Whole Exome Sequencing Reveals Multiple Mutations in Uncommon Genes of Familial Hypercholesterolaemia

Mohammad Athar1,2, Faisal Ahmed Al-Allaf1,3, Fahad Alnouri1, Mohiuddin Mohammed Taher1,2, Abdellatif Bouazzaoui1,2, Neda Mustafa Bogari, Wafa Mohammed Elbjeirami1, Hani Saleh Faidah5, Zainularifeen Abduljaleel1,2,*

1Department of Medical Genetics, Faculty of Medicine, Umm Al-Qura University, Makkah, SAUDI ARABIA.
2Science and Technology Unit, Umm Al-Qura University, Makkah, SAUDI ARABIA.
3Department of Laboratory and Blood Bank, Molecular Diagnostics Unit, King Abdullah Medical City, Makkah, SAUDI ARABIA.
4Cardiovascular Prevention and Rehabilitation Unit, Prince Sultan Cardiac Center, Riyadh, SAUDI ARABIA.
5Department of Medical Microbiology, Faculty of Medicine, Umm Al-Qura University, Makkah, SAUDI ARABIA.

ABSTRACT

Whole exome sequencing (WES) is an innovative, precise procedure established on the next generation sequencing and has rapidly become esteeemed in medical genetics. We have used WES to study patient with suspected genetic conditions with no apparent recognizable familial hypercholesterolemia (FH). In the tested proband, multiple causative mutations were identified in LPL, EPHX2, PCSK9, APOB, ABCA1 and ABCG8 genes. The identified mutations in the LPL gene were heterozygous mutation [c.106G>A, p. (Asp36Asn)] in exon 2 and another heterozygous mutation [c.1421C>G, p.(Ser474*)] in exon 9. EPHX2 gene of the proband showed a novel heterozygous mutation [c.1644delC, p.(Pro549fs)] in exon 19. In addition, heterozygous mutations were identified in exon 1 of PCSK9 gene [c.158C>T, p.(Ala53Val)] and exon 4 of APOB [c.293C>T, p.(Thr98Ile)]. Likewise, homozygous mutations were found in exon 35 of ABCA1 gene [c.4760A>G, p.(Lys1587Arg)] and exon 13 of ABCG8 gene [c.1895T>C, p.(Val632Ala)], a common mutation highly associated with hypercholesterolemia based on the p-value by phenolyzer prioritization method. All causative mutations were validated by capillary sequencing method and confirmed the presence of homozygous and heterozygous mutations detected in the proband. All the causative mutations were functionally predicted to be deleterious including a novel frameshift and common missense mutation. Further investigations of the common mutation using Molecular dynamic (MD) simulation with solvated condition confirmed the functional and structural impact of the changes on the modified proteins.

Key words: Familial hypercholesterolemia (FH), Functional variants, Whole Exome Sequencing (WES), Genotypic and phenotypic analysis, Low-density lipoprotein cholesterol (LDL-C).

INTRODUCTION

The genetic run through conventional molecular diagnostic systems has limitations for efficient FH diagnosis due to the wide variety of types and locations of mutations in known genes, as well as due to the existence of undiscovered or potential FH-causing genes. Next-generation sequencing (NGS) is a remarkable technique for identifying genetic alterations in considerably large genomic areas and among novel disease-related genes. Several studies have demonstrated the utility of NGS in the diagnosis of FH. Whole exome sequencing (WES) is a type of NGS, used for sequencing of the whole coding region of the genome (whole exome) to detect variations that could cause a genetic disorder. Familial hypercholesterolemia (FH) is a genetic disorder, characterized by high levels of serum low-density lipoprotein cholesterol (LDL-C) and an increased risk of premature coronary artery disease.

