**Supplementary Materials and Methods**

The cleaned data were then aligned to the reference human genome (hg19) using the Burrows-Wheeler Aligner (version 0.7.12) ([1](#_ENREF_1)). The alignment information was stored in \*.sam format and Samtools software (version 1.8) ([2](#_ENREF_2)) was used to create binary sequence alignment/map files for each sample. Duplicate reads were removed using Picard Tools (version 1.119) ([3](#_ENREF_3)) according to the standard data pipeline from the Broad Institute. Local realignments and base quality score recalibrations were performed using the Genome Analysis Toolkit (version 3.8.1) ([4](#_ENREF_4)).

**Somatic Mutation Calling**

Somatic point mutations (single-nucleotide variants, SNVs) and small insertions and/or deletions (indels) were identified by comparing tumour and matched normal BAM files using Mutect2 (v 4.0.8.0) ([5](#_ENREF_5)) and Strelka (v 2.9.7) ([6](#_ENREF_6)), and only variants that passed both quality filters were used in the follow-up analysis. All analyses were performed with default parameters. SNVs and indels falling within the target regions defined by VariantBaits™ Human All Exon Kit were included in further analyses. The Variant Effect Predictor (VEP) ([7](#_ENREF_7)) tool was utilized to add biological information to the variant set. Variants were removed if present in either the dbSNP or 1000 Genomes database. Variants were filtered to exclude intronic and silent changes and to retain mutations resulting in missense mutations, nonsense mutations, frameshift alterations, or splice site alterations. Candidate variants were further annotated using ANNOVAR ([8](#_ENREF_8))to add location, function, previously reported data and sequencing data supporting the status of the mutation, and the files were converted to Mutation Annotation Format (MAF) files with vcf2maf (<https://github.com/mskcc/vcf2maf>).

**Identification of Somatic Copy Number Alterations (SCNAs)**

Control-FREEC (v 11.5) ([9](#_ENREF_9), [10](#_ENREF_10)) was used to detect copy number variations. The GC content of the sequences was applied to normalize the read distribution, and the normalized distribution of aligned reads in slide windows was employed to calculate the copy number difference between tumour and normal samples.

**Identification of Driver Genes**

MutSigCV (v 1.4.1) ([10](#_ENREF_10)), which identifies significantly mutated genes, was used to detect driver genes. Both somatic point mutations and indels were used for this analysis. The required coverage file and covariate file were downloaded from the MutSigCV website, and the workflow described on the website was applied. MAF files were used as input files. An FDR-adjusted P-value (q-value) cut-off of 0.1 was applied to obtain the driver genes.

**Extraction of Mutation Signatures**

Somatic point mutation signatures were identified with the Bioconductor package deconstructSigs ([11](#_ENREF_11)), which identifies the linear combination of predefined signatures that most accurately reconstruct the mutational profile of a tumour sample. These candidate signatures were compared with COSMIC signatures. Each mutational signature was assigned a calculated weight representing its contribution to the case samples, with a higher weight indicating a greater relative contribution of the signature.

**Prediction of HLA Types and Neoantigens**

The human leukocyte antigen (HLA) types of 40 patients were identified computationally using HLA-HD (v 1.2.1) ([12](#_ENREF_12)) and HLAscan (v 2.1.2) ([13](#_ENREF_13)) with default parameters, and BAM files were input to determine genotypes at four-digit resolution. To predict the number of neoepitopes from nonsynonymous mutations, each nonsynonymous SNV was translated into a 22-mer peptide sequence centred on the mutated amino acid. Subsequently, the 22-mer was separated into 8-11-mers, which is a common length for peptides presented by human MHC class I molecules, via a sliding window to detect MHC class I binding. We then predicted the MHC binding affinity for each peptide, as described previously ([14](#_ENREF_14)). The NetMHCpan 4.0 ([15](#_ENREF_15)) tool was used to determine the binding affinity strength between each mutated peptide and patient-specific HLA allele. To cover most of the potential immune binding interactions, neoantigens with a predicted binding affinity % rank < 0.5 were considered binders.

To verify the accuracy of the predicted neoantigens in this study, we searched for data related to these candidate neoantigens in three neoantigen-related databases: Tumour-Specific Neoantigen Database (TSNAdb, http://biopharm.zju.edu.cn/tsnadb/ ([16](#_ENREF_16))), Immune Epitope Database (IEDB, <http://www.iedb.org/> ([17](#_ENREF_17))) and CTDatabase ([http:/s/www.cta.lncc.br](http://www.cta.lncc.br)) ([18](#_ENREF_18)). The methods for somatic mutation calling, identification of somatic copy number alterations (SCNAs), identification of driver genes, extraction of mutation signatures, prediction of HLA types and neoantigens are shown in supplementary materials.

**Bioinformatics analysis**

The cBioPortal database was used to evaluate relationships between survival outcomes and gene mutation frequency, smoking status, and TMB. The relationships between TIL characteristics and mutations in *TP53* and complement C1q B chain (*C1QB*) were evaluated using the TIMER database (versions 1 and 2).

**Supplementary Result**

**Potential driver genes in patients with LC&TB**

A list of potential driver genes is shown in **Supplementary Table S10**. The cBioPortal for Cancer Genomics database (<https://www.cbioportal.org/>) was used to evaluate whether the potential driver genes were related to clinicopathological characteristics and survival outcomes using data from 12 studies of patients with NSCLC (**Supplementary Table S7**) and one pan-cancer study of 4,075 patients ([19](#_ENREF_19)). The greatest difference in mutation frequency between the LC&TB and LC groups was observed for *TP53*, which was more frequently mutated in the LC group than in the LC&TB group (*P* = 0.003, **Supplementary Figure S9A**). A lollipop plot showing the *TP53* mutation frequencies, hotspots, and affected protein domains based on the WES results is shown in **Supplementary Figure S10A**. The previous studies revealed *TP53* mutation frequencies of 37.9–67.8%. We further analysis the data of the MSKCC study from the cBioPortal and revealed that *TP53* mutations were associated with significantly better outcomes (vs. wild-type *TP53*) (**Supplementary Figure S10 B-C**). There was a clear correlation between *CD274* and *TP53* mutations (*P* < 0.001, **Supplementary Figure S10D**), and a high *TP53* mutation rate was associated with a high TMB and cigarette smoke exposure based on the pan-cancer data from the MSKCC study ([19](#_ENREF_19)). The TIMER analysis revealed that the *TP53* CNV was closely related to macrophage and dendritic cell infiltration (**Supplementary Figure S8C**), while *TP53* mutations were significantly associated with dendritic cell infiltration (**Supplementary Figure S8D**).

The LC&TB group included 4 patients (4/19, 21.1%) harbouring *C1QB* c.274\_311del, while the LC group included no patients harbouring *C1QB* mutations (**Supplementary Figure S11B**). Significant differences in *CIQB* mutations were observed between the LC&O/CTB, LC&ATB, and LC groups (*P* = 0.028). A lollipop plot showing the *C1QB* mutation frequencies is shown in **Supplementary Figure S11C**. Variant Effect Predictor was used to predict the effect of each mutation, which revealed that *C1QB* c.274\_311del was predicted to substantially influence protein function*.* Interestingly, the previous studies revealed a low frequency of *C1QB* mutations in patients with LC (26/3, 673 cases, 0.7%) (**Supplementary Figure S11E)**, and the pan-cancer analysis from the MSKCC study ([19](#_ENREF_19)) revealed no *C1QB* mutations in 350 patients with NSCLC. The present study revealed that *C1QB* mutations in lung adenocarcinoma (LUAD) and lung squamous cell carcinoma (LUSC) were significantly associated with tumour infiltration by macrophages, M2 macrophages, and T follicular helper cells (all *P* < 0.05). Furthermore, *C1QB* mutations were marginally associated with dendritic cell infiltration (*P* = 0.065). Combined data from the 12 previous studies revealed no significant difference in median survival times according to *C1QB* mutation status among patients with NSCLC (mutated: 47.47 months vs. wild-type: 54.34 months, *P* = 0.514) (**Supplementary Figure S11D)**, although *C1QB* mutations were associated with increased exposure to cigarette smoke (*P* < 0.003, **Supplementary Figure S11F**). The mRNA expression of *C1QB* was decreased in LUAD and LUSC (**Supplementary Figure S12A)**, and mRNA expression was positively correlated with NSCLC infiltration of CD4+ T-cells, CD8+ T-cells, macrophages, and dendritic cells (**Supplementary Figure S12B)**. These results suggest that *C1QB* may play an important role in the tumour microenvironment of LC&TB.

[In](javascript:;) [addition](javascript:;), 3 of 4 tested patients (75%) harboured *POU3F3* c.737\_739del and 1 of 4 tested patients (25%) harboured *POU3F3* c.829\_831del. Only 1 patient in the LC group harboured *POU3F3* c.609\_611del. We did not perform additional bioinformatics analysis because there were no significant differences in the *POU3F3* mutation frequencies between the LC&O/CTB, LC&ATB, and LC groups.

**REFERENCES**

1. Li H, Durbin R. Fast and accurate long-read alignment with Burrows-Wheeler transform. Bioinformatics. 2010;26(5):589-95.

2. Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, et al. The Sequence Alignment/Map format and SAMtools. Bioinformatics. 2009;25(16):2078-9.

3. Picard Toolkit. 2019. Broad Institute, GitHub Repository. <http://broadinstitute.github.io/picard/>; Broad Institute.

4. DePristo MA, Banks E, Poplin R, Garimella KV, Maguire JR, Hartl C, et al. A framework for variation discovery and genotyping using next-generation DNA sequencing data. Nat Genet. 2011;43(5):491-8.

5. Cibulskis K, Lawrence MS, Carter SL, Sivachenko A, Jaffe D, Sougnez C, et al. Sensitive detection of somatic point mutations in impure and heterogeneous cancer samples. Nat Biotechnol. 2013;31(3):213-9.

6. Saunders CT, Wong WS, Swamy S, Becq J, Murray LJ, Cheetham RK. Strelka: accurate somatic small-variant calling from sequenced tumor-normal sample pairs. Bioinformatics. 2012;28(14):1811-7.

7. McLaren W, Gil L, Hunt SE, Riat HS, Ritchie GR, Thormann A, et al. The Ensembl Variant Effect Predictor. Genome Biol. 2016;17(1):122.

8. Wang K, Li M, Hakonarson H. ANNOVAR: functional annotation of genetic variants from high-throughput sequencing data. Nucleic Acids Res. 2010;38(16):e164.

9. Boeva V, Popova T, Bleakley K, Chiche P, Cappo J, Schleiermacher G, et al. Control-FREEC: a tool for assessing copy number and allelic content using next-generation sequencing data. Bioinformatics. 2012;28(3):423-5.

