DNA 抽提和质检

本实验已进行伦理审查并通过（如需要）

全血或组织样本使用 DNeasy Blood & Tissue Kit (QIAGEN) 试剂盒进行抽提；

FFPE 样本使用 QIAamp DNA FFPE tissue kit (QIAGEN)进行抽提。

DNA isolation

Ethical approval was granted by…(if needed)

Blood / Tissue：Genomic DNA was extracted from … using DNeasy Blood & Tissue Kit (QIAGEN)

FFPE： Genomic DNA was extracted from … using QIAamp DNA FFPE tissue kit (QIAGEN)

文库构建和测序

按照 Illumina 测序文库构建流程对基因组 DNA 进行片段化，末端修复、3’末端加 A、连接接头，文库扩增等步骤，完成测序样本文库构建。文库使用 Qubit® 2.0 Fluorometer 检测浓度，并使用 Agilent2100 检测文库的大小。将构建好的文库使用 Illumina Hiseq X 进行 2*150bp 测序。

Library preparations and Sequencing

The genomic DNA was sent to Shanghai Biotechnology Corporation for the de novo wholegenome sequencing. Library was prepared according to the TruSeq DNA Sample Preparation Guide（Illumina，15026486 Rev.C）. The quantity of libraries were assessed by Qubit® 2.0 Fluoromete. The quality and size of libraries were measured by 2100 Bioanalyzer High Sensitivity

DNA Assay according to the reagent kit guide. For Illumina sequencing, the qualified libraries were applied to 2×150 bp paired-end sequencing on Illumina HiSeq X-ten platform (Illumina).

基因组比对

用 BWA 工具把 fastq 文件中的 reads 比对到人的基因组上（hg19/GRCh37 版本），经过 samtools 排序和 picard 标记 duplicate 重复。随后用 GATK 工具进行插入缺失的部位的局部重新比对和碱基质量矫正。最终得到预处理 bam 文件，使用 BEDtools 和 perl/python 脚本统计比对信息，包括覆盖度和测序深度等信息。

Reads Mapping

FASTQ files were aligned to human reference genome (hg19/GRCh37) by BWA v0.7.13. The aligned files (sam/bam format files) were sorted by samtools (1.3) firstly, then duplicates were flagged by using Picard (v2.2.4). By using GATK v3.5, reads were locally realigned and base qualities were recalibrated. Finally, mapping statistics include coverage and depth were generated from recalibrated files by BEDTools (v2.16.1) and in-house perl/python scripts.

变异检测和注释

使用 GATK 的 HaplotypeCaller 检测 SNP 和 INDEL 变异，并根据 GATK 的 VQSR 算法进行假阳性判断，对可信变异标记为 PASS。使用 CNVnator（v0.2.7）和 LUMPY(v0.2.13)分别进行拷贝数变异（Copy Number Variation,

CNV）和结构变异（Structural Variation,SV）的分析。使用 ANNOVAR 工具进行变异位点的注释，包括变异标准命名、人群数据库频率、疾病数据库注释和突变功能性预测。

Variant calling and annotation

Variants (SNVs and indels) were genotyped from recalibrated BAM files using the multi-sample processing mode of the Unified Genotyper tool from the GATK. Then VQSR (Variant Quality Score Recalibration) was used to reduce false positive of variant calling. Copy Number Variation and Structural Variation were identified by CNVnator v0.2.7 and LUMPY v0.2.13, respectively. SNVs and indels were annotated using ANNOVAR software against multiple databases including HGVS variant description, population frequency, disease or phenotype and variant functional prediction.