FH is commonly caused by loss of function mutations in LDLR, mutations in APOB or less-frequent gain of function mutations within PCSK9. It has recently been reported that patients with severe, primary hypercholesterolemia with (LDL) cholesterol (≥190 mg/dL) are often assumed to have FH. FH is associated with dysfunctional clearance of LDL-C particles. It is however unknown that how much FH contributes to severe hypercholesterolemia in the general population and its impact on coronary artery disease (CAD) risk beyond LDL-C levels. Khera et al. recently utilized the whole exome sequencing for three causative genes i.e., PCSK9, APOB and LDLR on five prospective cohorts (n=11,908 participants) and seven case-control cohorts (n=8577 without CAD and n=5540 with CAD). Many clinicians believe that individuals with very high levels of LDL-C probably have a FH mutation, which is not true. Only about 2 % of such individuals will have a FH mutation. However, CAD risk is much higher in FH mutation carriers than non-carriers.

In Saudi Arabia, there are inadequate numbers of reports available for prevalence and molecular characteristics of FH. The disease may have a higher prevalence in Saudi Arabia than in other neighboring countries because consanguineous marriages exceeding 55% and due to the lack of national registries and genetic screening programmes for FH, the disease is under diagnosed and under estimated.

In the present study, we performed WES of FH patient, shown to be negative for LDLR mutation. Further this paper also deals with the MD simulation prediction and the fate of deleterious mutations in the target domains. All the possible non-synonymous SNP were analyzed and MD simulations were generated to assess the distortions caused by single-amino acid alteration. The results discussed are the probable pathological outcome of mutations, which will assist an initial diagnosis and FH therapy.

MATERIALS AND METHODS

Patient and Ethics statement

This study was part of an in-depth study focusing on the screening of FH mutations using targeted exome sequencing in the Saudi population. A total of two unrelated subjects (patient and another control with no lipid abnormalities) samples were analyzed using WES. The patient was a 19-year old female, clinically diagnosed as FH using Simon Broome Criteria, the value of total and LDL-C were found to be 6.82 and 4.61
mmol/l respectively, considered high compared with the optimal level [<2.59 mmol/l (LDL-C)] and also had family history of elevated total cholesterol and myocardial infarction. She is taking lipid-lowering drugs Atorvastatin and Ezetimibe. She was initially screened and shown to be negative for mutations in known highly FH associate gene LDLR. The sample collection and study were accomplished in accordance with the Research Ethics Committees regulation at Prince Sultan Cardiac Centre, Riyadh, Saudi Arabia. Informed consent was obtained from all individual participants included in the study.

**Whole exome sequencing**

Whole exome sequencing was performed on Ion proton system (Ion Torrent) using the Ion AmpliSeq Exome RDY Kit, according to the manufacturer's protocol. The DNA libraries and exome enrichment were attained using Ion AmpliSeq Exome RDY panel 1x8, Ion AmpliSeq Library Kit Plus and Ion Xpress bar-code adapters 1–16 kit (Life Technologies). Emulsion PCR based template preparation and high throughput sequencing was done using Ion PI Hi-Q OT2 200 Kit and Ion PI Hi-Q sequencing 200 Kit (Life Technologies), respectively.

**Whole exome sequencing data analysis**

NGS data assessment was accomplished with CLC Genomics Workbench v9, USA (http://www.clcbio.com). The Binary Alignment/Map (BAM) binary format sequence data raw reads was finalized using adapter trimming, i.e. reads shorter than 20 bp and duplicates were removed. The pre-processed reads were aligned with the reference human genome (Hg 19) sequences subsequent to the targeted genes, duplicate fragments were marked using Picard and GATK was then used to recalibrate base qualities. We used the human genome assembly Hg19 (GRCh37) as a reference. The single nucleotide variant (SNV) and indel were identified and all variants observed within the exons of the targeted genes were studied for succeeding evaluation through fixed ployid variant detection procedure.

**Phenotype-disease prioritization method**

Filtering of exonic candidate's variant and prioritization gene network was performed by wANNOVAR (http://phenolyzer.usc.edu). The procedure resulted in some candidate genes for hypercholesterolemia, which is depicted by high cholesterol levels as a result of abnormal levels of low-density lipoprotein (LDL) in the blood along with premature cardiovascular illness.

