Clinical Biochemistry 49 (2016) 166–171

Contents lists available at ScienceDirect

Clinical Biochemistry

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

Targeted next-generation sequencing of the *ATP7B* gene for molecular diagnosis of Wilson disease

Kok-Siong Poon [a], Karen Mei-Ling Tan [a], Evelyn Siew-Chuan Koay [a,b,*]

[a] Molecular Diagnosis Centre, Department of Laboratory Medicine, National University Health System, Singapore
[b] Department of Pathology, Yong Loo Lin School of Medicine, National University of Singapore, Singapore

ARTICLE INFO

Article history:
Received 26 July 2015
Received in revised form 9 September 2015
Accepted 14 October 2015
Available online 19 October 2015

Keywords:
Next-generation sequencing
Wilson disease
ATP7B
Molecular diagnosis
MiSeq

ABSTRACT

Objectives: In recent years, next-generation sequencing (NGS) technologies, which enable high throughput sample processing at relatively lower costs, are adopted in both research and clinical settings. A multiplex PCR-based NGS assay to identify mutations in the *ATP7B* gene for routine molecular diagnosis of Wilson disease was evaluated in comparison with the gold standard direct Sanger sequencing.

Design and methods: Five multiplex PCRs to amplify the partial promoter, 5′ untranslated and the entire coding regions of the *ATP7B* gene were designed. Indexed paired-end libraries were generated from the pooled amplicons using Nextera XT DNA Sample Preparation Kit and subjected to NGS on the MiSeq platform. DNA from the peripheral blood of 12 patients with Wilson disease, 2 B-lymphocyte cell lines and 3 external quality assurance samples were sequenced by the MiSeq and Sanger sequencing.

Results: Complete coverage was achieved across the targeted bases without any drop-out sequences. The observed read depth in a single run with 20 samples was >100X. Comparison of the NGS results against Sanger sequencing data on a panel of clinical specimens, cell lines and European Molecular Genetics Quality Networks (EMQN) quality assurance samples showed 100% concordance in identifying pathogenic mutations.

Conclusion: With the capability of generating relatively higher throughput in a short time period, the NGS assay is a viable alternative to Sanger sequencing for detecting *ATP7B* mutations causally linked to Wilson disease in the clinical diagnostic laboratory.

1. Introduction

The dideoxy sequencing method [1] which is commonly known as Sanger sequencing, continues to be the most widely used method for sequencing in molecular genetics laboratories [2]. Sanger sequencing is considered as the gold standard to identify germ-line mutations, including single nucleotide changes, small deletions and insertions, for the diagnosis of heritable genetic diseases. In recent years, next-generation sequencing (NGS) technologies, which dramatically increase the throughput of sequencing with reduced costs, have been increasingly adapted to both research and clinical settings. NGS had been successfully applied, for example, to interrogate the human exome for the identification of novel causal mutations of complex genetic disorders [3,4]. However, its implementation into routine diagnostic laboratories would require the establishment of appropriate NGS-specific guidelines by professional and regulatory bodies [8] and stringent adherence to such requirements prior to being put into clinical practice. Currently, most attempts to introduce NGS-based sequencing are transitions from Sanger sequencing-based methods that have been successfully applied to the same target genes.

Mutation screening of a large gene with multiple exons using Sanger sequencing is feasible, but is very costly and labor-intensive. Hence, molecular techniques including denaturing gradient gel electrophoresis (DGGE) [5], denaturing high performance liquid chromatography (dHPLC) [6], and high-resolution melting [7] are frequently utilized as screening methods prior to confirming the identity of variants using Sanger sequencing. The reduced cost of DNA sequencing offered by NGS has led to foregoing the pre-screening and the direct interrogation of mutations in the diagnosis of monogenic genetic diseases. For the detection of germ-line heterozygous variants, it is recommended to use 10-30X as a cut-off value for the read depth of NGS [8]. However, novel variants identified by NGS should be further confirmed by Sanger sequencing before they are reported.

