The Journal of Molecular Diagnostics, Vol. 17, No. 4, July 2015

ELSEVIER

the Journal of Molecular Diagnostics
jmd.amjpathol.org

A Variant Detection Pipeline for Inherited Cardiomyopathy—Associated Genes Using Next-Generation Sequencing

Théo G.M. Oliveira,* Miguel Mitne-Neto,† Louise T. Cerdeira,† Julia D.C. Marsiglia,* Edmundo Arteaga-Fernandez,‡ José E. Krieger,* and Alexandre C. Pereira*

From the Laboratory of Genetics and Molecular Cardiology and the Clinical Unit of Cardiomyopathies,‡ Heart Institute (InCor), University of São Paulo, São Paulo; and the Department of Research and Development,† Fleury Group, São Paulo, Brazil*

Accepted for publication February 26, 2015.

Address correspondence to Théo G.M. Oliveira, B.Sc., Av. Dr. Enéas de Carvalho Aguiar 44, 05403-900 São Paulo, SP, Brazil. E-mail: theo.gremen@usp.br.

In inherited cardiomyopathies, genetic testing is recognized as an enriching procedure in the diagnostic closure of a cardiac condition. Many genetic mutations have been described as pathogenically related to cardiomyopathies, turning next-generation sequencing into an extremely reliable scenario. Here we describe the validation process of a pipeline constructed with a target panel of 74 cardiomyopathy-related genes sequenced using a next-generation sequencing system. Fifty-two samples from a hypertrophic cardiomyopathy casuistic with previous molecular diagnostics (Sanger-sequenced for *MYH7*, *MYBCP3*, and *TNNT2*; 19 positives and 33 negatives) were processed in parallel with a HapMap reference sample (NA12878) applied for a complete panel assessment. Sequencing coverage values were satisfactory, with a mean of 250× (95% CI, 226.03—273.91) and 95.2% of target bases with a coverage of ≥10×. With a total of 567 variants, variant call sensitivity was tested in five scenarios of coverage and variant allele frequency cutoffs. Maximum achieved sensitivity was 96.7% for single-nucleotide variants and 28.5% for indels, and positive predictive values remained above 0.959 during the whole process. Inter- and intra-assay reproducibility values were 89.5% and 87.3%, respectively. After a careful assessment of analytical performance, we infer that the assay presents potential feasibility for application in diagnostic routines, with minimal time requirements and a simple bioinformatics structure. *(J Mol Diagn 2015, 17: 420—430; http://dx.doi.org/10.1016/j.jmoldx.2015.02.003)*

Inherited cardiomyopathies (ICs) are a group of cardiac diseases that include hypertrophic cardiomyopathy (HCM), dilated cardiomyopathy, restrictive cardiomyopathy, left ventricular noncompaction, and arrhythmogenic right ventricular cardiomyopathy.[1] These conditions have known genetic etiologies,[2] and the diagnostic procedures may include genetic testing, which, in the context of ICs, has the potential to determine the inheritance pattern of a given condition as well as to support the clarification of overlaps that might exist among all of these cardiac conditions. Additionally, the sooner a mutation is identified, the faster it can be used for the guidance of genetic counseling in affected families.[3]

Advances in DNA sequencing and target enrichment are allowing researchers and clinicians to assess genetic alterations in a more rapid and cost-effective manner, leading to an increasing trend in using molecular diagnostics tools for ICs.[4,5] Although genome and exome sequencing are becoming accessible for genetic investigation, disease-targeted panels are seen as a more feasible alternative for molecular screening regarding inherited diseases. Advantages of delimited panels are not only primarily related to the objectivity of diagnostic interpretation but also rely on their technical advantages, including better sequencing coverage and reliable analytical sensitivity and specificity.[6,7]

Supported by Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES—Brazil Federal Government) and Fleury Group no. 12618 (FINEP).

Disclosures: Fleury Group (São Paulo, Brazil) provided development and research infrastructure for this study.

http://dx.doi.org/10.1016/j.jmoldx.2015.02.003

Table 1 Sequencing Statistics from the 12 Runs Performed in the Experimental Design