**Finding coding SNPs prediction**

Single amino acid polymorphism database (SAAP) and dbSNPs were utilized to identify the protein and for the crystal structure analysis of ABCG8 gene and identification of single point mutation residue location. The particular mutation residue damaging effects were confirmed with PolyPhen2 and SIFT programs. Furthermore, deleterious variants correlation for genotypic to phenotypic analysis were in different population by Ensemble.

**Capillary sequencing analysis**

ABI 3500 Genetic analyzer was used for capillary sequencing (Sanger method) using BigDye 3.1® for identified deleterious variants in coding exons. PCR was done using a HotStarTaq Plus DNA Polymerase Kit (Qiagen, Cat. No. 203605) with genomic DNA (100 ng) in a 20μl reaction mixture as follows: at 95°C for 5 min Taq polymerase was activated, followed by denaturing at 95°C for 30s (40 cycles), annealing for 30s at 56°C, extension for 45s at 72°C and final extension for 5 min at 72°C. The multiplied products were resolved on agarose gel to confirm the size and quality of the band. The PCR products were cleaned using magnetic beads technique with the Agencourt AMPure XP kit (Beckman Coulter) and later exploited as templates for sequencing. The sequencing reaction products were cleaned with BigDye X-terminator purification kit (Applied Biosystems) then capillary sequencing was performed using ABI 3500 Genetic analyzer (Applied Biosystems) and the concluding assessment was completed with Sequence Analysis Software v5.4 (Applied Biosystems) US.

**Structural and functional analysis**

Structure analysis was performed using Have your protein defined (HOPE)-23 developed by the Centre for Molecular and Bio-molecular informatics (CMBI), Department of Bioinformatics, Radboud University. The parameters were very important to acquire factual data of actual protein structure resulting in annotated information in UniProtKB utilized by prediction by DAS servers. Moreover, the functional effects were determined by the SNAP. SNAP scores varies from -100 (highly neutral) to 100 (highly alter function); the distance is directly related to the binary determination boundary (0), which measures the reliability of the impact, which explains that the disease associated mutations will possibly have an impact on protein interactions. Protein function will typically be associated with evolutionarily conserved residues and a damaging signal will correspond to a mutation that is predicted as being stabilizing. Alterations in the folding free energy upon mutation (ΔΔG) support the concept that a mutation causes disease primarily because it impairs an important protein. SNAP scores are related to extra functional effects.

**Solvent accessibility of amino acid residues**

The protein stability was determined by Schrodinger (BioLuminate)-Maestro. The structure primarily depends on the mutant stability predictions. In the mutated form, the place of the mutated residue is specified, in addition to the wild type and mutant amino acids. Several known diseases with a known three-dimensional protein structure are associated with important residues missense mutations and sites, which structurally alter the protein and its relevant function. Disease involving mutations generally happen within the protein (buried) and at hydrogen bonding residues. In protein kinases, comparisons appear to cluster within the functionally important catalytic core, followed by residue scanning to fix polar and neutral residues for analysis of the protein stability and solvent accessibility.

**Molecular dynamics simulation**

The molecular dynamics (MD) simulations were utilized through biomolecular simulation application CHARMM-GUI (http://www. Charmm-gui.Org) that is a graphical user interface to analyze input data and molecular programs. Although, the simulation does not have the solvated system either minimized or equilibrated, still 0.15 M ions can also be adjusted in the simulation field through specifying ions (KCl) and concentration (C). Applying short Monte Carlo (MC) simulations with a basic model, for instance Van der Waals interactions, validated the initial arrangement of ions. Initially, the free energy simulations were carried out with few explicit solvent water molecules in adjacent neighborhood to the solute, while the effect of the rest of the solvent mass was once established fully as an effective solvent boundary potential (SSBP). Molecular dynamics simulations were achieved with a 2fs time period at a constant temperature of 300 K and a continuous pressure of 1-atm under periodic solvent boundary conditions.

