Atherosclerosis 277 (2018) 440–447

Contents lists available at ScienceDirect

Atherosclerosis

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

Identification of a novel *LDLR* disease-causing variant using capture-based next-generation sequencing screening of familial hypercholesterolemia patients in Taiwan

Yun-Chieh Hsiung[a,1], Po-Chih Lin[b,1], Chih-Shan Chen[a], Yi-Ching Tung[c,d], Wei-Shiung Yang[a,d,e,f,g], Pei-Lung Chen[a,d,e,f,h,**], Ta-Chen Su[b,i,*]

[a] Graduate Institute of Medical Genomics and Proteomics, National Taiwan University College of Medicine, Taipei, Taiwan
[b] Department of Internal Medicine and Cardiovascular Center, National Taiwan University Hospital, Taipei, Taiwan
[c] Department of Pediatrics, National Taiwan University Hospital and College of Medicine, National Taiwan University, Taipei, Taiwan
[d] Graduate Institute of Clinical Medicine, College of Medicine, National Taiwan University, Taipei, Taiwan
[e] Division of Endocrinology and Metabolism, Department of Internal Medicine, National Taiwan University Hospital, Taipei, Taiwan
[f] Research Center for Developmental Biology and Regenerative Medicine, National Taiwan University, Taipei, Taiwan
[g] Department of Medicine, College of Medicine, National Taiwan University, Taipei, Taiwan
[h] Department of Medical Genetics, National Taiwan University Hospital, Taipei, Taiwan
[i] Institute of Occupational Medicine and Industrial Hygiene, National Taiwan University College of Public Health, Taipei, Taiwan

HIGHLIGHTS

- This is the first capture-based NGS testing for FH to cover the whole *LDLR* genomic region.
- It can detect from small-size variants to large structural variants accounting for ~10% of disease-causing *LDLR* variants.
- Our genetic diagnosis rate was 75%.
- We identified *LDLR* c.1186 + 2T > G as a novel and common disease-causing variant in Taiwan.

ARTICLE INFO

Keywords:
Familial hypercholesterolemia
Next generation sequencing
Mutation spectrum
LDLR
Haplotype
Structural variation

ABSTRACT

Background and aims: Familial hypercholesterolemia (FH) is an autosomal dominant disorder with paramount health impacts. However, less than 1% FH patients in Taiwan were formally diagnosed, partly due to the lack of reliable cost-effective genetic testing. We aimed at using a next-generation sequencing (NGS) platform as the clinical genetic testing method for FH.

Methods: We designed probes to capture the whole *LDLR* gene and all coding sequences of *APOB* and *PCSK9*, and then sequenced with Illumina MiSeq platform (2 × 300 bps). The entire pipeline was tested on 13 DNA samples with known causative variants (including 3 large duplications and 2 large deletions). Then we enrolled a new cohort of 28 unrelated FH patients with Dutch Lipid Clinic Network score ≥5. Relatives were included in the cascade screening.

Results: From the 13 DNA samples, we correctly identify all the variants, including big duplications and deletions. From the new cohort, we identified the causative variants in 21 of the 28 unrelated probands; five of them carrying a novel splice site variant c.1186 + 2T > G in *LDLR*. Among the family members, the concentration of LDL cholesterol was 7.82 ± 2.13 mmol/l in *LDLR* c.1186 + 2T > G carrier group (n = 26), and was significantly higher than 3.18 ± 1.36 mmol/l in the non-carrier group (n = 25).

Conclusions: This is the first capture-based NGS testing for FH to cover the whole *LDLR* genomic region, and therefore making reliable structural variation detection. This panel can comprehensively detect disease-causing variants in *LDLR*, *APOB*, and *PCSK9* for FH patients.

