The Journal of Molecular Diagnostics, Vol. ■, No. ■, ■ 2017

ELSEVIER

the Journal of Molecular Diagnostics

jmd.amjpathol.org

Validation of a Next-Generation Sequencing Pipeline for the Molecular Diagnosis of Multiple Inherited Cancer Predisposing Syndromes

Q25 Paula Paulo,* Pedro Pinto,* Ana Peixoto,† Catarina Santos,† Carla Pinto,† Patrícia Rocha,† Isabel Veiga,† Gabriela Soares,‡ Catarina Machado,§ Fabiana Ramos,¶ and Manuel R. Teixeira*†‖

Q1
Q2
Q3 *From the Cancer Genetics Group,* IPO Porto Research Center, and the Department of Genetics,† Portuguese Oncology Institute of Porto, Porto; the Jacinto de Magalhães Medical Genetics Center,‡ Centro Hospitalar do Porto, Porto; the Department of Genetics,§ Hospital de Santa Maria, Centro Hospitalar Lisboa Norte, Lisboa; the Department of Genetics,¶ Hospital Pediátrico Carmona da Mota, Centro Hospitalar e Universitário de Coimbra, Coimbra; and the Biomedical Sciences Institute,‖ University of Porto, Porto, Portugal*

Address correspondence to Manuel R. Teixeira, M.D., Ph.D., Portuguese Oncology Institute of Porto, Rua Dr. António Bernardino de Almeida, 4200-072 Porto, Portugal. E-mail: manuel.teixeira@ipoporto.min-saude.pt.

Despite the growing knowledge of the genetic background behind the cancers that occur in a context of hereditary predisposition, personal or family cancer history may not be clear enough to support directional gene testing. Defined targeted next-generation sequencing gene panels allow identification of the causative disease mutations of multigene syndromes and differential diagnosis for syndromes with phenotypically overlapping characteristics. Herein, we established a next-generation sequencing analysis pipeline for the molecular diagnosis of multiple inherited cancer predisposing syndromes using the commercially available target sequencing panel TruSight Cancer. To establish the analysis pipeline, we included 22 control samples with deleterious mutations covering all genes currently analyzed at our institution by standard Sanger sequencing. We tested the pipeline using 51 samples from patients with a clinical diagnosis of neurofibromatosis type 1 (NF1), 10 of which without previous molecular characterization of the causative *NF1* mutations. We propose a thoroughly validated analysis pipeline that combines Isaac Enrichment, Burrows-Wheeler Aligner Enrichment, and NextGENe for the alignment and variant calling, and GeneticistAssistant for variant annotation and prioritization. This pipeline allowed the identification of disease-causing mutations in all 73 patients, including a large duplication of 37 bp in *NF1*. We show that high sensitivity and specificity can be achieved by using multiple bioinformatic tools for alignment and variant calling and careful variant filtering, having in mind the clinical question. *(J Mol Diagn 2017, ■: 1–12; http://dx.doi.org/10.1016/j.jmoldx.2017.05.001)*

Q8 Hereditary cancer syndromes may account for 5% to 10% of all cancers, being overrepresented in patients with early onset or family history of the disease.[1] Despite the growing knowledge of the genetic background behind the most common inherited cancer syndromes, personal or family cancer history may not be clear enough to support directional gene testing. Cumbersome molecular diagnosis may be because of genetic heterogeneity (the same syndrome being caused by several genes), the fact that some cancers may be a feature of different predisposition syndromes, and sometimes a striking personal or familial history may not fit any recognizable syndrome, be that because of phenotypic heterogeneity or lack of complete information regarding family history. Moreover, even when clinical diagnosis is straightforward and directional gene testing of a single gene is mandatory, molecular diagnosis may prove difficult with standard sequencing techniques because of the large size of the gene and/or the absence of mutational hot-spots, as

Supported by Portuguese Oncology Institute of Porto Research Center grant CI-IPOP-16-2012, the Portuguese television broadcasting channel TVI (solidary fundraising event), and Fundação para a Ciência e a Tecnologia (FCT) grant PEst-OE/SAU/UI0776/2014. P.Pa. and P.Pi. are research fellows from FCT (UID/DTP/00776/2013/POCI-01-0145-FEDER-006868 and SFRH/BD/73719/2010).

P.Pa. and P.Pi. contributed equally to this work.

