Gene 565 (2015) 76–84

Contents lists available at ScienceDirect

Gene

journal homepage: www.elsevier.com/locate/gene

Research paper

Next generation sequencing to identify *novel* genetic variants causative of autosomal dominant familial hypercholesterolemia associated with increased risk of coronary heart disease

Faisal A. Al-Allaf [a,b,c,*,1], Mohammad Athar [a,b,*,1], Zainularifeen Abduljaleel [a,b,*,1], Mohiuddin M. Taher [a,b], Wajahatullah Khan [d], Faisal A. Ba-hammam [a], Hala Abalkhail [e], Abdullah Alashwal [e,1]

[a] Department of Medical Genetics, Faculty of Medicine, Umm Al-Qura University, P.O. Box 715, Makkah 21955, Saudi Arabia
[b] Science and Technology Unit, Umm Al-Qura University, P.O. Box 715, Makkah 21955, Saudi Arabia
[c] Molecular Diagnostics Unit, Department of Laboratory and Blood Bank, King Abdullah Medical City, Makkah 21955, Saudi Arabia
[d] Department of Basic Sciences, College of Science and Health Professions, King Saud Bin Abdulaziz University for Health Sciences, PO Box 3124, Riyadh 11426, Saudi Arabia
[e] Department of Pediatrics, MBC 58, King Faisal Specialist Hospital and Research Centre, P.O. Box 3354, Riyadh 11211, Saudi Arabia

ARTICLE INFO

Article history:
Received 4 December 2014
Received in revised form 27 March 2015
Accepted 29 March 2015
Available online 1 April 2015

Keywords:
Familial hypercholesterolemia (FH)
Low-density lipoprotein cholesterol (LDLc)
LDLR
Coronary heart disease (CHD)
Functional changes

ABSTRACT

Familial hypercholesterolemia (FH) is an autosomal dominant inherited disease characterized by elevated plasma low-density lipoprotein cholesterol (LDL-C). It is an autosomal dominant disease, caused by variants in *Ldlr*, *ApoB* or *Pcsk9*, which results in high levels of LDL-cholesterol (LDL-C) leading to early coronary heart disease. Sequencing whole genome for screening variants for FH are not suitable due to high cost. Hence, in this study we performed targeted customized sequencing of FH 12 genes (*Ldlr*, *ApoB*, *Pcsk9*, *Abca1*, *Apoa2*, *Apoc3*, *Apon2*, *Arh*, *Ldlrap1*, *Apoc2*, *ApoE*, and *Lpl*) that have been implicated in the homozygous phenotype of a proband pedigree to identify candidate variants by NGS Ion torrent PGM. Only three genes (*Ldlr*, *ApoB*, and *Pcsk9*) were found to be highly associated with FH based on the variant rate. The results showed that seven deleterious variants in *Ldlr*, *ApoB*, and *Pcsk9* genes were pathological and were clinically significant based on predictions identified by SIFT and PolyPhen. Targeted customized sequencing is an efficient technique for screening variants among targeted FH genes. Final validation of seven deleterious variants conducted by capillary resulted to only one novel variant in *Ldlr* gene that was found in exon 14 (c.2026delG, p. Gly676fs). The variant found in *Ldlr* gene was a novel heterozygous variant derived from a male in the proband.

1. Introduction

Familial hypercholesterolemia (FH) is an autosomal dominant disorder that causes elevations in total cholesterol and low-density lipoprotein cholesterol (LDLc) (Asahina et al., 2012; Ueda, 2005). The inherited genetic variants occur most commonly in the *Ldlr* gene. Less common defects in apolipoprotein B 100 (*ApoB*-100), and the proprotein convertase subtilisin/kexin 9 (*Pcsk9*), may also cause FH phenotype (Al-Allaf et al., 2010, 2014). FH is among the most common inherited metabolic disorders, a well-recognized cause of premature coronary heart disease (CHD) (Hopkins et al., 2011) and the first genetic disorder shown to cause myocardial infarction (Asahina et al., 2012), which causes premature heart disease and death in affected individuals. If untreated, men have a 50% chance of CHD before the age of 50 years and women a 30% risk by the age of 60 years (The National Collaborating Centre for Primary Care, 2008). Atherosclerosis caused by FH starts in childhood and adolescence, highlighting the need to identify cases early and commence preventive measures (Descamps et al., 2011). The frequency of causative variants in the *Ldlr* gene is estimated to be one in 500 in the general population, and up to one in 100 in certain ethnic groups such as Ashkenazi Jews from South Africa, Christian Lebanese and Dutch Afrikaaners, due to a genetic founder effect (Austin et al., 2004).