**RESULTS**

Landscape of identified FH associated genes mutations

To identify genomic DNA mutations of disease-causing genes implicated in familial hypercholesterolemia, we considered a typical high throughput
density whole exome sequencing capture array applied to all exome at 100x coverage. The results identified mutations in disease-causing LPL, EPHX2, PCSK9, APOB, ABCA1 and ABCG8 genes in a FH patient. The patient’s sample upon whole exome sequencing identified a total of 543,366 functional variants in and along with high quality 2,009,135 maps reads. Subsequent mapping when compared with the reference human genome revealed about 90 % clean reads distinctively corresponding to the whole exome areas covering about 98 % of the section with at least 95 % folds mean depth coverage. Moreover, the maximum depth coverage was 99 %, sufficient to reliably identify variants within majority of the exome regions in the genomic DNA. Complete variants distribution of whole exome sequencing (Figure 1A-C). The results for six associated genes were also compared with the previous whole-genome sequencing results that were accessible from other SNPs reference databases such as dbSNP,13 cosmic,14 ensemble15 and HapMap.16 The variants were taken in to account as candidate variants only if a variant sequence was confined to more than 30 % from the total reads with a quality value more than 20-25 kb. Using filtering with the reference human genome, around 407 variants were detected among 2953 genes when compared with a healthy control. To establish an appropriate ratio of candidate variants in heterozygous genotypes, distribution ratio of candidate variants was also compared with the 960 in common variants. The results exhibited that the ratio of candidate variants in heterozygous genotypes was greater than 25-30 % of the common variants. According to the established criteria, a total of seven deleterious and two missense but neutral variants were found in seven different genes in the genomic DNA samples, out of which one was a novel variant (Table 1). The maximum predictive false positive value of the targeted genomic DNA was 3 % (1/960). Our result revealed a novel EPHX2 frameshift variant (c.1644delC, p.(Pro549fs) at exon 19 and a common significant variant c.1895T>G, p.(Val632Ala) at exon 13 has a significantly higher association with FH disease compared to the other four reported variants (Table 1). This high throughput study displayed extremely precise outcomes with a very low false-positive rate.

Prioritization and significance of disease associated genes

The FH samples were examined to elucidate with a known cardiovascular disease genes studied in the sequencing study (Figure 2A). The results identified 20794 variants from a FH Saudi patient. The number of exonic variants based on the wANNOVAR output following filtration showed 3925 variants that are notable. We found that 81.2 % of the genes were ranked as “top 1”, 90.7 % as “top 10”, whereas overall 93.4 % of the variants were ranked as “top 100” genes. The number of exonic variants based on the wANNOVAR output following filtration showed 3925 variants that are notable. We found that 81.2 % of the genes were ranked as “top 1”, 90.7 % as “top 10”, whereas overall 93.4 % of the candidate 3201 genes could be identified in its ranked list by Phenolyzer (Figure 2B). The wANNOVAR listed 3201 genes further analysis and the results demonstrated that phenolyzer was proficient of finding 1150 genes, known to be associated with cardiovascular diseases. Several other datasets were also examined to evaluate the ability of phenolyzer to prioritize candidate genes for FH complex diseases (Figure 2C and 2D). Furthermore, the FH disease genes observed identified 517 genes were cardiovascular genes from COSMIC (catalog of somatic mutations in cancer), 1150 genes strongly associated (false-discovery rate <0.05) with FH from exome sequencing. All the reported genes were used as positive genes, whereas all other genes were used as a negative set (although some may still be genuine disease genes). The results demonstrated that Phenolyzer adequately performs tasks compared to other tools, such as for cardiovascular-associated genes that were identified from an exome sequencing study (Figure 2E), phenolyzer achieved an AUC score above 0.85, but none of the other tools has AUC scores higher than 0.81. In addition, we also evaluated the phenolyzer with only the seed gene list (without the seed gene growth step) and found that

Table 1: Summary of whole exome sequencing data and variants filtration.

