Blood Cells, Molecules and Diseases 61 (2016) 10–15

Contents lists available at ScienceDirect

Blood Cells, Molecules and Diseases

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

Next-generation sequencing of hereditary hemochromatosis-related genes: Novel likely pathogenic variants found in the Portuguese population

Ricardo Faria [a], Bruno Silva [a], Catarina Silva [a], Pedro Loureiro [a], Ana Queiroz [b], Sofia Fraga [b], Jorge Esteves [c], Diana Mendes [d], Rita Fleming [e], Luís Vieira [a,f], João Gonçalves [a,f], Paula Faustino [a,g,*]

[a] Departamento de Genética Humana, Instituto Nacional de Saúde Doutor Ricardo Jorge, Lisboa, Portugal
[b] Serviço de Pediatria, Hospital Garcia de Orta, Almada, Portugal
[c] Serviço de Gastrenterologia, Hospital de Santo António dos Capuchos, Centro Hospitalar Lisboa Central, Lisboa, Portugal
[d] Serviço de Medicina Transfusional, Hospital São Francisco Xavier, Centro Hospitalar Lisboa Ocidental, Lisboa, Portugal
[e] Serviço de Imuno-Hemoterapia, Hospital de Santa Maria, Centro Hospitalar Lisboa Norte, Lisboa, Portugal
[f] ToxOmics, Faculdade de Ciências Médicas, Lisboa, Portugal
[g] Instituto de Saúde Ambiental, Faculdade de Medicina de Lisboa, Lisboa, Portugal

ARTICLE INFO

Article history:
Submitted 15 April 2016
Revised 20 July 2016
Accepted 21 July 2016
Available online 22 July 2016

Keywords:
Iron metabolism
Hereditary hemochromatosis
Next-generation sequencing
Novel mutations
Genetic variants

ABSTRACT

Hereditary hemochromatosis (HH) is an autosomal recessive disorder characterized by excessive iron absorption resulting in pathologically increased body iron stores. It is typically associated with common *HFE* gene mutation (p.Cys282Tyr and p.His63Asp). However, in Southern European populations up to one third of HH patients do not carry the risk genotypes.

This study aimed to explore the use of next-generation sequencing (NGS) technology to analyse a panel of iron metabolism-related genes (*HFE*, *TFR2*, *HJV*, *HAMP*, *SLC40A1*, and *FTL*) in 87 non-classic HH Portuguese patients.

A total of 1241 genetic alterations were detected corresponding to 53 different variants, 13 of which were not described in the available public databases. Among them, five were predicted to be potentially pathogenic: three novel mutations in *TFR2* [two missense (p.Leu750Pro and p.Ala777Val) and one intronic splicing mutation (c.967-1G>C)], one missense mutation in *HFE* (p.Tyr230Cys), and one mutation in the 5′-UTR of *HAMP* gene (c.-25G>A).

The results reported here illustrate the usefulness of NGS for targeted iron metabolism-related gene panels, as a likely cost-effective approach for molecular genetics diagnosis of non-classic HH patients. Simultaneously, it has contributed to the knowledge of the pathophysiology of those rare iron metabolism-related disorders.

1. Introduction

Hereditary hemochromatosis (HH; OMIM #235200) is an adult-onset autosomal recessive disorder, common among Caucasians of Northern European ancestry, which leads to progressive accumulation of iron in parenchymal cells of multiple organs that can lead, ultimately, to multi-systemic damage, if left untreated. HH is prevalently due to a founder missense mutation, p.Cys282Tyr (c.845G>A) in the *HFE* gene (6p21.3) [1]. This mutation, in the homozygous state, remains the most frequent HH patients' genotype. However, particularly in Southern European populations, around one third of the patients with primary iron overload do not present that genotype or the compound heterozygosity for the p.Cys282Tyr and the p.His63Asp mutations [2]. In fact, in addition to *HFE*, other iron metabolism-related genes may be involved in HH development. The systemic regulation of iron homeostasis is fundamentally achieved through the hepcidin/ferroportin axis. Under homeostasis, hepcidin, a liver-secreted hormone, regulates the efflux of iron from cells through its interaction with ferroportin, the only known cellular iron exporter. While hepcidin control is relatively well known, the regulation of hepcidin expression is a multifaceted mechanism [3]. Consequently, there are different forms of HH resulting from mutations in genes either involved in the hepcidin/ferroportin axis or in the regulation of hepcidin expression: the HH type 2a and 2b (OMIM#602390 and #61333), also known as Juvenile Hemochromatosis (JH), are due to mutation in *HJV* and *HAMP* genes, respectively; HH type 3 (OMIM#604250), which is a disease mainly of adult onset but where some juvenile forms can also be described, is characterized by genetic mutations in the *TFR2* gene, and HH type 4 (OMIM#606069) is due to mutations in the iron exporter ferroportin (*SLC40A1* gene).