Wilson disease (WD) is an autosomal recessive disorder of copper metabolism. The causative gene, *ATP7B*, is located at 13q14.3, spanning a genomic region of 78 kb and including 21 exons varying from 77 to 1234 bp. Current molecular diagnosis of WD uses Sanger sequencing for direct mutation analysis of the *ATP7B* gene [9]. Without any specific mutational hotspots, more than 300 mutations had been reported in the *ATP7B* gene including its promoter region, 5′ untranslated region (UTR), splice sites and coding exons [10]. While many of the mutations are

* Corresponding author at: Department of Laboratory Medicine, National University Health System, 5 Lower Kent Ridge Road,119074, Singapore .
E-mail address: evelyn_koay@nuhs.edu.sg (E.S.-C. Koay).

http://dx.doi.org/10.1016/j.clinbiochem.2015.10.003
0009-9120/

population- and family-specific, a high proportion of WD patients are compound heterozygotes for two different mutations [10]. Although it has been estimated that the prevalence of WD is between 1:30,000 and 1:100,000 [11], ceruloplasmin-based screening for WD in the East Asian population of Korea and Japan, suggested a frequency as high as 1 in 3000 [12] and 1 in 1500 [13], respectively.

We developed a multiplex PCR-based NGS assay, using the MiSeq platform (Illumina), to identify mutations in the *ATP7B* gene for the molecular diagnosis of WD. The analytical performance of the NGS assay was evaluated against the gold standard Sanger sequencing method using a panel of clinical specimens, cell lines and European Molecular Genetics Quality Networks (EMQN) quality assurance samples.

2. Materials and methods

2.1. Patients

A total of 12 de-identified DNA samples from WD patients in whom the *ATP7B* gene had been previously screened using Sanger sequencing and one DNA sample from a normal control, were analyzed in this study. Genomic DNA was extracted from 3 mL of peripheral blood samples collected in EDTA tubes using the Gentra Puregene Blood Kit (QIAGEN, Hilden, Germany) according to the manufacturer's instructions.

2.2. Cell line DNA and EMQN quality assurance samples

DNA from 2 B-lymphocyte cell lines (NA05258 and NA05761) with known mutations in the *ATP7B* gene (Coriell Institute, Camden, NJ), and 3 DNA samples from the EMQN quality assurance program (Glasgow, Scotland) were analyzed in this study.

2.3. Mutational analysis by Sanger sequencing

Twenty two singleplex PCR reactions were performed to amplify the part of the promoter, the 5′ untranslated region (UTR) and all 21 coding exons of the *ATP7B* gene for all the DNA samples. PCR amplification was performed in a 25 μL reaction consisting of 1X PCR buffer (QIAGEN), 1 mM $MgCl_2$, 0.2 mM of each dNTPs, 0.4 μM each of the forward and reverse primers, 1 unit of HotStarTaq DNA polymerase (QIAGEN), and 100 ng of extracted genomic DNA. Thermal cycling was as follows: 95 °C for 15 min, followed by 40 cycles of 95 °C for 30 s, 59 °C for 30 s, 72 °C for 60 s, and 72 °C for 10 min. All PCR and sequencing primers were synthesized according to published primer sequences [14,15]. One additional sequencing primer (5′-cctgccaggaggctgtggtcaa-3′) was designed to supplement sequencing of exon 2. Gel electrophoresis was performed to confirm the 22 PCR products on a 2% (w/v) agarose gel. The PCR products were extracted from the gel using the QIAquick® gel extraction kit (Qiagen). Cycle sequencing was performed using the BigDye Terminator kit version 3.1 (Applied Biosystems, Austin, TX), according to the manufacturer's instructions. Sequencing products were analyzed on a 3100 Genetic Analyzer (Applied Biosystems). The Assign™ ATF (Conexio Genomics, Fremantle, Australia) software package was used to align the sequences against the NCBI reference sequence, NG_008806.1.