10. Lawrence MS, Stojanov P, Polak P, Kryukov GV, Cibulskis K, Sivachenko A, et al. Mutational heterogeneity in cancer and the search for new cancer-associated genes. Nature. 2013;499(7457):214-8.

11. Rosenthal R, McGranahan N, Herrero J, Taylor BS, Swanton C. DeconstructSigs: delineating mutational processes in single tumors distinguishes DNA repair deficiencies and patterns of carcinoma evolution. Genome Biol. 2016;17:31.

12. Kawaguchi S, Higasa K, Shimizu M, Yamada R, Matsuda F. HLA-HD: An accurate HLA typing algorithm for next-generation sequencing data. Hum Mutat. 2017;38(7):788-97.

13. Ka S, Lee S, Hong J, Cho Y, Sung J, Kim HN, et al. HLAscan: genotyping of the HLA region using next-generation sequencing data. BMC Bioinformatics. 2017;18(1):258.

14. Gubin MM, Zhang X, Schuster H, Caron E, Ward JP, Noguchi T, et al. Checkpoint blockade cancer immunotherapy targets tumour-specific mutant antigens. Nature. 2014;515(7528):577-81.

15. Hoof I, Peters B, Sidney J, Pedersen LE, Sette A, Lund O, et al. NetMHCpan, a method for MHC class I binding prediction beyond humans. Immunogenetics. 2009;61(1):1-13.

16. Wu J, Zhao W, Zhou B, Su Z, Gu X, Zhou Z, et al. TSNAdb: A Database for Tumor-specific Neoantigens from Immunogenomics Data Analysis. Genomics Proteomics Bioinformatics. 2018;16(4):276-82.

17. Vita R, Mahajan S, Overton JA, Dhanda SK, Martini S, Cantrell JR, et al. The Immune Epitope Database (IEDB): 2018 update. Nucleic Acids Res. 2019;47(D1):D339-D43.

18. Almeida LG, Sakabe NJ, deOliveira AR, Silva MC, Mundstein AS, Cohen T, et al. CTdatabase: a knowledge-base of high-throughput and curated data on cancer-testis antigens. Nucleic Acids Res. 2009;37(Database issue):D816-9.

19. Samstein RM, Lee CH, Shoushtari AN, Hellmann MD, Shen R, Janjigian YY, et al. Tumor mutational load predicts survival after immunotherapy across multiple cancer types. Nat Genet. 2019;51(2):202-6.

**Supplementary Tables**

**Supplementary Table S1. The clinicopathological characteristics of entire patient cohort (divided into three groups)**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Patient Characteristics** | **LC with Active TB**（N=33） | **LC with Old TB**（N = 26\*） | **LC**（N=28） | ***P*** value |
| **Sex** |  |  |  | 0.113 |
| **Male** | 24 (72.7%) | 24 (92.3%) | 20 (71.4%) |  |
| **Female** | 9 (27.3%) | 2 (7.7%) | 8 (28.6%) |  |
| **Median Age  (Range)** | 58 (36-75) years | 66 (52-79) years | 62 (37-75) years |  |
| **Smoking status** |  |  |  | 0.743 |
| **Non-smoker** | 12 (36.4%) | 7 (26.9%) | 9 (32.1%) |  |
| **Smoker** | 21 (63.6%) | 19 (73.1%) | 19 (67.9%) |  |
| **Histology** |  |  |  | 0.720 |
| **adenocarcinoma** | 17 (51.5%) | 11 (42.3%) | 12 (42.9%) |  |
| **Non-adenocarcinoma** | 16 (48.5%) | 15 (57.7%) | 16 (57.1%) |  |
| **Pathologic T stage** |  |  |  | 0.242 |
| **T1+T2** | 24 (72.7%) | 16 (64.0%) | 22 (84.6%) |  |
| **T3+T4** | 9 (27.3%) | 9 (36.0%) | 4 (15.4%) |  |
| **Lymph node metastasis** |  |  |  | 0.607 |
| **N0** | 14 (42.4%) | 10 (41.7%) | 15 (53.6%) |  |
| **N1-3** | 19 (57.6%) | 14 (58.3%) | 13 (46.4%) |  |
| **Distant** metastasis |  |  |  | 0.813 |
| **M0** | 24 (72.7%) | 20 (80.0%) | 21 (75.0%) |  |
| **M1** | 9 (27.3%) | 5 (20.0%) | 7 (25.0%) |  |
| **Clinical stage** |  |  |  | 0.101 |
| **I+II** | 15 (45.5%) | 7 (28.0%) | 16 (57.1%) |  |
| **III+IV** | 18 (54.5%) | 18 (72.0%) | 12 (42.9%) |  |
|  |  |  |  |  |
| **Differentiation degree** |  |  |  | 0.974 |
| **Poor or undifferentiated** | 11 (50.0%) | 11 (52.4%) | 7 (53.8%) |  |
| **Well-intermediate** | 11 (50.0%) | 10 (47.6%) | 6 (46.2%) |  |
|  |  |  |  |  |
| **EGFR** |  |  |  | 0.283 |
| **Positive** | 1 (7.7%) | 0 (0%) | 6 (21.4%) |  |
| **Negative** | 12 (92.3%) | 6 (100%) | 22 (78.6%) |  |
| **PD-L1** |  |  |  | **0.001** |
| **Positive** | 9 (30.0%) | 4 (15.4%) | 18 (64.3%) |  |
| **Negative** | 21 (70.0%) | 22 (84.6%) | 10 (35.7%) |  |
| **CD3** |  |  |  | **0.004** |
| **Positive** | 4 (16.7%) | 3 (12.5%) | 11 (52.4%) |  |
| **Negative** | 20 (83.3%) | 21 (87.5%) | 10 (47.6%) |  |
| **CD4** |  |  |  | **<0.001** |
| **Positive** | 2 (8.3%) | 1 (4.2%) | 14 (66.7%) |  |
| **Negative** | 22 (91.7%) | 23 (95.8%) | 7 (33.3%) |  |
| **CD8** |  |  |  | **<0.001** |
| **Positive** | 0 (0%) | 3 (12.5%) | 15 (53.6%) |  |
| **Negative** | 24 (100%) | 21 (87.5%) | 13 (46.4%) |  |
| **Venous/lymphatic/Perineural invasion** |  |  |  | 0.149 |
| **No** | 11 (64.7%) | 7 (100%) | 17 (81.0%) |  |
| **Yes** | 6 (35.3%) | 0 (0%) | 4 (19.0%) |  |
| **Adjuvant radiotherapy** |  |  |  | 0.562 |
| **No** | 10 (43.5%) | 11 (57.9%) | 12 (57.1%) |  |
| **Yes** | 13 (56.5%) | 8 (26.7%) | 9 (30.0%) |  |
| **Adjuvant chemotherapy** |  |  |  | 0.268 |
| **No** | 27 (81.8%) | 22 (84.6%) | 19 (67.9%) |  |
| **Yes** | 6 (18.2%) | 4 (15.4%) | 9 (32.1%) |  |
| **Family history of cancer** |  |  |  | 0.187 |
| **No** | 31 (93.9%) | 21 (80.8%) | 22 (78.6%) |  |
| **Yes** | 2 (6.1%) | 5 (19.2%) | 6 (21.4%) |  |

\*NOTE: One of the patients had no staging information.

**Supplementary Table S2. The PD-L1 expression and T lymphocytes infiltrate in tumor foci and tuberculosis foci of 10 patients**

|  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Patients** | **Whether active** | **Tumor** | | | | |  | **Tuberculosis** | | | | |
| **PD-L1 in TC** | **PD-L1 in IC** | **CD3** | **CD4** | **CD8** |  | **PD-L1 in TC** | **PD-L1 in IC** | **CD3** | **CD4** | **CD8** |
| 1 | ATB | 0% | 0% | 0% | 0% | 1% |  | 0% | 0% | NA | 0% | 0% |
| 2 | OTB | 0% | 0% | 5% | 0% | 5% |  | 0% | 0% | 5% | 5% | 0% |
| 3 | OTB | 0% | 0% | 1% | 0% | 5% |  | 0% | 0% | 5% | 5% | 0% |
| 4 | ATB | 0% | 0% | 0% | 0% | 5% |  | 0% | 10% | 10% | 20% | 10% |
| 5 | OTB | 0% | 0% | 10% | 0% | 15% |  | 0% | 0% | 0% | 0% | 0% |
| 6 | ATB | <1% | 40% | 3% | 0% | 15% |  | 0% | 0% | 30% | 30% | 0% |
| 7 | ATB | <1% | 0% | 0% | 0% | 0% |  | <1% | 5% | 20% | 40% | 10% |
| 8 | ATB | <1% | 0% | 0% | 0% | 0% |  | 0% | 55% | 20% | 30% | 15% |
| 9 | ATB | <1% | 0% | 0% | 0% | 0% |  | 0% | 2% | 40% | 25% | 10% |
| 10 | ATB | 0% | 0% | 40% | 1% | 5% |  | 0% | 0% | 5% | 0% | 0% |

**Supplementary Table S3. Univariate analysis of factors that influence overall survival of lung cancer patients with tuberculosis.**

|  |  |
| --- | --- |
| **Factors** | **P value of univariate survival analysis in LC&TB**  (N = 59) |
| **Sex** |  |
| **Male** | 0.450 |
| **Female** |  |
| **Age** |  |
|  | 0.201 |
| **Smoking status** |  |
| **Non-smoker** | 0.669 |
| **Smoker** |  |
| **Histology** |  |
| **adenocarcinoma** | 0.051 |
| **Non-adenocarcinoma** |  |
| **Pathologic T stage** |  |
| **T1+T2** | **0.049** |
| **T3+T4** |  |
| **Lymph node metastasis** |  |
| **N0** | **0.007** |
| **N1-3** |  |
| **Distant metastasis** |  |
| **M0** | 0.138 |
| **M1** |  |
| **Clinical stage** |  |
| **I+II** | **0.006** |
| **III+IV** |  |
| **Differentiation** |  |
| **Poor or undifferentiated** | 0.053 |
| **Well-intermediate** |  |
| **Venous/lymphatic/Perineural invasion** |  |
| **No** | 0.239 |
| **Yes** |  |
| **Adjuvant radiotherapy** |  |
| **No** | 0.586 |
| **Yes** |  |
| **Adjuvant chemotherapy** |  |
| **No** | 0.519 |
| **Yes** |  |
| **Family history of cancer** |  |
| **No** | 0.762 |
| **Yes** |  |
| **EGFR** |  |
| **Positive** | 0.520 |
| **Negative** |  |
| **PD-L1** |  |
| **Positive** | 0.132 |
| **Negative** |  |
| **CD3** |  |
| **Positive** | 0.229 |
| **Negative** |  |
| **CD4** |  |
| **Positive** | 0.140 |
| **Negative** |  |
| **CD8** |  |
| **Positive** | 0.111 |
| **Negative** |  |