* Corresponding author at: Departamento de Genética Humana, Instituto Nacional de Saúde Doutor Ricardo Jorge, Avenida Padre Cruz, 1649-016 Lisboa, Portugal.
E-mail address: paula.faustino@insa.min-saude.pt (P. Faustino).

http://dx.doi.org/10.1016/j.bcmd.2016.07.004
1079-9796/

Furthermore, in addition to the common *HFE* mutations, several rare variants of this gene have also been associated with HH [4–9].

With the advent of massive parallel DNA sequencing, such as the next-generation sequencing (NGS), a number of challenges in the molecular genetics diagnosis of non-classic HH can be addressed. This study aimed to explore the use of a fully customized amplicon-based assay for targeted resequencing by NGS a panel of six HH-related genes in patients presenting clinical and biochemical features compatible with HH but in which the classic HH-associated *HFE* genotypes were not found. This way, several rare and novel variants were detected. Subsequently, *in silico* as well as genotype/phenotype correlation studies have facilitated the interpretation of the likely etiologic significance of some of the previously undescribed variants. Therefore, it was possible to propose the novel variants play a pathogenic role in different types of HH.

The results gathered in this study allowed to conclude that the merging of TruSeq Custom Amplicon (TSCA) methodology and NGS technology can be applied to facilitate the detection of rare variants associated with non-classic HH and, ultimately, to contribute to the understanding of the pathophysiology of those clinical conditions.

2. Materials and methods

2.1. Sample characterization

Patients enrolled in this study were selected by clinicians due to their persistent increased iron status biomarkers, including serum ferritin >300 μg/L and transferrin saturation > 60%, absence of evident environmental risk factors for secondary iron overload (such as, alcohol and hepatitis), and for having a negative first level HH genetic test (which means, absence of homozygosity for the *HFE* p.Cys282Tyr mutation or of compound heterozygosity for p.Cys282Tyr and p.His63Asp).

Following appropriate informed consent, the study was performed in 87 Portuguese patients (69 male and 18 female), presenting a mean age of 51 years and a mean serum levels of iron-related parameters of: iron 185 μg/dL, ferritin 940 μg/L, and transferrin saturation 76%.

Peripheral blood samples were collected in EDTA and used for DNA extraction in a MagNA Pure nucleic acid extractor (Roche). DNA fluorometric quantitation assays were performed in a Qubit™ equipment (Invitrogen) as recommended by the manufacturer. DNA sample concentrations were normalised to 25 ng/μL using purified bidistilled water.

An extra set of five DNA samples previously analysed by Sanger sequencing and known to present a total 43 genetic alterations corresponding to 14 different genetic variants in *HFE*, *TFR2*, *HAMP*, and *SLC40A1* [8] were used as positive controls: *HFE* (rs1799945, rs2071303, rs200706856, rs1572982, and rs1800758); *TFR2* (rs41295912, rs148902192, rs80338885, and rs2075674); *HAMP* (c.-25G>A); *SLC40A1* (rs2304704, rs4287798, rs1156835, and rs13008848).