Runs	No. of samples (positive/negative)	Throughput (Mb)	Reads	Mean Mb per sample	Mean Q ≥ 20 Mb per sample	Mean coverage
1	5 (2/3)	834	6,211,039	173.0	144.5	252.6
2	5 (1/4)	1000	7,558,913	212.4	186.8	326.1
3	5 (2/3)	885	5,671,445	173.3	156.6	252.3
4	5 (5/0)	1100	7,316,262	235.4	211.8	344.6
5	5 (3/2)	961	6,419,316	191.3	168.2	293.4
6	5 (1/4)	824	5,322,234	163.8	146.7	236.6
7	5 (3/2)	937	7,773,122	181.3	155.7	250.9
8	5 (2/3)	794	5,311,165	157.5	137.2	232
9	5 (0/4 + NA12878_1)*	608	3,929,785	120.4	111.1	184.2
10	5 (0/4 + NA12878_2)*	578	3,675,938	114.1	104.6	170.7
11	5 (0/4 + NA12878_3)*	871	5,512,872	172.1	157.3	258.2
12	5 (2/0 + NA12878_1.1, _2.1 and _3.1)†	704	4,521,446	140.0	123.9	197.6
	Mean	841.3	5,768,628	169.5	150.4	250
	95% CI	753.75–928.92	5,005,008.49–6,532,247.51	153.23–186.34	135.75–164.99	226.03–273.91

*Interassay test using HapMap reference sample NA12878.
†Intra-assay test using HapMap reference sample NA12878, in which the same three replicates were sequenced along with two other repeated HCM-casuistic samples.
Q, Phred value (Q score).

Particularly for ICs, the indication of genetic panels is noticeable in the literature, with the number of interrogated genes ranging between 23[8] and >200.[4] This variability can be attributed to an increasing number of genes indicated as related to cardiomyopathies and the comprehensiveness and objectivity designed for each panel. Attention is being given to dilated cardiomyopathies and HCM,[8,9] the latter presenting the higher prevalence (1:500 population)[2] and a more extensive evidence-based genetic background among all ICs.[10] Both conditions are more commonly related to mutations in sarcomeric genes, although mutations in non-sarcomeric genes are also described.

Beyond its primary scope, cardiomyopathy panels can also include genes for differential diagnostics of other syndromes in which there is cardiac hypertrophy. In cases such as Noonan spectrum disorders, genetic investigation of known related genes along with sarcomeric genes is indicated, especially in cases of severe hypertrophy.[11] Another issue concerns the sequencing of genes related to HCM phenocopies, such as glycogen-storage diseases (eg, amyloidosis, Fabry disease, and Pompe disease)[3,12] as these conditions might present with cardiac hypertrophy but with a genetic etiology different from that of the HCM.

Efforts are being made to guide laboratories and researchers in the implementation of next-generation sequencing (NGS) assays into clinical routine, with best-practice use of these tools, and to validate the practicability of NGS assays.[7,13,14] Of all of the steps involved in the analysis of NGS data, computational pipelines appear as a fundamental procedure, with crucial points that have influence in the final result.[15] For an accurate analysis, the validation process is an important step in defining the use of filters and other parameters of data processing.[16]

Specifically for NGS, guidelines suggest that a set of reference samples previously analyzed with a gold-standard method (ie, Sanger sequencing) be used in the validation process of the newly implemented technology.[7] The confirmation of NGS findings via Sanger sequencing is a laborious and costly practice; the use of reference materials is becoming a widely adopted strategy for the validation of pipelines. A recently published article by the National Institute for Standards and Technology[17] introduced a high-resolution list of variants from a HapMap sample (NA12878) that now can be used as a benchmark for the validation of NGS assays.

Here, we delineate a validation process to test the accuracy of a target panel that includes 74 genes accounting for several genetic classes, such as sarcomeric, Z-disc, desmosomal, cytoskeletal, and calcium homeostasis. Noonan spectrum disorder–related genes (*PTPN11, KRAS, HRAS, SOS, RAF1,* and *SPRED1*) were also included. Additionally, we evaluated genes from glycogen-storage diseases such as amyloidosis, Pompe disease, and Fabry disease (*TTR*, *GAA,* and *GLA,* respectively), for possible differential diagnostics of the HCM. An NGS pipeline was designed for clinical application in the ICs, using a set of molecular positive and negative samples from a previously genotyped casuistic of HCM patients[18] and the reference sample (NA12878) from the National Institute for Standards and Technology as a benchmark.

Materials and Methods

Panel Design

The enrichment of 74 target genes (Supplemental Table S1) was performed with the HaloPlex Target Enrichment System (Agilent Technologies, Santa Clara, CA). First, the probes were designed via a SureDesign web tool version 3.0.1.4 (Agilent Technologies), with genes being indicated by its respective RefSeq symbols. The genome reference used was University of California, Santa Cruz hg19 (*http://hgdownload.soe.ucsc.edu/goldenPath/hg19/bigZips*, last accessed May