* Corresponding author. Department of Internal Medicine, National Taiwan University Hospital, 7 Chung-Shan South Road, Taipei, 10002, Taiwan.
** Corresponding author. Graduate Institute of Medical Genomics and Proteomics, National Taiwan University College of Medicine, 2 Xuzhou Rd, Zhongzheng Dist, Taipei, 10055, Taiwan.
E-mail addresses: paylong@ntu.edu.tw (P.-L. Chen), tachensu@ntu.edu.tw (T.-C. Su).
[1] The authors contributed equally to this work.

https://doi.org/10.1016/j.atherosclerosis.2018.08.022
Received 31 March 2018; Received in revised form 10 August 2018; Accepted 21 August 2018

1. Introduction

Familial hypercholesterolemia (FH, OMIM #143890) is primarily an autosomal dominant disorder, characterized by very high serum low-density lipoprotein-cholesterol (LDL-C) [1], which results in excess deposition of cholesterol in tissues, leading to accelerated atherosclerosis and increased risk of premature coronary heart disease (CHD) [2]. FH is commonly caused by the disease-causing variants in the low-density lipoprotein receptor (*LDLR*), ApoB (*APOB*), or proprotein convertase subtilisin-kexin type 9 (*PCSK9*) genes [3–5]. Recently, apolipoprotein E (*APOE*) variant was also found to be associated with dominant FH [6,7].

Positive genetic testing result is one of the most important criteria for FH diagnosis along with total cholesterol levels (especially, LDL-C levels), tendon xanthomata, corneal arcus, and family history. More than 1700 *LDLR* variants have been reported in the Human Gene Mutation Database (http:/www.hgmd.cf.ac.uk/ac/index.php) [8]. Large scale genetic screening for FH has been performed in the Netherlands, Spain, Norway, Denmark, UK, USA and Wales [9,10], but most countries (including Taiwan) lack valid nationwide registries for FH. FH patients are usually underdiagnosed and undertreated in Taiwan [11].

The traditional genetic testing used for FH is a combination of Sanger sequencing and multiplex ligation-dependent probe amplification (MLPA) [12]. In the past, we used an FH resequencing array (designed by Vita Genomics, Taipei, Taiwan) to detect disease-causing variants in 3 FH-causing genes (*LDLR*, *APOB* and *PCSK9*). It should be noted that FH resequencing array was accurate and efficient for identifying SNV and small indel, in spite of the limitations on detecting large structural variations [13]. According to previous studies, about 8% of patients with FH in Taiwan have large structural variations [14]. Therefore, for those without any causative variant after FH resequencing array, we still needed to perform MLPA. This workflow was complicated and may miss patients with hemizygous mutation. Moreover, the cost of microarray is expensive. For these reasons, we aimed to using a probe captured-based NGS genetic testing that can combine the benefits of resequencing array and MLPA while reduce costs, resources, and analysis time. Several reports have applied next-generation sequence (NGS) for genetic screening of patients with clinically suspected FH [15–17]. In terms of detecting large structural variation, some of these reports have used NGS bypassing MLPA [18,19]. The difficulty for large structural variation detection using coding-region-only targeted NGS approach is due to the uneven distribution of target sequencing reading depth and the variation in capture efficiency [20]. Furthermore, the intron sequences of the *LDLR* gene are rich in Alu repeats [21], making the probes design in this region troublesome. Thus, in this study we planned to investigate the causative variants in Taiwan FH families using probe capture-based NGS approach while avoiding the above problems and simultaneously detecting small-size variants and large structural variations without using MLPA.

