**Supplementary Material**

**DNA extraction and next-generation sequencing (NGS)**

DNA extraction from 12 pieces of tissue sections of the FFPE tumor specimens, 8 ml of cerebrospinal fluid, plasma and peripheral blood lymphocytes from 10 ml of whole blood was performed using the MagPure FFPE DNA Isolation Kit (#IVD3101-TL-06, Magen, China), the Magmax CF DNA Elution Solution EA (#A33602, ThermoFisher, USA), and the Whole Blood DNA kit (#DP601-T5C, Tiangen, Beijing, China), respectively. Targeted gene capture sequencing for the TMB test were performed using the standard protocols mentioned previously. DNA from the peripheral blood lymphocytes was extracted as a normal control. DNA concentration was measured and normalized using the Qubit dsDNA HS Assay Kit or Qubit dsDNA BR Assay Kit (Life Technologies, CA, USA). Genomic DNA was sheared into 150-200-bp fragments with Covaris M220 Focused-ultrasonicatorTM Instrument (Covaris, MA, USA). Fragmented DNA libraries were constructed by a KAPA HTP Library Preparation Kit (Illumina platforms) (KAPA Biosystems, MA, USA) following the manufacturer’s instruction. DNA libraries were sequenced by a custom-designed assay that comprised a hybridization capture-based Genescope panel of 543 genes (Genecast Biotechnology Co., Wuxi) which cover 1.7 Mb of the entire genome. This sequencing was mostly performed using ‘hotspot’ or targeted panels of known cancer-associated genes. This non-uniformity of coverage is mostly local (focused on a given exon) and partly global (focused on some exons across the genome). The paired-end sequencing was performed by Illumina NovaSeq6000.

**Bioinformatics pipeline**

Paired-end reads generated from the NovaSeq6000 platform were sorted, filtered and indexed with SAM tools. The hg19 reference genome was used for read mapping with BWA 0.7.12 (default parameters). To identify somatic SNP and indel mutations, the obtained BAM files from both tumor tissue samples, cerebrospinal fluid, plasma and peripheral blood lymphocytes for each patient were processed for pairwise variant calling using VarScan (v2.4.2) according to the following parameters. i) The minimum coverage for calling somatic variants in the peripheral blood lymphocyte samples was either 8×, or 6× for calling in tumor tissue samples; the *P* value threshold to call a somatic site was 0.05. ii) Variants with <90% strand bias were kept for further study. The generated candidate mutations were annotated using Annovar software tools, and the dbNSFP and Exome Aggregation Consortum (ExAC) database was used to filter out either the benign mutations with pp2\_hdiv score <0.452 or the population polymorphic sites. Finally, the resulting nonsynonymous mutations at the exonic regions were kept. During the software working procedure, three main sources of bias that induce the extraneous variability of the sequencing read depth, which included the GC content, target footprint size and spacing, and the repetitive sequences, were also evaluated and corrected. We used all of the blood cell samples from patients to construct the copy number baseline as the negative control and used the CNV kit to call copy number variation from the FFPE samples, cerebrospinal fluid and plasma for each patient. Target region sequencing induces some poor coverage homogeneity between different enrichment regions due to biases related to the efficiency of target capture and library preparation, which is a limitation for precise copy number detection. We used a software package CNV kit that applies both the sequencing target region reads and non-specifically captured off-target reads to help reduce bias and therefore improved the somatic CNV detection resolution. During the software working procedure, three main sources of bias that induce the extraneous variability of sequencing read depth, which included GC content, target footprint size and spacing, and repetitive sequences, were also evaluated and corrected. Thresholds of copy number ≥2.8 and ≤1.2 were used to categorize altered regions as CNV gains (amplification) and copy number losses (deletions).

**CNI score calculation**

After correction for GC content and target region length using proprietary algorithms for each region, the read counts were transformed into log2 ratios and converted into Z-scores based on Gaussian transformations versus a normal control group (n=30). The target regions that satisfied the Z-score greater than the 95th percentile plus two times the absolute standard deviation of the normal control group were retained, and these Z-score was summed as the CNI score.

