Biomolecular Detection and Quantification 6 (2016) 22–26

Contents lists available at ScienceDirect

Biomolecular Detection and Quantification

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

Research paper

Targeted resequencing and variant validation using pxlence PCR assays

Frauke Coppieters[a,b], Kimberly Verniers[a], Kim De Leeneer[a], Jo Vandesompele[a,b], Steve Lefever[a,b,*]

[a] Center for Medical Genetics, Ghent University, Ghent, Belgium
[b] pxlence, Dendermonde, Belgium

ARTICLE INFO

Article history:
Received 21 June 2015
Received in revised form 14 August 2015
Accepted 6 September 2015
Available online 9 October 2015

Keywords:
PCR
Next-generation sequencing
Sanger sequencing
Amplification specificity

ABSTRACT

The advent of next-generation sequencing technologies had a profound impact on molecular diagnostics. PCR is a popular method for target enrichment of disease gene panels. Using our proprietary primer-design pipeline, primerXL, we have created almost one million assays covering over 98% of the human exome. Here we describe the assay specification and both *in silico* and wet-lab validation of a selected set of 2294 assays using both next-generation sequencing and Sanger sequencing. Using a universal PCR protocol without optimization, these assays result in high coverage uniformity and limited non-specific coverage. In addition, data indicates a positive correlation between the predictive *in silico* specificity score and the amount of assay non-specific coverage.

1. Introduction

The Online Mendelian Inheritance in Man (OMIM) database currently contains over 4400 human inherited diseases with a known genetic cause [1]. Over 300 new disease genes are being identified yearly, with novel mutations accumulating at a rate of 10,000 per annum [2]. The establishment of a molecular diagnosis in a family confirms the clinical diagnosis, enables reproductive options, and, more recently, is a prerequisite for gene-specific therapies. Indeed, several ongoing clinical trials for human gene and gene-specific therapy emphasize the advent of personalized genomic medicine [3].

Over the past decade, molecular diagnostic testing has faced an exponential growth due to the replacement of laborious gene-by-gene Sanger sequencing by parallel resequencing of multiple genes with massively parallel or so-called next-generation sequencing (NGS) technologies. Various target enrichment strategies are available, enabling the customer to resequence any region of interest. Molecular diagnostic laboratories often develop customized NGS platforms, offering a specific diagnostic portfolio. In addition to gene centric analyses, both exome and genome sequencing are appealing NGS approaches because of the greatly decreased sequencing cost per base [4]. Thus far, these are not yet routinely used in diagnostics because of data quality and ethical reasons, *i.e.* insufficient coverage for relevant genes and incidental findings, respectively. As most genetic centres are accredited, strict regulations are applicable regarding variant reporting [5,6]. Variants identified through NGS generally require confirmation using either NGS or Sanger sequencing. Based on a survey we did in September 2014, almost 70% of 178 respondents from Europe, USA and Asia indicated they are currently validating their NGS findings using either NGS or Sanger sequencing (Fig. 1A). According to this survey, PCR amplification is the most commonly used target enrichment method, followed by hybridization (Fig. 1B).

So far, the currently available NGS enrichment methods for gene panels are hampered by technical limitations. Capture based enrichment for instance often struggles with GC content or repeat rich regions. On the other hand, major advantages of PCR-based enrichment include high flexibility when using singleplex PCR and cost-effectiveness in case of automation or multiplex PCR [7–9]. Of note, PCR based enrichment is sensitive to allelic dropout and/or lower amplification efficiency caused by single nucleotide polymorphisms (SNP) in primer annealing sites and requires more optimization in case of less efficient PCR assays [10]. We recently developed a primer design pipeline for targeted resequencing PCR assays, called primerXL, tackling these issues [11]. PrimerXL makes use of the third-party software packages primer3 v3.2.2 (primer design), UNAFold v3.8 (secondary structures) and Bowtie v0.12.7 (specificity) and includes optimized settings to maximize target

* Corresponding author at: Center for Medical Genetics, Ghent University, Ghent, Belgium.
E-mail address: steve.lefever@ugent.be (S. Lefever).

http://dx.doi.org/10.1016/j.bdq.2015.09.001

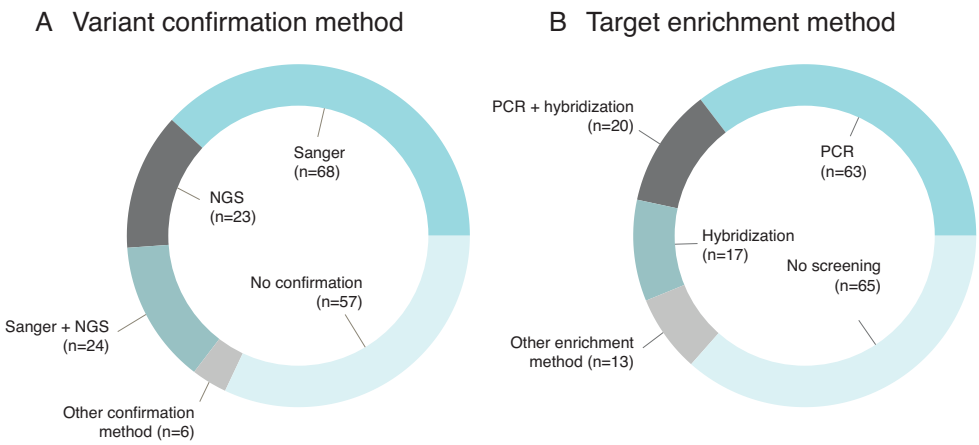

Fig. 1. Distribution of respondents' (n = 178) answers on two survey questions, (A) do you validate your NGS findings and if yes, using what method, (B) do you perform screening experiments, if so how do you enrich your targets?