Disclosures: None declared.

http://dx.doi.org/10.1016/j.jmoldx.2017.05.001

exemplified by the analysis of the *NF1* gene in patients clinically diagnosed with neurofibromatosis type 1 (NF1).[2]

Targeted next-generation sequencing (NGS) of defined gene packages offers the opportunity to cover the genomic heterogeneity of multigene syndromes, while allowing identification of causative disease mutations in genes associated with phenotypically overlapping syndromes. Some groups have established targeted NGS pipelines for the molecular diagnosis of specific inherited cancer syndromes, covering genes predisposing to hereditary breast/ovarian cancer,[3–8] NF1,[9,10] pheochromocytomas and paragangliomas,[11,12] or retinoblastomas.[13] Others reported NGS pipelines aiming to cover a panel of hereditary cancer syndromes: whereas Kurian et al[14] evaluated a panel of 42 genes associated with breast cancer predisposition in various syndromes, Guan et al[15] described a pipeline covering genes associated with most hereditary cancers, and Judkins et al[16] established a pipeline covering 25 genes involved in the eight most frequent cancers associated with inherited cancer syndromes. These studies, however, either included only control samples for *BRCA1*/*BRCA2* mutations and/or reported HiSeq pipelines that may not be applicable to the more recent, affordable, NGS benchtop equipment. Considering the high-input requirements of high-throughput NGS sequencers to reduce costs per sample, benchtop NGS sequencers may constitute the best alternative for laboratories with average case load numbers.

Herein, we validated a benchtop NGS pipeline for the molecular diagnosis of multiple inherited cancer predisposing syndromes using a commercially available sequencing panel that targets the coding and flanking regions of 94 genes associated with an increased risk for cancer development.

Materials and Methods

Patient Samples

To establish the pipeline, we included control samples for the 22 genes associated with inherited cancer predisposition for which we have deleterious germline mutations at the Department of Genetics of the Portuguese Oncology Institute of Porto (IPO Porto) (detected by standard Sanger sequencing). To test the pipeline, we used two test series, the first (test-I) comprising 10 samples from patients with clinical diagnosis of NF1 at our institution (NIH criteria[17]), but for whom molecular diagnosis was not available, and the second (test-II) comprising 41 DNA samples from patients with clinical and molecular diagnosis (point mutations) of NF1 contributed by three collaborative national Portuguese institutions (16 samples from Centro Hospitalar do Porto, Porto; 19 samples from Hospital de Santa Maria, Lisboa; and 6 samples from Hospital Pediátrico Carmona da Mota, Coimbra).

DNA samples from the 22 controls (control series) and 10 NF1 patients (test-I series) were extracted from peripheral blood leukocytes by standard techniques. DNA samples from the 41 NF1 patients comprising the test series with known *NF1* point mutations (test-II series) were provided by the three collaborating institutions and were blind tested. Samples were quantified using Qubit Fluorometer (Life Technologies, Darmstadt, Germany). To establish the NGS analysis pipeline, the mutational landscape of the control samples was selected to include a large diversity of mutation types (all deleterious mutations): single-nucleotide variants (SNVs) and insertions/deletions (INDELs; short and long) and coding and intronic, encompassing missense, nonsense, splice, and frameshift mutations (Table 1).

Library Preparation and Sequencing

For library preparation, we followed the manufacturer's instructions (TruSight Rapid Capture Kit; Illumina Inc., San Diego, CA). Briefly, 50 ng of DNA was subjected to tagmentation (fragmentation and addition of tags) at 58°C for 10 minutes, followed by capture and cleaning of tagged fragments. Index primers were added and samples were amplified in a 10-cycle PCR. After PCR cleanup, samples were pooled and enriched by hybridization of the TruSight Cancer Content Set Oligos to the specific target regions using an 18-cycle touch-down PCR amplification program, followed by incubation at 58°C for 2 hours. Hybridized probes were captured using streptavidin magnetic beads, washed, and eluted for a second round of hybridization at the same conditions, with a final incubation at 58°C for 15 to 17 hours. After streptavidin-based capture and washing, enriched samples were eluted and purified for a second round of PCR amplification in a 12-cycle program. Libraries were purified, quantified using Qubit HS Assay, and run in QIAxel (Qiagen GmbH, Hilden, Germany) for fragment size evaluation. A final 24-plex pool of 10 pmol/L denatured libraries was loaded in the reagent cartridge (MiSeq Reagent Kit V2; 300-cycle kit) and cluster generation and paired-end sequencing took place in a standard flow cell in a MiSeq platform (Illumina Inc.). Primary analysis was performed in MiSeq Reporter (v2.5.1; Illumina Inc.). As an acceptance quality score, a cutoff value of 30 was considered. Duplicate reads were discarded for further analysis.

Variant Analysis

For alignment to the human genome build 19 and variant calling, we used four different software programs: MiSeq Reporter version 2.5.1 (Illumina Inc.), Isaac Enrichment version 2.1.0, Burrows-Wheeler Aligner (BWA) Enrichment version 2.1.0, and NextGENe version 2.3.4.4 (Softgenetics, State College, PA), according to the manufacturer's protocols. A manifest file, containing the coordinates of the genomic regions targeted by the TruSight Cancer panel, was used for variant calling in the four analysis software programs. For variant annotation and

FLA 5.4.0 DTD ■ JMDI603_proof ■ 18 May 2017 ■ 11:00 am ■ EO: 2017_68