**Supplementary Table S4. The significantly difference gene in three groups.**

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| No | Gene | Frequency in LC | Frequency in LC&OTB | Frequency in LC&ATB | Frequency in LC&TB | P-value (LC&ATB vs. LC&OTB vs. LC) | P-value (LC&TB vs. LC) |
| 1 | ZNF208 | 4.76% | 83.33% | 69.23% | 73.68% | 0.000 | 0.000 |
| 2 | HRNR | 4.76% | 66.67% | 53.85% | 57.89% | 0.001 | 0.002 |
| 3 | TP53 | 76.19% | 66.67% | 23.08% | 36.84% | 0.010 | 0.003 |
| 4 | USP17L2 | 0.00% | 50.00% | 38.46% | 42.11% | 0.001 | 0.004 |
| 5 | ZNF626 | 0.00% | 0.00% | 46.15% | 31.58% | 0.001 | 0.007 |
| 6 | IGFN1 | 0.00% | 33.33% | 38.46% | 36.84% | 0.003 | 0.009 |
| 7 | KIF13A | 0.00% | 16.67% | 38.46% | 31.58% | 0.006 | 0.009 |
| 8 | LRRC37A3 | 0.00% | 33.33% | 38.46% | 36.84% | 0.003 | 0.009 |
| 9 | TRIOBP | 0.00% | 16.67% | 38.46% | 31.58% | 0.006 | 0.009 |
| 10 | ZNF729 | 9.52% | 100.00% | 46.15% | 63.16% | 0.000 | 0.009 |
| 11 | LRRC37A2 | 0.00% | 50.00% | 30.77% | 36.84% | 0.002 | 0.009 |
| 12 | ZNF91 | 0.00% | 50.00% | 30.77% | 36.84% | 0.002 | 0.009 |
| 13 | FLG | 19.05% | 50.00% | 76.92% | 68.42% | 0.003 | 0.012 |
| 14 | ZNF486 | 0.00% | 0.00% | 38.46% | 26.32% | 0.003 | 0.018 |
| 15 | ZNF724P | 0.00% | 0.00% | 38.46% | 26.32% | 0.003 | 0.018 |
| 16 | ABCA13 | 0.00% | 16.67% | 30.77% | 26.32% | 0.025 | 0.021 |
| 17 | CTNND2 | 0.00% | 16.67% | 30.77% | 26.32% | 0.025 | 0.021 |
| 18 | NPIPB7 | 0.00% | 16.67% | 30.77% | 26.32% | 0.025 | 0.021 |
| 19 | ZNF117 | 0.00% | 16.67% | 30.77% | 26.32% | 0.025 | 0.021 |
| 20 | ZNF254 | 0.00% | 16.67% | 30.77% | 26.32% | 0.025 | 0.021 |
| 21 | ZNF721 | 0.00% | 16.67% | 30.77% | 26.32% | 0.025 | 0.021 |
| 22 | ZNF93 | 0.00% | 50.00% | 23.08% | 31.58% | 0.004 | 0.021 |
| 23 | RANBP2 | 4.76% | 50.00% | 38.46% | 42.11% | 0.011 | 0.021 |
| 24 | GPR112 | 0.00% | 66.67% | 15.38% | 31.58% | 0.000 | 0.022 |
| 25 | AHNAK | 9.52% | 16.67% | 53.85% | 42.11% | 0.011 | 0.032 |
| 26 | ANKRD36C | 9.52% | 50.00% | 46.15% | 47.37% | 0.021 | 0.034 |
| 27 | ANKRD26 | 0.00% | 0.00% | 30.77% | 21.05% | 0.015 | 0.042 |
| 28 | KIAA1462 | 0.00% | 0.00% | 30.77% | 21.05% | 0.015 | 0.042 |
| 29 | MRC2 | 0.00% | 0.00% | 30.77% | 21.05% | 0.015 | 0.042 |
| 30 | OTOP2 | 0.00% | 0.00% | 30.77% | 21.05% | 0.015 | 0.042 |
| 31 | PCNXL3 | 0.00% | 0.00% | 30.77% | 21.05% | 0.015 | 0.042 |
| 32 | TRANK1 | 0.00% | 0.00% | 30.77% | 21.05% | 0.015 | 0.042 |
| 33 | ZNF141 | 0.00% | 0.00% | 30.77% | 21.05% | 0.015 | 0.042 |
| 34 | ZNF431 | 0.00% | 0.00% | 30.77% | 21.05% | 0.015 | 0.042 |
| 35 | ANKRD18B | 4.76% | 16.67% | 38.46% | 31.58% | 0.037 | 0.045 |
| 36 | FLG2 | 4.76% | 33.33% | 38.46% | 36.84% | 0.030 | 0.045 |
| 37 | MKI67 | 4.76% | 33.33% | 38.46% | 36.84% | 0.030 | 0.045 |
| 38 | NPIPB11 | 4.76% | 33.33% | 38.46% | 36.84% | 0.030 | 0.045 |
| 39 | RGPD3 | 4.76% | 16.67% | 38.46% | 31.58% | 0.037 | 0.045 |
| 40 | ZNF680 | 4.76% | 16.67% | 38.46% | 31.58% | 0.037 | 0.045 |
| 41 | AKAP9 | 0.00% | 16.67% | 23.08% | 21.05% | 0.047 | 0.048 |
| 42 | C5orf42 | 0.00% | 16.67% | 23.08% | 21.05% | 0.047 | 0.048 |
| 43 | CELSR1 | 0.00% | 16.67% | 23.08% | 21.05% | 0.047 | 0.048 |
| 44 | CEP290 | 0.00% | 33.33% | 23.08% | 26.32% | 0.025 | 0.048 |
| 45 | CHD7 | 0.00% | 16.67% | 23.08% | 21.05% | 0.047 | 0.048 |
| 46 | CTB-134H23.2 | 0.00% | 16.67% | 23.08% | 21.05% | 0.047 | 0.048 |
| 47 | HOXA3 | 0.00% | 16.67% | 23.08% | 21.05% | 0.047 | 0.048 |
| 48 | JAK1 | 0.00% | 16.67% | 23.08% | 21.05% | 0.047 | 0.048 |
| 49 | KCNH1 | 0.00% | 16.67% | 23.08% | 21.05% | 0.047 | 0.048 |
| 50 | LAMB2 | 0.00% | 16.67% | 23.08% | 21.05% | 0.047 | 0.048 |
| 51 | MYO3B | 0.00% | 33.33% | 23.08% | 26.32% | 0.025 | 0.048 |
| 52 | NBPF9 | 0.00% | 33.33% | 23.08% | 26.32% | 0.025 | 0.048 |
| 53 | NF1 | 0.00% | 33.33% | 23.08% | 26.32% | 0.025 | 0.048 |
| 54 | NPIPB15 | 0.00% | 16.67% | 23.08% | 21.05% | 0.047 | 0.048 |
| 55 | OR2L5 | 0.00% | 16.67% | 23.08% | 21.05% | 0.047 | 0.048 |
| 56 | RGPD4 | 0.00% | 16.67% | 23.08% | 21.05% | 0.047 | 0.048 |
| 57 | SVIL | 0.00% | 16.67% | 23.08% | 21.05% | 0.047 | 0.048 |
| 58 | TRIM42 | 0.00% | 16.67% | 23.08% | 21.05% | 0.047 | 0.048 |
| 59 | ZNF28 | 0.00% | 33.33% | 23.08% | 26.32% | 0.025 | 0.048 |
| 60 | ZNF678 | 0.00% | 33.33% | 23.08% | 26.32% | 0.025 | 0.048 |
| 61 | ZNF736 | 0.00% | 16.67% | 23.08% | 21.05% | 0.047 | 0.048 |
| 62 | ZNF85 | 0.00% | 16.67% | 23.08% | 21.05% | 0.047 | 0.048 |
| 63 | CACNA1C | 0.00% | 50.00% | 15.38% | 26.32% | 0.005 | 0.048 |
| 64 | DPCR1 | 0.00% | 50.00% | 15.38% | 26.32% | 0.005 | 0.048 |
| 65 | POTEE | 0.00% | 50.00% | 15.38% | 26.32% | 0.005 | 0.048 |
| 66 | RP1L1 | 0.00% | 50.00% | 15.38% | 26.32% | 0.005 | 0.048 |
| 67 | ZNF90 | 0.00% | 50.00% | 15.38% | 26.32% | 0.005 | 0.048 |
| 68 | ZNF107 | 4.76% | 0.00% | 38.46% | 26.32% | 0.023 | 0.085 |
| 69 | ZNF493 | 4.76% | 0.00% | 38.46% | 26.32% | 0.023 | 0.085 |
| 70 | AHNAK2 | 14.29% | 33.33% | 53.85% | 47.37% | 0.041 | 0.085 |
| 71 | BDP1 | 4.76% | 33.33% | 30.77% | 31.58% | 0.049 | 0.093 |
| 72 | HERC1 | 4.76% | 33.33% | 30.77% | 31.58% | 0.049 | 0.093 |
| 73 | ERVW-1 | 4.76% | 50.00% | 23.08% | 31.58% | 0.028 | 0.095 |
| 74 | C1QB | 0.00% | 33.33% | 15.38% | 21.05% | 0.028 | 0.107 |
| 75 | DCHS1 | 0.00% | 33.33% | 15.38% | 21.05% | 0.028 | 0.107 |
| 76 | FAT2 | 0.00% | 33.33% | 15.38% | 21.05% | 0.028 | 0.107 |
| 77 | GPR179 | 0.00% | 33.33% | 15.38% | 21.05% | 0.028 | 0.107 |
| 78 | IGHV3-20 | 0.00% | 33.33% | 15.38% | 21.05% | 0.028 | 0.107 |
| 79 | LRRTM4 | 0.00% | 33.33% | 15.38% | 21.05% | 0.028 | 0.107 |
| 80 | LYST | 0.00% | 33.33% | 15.38% | 21.05% | 0.028 | 0.107 |
| 81 | POLQ | 0.00% | 33.33% | 15.38% | 21.05% | 0.028 | 0.107 |
| 82 | PRR14L | 0.00% | 33.33% | 15.38% | 21.05% | 0.028 | 0.107 |
| 83 | RIMS2 | 0.00% | 33.33% | 15.38% | 21.05% | 0.028 | 0.107 |
| 84 | ZMYND8 | 0.00% | 33.33% | 15.38% | 21.05% | 0.028 | 0.107 |
| 85 | ZNF479 | 0.00% | 33.33% | 15.38% | 21.05% | 0.028 | 0.107 |
| 86 | ZNF595 | 0.00% | 33.33% | 15.38% | 21.05% | 0.028 | 0.107 |
| 87 | ZNF737 | 0.00% | 33.33% | 15.38% | 21.05% | 0.028 | 0.107 |
| 88 | ZNF845 | 0.00% | 33.33% | 15.38% | 21.05% | 0.028 | 0.107 |
| 89 | ZNF92 | 0.00% | 33.33% | 15.38% | 21.05% | 0.028 | 0.107 |
| 90 | KANSL1 | 0.00% | 50.00% | 7.69% | 21.05% | 0.003 | 0.108 |
| 91 | KRAS | 0.00% | 50.00% | 7.69% | 21.05% | 0.003 | 0.108 |
| 92 | PPIAL4G | 0.00% | 50.00% | 7.69% | 21.05% | 0.003 | 0.108 |
| 93 | TENM4 | 0.00% | 50.00% | 7.69% | 21.05% | 0.003 | 0.108 |
| 94 | ZNF429 | 0.00% | 50.00% | 7.69% | 21.05% | 0.003 | 0.108 |
| 95 | ZNF681 | 0.00% | 50.00% | 7.69% | 21.05% | 0.003 | 0.108 |