2.2. Panel design

A fully customizable, amplicon-based assay for targeted resequencing was designed using the Design Studio 1.7 software (Illumina). It is a user-friendly online tool that provides dynamic feedback to optimize design and region coverage. Automated data analysis using the Illumina Amplicon Viewer software allowed to easily reviewing the project data. The panel consisted of ninety-seven amplicons with an average of 250 bp in length, covering a cumulative target sequence of 12,115 bp including coding regions, exon/intron junctions, promoters and untranslated regions (UTRs) of *HFE*, *TFR2*, *HJV*, *HAMP*, *SLC40A1* and *FTL* genes. The location of the designed amplicons and their corresponding length are summarized in Supplementary Table I.

2.3. Next-generation sequencing

TSCA DNA libraries were prepared according to the manufacturer's instructions and sequenced on a MiSeq instrument (Illumina) using paired-end reads. Firstly, a training set of 5 samples (the positive controls) was used to optimize the entire process of target amplification and sequencing. In this way it was verified that no true variant was missed and that no other genetic alterations undetected by Sanger sequencing were observed. Then, the 87 uncharacterized DNA samples were analysed. Sequence variants were identified using the built-in data analysis workflow of the *MiSeqReporter* software (Illumina). Variants were called against the human genome reference hg19 for genomic coordinates contained in the manifest file. Each variant was analysed in respect to the total coverage (TC), quality (QUAL), variant frequency (VF) and genotype (GT). A TC of 20 was used as a threshold for the acceptance of variants. Also an analysis of individual reads was made using the *Integrative Genomics Viewer* software (http://www.broadinstitute.org/igv/). The identified variants classified as likely pathogenic were confirmed by Sanger sequencing using the BigDye Terminator v1.1 Cycle sequencing kit (Applied Biosystems) and a 3130XL Genetic Analyser (Applied Biosystems).

2.4. Bioinformatics studies

Genetic variants were considered novel when not annotated in any of the following public databases: *ENSEMBL* (http: www.ensembl.org/index.html), *dbSNPs* (http://www.ncbi.nlm.nih.gov/), *GeneCards* (http://www.genecards.org/), and *1000 Genomes* (http://www.1000genomes.org).

Novel variants were named according to the *Human Genome Variation Society* (HGVS) recommendations (http://www.hgvs.org/mutnomen/recs.html). Bioinformatics tools were used to predict their deleterious effect. In the case of missense mutations the putative effect on protein structure and function, as well as the study of multiple sequence alignment were analysed using the *Polyphen-2* software (http://genetics.bwh.havard.edu/pph2) and the *HumDiv* and *HumVar* models. *Polyphen-2* prediction is based on a number of features comprising the sequence, phylogenetic and structural information characterizing the substitution. Also, the software *Sorting Intolerant From Tolerant* (SIFT) which predicts whether an amino acid substitution affects protein function based on sequence homology and the physical properties of amino acids, was applied [10]. In what concerns splicing variants their putative effect on the corresponding pre-mRNA splicing was studied using the *Human Splicing Finder, vs 3.0* software (http://www.umd.be/HSF/). The *MaxEntScan* matrice based on the *Maximum Entropy Principle* was also used. It is based on the approach for modelling the sequences of short sequence motifs such as those involved in RNA splicing which simultaneously accounts for non-adjacent as well as adjacent dependencies between positions. This method is based on the most previous probabilistic models of sequence motifs such as weight matrix models and inhomogeneous Markov models [11].

3. Results

3.1. Validation of the amplicon panel for next-generation sequencing studies

A gene panel was designed in order to generate 97 amplicons, including exons, intron/exon junctions, promoters and UTRs of of six genes that we routinely screen by Sanger sequencing: *HFE*, *TFR2*, *HJV*, *HAMP*, *SLC40A1* and *FTL* (Supplementary Table I). The amplified targets of approximately 250 bp were sequenced on the MiSeq equipment (Illumina) and the resultant paired-end reads were analysed through the built-in analysis pipeline. The sequences were aligned against human genome reference hg19 using alignment and variant caller algorithms in the *MiSeqReporter* software. Firstly, five known DNA positive