Appendix S1.

METHODS

Whole exome sequencing (WES).

Genomic DNA (2 µg) was fragmented and enriched for human exonic sequences using the Human All Exon V5 Agilent SureSelect kit (Agilent Technologies, USA) using the manufacturers protocol, and sequenced on the Illumina HiSeq 2000 platform using Truseq (v3 Chemistry) (Illumina, USA) to generate 100 base paired-end reads. Fastq files were mapped to the reference human genome (hg19/GRCh37) using the Burrows-Wheeler Aligner (BWA) package (v0.6.2)[1]. Local realignment of the mapped reads around potential insertion/deletion (indel) sites was carried with the Genome Analysis Tool Kit (GATK)[2] v1.6. Duplicate reads were marked using Picard v1.8 and BAM files were sorted and indexed with SAMtools v0.1.18[3]. We filtered the variants for high-quality unknown variants [dbSNP132, the 1,000 Genomes database and ESP6500 dataset (http://evs.gs.washington.edu/EVS/)] in the linkage intervals. Approximately 5.8 GB of sequence data was generated for this sample. A minimum of 92.06% of the targeted exome was covered to a depth of at least 20× coverage, and 92,870 variants were identified, including 8,859 novel changes (defined against dbSNP 132 inclusion). The pathological characteristic of substitutions was analysed by protein prediction tool, PolyPhen-2 (http://genetics.bwh.harvard.edu/pph2/).

---

[1]BWA: Li H. and Durbin R. (2009) Fast and accurate short read alignment with Burrows-Wheeler Transform. Bioinformatics, 25:1754-60. [PMID: 19451168].

[2]GATK: McKenna A, Hanna M, Banks E, Sivachenko A, Cibulskis K, Kernytsky A, Garimella K, Altshuler D, Gabriel S, Daly M, DePristo MA (2010). The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. Genome Res. 20:1297-303. [Pubmed].

[3]http://samtools.sourceforge.net/; Li H.*, Handsaker B.*, Wysoker A., Fennell T., Ruan J., Homer N., Marth G., Abecasis G., Durbin R. and 1000 Genome Project Data Processing Subgroup (2009) The Sequence alignment/map (SAM) format and SAMtools. Bioinformatics, 25, 2078-9. [PMID: 19505943].