<table>
<thead>
<tr>
<th>Chr</th>
<th>Type</th>
<th>Ref</th>
<th>Zygosity</th>
<th>DbSNP</th>
<th>Func dbSNP</th>
<th>Amino acid change</th>
<th>Coding region change</th>
<th>Exon</th>
<th>Gene</th>
<th>SNP effect</th>
<th>SIFT</th>
<th>PolyPhen2</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>C</td>
<td>T</td>
<td>Heter</td>
<td>rs11583680</td>
<td>Missense</td>
<td>p.Ala53Val</td>
<td>c.158C&gt;T</td>
<td>T</td>
<td>PCSK9</td>
<td>Deleterious</td>
<td>Tolerated</td>
<td>Benign</td>
</tr>
<tr>
<td>2</td>
<td>T</td>
<td>C</td>
<td>Homo</td>
<td>rs6544718</td>
<td>Missense</td>
<td>p.Val632Ala</td>
<td>c.1895T&gt;C</td>
<td>13</td>
<td>ABCG8</td>
<td>Deleterious</td>
<td>Tolerated</td>
<td>Benign</td>
</tr>
<tr>
<td>2</td>
<td>G</td>
<td>A</td>
<td>Heter</td>
<td>rs1367117</td>
<td>Missense</td>
<td>p.Thr98Ile</td>
<td>c.293C&gt;T</td>
<td>4</td>
<td>APOB</td>
<td>Deleterious</td>
<td>Tolerated</td>
<td>Benign</td>
</tr>
<tr>
<td>9</td>
<td>T</td>
<td>C</td>
<td>Homo</td>
<td>rs2230808</td>
<td>Missense</td>
<td>p.Lys1587Arg</td>
<td>c.4760A&gt;G</td>
<td>35</td>
<td>ABCA1</td>
<td>Deleterious</td>
<td>Tolerated</td>
<td>Benign</td>
</tr>
<tr>
<td>9</td>
<td>C</td>
<td>T</td>
<td>Heter</td>
<td>rs2220806</td>
<td>Missense</td>
<td>p.Arg219Lys</td>
<td>c.656G&gt;A</td>
<td>7</td>
<td>ABCA1</td>
<td>Benign</td>
<td>Tolerated</td>
<td>Benign</td>
</tr>
<tr>
<td>8</td>
<td>C</td>
<td>-</td>
<td>Heter</td>
<td>Novel</td>
<td>Frameshift</td>
<td>p.Pro549fs</td>
<td>c.1644delC</td>
<td>19</td>
<td>EPHX2</td>
<td>Truncated</td>
<td>protein</td>
<td>N/A</td>
</tr>
<tr>
<td>8</td>
<td>G</td>
<td>A</td>
<td>Heter</td>
<td>rs1801177</td>
<td>Missense</td>
<td>p.Asp36Asn</td>
<td>c.106G&gt;A</td>
<td>2</td>
<td>LPL</td>
<td>Deleterious</td>
<td>Damaging</td>
<td>N/A</td>
</tr>
<tr>
<td>8</td>
<td>C</td>
<td>G</td>
<td>Heter</td>
<td>rs328</td>
<td>Nonsense</td>
<td>p.Ser474*</td>
<td>c.1421C&gt;G</td>
<td>9</td>
<td>LPL</td>
<td>N/A</td>
<td>N/A</td>
<td>N/A</td>
</tr>
<tr>
<td>16</td>
<td>G</td>
<td>A</td>
<td>Heter</td>
<td>rs5882</td>
<td>Missense</td>
<td>p.Val422Ile</td>
<td>c.1264G&gt;A</td>
<td>14</td>
<td>CETP</td>
<td>Benign</td>
<td>Tolerated</td>
<td>Benign</td>
</tr>
</tbody>
</table>

N/A: not applicable
its performance was greatly reduced, suggesting the importance of seed gene growth to find new genes not documented in phenolyzer disease-gene.