| 96 | ZNF726 | 0.00% | 50.00% | 7.69% | 21.05% | 0.003 | 0.108 |
| 97 | MUC17 | 28.57% | 83.33% | 61.54% | 68.42% | 0.031 | 0.131 |
| 98 | MGA | 4.76% | 50.00% | 15.38% | 26.32% | 0.031 | 0.185 |
| 99 | ZNF578 | 4.76% | 50.00% | 15.38% | 26.32% | 0.031 | 0.185 |
| 100 | ANO6 | 0.00% | 33.33% | 7.69% | 15.79% | 0.022 | 0.232 |
| 101 | ATP7B | 0.00% | 33.33% | 7.69% | 15.79% | 0.022 | 0.232 |
| 102 | BRWD3 | 0.00% | 33.33% | 7.69% | 15.79% | 0.022 | 0.232 |
| 103 | C4orf51 | 0.00% | 33.33% | 7.69% | 15.79% | 0.022 | 0.232 |
| 104 | COBLL1 | 0.00% | 33.33% | 7.69% | 15.79% | 0.022 | 0.232 |
| 105 | CSGALNACT1 | 0.00% | 33.33% | 7.69% | 15.79% | 0.022 | 0.232 |
| 106 | DCN | 0.00% | 33.33% | 7.69% | 15.79% | 0.022 | 0.232 |
| 107 | FAM221A | 0.00% | 33.33% | 7.69% | 15.79% | 0.022 | 0.232 |
| 108 | IGKV1-16 | 0.00% | 33.33% | 7.69% | 15.79% | 0.022 | 0.232 |
| 109 | IGKV1D-8 | 0.00% | 33.33% | 7.69% | 15.79% | 0.022 | 0.232 |
| 110 | KCNT2 | 0.00% | 33.33% | 7.69% | 15.79% | 0.022 | 0.232 |
| 111 | MAP4K2 | 0.00% | 33.33% | 7.69% | 15.79% | 0.022 | 0.232 |
| 112 | MYH8 | 0.00% | 33.33% | 7.69% | 15.79% | 0.022 | 0.232 |
| 113 | PABPC1L | 0.00% | 33.33% | 7.69% | 15.79% | 0.022 | 0.232 |
| 114 | PJA1 | 0.00% | 33.33% | 7.69% | 15.79% | 0.022 | 0.232 |
| 115 | PPARGC1A | 0.00% | 33.33% | 7.69% | 15.79% | 0.022 | 0.232 |
| 116 | PPARGC1B | 0.00% | 33.33% | 7.69% | 15.79% | 0.022 | 0.232 |
| 117 | RTTN | 0.00% | 33.33% | 7.69% | 15.79% | 0.022 | 0.232 |
| 118 | SALL1 | 0.00% | 33.33% | 7.69% | 15.79% | 0.022 | 0.232 |
| 119 | SIPA1L2 | 0.00% | 33.33% | 7.69% | 15.79% | 0.022 | 0.232 |
| 120 | SORCS3 | 0.00% | 33.33% | 7.69% | 15.79% | 0.022 | 0.232 |
| 121 | STEAP2 | 0.00% | 33.33% | 7.69% | 15.79% | 0.022 | 0.232 |
| 122 | UGGT2 | 0.00% | 33.33% | 7.69% | 15.79% | 0.022 | 0.232 |
| 123 | CASP9 | 0.00% | 50.00% | 0.00% | 15.79% | 0.002 | 0.233 |
| 124 | VPS13D | 4.76% | 50.00% | 7.69% | 21.05% | 0.033 | 0.345 |
| 125 | AAAS | 0.00% | 33.33% | 0.00% | 10.53% | 0.019 | 0.488 |
| 126 | ACE | 0.00% | 33.33% | 0.00% | 10.53% | 0.019 | 0.488 |
| 127 | AHI1 | 0.00% | 33.33% | 0.00% | 10.53% | 0.019 | 0.488 |
| 128 | ANO4 | 0.00% | 33.33% | 0.00% | 10.53% | 0.019 | 0.488 |
| 129 | AQR | 0.00% | 33.33% | 0.00% | 10.53% | 0.019 | 0.488 |
| 130 | BTNL3 | 0.00% | 33.33% | 0.00% | 10.53% | 0.019 | 0.488 |
| 131 | BUD13 | 0.00% | 33.33% | 0.00% | 10.53% | 0.019 | 0.488 |
| 132 | C21orf62 | 0.00% | 33.33% | 0.00% | 10.53% | 0.019 | 0.488 |
| 133 | CCDC138 | 0.00% | 33.33% | 0.00% | 10.53% | 0.019 | 0.488 |
| 134 | CCDC154 | 0.00% | 33.33% | 0.00% | 10.53% | 0.019 | 0.488 |
| 135 | CCDC87 | 0.00% | 33.33% | 0.00% | 10.53% | 0.019 | 0.488 |
| 136 | CCNL1 | 0.00% | 33.33% | 0.00% | 10.53% | 0.019 | 0.488 |
| 137 | CD38 | 0.00% | 33.33% | 0.00% | 10.53% | 0.019 | 0.488 |
| 138 | CENPT | 0.00% | 33.33% | 0.00% | 10.53% | 0.019 | 0.488 |
| 139 | CIC | 0.00% | 33.33% | 0.00% | 10.53% | 0.019 | 0.488 |
| 140 | CLCA4 | 0.00% | 33.33% | 0.00% | 10.53% | 0.019 | 0.488 |
| 141 | CLIP2 | 0.00% | 33.33% | 0.00% | 10.53% | 0.019 | 0.488 |
| 142 | CLN6 | 0.00% | 33.33% | 0.00% | 10.53% | 0.019 | 0.488 |
| 143 | COL17A1 | 0.00% | 33.33% | 0.00% | 10.53% | 0.019 | 0.488 |
| 144 | CRBN | 0.00% | 33.33% | 0.00% | 10.53% | 0.019 | 0.488 |
| 145 | DBN1 | 0.00% | 33.33% | 0.00% | 10.53% | 0.019 | 0.488 |
| 146 | EIF2B3 | 0.00% | 33.33% | 0.00% | 10.53% | 0.019 | 0.488 |
| 147 | FAM214A | 0.00% | 33.33% | 0.00% | 10.53% | 0.019 | 0.488 |
| 148 | FANCI | 0.00% | 33.33% | 0.00% | 10.53% | 0.019 | 0.488 |
| 149 | FLII | 0.00% | 33.33% | 0.00% | 10.53% | 0.019 | 0.488 |
| 150 | GAPVD1 | 0.00% | 33.33% | 0.00% | 10.53% | 0.019 | 0.488 |
| 151 | GSPT1 | 0.00% | 33.33% | 0.00% | 10.53% | 0.019 | 0.488 |
| 152 | HNRNPC | 0.00% | 33.33% | 0.00% | 10.53% | 0.019 | 0.488 |
| 153 | ISPD | 0.00% | 33.33% | 0.00% | 10.53% | 0.019 | 0.488 |
| 154 | LAMC2 | 0.00% | 33.33% | 0.00% | 10.53% | 0.019 | 0.488 |
| 155 | LMBR1 | 0.00% | 33.33% | 0.00% | 10.53% | 0.019 | 0.488 |
| 156 | MACC1 | 0.00% | 33.33% | 0.00% | 10.53% | 0.019 | 0.488 |
| 157 | MCTP2 | 0.00% | 33.33% | 0.00% | 10.53% | 0.019 | 0.488 |
| 158 | MTTP | 0.00% | 33.33% | 0.00% | 10.53% | 0.019 | 0.488 |
| 159 | MUC3A | 0.00% | 33.33% | 0.00% | 10.53% | 0.019 | 0.488 |
| 160 | NALCN | 0.00% | 33.33% | 0.00% | 10.53% | 0.019 | 0.488 |
| 161 | NCDN | 0.00% | 33.33% | 0.00% | 10.53% | 0.019 | 0.488 |
| 162 | NIN | 0.00% | 33.33% | 0.00% | 10.53% | 0.019 | 0.488 |
| 163 | OR2G3 | 0.00% | 33.33% | 0.00% | 10.53% | 0.019 | 0.488 |
| 164 | OR51G1 | 0.00% | 33.33% | 0.00% | 10.53% | 0.019 | 0.488 |
| 165 | OR5K1 | 0.00% | 33.33% | 0.00% | 10.53% | 0.019 | 0.488 |
| 166 | PDCD11 | 0.00% | 33.33% | 0.00% | 10.53% | 0.019 | 0.488 |
| 167 | PIGK | 0.00% | 33.33% | 0.00% | 10.53% | 0.019 | 0.488 |
| 168 | PKP1 | 0.00% | 33.33% | 0.00% | 10.53% | 0.019 | 0.488 |
| 169 | PNLIPRP2 | 0.00% | 33.33% | 0.00% | 10.53% | 0.019 | 0.488 |
| 170 | PRRC2B | 0.00% | 33.33% | 0.00% | 10.53% | 0.019 | 0.488 |
| 171 | PTPRH | 0.00% | 33.33% | 0.00% | 10.53% | 0.019 | 0.488 |
| 172 | RAB3IP | 0.00% | 33.33% | 0.00% | 10.53% | 0.019 | 0.488 |
| 173 | REG3G | 0.00% | 33.33% | 0.00% | 10.53% | 0.019 | 0.488 |
| 174 | REV3L | 0.00% | 33.33% | 0.00% | 10.53% | 0.019 | 0.488 |
| 175 | SDHB | 0.00% | 33.33% | 0.00% | 10.53% | 0.019 | 0.488 |
| 176 | SESN3 | 0.00% | 33.33% | 0.00% | 10.53% | 0.019 | 0.488 |
| 177 | SLC6A16 | 0.00% | 33.33% | 0.00% | 10.53% | 0.019 | 0.488 |
| 178 | SMAD6 | 0.00% | 33.33% | 0.00% | 10.53% | 0.019 | 0.488 |
| 179 | SNTG2 | 0.00% | 33.33% | 0.00% | 10.53% | 0.019 | 0.488 |
| 180 | STK11 | 0.00% | 33.33% | 0.00% | 10.53% | 0.019 | 0.488 |
| 181 | SUZ12 | 0.00% | 33.33% | 0.00% | 10.53% | 0.019 | 0.488 |
| 182 | THSD7B | 0.00% | 33.33% | 0.00% | 10.53% | 0.019 | 0.488 |
| 183 | TIGD1 | 0.00% | 33.33% | 0.00% | 10.53% | 0.019 | 0.488 |
| 184 | TRIM41 | 0.00% | 33.33% | 0.00% | 10.53% | 0.019 | 0.488 |
| 185 | UNC5D | 0.00% | 33.33% | 0.00% | 10.53% | 0.019 | 0.488 |
| 186 | WDFY3 | 0.00% | 33.33% | 0.00% | 10.53% | 0.019 | 0.488 |
| 187 | WDTC1 | 0.00% | 33.33% | 0.00% | 10.53% | 0.019 | 0.488 |
| 188 | ZAN | 0.00% | 33.33% | 0.00% | 10.53% | 0.019 | 0.488 |
| 189 | ZC3H12A | 0.00% | 33.33% | 0.00% | 10.53% | 0.019 | 0.488 |
| 190 | ZC3HC1 | 0.00% | 33.33% | 0.00% | 10.53% | 0.019 | 0.488 |
| 191 | ZFR | 0.00% | 33.33% | 0.00% | 10.53% | 0.019 | 0.488 |
| 192 | ZNF491 | 0.00% | 33.33% | 0.00% | 10.53% | 0.019 | 0.488 |
| 193 | ZNF585A | 0.00% | 33.33% | 0.00% | 10.53% | 0.019 | 0.488 |
| 194 | RP11-166B2.1 | 14.29% | 66.67% | 7.69% | 26.32% | 0.016 | 0.701 |
| 195 | ZNF804A | 14.29% | 50.00% | 0.00% | 15.79% | 0.016 | 1.000 |

