Supplement 2

Methods for genomic sequencing and analysis

Patient enrollment

5 to 10 mL of peripheral blood was collected from each patient in EDTA-coated tubes (BD Biosciences). Plasma was extracted within 2 hours of blood collection and shipped to the central testing laboratory within 48 hours.

Targeted NGS and Data Processing

Genomic DNA 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. Circulating cell-free DNA (cfDNA) from plasma was extracted using the QIAamp Circulating Nucleic Acid kit (Qiagen). Sequencing libraries were prepared using the KAPA Hyper Prep Kit (KAPA Biosystems) according to manufacturer's instructions for different sample types. Customized xGen lockdown probes (Integrated DNA Technologies) targeting 425 cancer-relevant genes were used for hybridization enrichment. Captured libraries were on-beads PCR amplified with Illumina p5 (5' AAT GAT ACG GCG ACC ACC GA 3') and p7 primers (5' CAA GCA GAA GAC GGC ATA CGA GAT 3') in KAPA HiFi HotStart ReadyMix (KAPA Biosystems), followed by purification using Agencourt AMPure XP beads. Libraries were quantified by qPCR using KAPA Library Quantification kit (KAPA Biosystems). The target-enriched library was then sequenced on the HiSeq4000 NGS platform(Illumina) according to the manufacturer’s instructions. The mean coverage depth was 143X for the whole blood control samples, and 1200X for tumor tissues. For cfDNA samples, the mean coverage sequencing depth was 4000X.

Sequence Alignment and Processing

Base calling was performed on bcl2fastq v2.16.0.10 (Illumina, Inc.) to generate sequence reads in FASTQ format (Illumina 1.8+ encoding). Quality control (QC) was performed with Trimmomatic. High quality reads were mapped to the human genome (hg19, GRCh37 Genome Reference Consortium Human Reference 37) using the BWA aligner 0.7.12 with BWA-MEM algorithm and default parameters to create SAM files. Picard 1.119 was used to convert SAM files to compressed BAM files which were then sorted according to chromosome coordinates. The Genome Analysis Toolkit (GATK, version 3.4-0) was used to locally realign the BAMs files at intervals with indel mismatches and recalibrate base quality scores of reads **in BAM files.**

SNVs / Indels / CNVs Detections

Single nucleotide variants (SNVs) and short insertions/deletions (indels) were identified by VarScan2 2.3.9 with minimum variant allele frequency threshold set at 0.01, and p-value threshold for calling variants set at 0.05 to generate Variant Call Format (VCF) files. All SNVs/indels were annotated with ANNOVAR, and each SNV/indel was manually checked on the Integrative Genomics Viewer (IGV). Copy number variations (CNVs) were detected using in-house-developed software.

Whole exome sequencing (WES) and data analysis

DNA from fresh tumor tissue and whole blood control were fragmented into ~250bp by M220 Focused-ultrasonicator (Covaris), followed by whole genome library preparation using KAPA Hyper Prep Kit (KAPA Biosystems). Exome capture was performed using the xGen Exome

Research Panel (Integrated DNA Technologies). Enriched libraries were sequenced using the Illumina HiSeq 4000 platform to reach the mean coverage depth of ~80X for the normal control and ~250X for the tumor samples.

The same data processing methods were applied for the WES sequencing data. MuTect [1] with default parameters was applied to paired normal and tumor BAM files for identification of somatic single nucleotide variants (SNVs). SNVs in the 1000 Genomes project and dbSNP with frequency>1% were excluded. Small insertions and deletions (indels) were detected using SCALPEL [2]. SNV and indel annotation was performed by ANNOVAR [3] using the hg19 reference genome and 2014 versions of standard databases and functional prediction programs. Gene-level copy ratios were calculated by CNVKit [4]. Using CNVKit algorithm, relative copy-ratios for each exon were calculated by correcting for imbalanced library size, GC bias, sequence repeats and target density.

Neoantigen prediction

HLA typing of the paired peripheral blood and tumor samples was performed from WES data using OptiType, and the overall results were used for further neoantigen prediction. All nonsynonymous mutations and indels were translated into 21-mer peptide sequences using in-house software centered on mutated amino acid. Then, the 21-mer peptide was used to create a 9- to 11-mer peptide via a sliding window approach for prediction of MHC class I binding affinity. Neopredpipe was used to predict the binding strength of mutated peptides to patient-specific HLA alleles. A peptide with predicted binding affinity to any HLA allele with IC50 < 500 nM was selected. If several selected peptides were generated from the same mutation, it was only counted as one neoantigen. Tumor neoantigen burden (TNB) was

determined as the number of all putative neoantigens per megabase of genome.

Intratumoral heterogeneity (ITH) analysis

FACETS (v0.6.1) [5] was used to analyze copy number variations (CNVs). ABSOLUTE (v1.0.6) [6] was used to estimate tumor purities. PyClone (v0.13.1) [7] was then used to calculate cancer cell fractions (CCFs) of mutations and infer tumor clones, utilizing CNV information adjusted with tumor purities. The mutation cluster with the highest CCF in each tumor was defined as the major tumor clone. The cluster with the second highest CCF was also considered clonal if the highest one contained only one mutation. The rest clusters were termed as subclones. As also described by Fang et al. [8], the intratumor heterogeneity (ITH) was evaluated by the proportion of subclonal mutations:

$$ITH = \frac{m_{subclone}}{m_{clone} + m_{subclone}}$$

where mclone and msubclone are the numbers of mutations in clones and subclones, respectively.

Statistical analysis

Comparisons of proportion between groups were done using the Fisher's exact test. For survival analyses, Kaplan–Meier curves were compared using the log-rank test, and hazard ratios (HR) were calculated by Cox proportional hazards model. A two-sided P value of less than 0.01 was considered significant for all tests unless indicated otherwise. All statistical analyses were done in R.