Atherosclerosis 240 (2015) 324–329

Contents lists available at ScienceDirect

Atherosclerosis

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

Whole exome sequencing combined with integrated variant annotation prediction identifies asymptomatic Tangier disease with compound heterozygous mutations in ABCA1 gene

Hayato Tada [a, *], Masa-aki Kawashiri [a], Atsushi Nohara [b], Reina Saito [c], Yoshihiro Tanaka [a], Akihiro Nomura [a], Tetsuo Konno [a], Kenji Sakata [a], Noboru Fujino [a], Toshinari Takamura [c], Akihiro Inazu [d], Hiroshi Mabuchi [b], Masakazu Yamagishi [a], Kenshi Hayashi [a]

[a] Division of Cardiovascular Medicine, Kanazawa University, Graduate School of Medicine, Kanazawa, Japan
[b] Department of Lipidology, Graduate School of Medical Science, Kanazawa University, Kanazawa, Japan
[c] Department of Disease Control and Homeostasis, Kanazawa University, Graduate School of Medical Sciences, Kanazawa, Japan
[d] Department of Laboratory Science, Molecular Biochemistry and Molecular Biology, Graduate School of Medical Science, Kanazawa University, Kanazawa, Japan

ARTICLE INFO

Article history:
Received 30 January 2015
Received in revised form
8 March 2015
Accepted 2 April 2015
Available online 7 April 2015

Keywords:
Tangier disease
ABCA1
HDL
Exome
Genetics

ABSTRACT

Objective: Molecular diagnosis for subjects with extremely low HDL-C through candidate-gene approaches requires huge effort. Whole exome-sequencing (WES) has already shown approximately ~30% success in the diagnosis of Mendelian disorders. Moreover, novel *in silico* prediction software for the pathogenicity of novel missense variants named Combined Annotation Dependent Depletion (CADD) has recently been developed, enabling the objective integration of many diverse annotations into a single measure (C-score) for each variant. Here, we investigated whether WES combined with integrated variant annotation prediction could facilitate the molecular diagnosis of this rare condition.
Methods: WES was performed on 8 individuals including 2 individuals exhibiting extremely low HDL-C (2 mg/dl and 6 mg/dl), 2 unaffected family members, and 4 unrelated individuals as controls. We filtered out the following variants: 1) Benign variants predicted by SnpEff; 2) Minor allele frequency (MAF) > 1%; 3) Segregation unmatched for the recessive form of inheritance; 4) C-score < 10.
Results: Among 305,202 variants found in those individuals, we found 21,708 nonsense, missense, or splice site variants, of which 5192 were rare (MAF ≤ 1% or not reported). Filtering assuming a recessive pattern of inheritance, combined with the use of the C-score, successfully narrowed down the candidates to compound heterozygous mutations in the ABCA1 gene (c.6230C > A or p.P2077H/c.6137G > A or p.S2046N, and c.2842G > A or p.G948R/c.1130C > T or p.P377L).
Conclusions: WES combined with integrated variant annotation prediction successfully identified asymptomatic Tangier disease with novel ABCA1 mutations. This comprehensive approach is useful to determine causative variants, especially in recessive inherited diseases.

1. Introduction

Decreased HDL-C levels are the most common lipoprotein abnormality in patients with premature coronary artery disease (CAD) [1]. Indeed, epidemiological studies have demonstrated that plasma HDL-C concentration is inversely correlated with the incidence of CAD both in Caucasians and Japanese [2,3]. Under these conditions, it has been shown that plasma lipoprotein levels, including the level of HDL-C, are highly heritable [4], and that both common and rare genetic variants play a key role in regulating HDL-C levels [5–10].

Uncommon encounters with subjects exhibiting extremely low HDL-C with mutations in a specific gene provide an opportunity to directly observe the role of certain molecules in HDL metabolism and atherosclerosis. In order to complete a molecular diagnosis for such subjects, genetic screening for Apolipoprotein A-1 (APOA1), ATP-binding cassette sub-family A member 1 (ABCA1), and lecithin

* Corresponding author. Division of Cardiovascular Medicine, Kanazawa University, Graduate School of Medicine, 13-1 Takara-machi, Kanazawa 920-8641, Japan.

http://dx.doi.org/10.1016/j.atherosclerosis.2015.04.003
0021-9150/

cholesterol acyltransferase (LCAT) genes is currently recommended [11]. However, it is administratively onerous and time-consuming to genotype them through conventional Sanger sequencing because those genes are rather large. It is also possible to miss the true causative variants (genes) in such a targeted strategy, motivating researchers to perform whole exome-sequencing (WES) using comprehensive next generation sequencing techniques. Moreover, a novel *in silico* prediction software for the pathogenicity of novel missense variants named Combined Annotation Dependent Depletion (CADD), which can objectively integrate many diverse annotations into a single measure (C-score) for each variant, has recently been developed [12]. Therefore, in this study, we investigated whether WES combined with integrated variant annotation prediction could facilitate the molecular diagnosis of this rare condition.