**Supplementary Table S5. TMB, HLA types and the number of nonsynomous and neoantigens in 40 patients.**

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **Group** | **TMB** | **Number of nonsynomous** | **The number of neoantigens** | **HLA-A\*** | **HLA-B\*** | **HLA-C\*** |
| LC&ATB | 12.69 | 263 | 877 | 02\*01/02\*07 | 54\*01/46\*01 | 01\*02/01\*03 |
| LC&ATB | 28.24 | 650 | 1518 | 11\*01/02\*07 | 40\*01/57\*01 | 07\*02/03\*04 |
| LC&ATB | 13.72 | 318 | 761 | 02\*01/02\*07 | 46\*01/15\*02 | 01\*02/08\*01 |
| LC&ATB | 26.61 | 673 | 2041 | 01\*01/24\*02 | 40\*01/13\*01 | 06\*02/03\*04 |
| LC&ATB | 26.03 | 698 | 1924 | 24\*02/02\*01 | 54\*01/51\*01 | 14\*02/01\*02 |
| LC&ATB | 16.77 | 438 | 1060 | 24\*02/23\*01 | 54\*01/45\*01 | 01\*02/06\*02 |
| LC&ATB | 15.32 | 373 | 1203 | 11\*01/11\*02 | 15\*18/40\*01 | 07\*02/08\*01 |
| LC&ATB | 22.5 | 540 | 1107 | 11\*01/26\*01 | 46\*01/40\*02 | 07\*02/01\*02 |
| LC&ATB | 10.63 | 239 | 1743 | 02\*01/02\*07 | 55\*02/15\*01 | 01\*02/04\*01 |
| LC&ATB | 24.71 | 563 | 998 | 11\*01/02\*01 | 46\*01/40\*01 | 01\*02/07\*02 |
| LC&ATB | 10.07 | 200 | 672 | 24\*02/02\*07 | 46\*01/54\*01 | 01\*02/07\*02 |
| LC&ATB | 25.78 | 583 | 609 | 24\*02/11\*01 | 55\*02/15\*27 | 03\*03/04\*01 |
| LC&ATB | 9.33 | 204 | 386 | 02\*07/02\*01 | 46\*01/39\*05 | 07\*02/01\*02 |
| LC&OTB | 18.42 | 480 | 877 | 02\*03/02\*01 | 40\*01/13\*01 | 03\*04/01\*02 |
| LC&OTB | 15.98 | 366 | 803 | 02\*06/02\*01 | 40\*01/57\*01 | 03\*02/03\*04 |
| LC&OTB | 30.57 | 806 | 1665 | 02\*01/02\*07 | 46\*01/51\*01 | 14\*02/01\*02 |
| LC&OTB | 23.3 | 516 | 725 | 24\*02/02\*01 | 40\*02/40\*01 | 03\*03/15\*02 |
| LC&OTB | 7.5 | 174 | 623 | 11\*01/02\*07 | 46\*01/40\*01 | 01\*02/07\*02 |
| LC&OTB | 17.39 | 406 | 637 | 24\*02/11\*01 | 15\*01/40\*02 | 03\*04/04\*01 |
| LC | 10.9 | 282 | 403 | 11\*01/33\*03 | 58\*01/15\*27 | 03\*02/04\*01 |
| LC | 8.66 | 195 | 178 | 24\*02/11\*01 | 07\*02/27\*04 | 07\*02/12\*02 |
| LC | 15.5 | 288 | 438 | 24\*02/11\*01 | 48\*01/40\*01 | 03\*04/08\*01 |
| LC | 19.12 | 481 | 631 | 11\*01/33\*03 | 39\*05/58\*01 | 07\*02/03\*02 |
| LC | 10.67 | 216 | 231 | 24\*02/02\*01 | 39\*01/51\*02 | 07\*02/15\*02 |
| LC | 8.34 | 170 | 181 | 11\*01/02\*01 | 55\*02/40\*02 | 07\*02/03\*04 |
| LC | 10.96 | 189 | 365 | 11\*01/24\*02 | 15\*02/51\*01 | 15\*02/08\*01 |
| LC | 18.71 | 462 | 2279 | 02\*03/02\*07 | 48\*01/46\*01 | 01\*02/08\*01 |
| LC | 3.13 | 49 | 13 | 24\*02/11\*01 | 40\*01/46\*01 | 07\*02/01\*02 |
| LC | 2.67 | 51 | 288 | 11\*01/02\*01 | 46\*01/15\*01 | 01\*02/04\*01 |
| LC | 12.91 | 259 | 205 | 33\*03/02\*07 | 46\*01/44\*03 | 01\*02/14\*03 |
| LC | 25.08 | 463 | 429 | 24\*02/11\*01 | 35\*05/15\*27 | 04\*01/07\*02 |
| LC | 7.98 | 169 | 190 | 02\*06/02\*07 | 39\*01/46\*01 | 07\*02/01\*02 |
| LC | 1.99 | 19 | 0 | 24\*02/30\*01 | 13\*02/40\*01 | 06\*02/01\*02 |
| LC | 5.67 | 129 | 69 | 02\*01/02\*07 | 46\*01/40\*01 | 07\*66/01\*02 |
| LC | 11.62 | 276 | 253 | 02\*06/24\*02 | 40\*01/35\*01 | 03\*04/03\*03 |
| LC | 10.77 | 249 | 312 | 02\*01/02\*07 | 46\*01/15\*11 | 01\*02/03\*03 |
| LC | 7.58 | 182 | 486 | 11\*01/24\*02 | 54\*01/55\*02 | 03\*03/01\*02 |
| LC | 9.25 | 202 | 562 | 02\*03/24\*02 | 40\*01/38\*02 | 07\*02/04\*01 |
| LC | 11.8 | 299 | 416 | 24\*02/11\*01 | 54\*01/15\*01 | 01\*02/04\*01 |
| LC | 6.17 | 132 | 372 | 11\*01/02\*07 | 46\*01/15\*27 | 01\*02/04\*01 |