**Identification of candidate variants by WES**

In the examined patient, we identified 9 different variants in 7 genes (Table 1). From the revealed variants, 7 were found deleterious and others 2 variants are benign or tolerated (Table 1). WES reads alignment showing the deleterious variants detected in the patient (Figure 3). Out of the seven deleterious variants, there was one novel frameshift variant c.1644delC, p.(Pro549fs) detected in EPHX2. In the LPL gene, one missense variant i.e. c.106G>A, p.(Asp36Asn) and a common non-sense variant c.1421C>G, p.(Ser474∗) were identified. The ABCA1 gene showed two missense variants c.44760A>G, p.(Lys1587Arg) and c.1895T>C p.(Val632Ala) which was highly associated with FH and was predicted as functionally deleterious from SNP effect program (Table 1). From the revealed variants, 7 were found deleterious and others 2 variants are benign or tolerated (Table 1). WES reads alignment showing the deleterious variants detected in the patient (Figure 3). Out of the seven deleterious variants, there was one novel frameshift variant c.1644delC, p.(Pro549fs) detected in EPHX2. In the LPL gene, one missense variant i.e. c.106G>A, p.(Asp36Asn) and a common non-sense variant c.1421C>G, p.(Ser474∗) were identified. The ABCA1 gene showed two missense variants c.44760A>G, p.(Lys1587Arg) and c.1895T>C p.(Val632Ala) which was highly associated with FH and was predicted as functionally deleterious from SNP effect program (Table 1). The PCSK9 showed one variant c.158C>T, p.(Ala53Val) that was functionally deleterious. Lastly, the APOB showed a missense variant c.293C>T, p.(Thr98Ile) that was found to be functionally deleterious. Capillary sequencing validated all the deleterious variants.

**Functional and structural analysis by MD simulation**

The results obtained were further confirmed based on in-silico MD simulation studies to observe whether a mutation in FH associated genes relates to the changes in the translated regions. The results indicated that mutations in the six genes were all deleterious mutations, because all of them were functionally deleterious however, we focused on only one mutation that was common and highly associated with FH from the phenolyzer i.e., Val632Ala in ABCG8. The wild and mutant protein structures were also compared using MD simulation. The common splice variant Val632Ala in ABCG8 showed a change in the translated protein and regardless of whether it had an impact, there was a product of a missense mutation in the protein. Besides, associating the results within the coding regions of the protein structure of a FH gene, an annotation on the missense mutations retrieved from the protein data bank (PDB) within the gene on chromosome 2p21 was also performed to determine the structure from the UniProtKB - Q9H221 (ABCG8_HUMAN) (PDB ID: 5do7_B) (Figure 4A-E).

The predicted protein structure had four chains A, B, C and D however we focused on the “Chain B” for alignment parameters and for MD simulations. The mutation Val632Ala was situated at beta helix of the ABCG8 protein domain structure. The mutations connected to each position were accomplished using CCP4 (QtMG), which exclusively identify to examine the altered model structures. The altered residues as well as a helix mutation analysis for ABCG8 protein were also studied using another method called the consensus, which utilized the extra basic residue with a hot residue to expand the thermodynamic strength of ABCG8 protein. It is observed that such a scheme is straightforward in contrast to already depicted methodologies. The variant Val632 in ABCG8 and showed another thermostable site (632Val) and was a critical basic residue with a hot residue to expand the thermodynamic strength of ABCG8 protein. It is observed that such a scheme is straightforward in contrast to already depicted methodologies. The variant Val632 in ABCG8 and showed another thermostable site (632Val) and was a critical basic residue with a hot residue to expand the thermodynamic strength of ABCG8 protein. It is observed that such a scheme is straightforward in contrast to already depicted methodologies. The variant Val632 in ABCG8 and showed another thermostable site (632Val) and was a critical basic residue with a hot residue to expand the thermodynamic strength of ABCG8 protein. It is observed that such a scheme is straightforward in contrast to already depicted methodologies.
The deviation of mutation Val632Ala was retained at 1.2 Å until the simulation was completed (t=10 ns). This demonstrated that Val632Ala had achieved its folded state while the little peak at 1.2 ns showed that Val632Ala settled the protein structure. The resulting evidence demonstrated that the Val632Ala mutant structure was steady and could keep up its adaptation at 300 K, at a pressure bar of 1.00047. Moreover, the mutation Val632Ala also demonstrated a higher dissolvable and accessible surface area than the mutant 632Ala within the protein structure.

