WES library preparation and sequencing

Whole exome sequencing (WES) was performed on tumor tissue samples and matched normal blood controls. Genomic DNAs from FFPE sections or biopsy samples and the whole blood control samples were extracted with QIAamp DNA FFPE Tissue Kit and DNeasy Blood and tissue kit (Qiagen, USA), respectively, and quantified by Qubit 3.0 using the dsDNA HS Assay Kit (ThermoFisher Scientific, USA). Library preparations were performed with KAPA Hyper Prep Kit (KAPA Biosystems, USA). Target enrichment was performed using the xGen Exome Research Panel and Hybridization and Wash Reagents Kit (Integrated DNA Technology, USA) according to manufacturer's protocol. Sequencing was performed on Illumina HiSeq4000 platform using PE150 sequencing chemistry (Illumina, USA).

Mutation calling

Trimmomatic was used for FASTQ file quality control. Leading/trailing low quality (quality reading below 20) or N bases were removed. Paired-end reads were then aligned to the reference human genome (build hg19), using the Burrows-Wheeler Aligner (BWA) with the parameters. PCR deduplication was performed using Picard. GATK3 was utilized for local realignment around indels and base quality score recalibration. Matched tumor and normal sample pairs were first checked to have the same SNP fingerprint using VCF2LR (GeneTalk) and nonmatching samples were removed from analysis. Further, samples with mean dedup depth <30X were removed. Cross-sample contamination was estimated using ContEst (Broad Institute). Briefly, ContEst quantifies contamination in next-generation sequencing data by identifying homozygous non-reference SNPs in the 1000g database and assessing the likelihood of observing alternate alleles at these genomic locations in the sequencing data. Somatic Single Nucleotide Variant (SNV) calling was performed using Mutect and insertion/deletions (INDELs)

were called running Scalpel (scalpel-discovery in –somatic mode). SNVs and INDELs called were further filtered using the following criteria: i) minimum ≥4 variant supporting reads and ≥2% variant allele frequency (VAF) supporting the variant, ii) filtered if present in > 1% population frequency in the 1000g or ExAC database, iii) filtered through an internally collected list of recurrent sequencing errors (≥3 variant reads and ≤20% VAF in at least 30 out of ~2000 normal samples) on the same sequencing platform. Final list of mutations was annotated using vcf2maf (call VEP for annotation). Tumor mutation burden (TMB) was defined as the total number of missense mutations.

Copy number alteration analysis

Copy number alteration (CNA) analysis from sequencing data was performed using CNVKit. Focal level gain and loss were defined as normalized log2 depth ratio ≥ 1 or ≤ -0.6, respectively. Several thresholds for number of genes affected were used to evaluate CNA in corresponding immune pathways.