**Supplementary Table S5. Driver gene and high frequency genes worth studying in LC&TB and LC group**

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **NO** | **Gene** | **Class** | **Significantly different genes** | **Frequency in LC&ATB** | **Frequency in LC&OTB** | **Frequency in LC** | **Frequency in LC&TB** | **P-value (LC&ATB vs. LC&OTB vs. LC)** | **P-value (LC&TB vs. LC)** |
| 1 | TP53 | DG3/SG3 | Yes | 23.08% | 66.67% | 76.19% | 36.84% | 0.010 | 0.003 |
| 2 | AGAP3 | DG3/SG3 |  | 7.69% | 16.67% | 33.33% | 10.53% | 0.232 | 0.130 |
| 3 | ANTXR2 | DG3 |  | 0.00% | 0.00% | 19.05% | 0.00% | 0.192 | 0.108 |
| 4 | NFE2L2 | DG3/SG3 | Yes | 7.69% | 0.00% | 33.33% | 5.26% | 0.090 | 0.046 |
| 5 | DEAF1 | DG3 |  | 0.00% | 16.67% | 19.05% | 5.26% | 0.209 | 0.343 |
| 6 | RB1 | DG3 |  | 7.69% | 16.67% | 9.52% | 10.53% | 0.811 | 1.000 |
| 7 | ABHD17A | DG3 |  | 0.00% | 16.67% | 19.05% | 5.26% | 0.209 | 0.343 |
| 8 | C1QB | DG12 | Yes | 15.38% | 33.33% | 0.00% | 21.05% | 0.028 | 0.107 |
| 9 | CDKN2A | DG12 |  | 7.69% | 16.67% | 9.52% | 10.53% | 0.810 | 1.000 |
| 10 | POU3F3 | DG12 |  | 15.38% | 33.33% | 4.76% | 21.05% | 0.154 | 0.343 |
| 11 | RANBP2 | DG12 | Yes | 38.46% | 50.00% | 4.76% | 42.11% | 0.011 | 0.021 |
| 12 | USP17L2 | DG12 | Yes | 38.46% | 50.00% | 0.00% | 42.11% | 0.001 | 0.004 |
| 13 | ZNF208 | SG12 | Yes | 69.23% | 83.33% | 4.76% | 73.68% | 0.000 | 0.000 |
| 14 | FLG | SG12 | Yes | 76.92% | 50.00% | 19.05% | 68.42% | 0.003 | 0.012 |
| 15 | MUC17 | SG12 | Yes | 61.54% | 83.33% | 28.57% | 68.42% | 0.031 | 0.131 |
| 16 | ZNF729 | SG12 | Yes | 46.15% | 100.00% | 9.52% | 63.16% | 0.000 | 0.009 |
| 17 | HRNR | SG12 | Yes | 53.85% | 66.67% | 4.76% | 57.89% | 0.001 | 0.002 |
| 18 | AHNAK2 | SG12 | Yes | 53.85% | 33.33% | 14.29% | 47.37% | 0.041 | 0.086 |
| 19 | ANKRD36C | SG12 | Yes | 46.15% | 50.00% | 9.52% | 47.37% | 0.021 | 0.034 |
| 20 | MUC16 | SG12 |  | 46.15% | 50.00% | 14.29% | 47.37% | 0.056 | 0.088 |
| 21 | AHNAK | SG12 |  | 53.85% | 16.67% | 9.52% | 42.11% | 0.011 | 0.033 |
| 22 | KMT2C | SG12 |  | 38.46% | 50.00% | 19.05% | 42.11% | 0.289 | 0.310 |
| 23 | MUC12 | SG123 |  | 46.15% | 83.33% | 33.33% | 57.89% | 0.105 | 0.543 |
| 24 | TTN | SG123 |  | 69.23% | 33.33% | 66.67% | 57.89% | 0.296 | 0.530 |
| 25 | RYR2 | SG123 |  | 46.15% | 66.67% | 42.86% | 52.63% | 0.624 | 1.000 |
| 26 | MUC4 | SG123 |  | 46.15% | 50.00% | 38.10% | 47.37% | 0.830 | 1.000 |
| 27 | MUC19 | SG123 |  | 38.46% | 50.00% | 33.33% | 42.11% | 0.824 | 1.000 |
| 28 | KMT2D | SG3 |  | 23.08% | 0.00% | 38.10% | 15.79% | 0.183 | 0.163 |
| 29 | CSMD3 | SG3 |  | 23.08% | 16.67% | 33.33% | 21.05% | 0.704 | 0.484 |
| 30 | DNAH5 | SG3 |  | 7.69% | 16.67% | 33.33% | 10.53% | 0.232 | 0.130 |
| 31 | RELN | SG3 |  | 23.08% | 16.67% | 33.33% | 21.05% | 0.704 | 0.484 |
| 32 | USH2A | SG3 |  | 30.77% | 33.33% | 33.33% | 31.58% | 1.000 | 1.000 |
| 33 | ZFHX4 | SG3 |  | 30.77% | 16.67% | 33.33% | 26.32% | 0.896 | 0.734 |
| 34 | C1orf173 | SG3 |  | 0.00% | 16.67% | 28.57% | 5.26% | 0.098 | 0.093 |

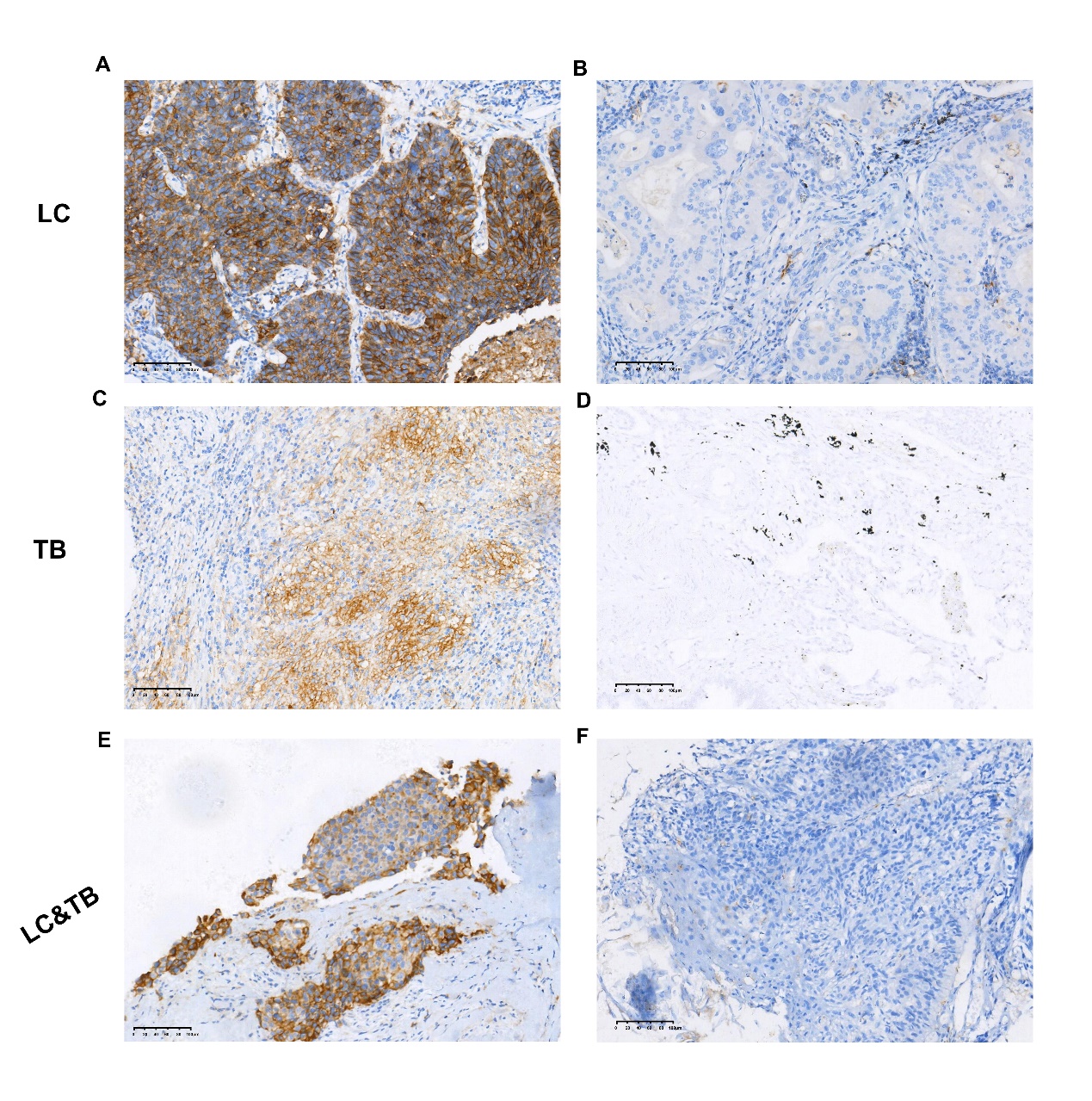
Note: DG3: Driver gene of LC; DG12: Driver gene of LC&TB; SG3: High frequency genes; SG12: High frequency genes of LC&TB; SG123: High frequency genes of LC&TB and LC.

**Supplementary Table S7. 12 studies selected in cBioPortal database**

|  |  |  |
| --- | --- | --- |
| **NO** | **Study name** | **PMID** |
| 1 | [Non-Small Cell Lung Cancer (MSK, Cancer Cell 2018)](https://www.cbioportal.org/study?id=nsclc_mskcc_2018) | 29657128 |
| 2 | [Non-Small Cell Lung Cancer (MSKCC, J Clin Oncol 2018)](https://www.cbioportal.org/study?id=nsclc_pd1_msk_2018) | 29337640 |
| 3 | [Non-Small Cell Lung Cancer (TRACERx, NEJM & Nature 2017)](https://www.cbioportal.org/study?id=nsclc_tracerx_2017) | 28445112 |
| 4 | [Non-Small Cell Lung Cancer (University of Turin, Lung Cancer 2017)](https://www.cbioportal.org/study?id=nsclc_unito_2016) | 27346245 |
| 5 | [Non-small cell lung cancer (MSK, Science 2015)](https://www.cbioportal.org/study?id=nsclc_mskcc_2015) | 25765070 |
| 6 | [Pan-Lung Cancer (TCGA, Nat Genet 2016)](https://www.cbioportal.org/study?id=nsclc_tcga_broad_2016) | 27158780 |
| 7 | [Lung Adenocarcinoma (Broad, Cell 2012)](https://www.cbioportal.org/study?id=luad_broad) | 22980975 |
| 8 | [Lung Adenocarcinoma (MSKCC, Science 2015)](https://www.cbioportal.org/study?id=luad_mskcc_2015) | 25765070 |
| 9 | [Lung Adenocarcinoma (OncoSG, Nat Genet 2020)](https://www.cbioportal.org/study?id=luad_oncosg_2020) | 32015526 |
| 10 | [Lung Adenocarcinoma (TCGA, PanCancer Atlas)](https://www.cbioportal.org/study?id=luad_tcga_pan_can_atlas_2018) | 18948947 |
| 11 | [Lung Adenocarcinoma (TSP, Nature 2008)](https://www.cbioportal.org/study?id=luad_tsp) | 28336552 |
| 12 | [Non-Small Cell Cancer (MSKCC, Cancer Discov 2017)](https://www.cbioportal.org/study?id=lung_msk_2017) | 29625055 |

**Supplementary Figures**

**Supplementary Figure S1. Immunohistochemical staining showing positive PD-L1 expression (A, C, and E) and negative expression (B, D, and F) in tissues from patients with lung cancer and tuberculosis (original magnification, ×200).**A tumour proportion score (TPS) of >25% was used to identify positive expression of PD-L1.

