Supplementary Methods

1. Targeted gene sequencing

For peripheral blood samples, plasma was isolated by centrifugation at 1,600 ×g for 10 minutes, and then at 16,000 ×g for 10 minutes. A QIAamp Circulating Nucleic Acid Kit (Qiagen, Hilden, Germany) was used to extract cell free DNA (cfDNA) from 3 to 5 mL of plasma. Tumor DNA was extracted from 5 to 10 formalin-fixed, paraffin-embedded (FFPE) sections (5 μm thick) using a QIAamp DNA FFPE Tissue Kit (Qiagen), and Blood Genomic DNA Mini Kit (Cwbiotech, Beijing, China) was used to extract Genomic DNA (gDNA) from white blood cells (WBCs). Twenty to 100 ng of cfDNA, 200 to 500 ng of FFPE DNA, or 500 ng of gDNA were then used for library preparation and quantification, guided by the protocols of the KAPA Hyper Prep kit (Roche, Basel, Switzerland). Final libraries were sequenced on Illumina Nextseq500 (PE 75) or Novoseq 6000 (PE 150) sequencers (Illumina, San Diego, CA).

2. Quality control and variant calling

Trimmomatic [1] was used to trim the sequencing adapters from the raw data. The reads were aligned with the human reference genome (hg19) using BWA [2]. Duplicated reads were removed using Picard (http://broadinstitute.github.io/picard/). Mapped reads were realigned to the genome using Genome Analysis Tool Kit [3]. Somatic mutations and germline mutations were called using Mutect2 and GATK's Haplotype Caller [3] with a paired workflow, respectively. Variants were then annotated using ANNOVAR [4] and in-house developed code. An in-house script was used to verify the human identity concordance of paired samples. Somatic copy number alterations were also detected using GATK [3].

3. Germline and somatic variant filtering

With a threshold of minimum coverage of 50× and an allele frequency of over 30%, germline variants called by GATK on WBC samples were filtered initially. Then, variants not on coding regions and synonymous mutations annotated using ANNOVAR [4] were filtered out. Furthermore, variants with a population minor allele frequency over 0.1% (annotated using the ExAC database) were considered less functional. The somatic alterations filtering rules for ctDNA samples were as follows: (1) 10 allele reads support; (2) 1% allele frequency; (3) supporting reads should be below 4 in the WBC control; (4) mutation frequency should be five times higher than that in the WBC control; (5) mutations should not occur more than two times in the panel of normals (PoN); (6) no significant strand bias (GATK parameter FisherStrand (FS) > 60 for single nucleotide polymorphisms and FS > 200 for insertion-deletions (indels). Similar filtering rules were applied for somatic alterations from FFPE samples, except that the allele frequency was required to be over 5% and the alteration frequency was required to be eight times higher than that in the WBC control. Functional filtering removed variants located in non-coding regions and synonymous mutations were removed for

downstream analysis. A log2 ratio more than 0.6 was considered a copy gain. A log2 ratio less than −0.7 was considered a copy loss.

References

1.   Bolger AM, Lohse M, Usadel B. Trimmomatic: a flexible trimmer for Illumina sequence data. Bioinformatics. 2014;30:2114-20.

2.   Li H, Durbin R. Fast and accurate short read alignment with Burrows-Wheeler transform. Bioinformatics. 2009;25:1754-60.

3.   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.

4.   Wang K, Li M, Hakonarson H. ANNOVAR: functional annotation of genetic variants from high-throughput sequencing data. Nucleic Acids Res. 2010;38:e164.