2. Materials and methods

2.1. Patient recruitment

In order to compare the difference between previous method and NGS panel performance, we selected 13 DNA samples with known causative variants in *LDLR* and 3 FH patients without causative variant detected by the previous combined FH array and MLPA methods. Then we recruited 28 new families from our lipid clinic of National Taiwan University Hospital. All these patients met the criteria with LDL-C $\geq$230 mg/dl (5.95 mmol/l) and total cholesterol $\geq$320 mg/dl (8.28 mmol/l), which also met the adult criteria for FH screening recommended by the European Atherosclerosis Society [11]. At the same time, for each proband, we ascertained the family history, physical examination and clinical history of premature CHD or other cardiovascular diseases. Comprehensive consideration of all conditions up to Dutch Lipid Clinic Network score $\geq$5 as the acceptance criteria. We excluded individuals with possible causes of secondary hypercholesterolemia, including hypothyroidism, nephrotic syndrome and cholestatic jaundice. Family members were also included in cascade screening programs. Disease-causing variants were confirmed by Sanger sequencing. All procedures in this study were approved by the Research Ethics Committee of National Taiwan University Hospital and informed consents were obtained from those participants.

2.2. Lipids and lipoprotein measurements

Lipid concentrations, including total cholesterol, triglyceride (TG), high-density lipoprotein cholesterol (HDL-C), and LDL-C, were analyzed using a homogeneous enzymatic method on a Toshiba FR-200 automatic chemistry analyzer (Toshiba, Tokyo, Japan) as previous described [22]. LDL-C was measured by a direct homogenous method (LDL-EX (N) Seiken, Denka Seiken, Tokyo, Japan). Coefficients of variation (CVs) on total cholesterol, triglyceride, HDL-C and LDL-C were under 3%.

2.3. DNA capture probes

One hundred and twenty-eight xGen Lockdown probes (Integrated DNA Technologies, Coraville, IA) were custom-designed based on the DNA sequences, which capture the whole *LDLR* genomic region (including all exons and introns) and all protein-coding sequence of *APOB* and *PCSK9*. Each DNA probe was 120 bps in length. We made great effort to design probes covering all introns of the *LDLR* gene, but to avoid overlaps with the Alu sequences. The probes were dissolved in a double distilled water solution to a final concentration of 0.75 pmol/μl and stored at −20 °C.

2.4. Target enrichment, massively parallel sequencing and variant calling

The whole pipeline included target enrichment, massively parallel sequencing, variant calling, and data filtering. Genomic DNA (gDNA) was extracted from the probands of the 28 families for target capture and sequencing. Briefly, gDNA extraction was performed using the Gentra Puregene Blood Kit (Qiagen, Hilden, Mettmann, Germany), and subjected to agarose gel and O.D. ratio tests to confirm its purity and concentration prior to Covaris fragmentation (Covaris, Woburn, MA). Fragmented gDNAs, aiming at the peak length of 800 bp, were tested for size distribution and concentration using the Agilent Bioanalyzer 2100 (Agilent Technologies, Santa Clara, CA) and Qubit (Thermo Scientific, Waltham, MA). Illumina libraries were generated from gDNA using TruSeq Library Preparation Kit (Illumina, San Diego, CA), the resulting library using the SeqCap EZ Hybridization and Wash Kit (Roche NimbleGen, Madison, WI) for DNA probe hybridization and target capture from Integrated DNA technologies, and were sequenced using Illumina MiSeq to generate paired-end reads of 300 nucleotides. Raw sequencing data was aligned to the reference human genome (Feb. 2009, GRCh37/hg19) using BWA-MEM [23]. We used Picard (version 1.54) (http://picard.sourceforge.net) to perform the necessary data conversion, sorting, and indexing. The main variant calling process, for both indels (insertion/deletions) and single nucleotide variants, was operated by using the GATK software package (version 1.2–59-gd74367c1) [24,25]. We applied Variant Tools (version 1.0.2) to select and analyze the variants [26,27]. ANNOVAR (version 2016-02-01) was used to appropriately annotate the genetic variants; this included gene annotation, amino acid change annotation, SIFT scores [28], PolyPhen2scores [29], dbSNP identifiers, 1000 Genomes Project allele frequencies, NHLBI-ESP 6500 exome project allele frequencies, ExAC 65000 exome allele frequencies, and ClinVar database with variant clinical significance. Data filtering was conducted by an in-house python script with the following steps: (1) selection of variants located in the three targeted genes; (2) filtering out of variants with allele