coverage while minimizing off-target amplification [12–14]. Each of the assay designs is put through a series of stringent *in silico* tests to filter out primer pairs harbouring secondary structures or SNPs in their annealing sites, or having non-specific amplification potential (because of sequence homology to off-target regions). Together, this results in robust PCR assays amplifiable under uniform conditions. Using primerXL, two databases containing almost one million pre-designed sets of assays were developed, each covering ~98% of the exonic regions of all human protein coding genes. The main applications of these assays are the development of gene panels and validation of variants located in exonic regions. A selection of the assays has successfully been used in the development of NGS gene panels for congenital blindness, deafness and cancer [7,15,16]. In addition, these assays were used to replace Sanger-based sequencing with NGS for over 200 genes in an ISO15189-accreditated setting by our diagnostics department [8]. The goal of this study was to further determine the applications for these assays and their overall wet-lab success rate.

2. Material and methods

2.1. (Quantitative) PCR

Each reaction was performed in a 10 μL volume: 5 μL Kapa 2G mastermix, 2.5 μL primers (forward and reverse oligos mixed at a 1 μM concentration) and 2.5 μL DNA template (20 ng/μL). For sample 1, an additional 0.5 μL LCGreen Plus (Bioké) was added since this sample was assessed using qPCR. All (q)PCR reactions were run on a Roche LC480 instrument using the following protocol: (1) 180 s at 95 °C, (2) 15 s at 95 °C, (3) 10 s at 60 °C, (4) 15 s at 72 °C, (5) 60 s at 72 °C. Steps 2–4 were repeated 35 times. For sample 1, this was followed by running a melting curve starting at 65 °C up to 95 °C with 0.5 °C temperature increments each 5 s.

2.2. Library prep and sequencing

Following (q)PCR all reactions were pooled (no normalization was performed). Concentration measurement was performed with the Qubit Fluorometer (Life Technologies). A total of 2.5 μg of the pooled PCR product was used as input for the NEBNext DNA Library Prep Master Mix Set for Illumina (New England BioLabs). During each step, 2 μL was retained to assess the quality of the prep by means of a Bioanalyzer analysis. Both samples were sequenced on a single Illumina MiSeq run (2 × 150 cycles).

3. Results

3.1. Assay specifications

Two assay databases were created with different applications in mind. The first catalogue, with ~320,000 assays having amplicon lengths between 350–750 bp (with 65.2% between 350–450), is optimized for high-quality DNA samples. The second catalogue, suited for fragmented DNA (*e.g.* derived from FFPE samples), contains almost 550,000 short assays (amplicon lengths: 125–275 bp). The latter assays are also ideal candidates for multiplex PCR because of their uniform amplicon lengths. Both databases have been generated to cover all exons of all Ensembl canonical transcripts (Ensembl build 63). The exome coverage for the long and short dataset is 97.99% and 98.71%, respectively. Since then, the *in silico* SNP and specificity analysis was reassessed for the longest amplicons using a more recent genome build (Ensembl build 78). In 94.01% of these assays, no SNPs are present in the primer annealing sites. For the remaining assays, 85.74% contain SNP(s) outside the critical 5 bp 3′ region, whereas 92.47% contain only a single SNP. The *in silico* specificity analysis determined the likelihood of non-specific product generation for each assay. This was done by Bowtie-based alignment of an assay to the human genome (hg38), allowing up to 3 mismatches per primer (3 or more mismatches significantly impede the amplification process), and assigning each assay an *in silico* specificity score equal to the minimal number of mismatches across all predicted off-target hits [10,12]. The higher the resulting specificity level, the more specific the assays are predicted to be. A specificity level of 7 means that there are no non-specific hits, whereas a specificity level of 5 for *e.g.* refers to predicted off-target hits with three mismatches in one primer and two mismatches in the other primer. This analysis revealed that the majority (73.1%) of assays attain the most stringent specificity level (*i.e.* level 7) (Fig. 2). All assays are linked to their specificity level and SNP information, which is displayed to the user upon querying the database.

3.2. Wet-lab assay validation

From the 350–750 bp dataset, 2294 assays covering 169 diagnostically relevant diseases were randomly selected. Using these assays, singleplex amplification was performed on two pooled samples containing male and female DNA.

To assess assay performance and end-point equimolarity, quantitative PCR was used for sample 1, while classical PCR was performed for sample 2. Following amplification using a universal protocol (KAPA 2G Robust–spiked with LCGreen Plus for sample