2. Methods

2.1. Study population

A 28 year-old Japanese female (our individual ID = LHDL3) was referred to our lipid clinic due to her extremely low level of HDL-C (2 mg/dl) without any apparent secondary causes. Both her parents, who showed no evidence of consanguineous marriage, exhibited mildly low HDL-C levels (our individual ID = LHDL1 and LHDL2 for her father and mother, respectively) and were also included in this study. Another 35 year-old Japanese female (our individual ID = LHDL4), who was referred to our lipid clinic because for the same reason (HDL-C = 6 mg/dl), was also included in this study. Clinical examinations showed the absence of secondary causes. Another four individuals manifesting different types of extreme lipid phenotypes (three subjects with high HDL-C, and one subject with high LDL-C) were also included in this study as controls (our individual ID = HHDL1, HHDL2, HHDL3, and HLDL1, respectively).

2.2. Ethical considerations

This study was approved by the Ethics Committee of Kanazawa University. All procedures followed were in accordance with the ethical standards of the responsible committee on human experimentation (institutional and national) and with the Helsinki Declaration of 1975, as revised in 2008. Informed consents were obtained from all subjects for being included in the study.

2.3. Biochemical analysis

Blood samples were drawn for assays after overnight fasting. Serum levels of total cholesterol, triglycerides and HDL-C were determined enzymatically (Qualigent, Sekisui Medical, Tokyo, Japan) using automated instrumentation based on the assays previously described [13–15]. Serum cholesteryl ester transfer protein (CETP) levels were determined by a specific enzyme-linked immunosorbant assay [16]. LCAT mass was measured as described previously [17].

2.4. Exome sequencing

Genomic DNA was isolated from peripheral blood white blood cells using Genomic DNA Purification Kit (Gentra Systems, Minneapolis, MN). DNA was sheared, sizes were selected and ligated to sequencing adapters, and amplified to enrich for targets to be sequenced by the Agilent SureSelect^{XT} Target Enrichment System (Agilent Technologies Inc.). Exome capture was performed by the Agilent SureSelect^{XT} Human All Exon 50 Mb Kit (Agilent Technologies Inc.). Exome-enriched products were sequenced using the Illumina HiSeq 2000 (Illumina Inc.). One sample was sequenced per lane to obtain an average theoretical depth of 80 $\times$, using 2 $\times$ 100 bp sequencing.

2.5. Combined annotation dependent depletion (CADD)

The pathogenicities of novel missense variants were predicted using CADD prediction software [12]. This software integrates diverse genome annotations using conservation matrices as well as protein based matrices (~63 different tools) into a single measure (C-score) for each variant. The basis of CADD is to contrast the annotations of fixed alleles in humans with those of simulated variants. We determined the threshold of pathogenesis as having a C-score exceeding 10, since a variant could be predicted to be pathogenic if the scaled C-score calculated by the software was above 10, which indicates the top 10% deleterious state among possible substitutions.

2.6. Bioinformatics

Four independent filters were applied after standard variant quality controls in order to successfully discover causal variants among the low HDL-C families. Variants were filtered out as: 1) Benign variants predicted by SnpEff; 2) Minor allele frequency (MAF) > 1% in an Asian population; 3) Segregation unmatched under the assumption of a recessive form of inheritance; 4) C-score < 10 calculated using CADD prediction software.

For the samples, paired-end reads were aligned using the Burrows-Wheeler Aligner on the human reference genome build hg19 using quality score calibration, soft clipping and adapter trimming. Following the exclusion of PCR duplicate reads using Picard, insertion-deletions and SNPs were called using GATK [18,19]. Variants (SNPs/indels) were filtered on the basis of the Phred scaled genotype quality score. Re-alignment was performed and the calling algorithm merged the output of the GATK UnifiedGenotyper. All samples were annotated using SnpEff version 3.6 to classify variants (such as missense, nonsense, splice site, synonymous, intronic, or stop gain/loss) [20].

During functional filtering, missense, nonsense, and splice site variants were considered as candidate variants. The frequency filter adopted allele frequency estimates from the 1000 genomes project Asian cohort database, and a MAF > 1% was used as the cut-off. Segregation pattern matching was defined as narrowing the variants until affected subjects exhibited homozygous or compound heterozygous of alternative alleles in a particular gene. After the application of these filters, we filtered out such variants as C-score < 10. Candidate variants regardless of their C-scores were also evaluated to assess whether the genes including each variant were directly involved in lipoprotein metabolism. Following these evaluations, the putative variants identified by bioinformatics were confirmed using Sanger sequencing methods as previously described [21].

2.7. Phenotype assessments

We evaluated physical examinations, detailed laboratory tests, including determination of the lipoprotein fraction by ultracentrifugation as well as by agarose gel electrophoresis, carotid ultrasonography (Philips iE33 ultrasound system; Philips Inc., Eindhoven, The Netherlands, or Aplio 500; Toshiba Medical Systems, Tokyo, Japan), and abdominal ultrasound examination (Aplio 500; Toshiba Medical Systems, Tokyo, Japan) for the study subjects to establish a clinical diagnosis of those subjects with extremely low HDL-C. The presence of CAD, defined as >75% luminal stenosis, was evaluated either by coronary computed tomography, or by a coronary