco, C. (2013). Toll-like receptors in atherosclerosis. *Int J Mol Sci*, 14(7), 14008-14023. doi:10.3390/ijms140714008
10. Heusch, G., Libby, P., Gersh, B., Yellon, D., Bohm, M., Lopaschuk, G., & Opie, L. (2014). Cardiovascular remodelling in coronary artery disease and heart failure. *Lancet*, 383(9932), 1933-1943. doi:10.1016/S0140-6736(14)60107-0
11. Hyde, C. L., Macinnes, A., Sanders, F. A., Thompson, J. F., Mazzarella, R. A., Faergeman, O., . . . Paciga, S. A. (2010). Genetic association of the CCR5 region with lipid levels in at-risk cardiovascular patients. *Circ Cardiovasc Genet*, 3(2), 162-168. doi:10.1161/CIRCGENETICS.109.897793

12. Iqbal, J., Walsh, M. T., Hammad, S. M., & Hussain, M. M. (2017). Sphingolipids and Lipoproteins in Health and Metabolic Disorders. *Trends Endocrinol Metab*, 28(7), 506-518. doi:10.1016/j.tem.2017.03.005
13. Irizarry, R. A., Hobbs, B., Collin, F., Beazer-Barclay, Y. D., Antonellis, K. J., Scherf, U., & Speed, T. P. (2003). Exploration, normalization, and summaries of high density oligonucleotide array probe level data. *Biostatistics*, 4(2), 249-264. doi:10.1093/biostatistics/4.2.249
14. Jones, K. L., Maguire, J. J., & Davenport, A. P. (2011). Chemokine receptor CCR5: from AIDS to atherosclerosis. *Br J Pharmacol*, 162(7), 1453-1469. doi:10.1111/j.1476-5381.2010.01147.x
15. Li, J., & Ley, K. (2015). Lymphocyte migration into atherosclerotic plaque. *Arteriosclerosis, thrombosis, and vascular biology*, 35(1), 40-49. doi:10.1161/ATVBAHA.114.303227
16. Perek, B., Kowalska, K., Kempisty, B., Nowicki, A., Jankowski, M., Nawrocki, M. J., & Malinska, A. (2018). Role of macrophages in the pathogenesis of atherosclerosis and aortocoronary graft disease. *J Biol Regul Homeost Agents*, 32(5), 1055-1059.
17. Shannon, P., Markiel, A., Ozier, O., Baliga, N. S., Wang, J. T., Ramage, D., . . . Ideker, T. (2003). Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Research*, 13(11), 2498-2504. doi:10.1101/gr.1239303
18. Sharifi, M., Futema, M., Nair, D., & Humphries, S. E. (2017). Genetic Architecture of Familial Hypercholesterolaemia. *Curr Cardiol Rep*, 19(5), 44. doi:10.1007/s11886-017-0848-8
19. Smilde, T. J., van Wissen, S., Wollersheim, H., Kastelein, J. J., & Stalenhoef, A. F. (2001). Genetic and metabolic factors predicting risk of cardiovascular disease in familial hypercholesterolemia. *Neth J Med*, 59(4), 184-195.
20. Villa, G., Wong, B., Kutikova, L., Ray, K. K., Mata, P., & Bruckert, E. (2017). Prediction of cardiovascular risk in patients with familial hypercholesterolaemia. *Eur Heart J Qual Care Clin Outcomes*, 3(4), 274-280. doi:10.1093/ehjqcco/qcx011
21. Ye, Y., Li, K., Liu, J., Li, M., Wang, W., Wang, R., . . . Yuan, Z. (2013). Dysfunctional co-expression network analysis of familial hypercholesterolemia. *J Cardiol*, 62(1), 58-62. doi:10.1016/j.jjcc.2013.02.014
22. Yuan, J., Li, L. I., Wang, Z., Song, W., & Zhang, Z. (2016). Dyslipidemia in patients with systemic lupus erythematosus: Association with disease activity and B-type natriuretic peptide levels. *Biomed Rep*, 4(1), 68-72. doi:10.3892/br.2015.544