Supplementary Methods

1. Whole exome sequencing

(1) Experimental method (DNA QC/library construction/sequencing)

For molecular analysis, we extracted DNA from sections of formalin-fixed, paraffin-embedded tissues. The quality and quantity of purified DNA were assessed by fluorometry (Qubit, Invitrogen) and gel electrophoresis. Briefly, 500 ng of genomic DNA from each sample was fragmented by acoustic shearing on a Covaris S2 instrument. Fragments of 150-300 bp were ligated to Illumina's adapters and PCR-amplified. The samples were concentrated to 300 ng in 3.4 μL DW using a Speedvac (Thermo Scientific) and hybridized with RNA probes (SureSelect XT Human All Exon V5 Capture library) for 16-24 hours at 65℃. After hybridization, the captured targets were pulled down by biotinylated probe/target hybrids using streptavidin-coated magnetic beads (Dynabeads My One Streptavidine T1; Life Technologies Ltd.). The selected regions were then PCR-amplified using Illumina PCR primers. Libraries were quantified using the Agilent 2100 Bioanalyzer (Agilent Technologies) and KAPA Library Quantification Kit (KK4824, Kapa Biosystems). The resulting purified libraries were applied to an Illumina flow cell for cluster generation and sequenced using 100 bp paired-end reads on an Illumina Hiseq 2500 sequencer following the manufacturer's protocols. Image analysis was performed using the HiSeq control Software ver. 1.8.4.

(2) Sequencing quality control

The quality of the reads was checked using FastQC (v.0.10.1) [1], to determine the basic quality for sequence quality score, GC content, N content, length distribution, and duplication level. Low-quality reads below Q20 were trimmed using Cutadapt (v.1.8.1) [2].

(3) Sequence alignment

High quality reads were aligned to the human reference genome hg19 using Burrows Wheeler Aligner (BWA) (v.0.7.12) [3] with minimum seed length of 45. After aligning of the reads to the reference genome, we removed the duplicated reads using MarkDuplicates.jar in Picard Tools (v.1.98).

(4) Variant call and annotation

The suspicious intervals were examined for a more accurate realignment using the

RealignerTargetCreator tool in GenomeAnalysisTK (v.2.3.9) [4] and Mills-and-1000G-gold.standard-INDELs.hg19 was referred to as the known indel set. Base quality score recalibration (BQSR) adjusted the quality score using the BaseRecalibrator tool in GATK. For the realigned and recalibrated reads, variants were called using the UnifiedGenotyper tool in GATK. All variants were annotated using SnpEff (v.4.1) [5] to predict the damaging or clinical effects of the variants using dbNSFP [6], Cosmic [7], and ClinVar [8]. Allele frequencies were determined using the 1000 Genomes [9], ESP6500 [10], and ExAC database [11].

References

1. Leggett RM, Ramirez-Gonzalez RH, Clavijo BJ, Waite D, Davey RP. Sequencing quality assessment tools to enable data-driven informatics for high throughput genomics. Front Genet. 2013;4:288.
2. Martin M. Cutadapt removes adapter sequences from high-throughput sequencing reads. EMBnet J. 2011;17:10-2.
3. Li H, Durbin R. Fast and accurate short read alignment with Burrows-Wheeler transform. Bioinformatics. 2009;25:1754-60.
4. McKenna A, Hanna M, Banks E, Sivachenko A, Cibulskis K, Kernytsky A, et al. The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. Genome Res. 2010;20:1297-303.
5. Cingolani P, Platts A, Wang le L, Coon M, Nguyen T, Wang L, et al. A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff: SNPs in the genome of Drosophila melanogaster strain w1118; iso-2; iso-3. Fly (Austin). 2012;6:80-92.
6. Liu X, Wu C, Li C, Boerwinkle E. dbNSFP v3.0: a one-stop database of functional predictions and annotations for human nonsynonymous and splice-site SNVs. Hum Mutat. 2016;37:235-41.
7. Forbes SA, Bhamra G, Bamford S, Dawson E, Kok C, Clements J, et al. The Catalogue of Somatic Mutations in Cancer (COSMIC). Curr Protoc Hum Genet. 2008;Chapter 10:Unit-10.1.
8. Landrum MJ, Lee JM, Riley GR, Jang W, Rubinstein WS, Church DM, et al. ClinVar: public archive of relationships among sequence variation and human phenotype. Nucleic Acids Res. 2014;42:D980-5.
9. Genomes Project C, Auton A, Brooks LD, Durbin RM, Garrison EP, Kang HM, et al.

A global reference for human genetic variation. Nature. 2015;526:68-74.

10. NHLBI Exome Sequencing Project (ESP) Exome Variant Server [Internet]. St. Louis, MO: Washington University; [cited 2013 Mar 18]; Available from: https://evs.gs.washington.edu/EVS/.

11. Karczewski KJ, Weisburd B, Thomas B, Solomonson M, Ruderfer DM, Kavanagh D, et al. The ExAC browser: displaying reference data information from over 60 000 exomes. Nucleic Acids Res. 2017;45:D840-5.