The structural discharge of Val632 created less hydrophobic properties within the protein structure while the addition of 632Ala residue could expand minimization in the protein structure. The evaluation of annotated predicted solvent availability and pre-ascertained packing density revealed that Val632Ala was dense and reduced the internal cavities compared with the wild type. Consequently, the substitution of a residue Val632 with 632Ala enhanced the packing and compactness of the mutant structure and diminished the internal residue in the protein structure.

**DISCUSSION**

The NGS has been utilized to find out causative mutations in both known and novel genes associated with FH to be used as a competent tool.\(^3\) In this study, we utilized WES to identify FH causing mutations in a Saudi FH patient and for the diagnosis of FH. It was based on analyzing wide-ranging genetic variations that were not limited to SNVs, small insertions or deletions and copy number variations (CNV). Our results identified seven causative mutations in six known associated FH genes through WES analysis. Out of the seven causative mutations highly associated with FH, only one was a novel frameshift mutation Pro549fs, identified in EPHX2 gene, while the one mutation was a common missense mutation Val632Ala. The obtained results suggest that generated DNA sequences can be used for the discovery and genetic diagnosis of novel variants in genes associated with the FH without any additional costs. This challenge has also triggered an improvement in the development of next-generation sequencing technologies (NGS).

*In-silico* analysis of the seven missense mutations revealed that only five were predicted to be damaging, while the rest were benign. We discovered that one novel mutation is pathogenic and resulted in reduced EPHX2 expression and uptake. This novel frameshift mutation was likely to cause hypercholesterolemia because of the truncated transcript. One nonsense mutation p.(Ser474*) have identified in LPL gene that might be damaging because LPL gene encodes 475 amino acids and therefore only last amino acid affected. One study shows that this mutation p.(Ser474*) influence plasma lipid concentrations and interact with plasma n-6 PUFA to modulate lipid metabolism.\(^3\) Although, the novel mutation in EPHX2 showed precise association however, we were unable to elucidate the underlying mechanism between genetic mutation and high correlation with the CHD risk. It has previously been shown that the mutation in EPHX2 demonstrated a significant association between haplotype and CHD risk in Caucasian population,\(^3\) while the mutation was known for functional relevance *in vitro* and/or haplotype tagging properties.

Nonetheless, even a protein product would lead to impact function of the receptor and several receptor domains would be altered such as the crucial transmembrane region, C-terminal part of the epidermal growth factor precursor homology domain, O-linked sugar domain, membrane-spanning domain and the cytoplasmic domain. Apart from the mutations, lacking these domains, have been shown to reduce the activity (~2 %) and have been classified as null allele (class 1) mutations.\(^3\) All the causative mutations produced mRNA transcripts, however, they were in reduced concentration,\(^3\) which explains the increased risk of FH caused due to this mutation.

We studied, another common missense mutation Val632Ala highly associated with FH gene ABCG8 located at the extracellular domain (ABC transmembrane type-2). The mutation was annotated with severity polymorphism and located in a region with known splice variants, as
described in dbSNP (rs6544718). The ABCG8 gene has been linked with rare autosomal recessive disorder illustrated by enhanced intestinal absorption of all sterols including cholesterol. The sitosterolemia develops hypercholesterolemia due to increased amount of plant sterols in the plasma and often result in tendon and tuberous xanthomas, accelerated atherosclerosis and premature coronary artery disease.\(^4\) The A632V mutation in ABCG8 has previously been observed\(^4\) where the V allele was associated with a high plasma TC in Spanish children. It has also been reported that the heterozygotes had higher plasma TC and APOB levels than AA homozygotes, but only in the group with low cholesterol intake.\(^4\) However, no association of A632V and TC levels has been observed in other studies.\(^5,44\) Based on the functional studies, this common mutation probably lead to loss of interaction due to alteration in the domain and result in abolishing its function. The meta-analysis results consisting of 3,364 individuals from 16 investigations\(^2\) showed that the occurrence of the minor 632V allele was associated with reduced LDL-C concentrations compared with homozygosity for the 632A variant (\(−0.11\) mmol/l, \(P=0.01\)). The other SNPs (Q064E, D19H, Y54C and T400K) were not linked with plasma lipid levels. The result of ABCG8 gene association Prioritization method report showed to be clinically significant (p-value \(<0.01\)), which was based on the filtered genes and polymorphism as well functional analysis. The significantly associated variant A632V also exhibited an average 0.11 mmol/l decrease in plasma LDL-C levels, compared with carriers of the other variant. It has previously been shown that a 10\% reduction of plasma LDL-C levels is associated with a 20\% decrease in major cardiovascular events.\(^4\)