2.4. Mutational analysis by multiplex PCR-based NGS

In this study, Primer3 software [16,17] was used to design all the individual primer sets used in the five multiplex PCR reactions to amplify part of the promoter, the 5′ untranslated region (UTR) and all 21 coding exons of the *ATP7B* gene. Multiplex PCR amplification was performed in a 20 μL reaction consisting of 1X Multiplex PCR Master Mix (QIAGEN), 0.2 μM each of the forward and reverse primers (Table 1), 1X Q solution (QIAGEN), and 50 ng of extracted genomic DNA. The uniform thermal cycling conditions for all five multiplex PCR reactions were as follows: 95 °C for 5 min, followed by 40 cycles of 95 °C for 30 s, 61 °C for 30 s, 72 °C for 90 s, and 72 °C for 10 min. The multiplex PCR products were examined by agarose gel-electrophoresis to confirm successful amplification of individual amplicons. The size distribution of the amplicons ranged from 366 to 1528 bp.

For library generation for each individual sample, 2 μL of PCR amplicons from each of the 5 multiplex reactions was pooled into a single tube and purified using Agencourt AMPure XP Beads (Beckman Coulter, Pasadena, CA), according to manufacturer's instructions. The DNA concentration of the purified amplicon pool was measured using the Qubit dsDNA HS Assay kit on the Qubit 2.0 fluorometer (Life Technologies, Carlsbad, CA) and diluted to 0.2 ng/μL. Indexed paired-end libraries were generated from 5 μL of diluted amplicon pool using Nextera XT DNA Sample Preparation Kit (Illumina, San Diego, CA), according to manufacturer's protocol. Fragment length of the libraries was ascertained using the High Sensitivity DNA kit (Agilent Technologies, Santa Clara, CA) on the 2100 Bioanalyzer (Agilent Technologies). Normalized libraries were subjected to a 300 cycle sequencing run with MiSeq Reagent Nano Kit v2 (Illumina) on the MiSeq (Illumina). Variant call format (VCF) files generated from the MiSeq Reporter v2.3 software were further subjected to filtering and annotation using VariantStudio v2.1 software (Illumina). Integrative Genomics Viewer [18] (IGV version v1.0.0) was used to examine the read counts of the target amplicons and confirm the detected variants. All variants identified by NGS were compared with the Sanger sequencing results.

2.5. Nomenclature and classification for detected variants

All sequence variants are named according to the nomenclature guidelines published by the Human Genome Variation Society (HGVS) [19]. Nucleotide and amino acid numbering were based on GenBank reference sequences NM_000053.3 and NP_000044.2, respectively. All detected variants were searched for in the Wilson Disease Mutation Database (http://www.wilsondisease.med.ualberta.ca/database.asp; accessed on 18 July 2015) [10] and dbSNP (http://www.ncbi.nlm.nih.gov/snp/; accessed on 18 July 2015). All novel variants within coding exons were evaluated for sequence conservation, and the likelihood of pathogenicity using PolyPhen-2 (http://genetics.bwh.harvard.edu/pph2/; accessed on 18 July 2015) [20] and Sorting Intolerant From Tolerant (SIFT) (http://sift.jcvi.org/; accessed on 18 July 2015) [21], respectively.

3. Results

3.1. Coverage and sequencing depth of targets of multiplex PCR-based NGS

The NGS assay interrogates the entire protein coding regions and conserved splice sites of the *ATP7B* gene. The five multiplex PCRs resulted in the generation of 16 amplicons, equivalent to 11,749 bases of the gene sequence per individual sample. The Illumina MiSeq sequencing yielded a total of 0.22 Gbases, with 96.62% of bases having a Phred quality score (Q score) ≥Q30 (99.9% base call accuracy). The run had a cluster density of 519 ± 2 K/mm^2 and approximately 95.15 ± 0.24% of the clusters passed QC filters. Complete coverage (100%) was achieved across 100% of the targeted bases including the promoter, 5'UTR, all coding regions with 10 bp of immediate flanking intronic sequences, without any drop-out sequences detected. The observed read depth (Fig. 1) in all the 20 samples was >100X per base with Q ≥ 30.

3.2. Comparison of disease-causing mutations and variants of unknown significance (VUS) identified by the two methods

To evaluate the performance of the multiplex PCR-based NGS of the *ATP7B* gene, 12 clinical samples, for which Sanger sequencing of the entire *ATP7B* gene was performed, were analyzed. Both NGS and Sanger sequencing resulted in the detection of 14 identical mutations in these samples: 13 known mutations and 1 novel disease-causing mutation (Table 2). The novel deletion, c.1332delC