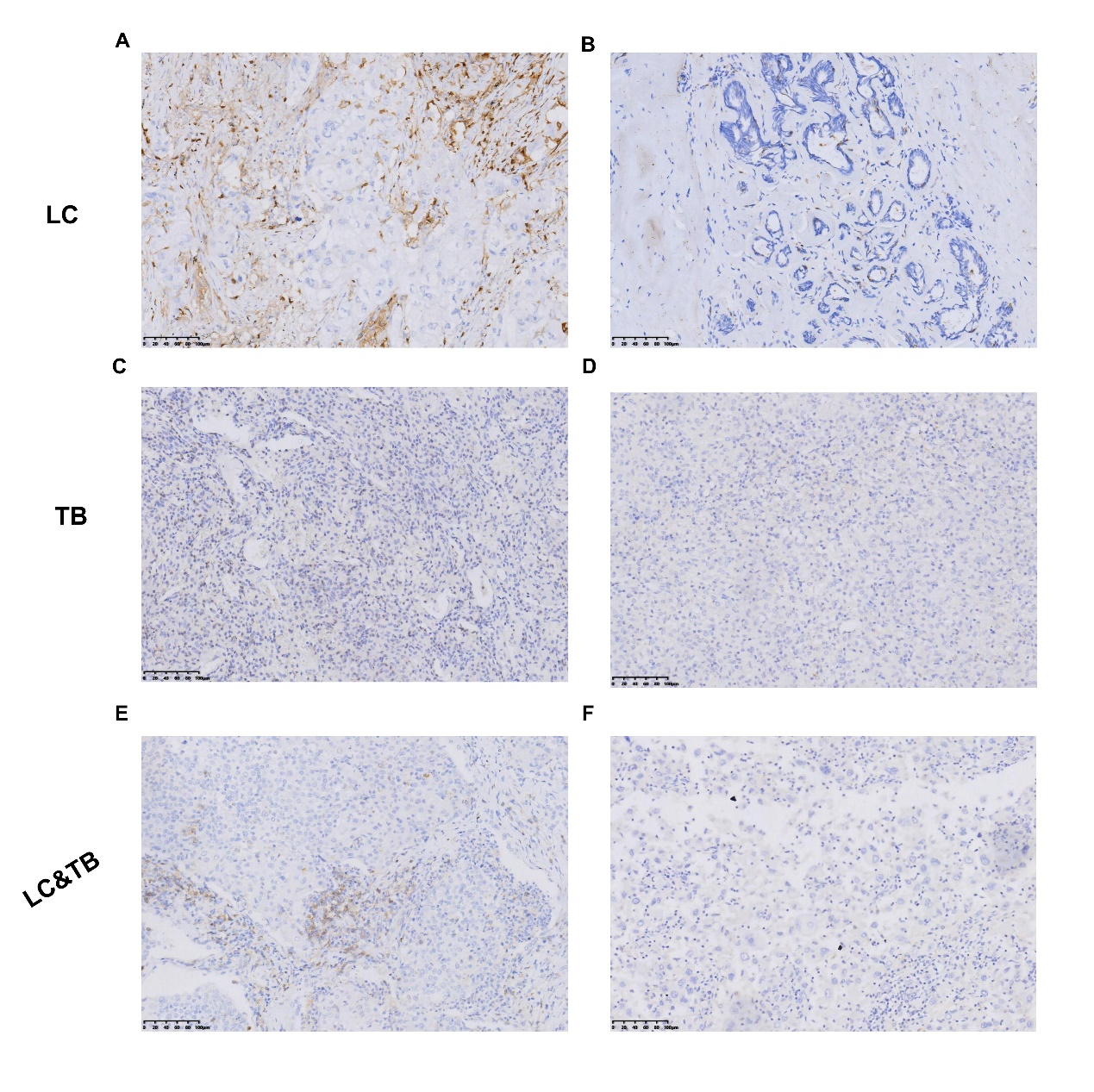


**Supplementary Figure S2. Immunohistochemical staining showing positive CD3 expression (A, C, and E) and negative expression (B, D, and F) in tissues from patients with lung cancer and tuberculosis (original magnification, ×200).**

A tumour proportion score (TPS) of >25% was used to identify positive expression of CD3.

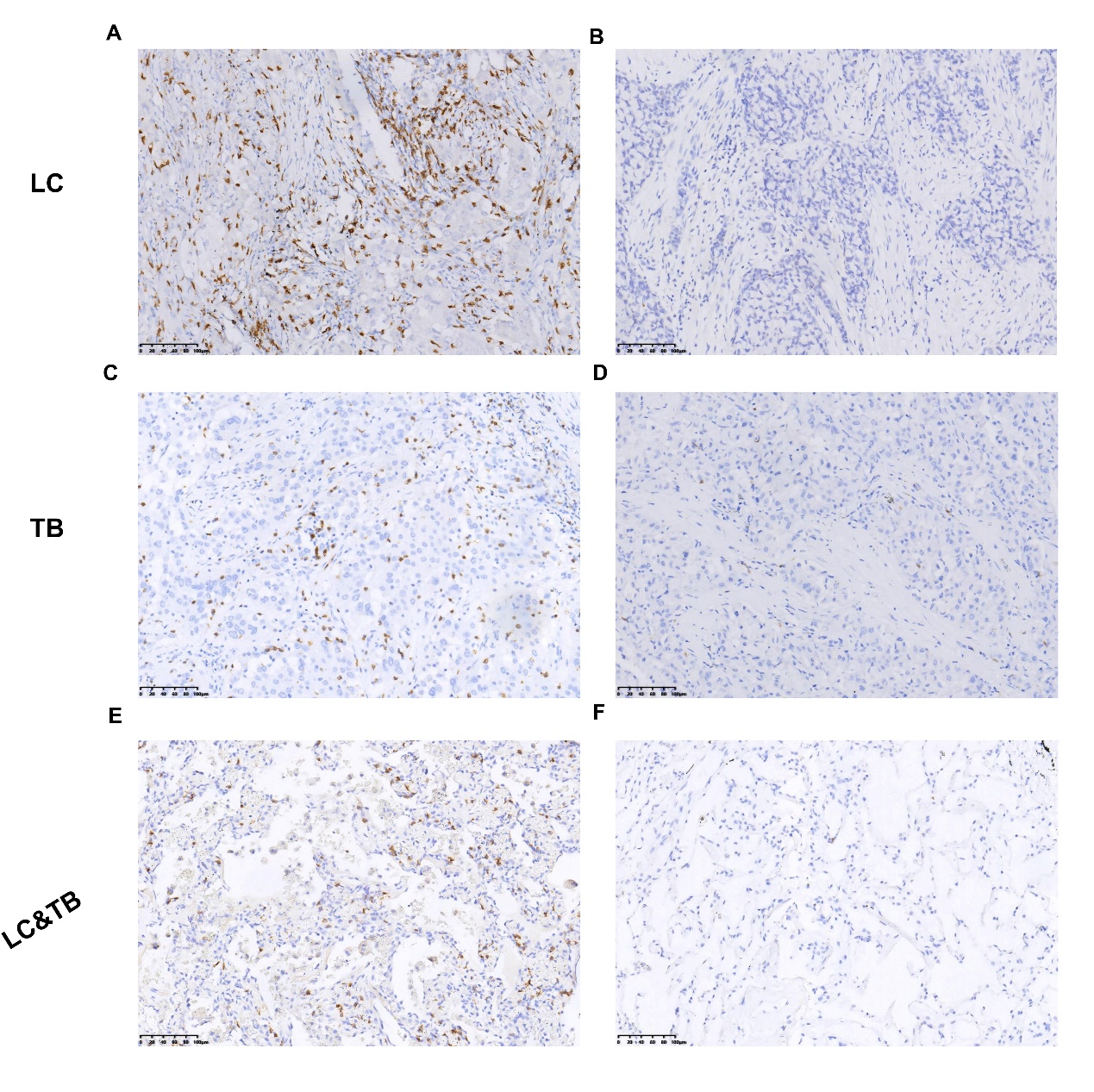


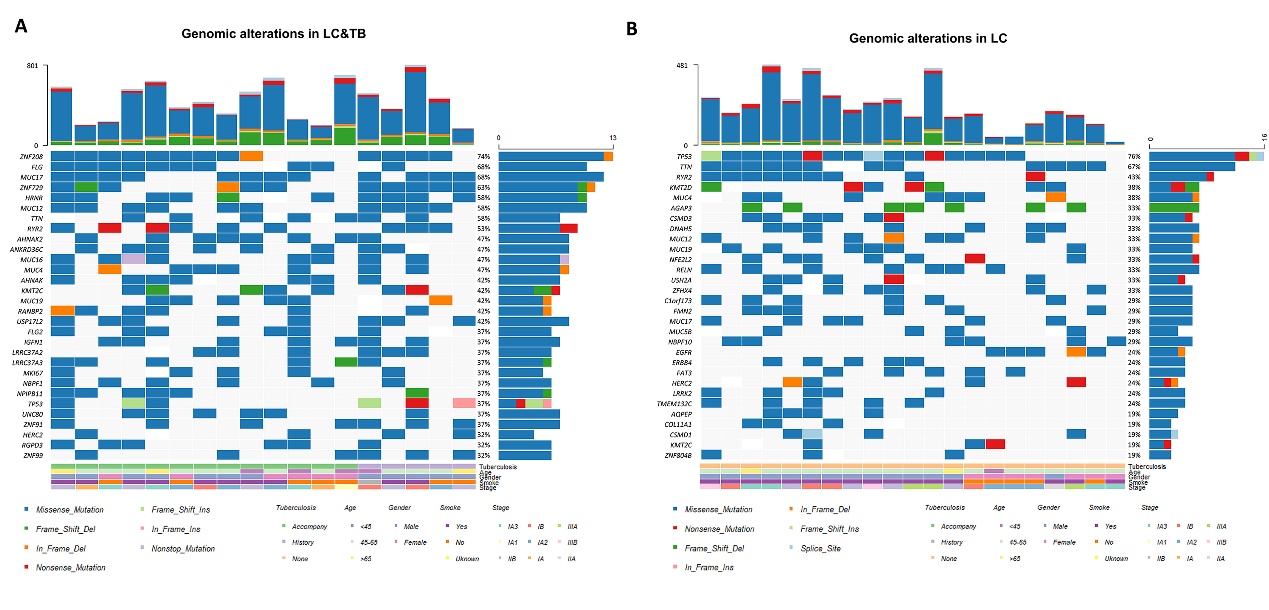
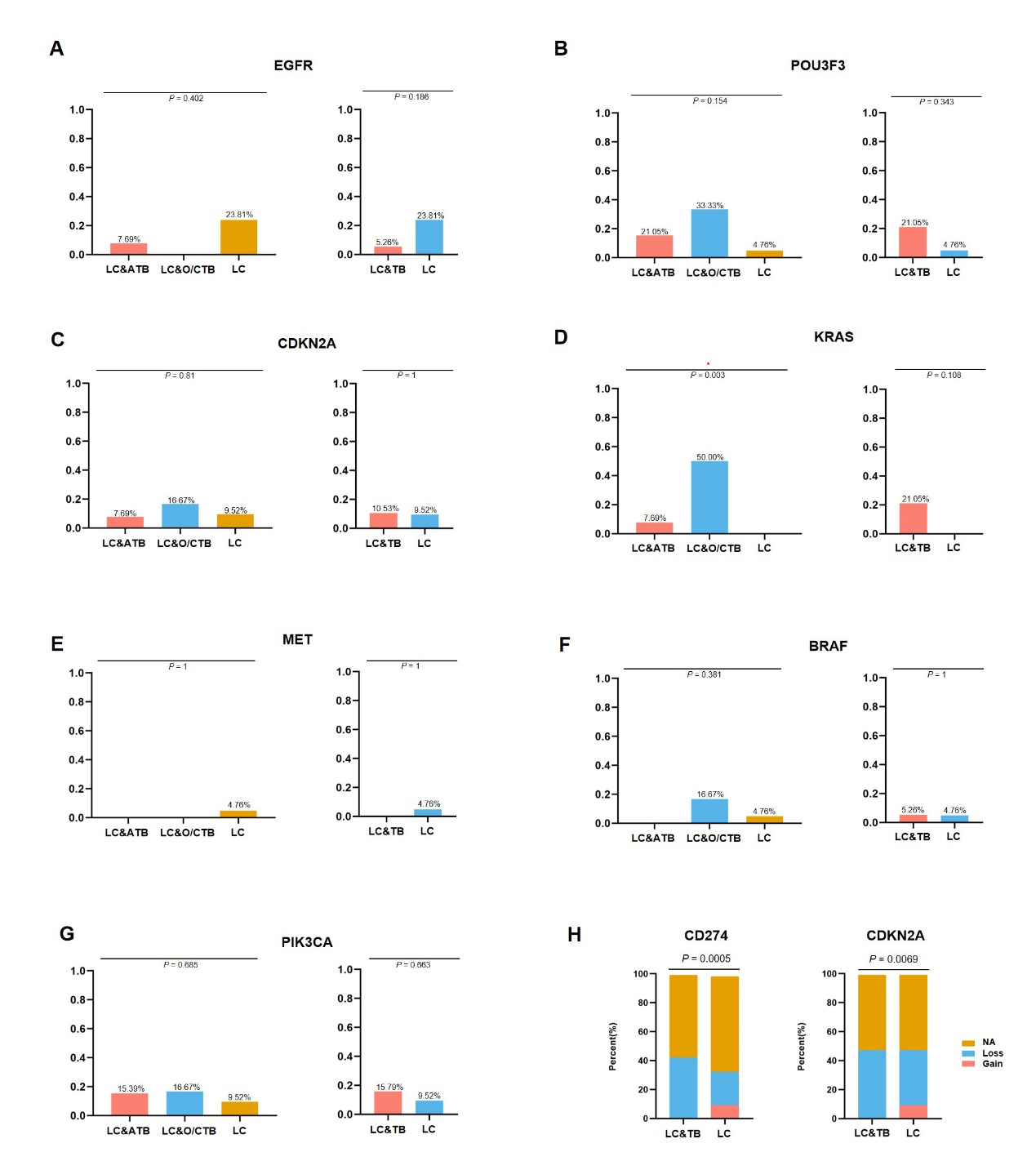
**Supplementary Figure S3. Immunohistochemical staining showing positive CD4 expression (A, C, and E) and negative expression (B, D, and F) in tissues from patients with lung cancer and tuberculosis (original magnification, ×200).**A tumour proportion score (TPS) of >25% was used to identify positive expression of CD4, although the CD4 expression rate was only 5% (C).