Currently, variants in the three genes *Ldlr*, apolipoprotein B (*ApoB*), and proprotein convertase subtilisin/kexin 9 (*Pcsk9*) have been established as causatives of the disease (Chiou and Charng, 2010). Reported variants of FH widely spread over the coding regions of these causative genes. Genetic screening through traditional approaches, such as direct sequencing is therefore difficult. A high-throughput and cost-effective method to detect the genetic defects is needed. Whole exome sequencing has been proven to be a powerful tool to discover novel disease-

Abbreviations: FH, familial hypercholesterolemia; LDL-C, low-density lipoprotein cholesterol; CHD, coronary heart disease; *ApoB*, apolipoprotein B; *Pcsk9*, proprotein convertase subtilisin/kexin 9; NGS, next-generation sequencing; ISP, Ion Sphere Particles; PCR, polymerase chain reaction; SIFT, sorting intolerant from tolerant; PolyPhen, Polymorphism Phenotyping v2; HG19, human genome.

* Corresponding authors at: Department of Medical Genetics, Faculty of Medicine, Umm Al-Qura University, Al-Abedia Campus, P.O. Box 715, Makkah 21955, Saudi Arabia.
E-mail addresses: fallaf@uqu.edu.sa (F.A. Al-Allaf), athar80@gmail.com (M. Athar), zainulbio@gmail.com (Z. Abduljaleel).

[1] Authors contributed equally to the manuscript.

http://dx.doi.org/10.1016/j.gene.2015.03.064
0378-1119/

related genes or genetic variants in large genomic regions (Kruglyak et al., 2014; Nair, 2013; Schekman, 2013; Zetterstrom, 2011). With the progresses on next-generation sequencing (NGS) and bioinformatics, it has been demonstrated to have higher efficiency but lower cost comparing with whole genome sequencing by using other method such as capillary sequencing. In this study, we utilized targeted customized gene sequencing to study genetic defects in Saudi patients, in order to establish a strategy feasible to genetic diagnosis of FH patients.

NGS platforms, this study, we have designed customized 12 genes involved in FH genetic disease. After enrichment, high-throughput sequencing using the Ion torrent PGM was performed to analyze the variants of disease-causing genes in FH patient sample in one sequencing lane using the sample barcoding method. Next, we used the capillary sequencing method followed by PCR amplification to confirm the results obtained from NGS. Our current findings suggest that this targeted DNA sequencing technology can be used to detect variants of FH disease-causing associated genes with high conformity.

2. Materials and methods

2.1. Library construction for Ion torrent PGM sequencing

The genomic DNA (gDNA) from proband and controls were extracted from EDTA treated whole blood samples using the MagNA Pure Compact Nucleic Acid Isolation Kit-I (Roche) according to the manufacturer's guidelines. DNA quantification was performed using a NanoDrop 2000 unit (Thermo Fisher Scientific, Wilmington, DE). The gDNA libraries were constructed using Ion Plus Fragment Library kit with BioRuptor sonication system and the Ion Xpress bar-code adapters 1–16 kit (Life Technologies). During library contraction, a Pippin Prep instrument (Sage Science, Beverly, MA) was used for size selection during library construction. DNA concentrations were quantified and size distributions were evaluated with an Agilent Bioanalyzer 2100 using the High sensitivity DNA kit (Agilent, Santa Clara, CA). A final library size of 310–340 bp, including adapter sequences, was selected. All these steps were carried out and done as per the manufacturer's instructions indicated in the kits.

2.2. Enrichment and sequencing of disease genes

Twelve FH associated genes (*Ldlr*, *ApoB*, *Pcsk9*, *Abca1*, *Apoa2*, *Apoc3*, *Apon2*, *Arh*, *Ldlrap1*, *Apoc2*, *ApoE*, and *Lpl*) were included in Ion TargetSeq custom enrichment kit design (Life Technologies). The primers for customized panel were designed to cover coding exons and flanking intron regions of the selected genes. Enrichment of targeted sequences of the genes was achieved using Ion TargetSeq custom enrichment kit. Emulsion PCR and Ion Sphere Particles' (ISP) enrichment was done using the Ion one touch template kit v2.0 (Life Technologies). Finally sequencing was performed on PGM using Ion PGM 200 sequencing kit (Life Technologies). All steps were performed as per the manufacturer's instructions of the kits used in the experiment.

