Supplement 2.

DNA extraction and sequence analysis

Total genomic DNA was extracted from the saliva samples using a bead-based method. DNA quality and quantity were assessed using electrophoretic methods. After assessment of DNA quality, the qualified genomic DNA sample was randomly fragmented using non-contact, isothermal sonochemistry processing. A sequencing library was prepared by ligating sequencing adapters to both ends of DNA fragments. Sequencing libraries were size-selected with a bead-based method to ensure optimal template size and amplified by polymerase chain reaction. Regions of interest (exons and intronic targets) were targeted using hybridization-based target capture method. The quality of the completed sequencing library was controlled by ensuring the correct template size and quantity and to eliminate the presence of leftover primers and adapter-adapter dimers. Ready sequencing libraries that passed the quality control were sequenced using the Illumina's sequencing-by-synthesis method using paired-end sequencing (150 by 150 bases). Primary data analysis converting images into base calls and associated quality scores was carried out by the sequencing instrument using Illumina's proprietary software, generating CBCL files as the final output.

Base called raw sequencing data was transformed into FASTQ format using Illumina's software (bcl2fastq). Sequence reads of each sample were mapped to the human reference genome (GRCh37/hg19). Burrows-Wheeler Aligner (BWA-MEM) software was used for read alignment. Duplicate read marking, local realignment around indels, base quality score recalibration and variant calling were performed using GATK algorithms (Sentieon) for nuclear DNA. Variant data was annotated using a collection of tools (VcfAnno and VEP) with a variety of public and private variant databases including the Genome Aggregation Database (gnomAD), ClinVar and HGMD. The median sequencing depth and coverage across the target regions for the tested sample were calculated based on MQ0 aligned reads. The sequencing run included in-process reference sample(s) for quality control, which passed our thresholds for sensitivity and specificity. The patient's sample was subjected to thorough quality control measures including assessments for contamination and sample mix-up.