To further evaluate through in-silico study, we selected one of the common splice variant Val632Ala through MD simulations. The results indicated that one variant represented structurally conserved mutation, which will not cause unfolding in the protein, whereas the variant may cause severe disruption in the protein. The protein structure having the variant Val632Ala was predicted and simulated. The wild type of Val632 residue was located in the alpha helix of the module, in a solvent exposed position and its side chain did not form hydrogen bonds. The in-silico study showed that the variant Val632Ala had small rearrangements at around 2 ns in the area of the mutated residue 632Val, otherwise it was quite stable until the end of the trajectory (5 ns). Notably, the hydrophobic bond network connecting the two lobes remains stable and the calcium cage was perfectly formed.

Although this study is based on single patient, but this WES approach would be helpful for the screening of the patient who shown to be negative for mutation/s in known FH associate genes. Similarly, using WES method, few other studies were also reported to analyze single patient sample belongs to other disorders.\(^5,41\) In our earlier study in a single FH proband, a LDLR frameshift variant was detected using targeted NGS.\(^5\) However, screening of that frameshift variant p.(G676Asfs*33) in several FH families revealed as a very common variant in Saudi FH population.\(^3\) Consequently, there is possibility, the results obtained in this study would be replicated to other patients.

In conclusion, we identified disease-associated variants in APOB, PCSK9, EPHX2, LPL, ABCG8 and ABCA1 genes in a Saudi FH patient. The results further indicated that seven of the variants were found to be deleterious mutations. Also, a novel frameshift mutation c.1644delC, p.(P549fs) was identified in EPHX2. Nevertheless, the characterization of a novel mutation and additional harmful mutations will not only assist to find out the important role of uncommon genes in FH, but also build up the range of FH mutations. These novel perceptions will augment the genetic diagnosis and counseling of FH individuals.

ACKNOWLEDGEMENT

The authors would like to thank the Science and Technology Unit at Umm Al-Qura University for the continuous support. This work was supported by the National Science, Technology and Innovative Plan (MAARIFAH) of the Kingdom of Saudi Arabia to Dr. Faisal A. Al-Allaf, (Grant Code: 08-BIO34–10).

CONFLICT OF INTEREST

The authors declare no conflict of interest.

ABBREVIATIONS

WES: Whole exome sequencing; FH: Familial hypercholesterolemia; MD: Molecular dynamic; NGS: Next-generation sequencing; LDL-C: Low-density lipoprotein cholesterol; CAD: Coronary artery disease; BAM: The Binary Alignment/Map; SAAP: Single amino acid polymorphism database; MC: Monte Carlo; CVN: Copy number variations.

SUMMARY

This study was part of an in-depth study focusing on the screening of FH variants using targeted exome sequencing in the Saudi population. Studied patient was initially screened and shown to be negative for causative variant/s in known highly FH associate gene LDLR. Using whole exome sequencing (WES) multiple causative variants were identified in LPL, EPHX2, PCSK9, APOB, ABCA1 and ABCG8 genes in a Saudi FH patient. Out of the seven deleterious variants identified in six genes, there was one novel frameshift variant c.1644delC, p.(Pro549fs) detected in EPHX2 gene. All the causative variants were functionally predicted to be deleterious. Further investigations of the common variant (p.Val632Ala) of ABCG8 gene, using Molecular dynamic (MD) simulation with solvated condition confirmed the functional and structural impact of the changes on the mutated protein.

REFERENCES