2.3. Mapping assembly and variant discovery from NGS data

The BAM binary format sequence data raw reads went through adapter trimming, hence removal of reads shorter than 20 bp and removal of exact duplicates, as well as quality trimming. The pre-processed reads were aligned using CLC Genomics Workbench v7 against reference sequences corresponding to the customized genes (ref Genome Hg19). This was followed by SNV and indel detection. Each variant within the exonic regions of targeted genes was confirmed by capillary sequencing. For all three samples, the Ion torrent PGM raw reads were aligned against the Hg19 using CLC genomics workbench to detect variants within aligned data with additional base quality score recalibration. All variants detected within the coding exons of customized genes were considered for subsequent analyses using probabilistic and quality based variant detection method.

2.4. Capillary sequencing

Polymerase chain reaction (PCR) amplification of FH three genes *Ldlr*, *ApoB*, and *Pcsk9* for identified pathogenic variants was performed. The PCR reaction was performed with 100 ng of genomic DNA as the template in a total 20 μl reaction mixture using a HotStarTaq Plus DNA Polymerase Kit (Qiagen, Germany) as follows: Taq polymerase was activated at 94 °C for 5 min, followed by 35 cycles of denaturing at 94 °C for 30 s, annealing at 61–64 °C for 30 s, extension at 72 °C for 45 s, and final extension at 72 °C for 5 min. The amplified products were also separated on agarose gel to ensure the size and quality of the band. The PCR products were purified based on magnetic beads method using Agencourt AMPure XP kit (Beckman Coulter). The purified PCR products were then used as templates for direct sequencing using BigDye Terminator v3.1 cycle sequencing ready reaction kit (Applied Biosystems, CA, USA). The sequencing reaction products were purified with BigDye X-terminator purification kit (Applied Biosystems) followed by capillary electrophoresis in an ABI 3500 Genetic analyzer (Applied Biosystems), and the final analysis was performed with Sequence Analysis Software v5.4 (Applied Biosystems). Furthermore *in silico* functional studies was determined by sorting intolerant from tolerant (SIFT and Provean) (Kumar et al., 2009; Sim et al., 2012; Ng and Henikoff, 2006), and Polymorphism Phenotyping v2 (PolyPhen) (Adzhubei et al., 2013); both are Nature protocols.

3. Results

3.1. Sequence enrichment, high-throughput sequencing, mapping and coverage

To find out the disease-causing variants in FH associated genes, we designed a unique high throughput density capture array to capture the 538,751 targets that covers exons and flanking intron regions of the selected genes known to be involved in dominant FH disease in general and with a higher incidence frequency in Saudi Arabian population as well. Each DNA probe in the designed array was approximately 60 bp long, and can hybridize with all of the targeted exon sequences. The enriched DNA was sequenced at a single-base resolution using the Ion-torrent PGM. To evaluate this technology, we collected clinically diagnosed heterozygous FH patient sample as well as two samples from normal human (one male, and another female) as a control. The NGS workflow and functional variants detection as shown (Fig. 1) using targeted customized genes obtained a total of 3,043,358–4,646,170 high-quality reads encompassing 613,053,599–629,131,774 high-quality bases per patient. After mapping to the reference human genome (HG19), 80.90% of the yielded clean reads could be uniquely matched to the target regions, and 98.60% of the targeted region was covered in at least 95.42% fold coverage of mean depth in each sample (Table 1). The average coverage depth for exons was 98.60%, and the highest coverage depth for exons was 99%, across one sample (Fig. 2A–C). Thus, the coverage should have been adequate to reliably detect DNA variants within the majority of the targeted regions (Fig. 3A–C).

3.2. Estimation of the accuracy of customized genes

To estimate the accuracy of our method, we detected variants of customized 12 genes using our targeted by Ion-torrent PGM and compared the results obtained with the results from whole-genome sequencing for the same 12 genes in the same sample deposited in dbSNP (Bhagwat, 2010; Sherry et al., 2001), Cosmic (Forbes et al., 2010; Bamford et al., 2004), Ensemble (Flicek et al., 2014; Chen et al., 2010), and HapMap (Abecasis et al., 2012; A Haplotype map of the human genome, 2005; The International HapMap Project, 2003). The