**Tumor mutational burden analysis**

The TMB was defined as the number of somatic, coding, base substitutions, and indel mutations per megabases of the genome examined. Only the regions with sequencing depth larger than 100× after deduplication were taken into TMB calculation. All base substitutions and indels in the coding region of the targeted genes, including synonymous alterations, were initially counted before filtering as described above. Synonymous mutations were counted to reduce sampling noise, while non-coding alterations, germline alterations occurring with two or more counts in the ExAC database, alterations that were predicted to be germline by the somatic-germline zygosity algorithm, and any known germline alterations in dbSNP were excluded. To calculate the TMB per megabase, the total number of mutations counted was divided by the size of the coding region of the targeted territory. The patients were stratified into 3 groups (high, moderate and low) according to the TMB level. Cut-off was selected to categorize patients into high and moderate groups (cut-off = median+SD), or moderate and low groups (cut-off = median-SD/2).

**Multiplexed immunofluorescence detection**

The multiplexed immunofluorescence staining was conducted at Genecast Biotechnology Co., Ltd. (Beijing, China). Breifly, 4μm thick section was cut from FFEP lung cancer tissues for each panel detection. The slides were deparaffinized, rehydrated, and subjected to epitope retrieval by boiling in Tris-EDTA buffer (pH=9; Klinipath #643901, Duiven, the Netherlands) for 20 minutes at 97°C. Endogenous peroxidase was then blocked by incubation in Antibody Diluent/Block (PerkinElmer #72424205, Massachusetts, USA) for 10 minutes, and protein was subsequently blocked in 0.05% Tween solution containing 0.3% bovine serum albumin for 30 minutes at room temperature. Only one antigen was detected in each round, including primary antibody incubation, secondary antibody incubation, tyramine signal amplification (TSA) visualization, followed by labeling of the next antibody after epitope retrieval and protein blocking as before.

Primary antibodies for PD-L1 (ZA-0629, clone sp142, Zsbio, 1:25), PD-1 (ZM-0381, clone UMAB199, Zsbio, 1:50), CD57 (ZM-0058, clone NK-1, Zsbio, 1:100), CD68 (ZM-0060, clone KP1, Zsbio, 1:500) were incubated for 1 h at room temperature, and CD8 antibody (ZA-0508, clone SP16, Zsbio, 1:100) was incubated for overnight at 4℃.

Anti-rabbit/mouse horseradish peroxidase (HRP) antibodies (Zsbio # PV-6002 and PV-8000) were used as the secondary antibody and incubated at 37℃ for 10 min. TSA visualization was then performed with the opal seven-color multiplex immunohistochemistry Kit (NEL797B001KT, PerkinElmer, Massachusetts, USA), containing fluorophores (DAPI), Opal 520 (CD8), Opal 570 (PD-L1), Opal 620 (CD57), Opal 650 (CD68), Opal 690 (PD-1) and TSA Coumarin system (NEL703001KT, PerkinElmer, Massachusetts, USA). After labeling for all of the 5 antigens for each panel, microwave treatment (MWT) was performed to remove the TSA-antibody complex with Tris-EDTA buffer (pH=9; Klinipath #643901, Duiven, the Netherlands) for 20 minutes at 97°C. All the slides were counterstained with 4’,6-Diamidino-2-Phenylindole (DAPI) for 5 min and were enclosed in Antifade Mounting Medium (NobleRyder #I0052, Beijing, China), prepared for imaging. Fresh whole-tissue section cuts from normal human tonsil were included in each staining batch as positive control and to assess the interexperimental reproducibility.

Slides were scanned using the PerkinElmer Vectra (Vectra 3.0.5; PerkinElmer, Massachusetts, USA). Multispectral images were unmixed with spectral libraries built from single stained tissue images for each antigen, using the inform Advanced Image Analysis software (inForm 2.3.0; PerkinElmer, Massachusetts, USA).

For batch analysis, an algorithm is acquired by training 10 to 15 representative multispectral images. Then tissue segmentation and cell segmentation were conducted with the algorithm. An experienced pathologist determined appropriate positive threshold X for each biomarker. We defined X, 2X, 3X as the threshold of 1+.2+,3+ respectively. Histochemistry score (H-score) was calculated with the formula of H-score= (3+)%×3+ (2+)%×2+ (1+)%×1.