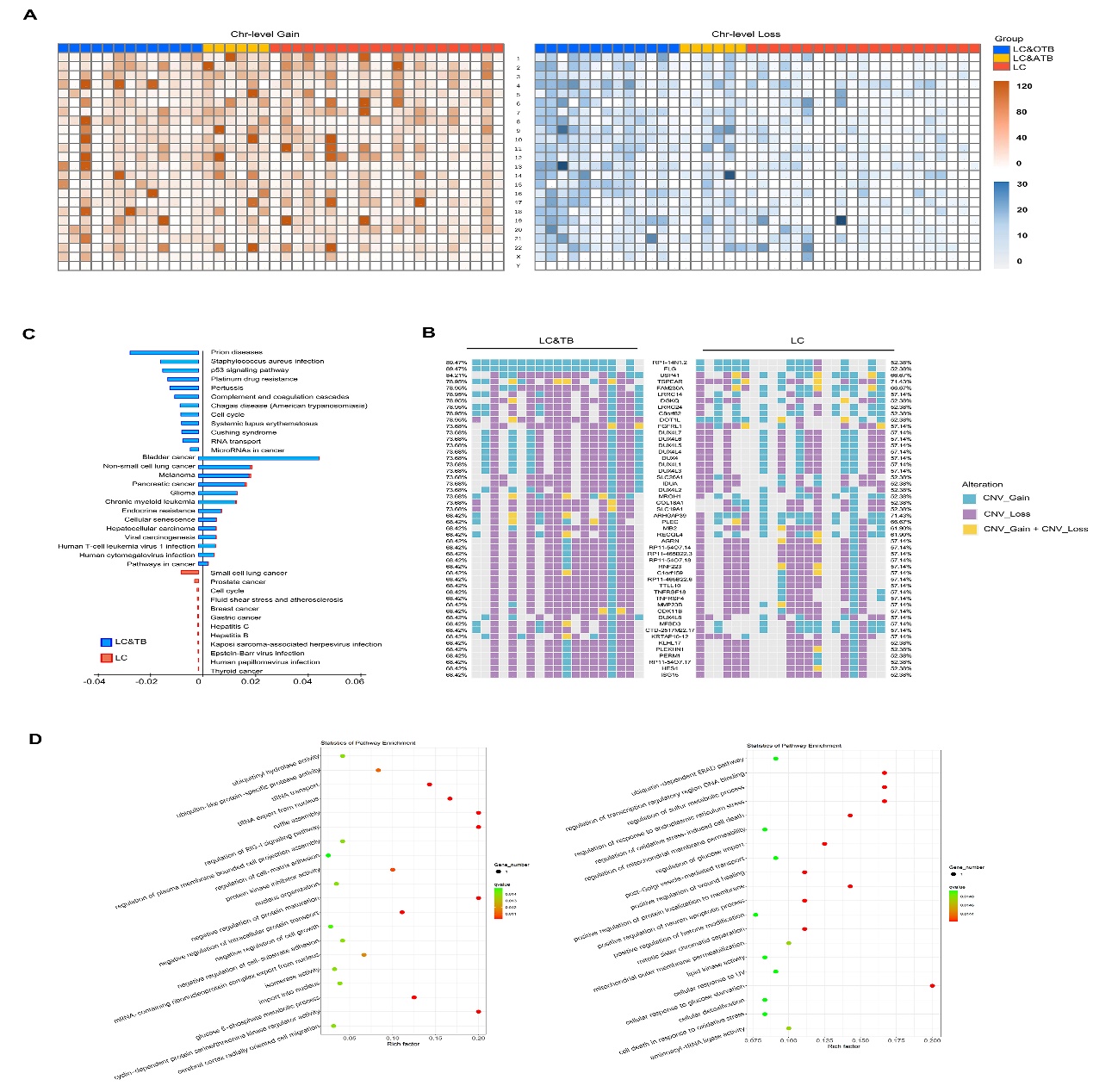
**Supplementary Figure S4. Immunohistochemical staining showing positive CD8 expression (A, C, and E) and negative expression (B, D, and F) in tissues from patients with lung cancer and tuberculosis (original magnification, ×200).**

A tumour proportion score (TPS) of >25% was used to identify positive expression of CD8.

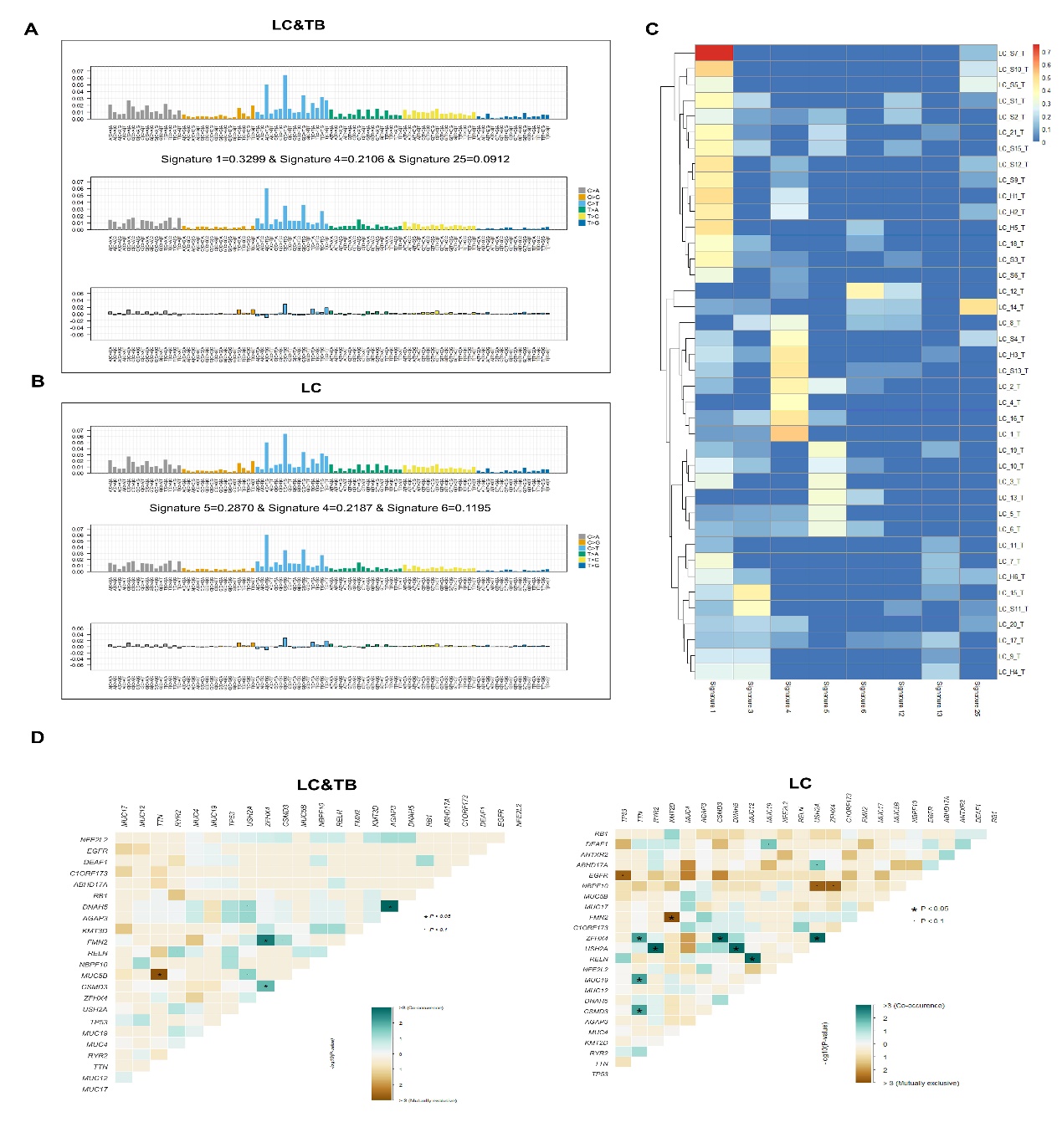


**Supplementary Figure S5. The 30 most common genomic alterations in patients with LC&TB or LC.****Supplementary Figure S6. Somatic mutation/copy number variations in candidate genes from the LC&O/CTB, LC&TB, and LC groups.**(A–G) Bar plots showing the somatic mutations in candidate genes from the LC&O/CTB, LC&TB, and LC groups. (H) The CD274 and CDKN2A copy number variations in the LC&TB and LC groups. 

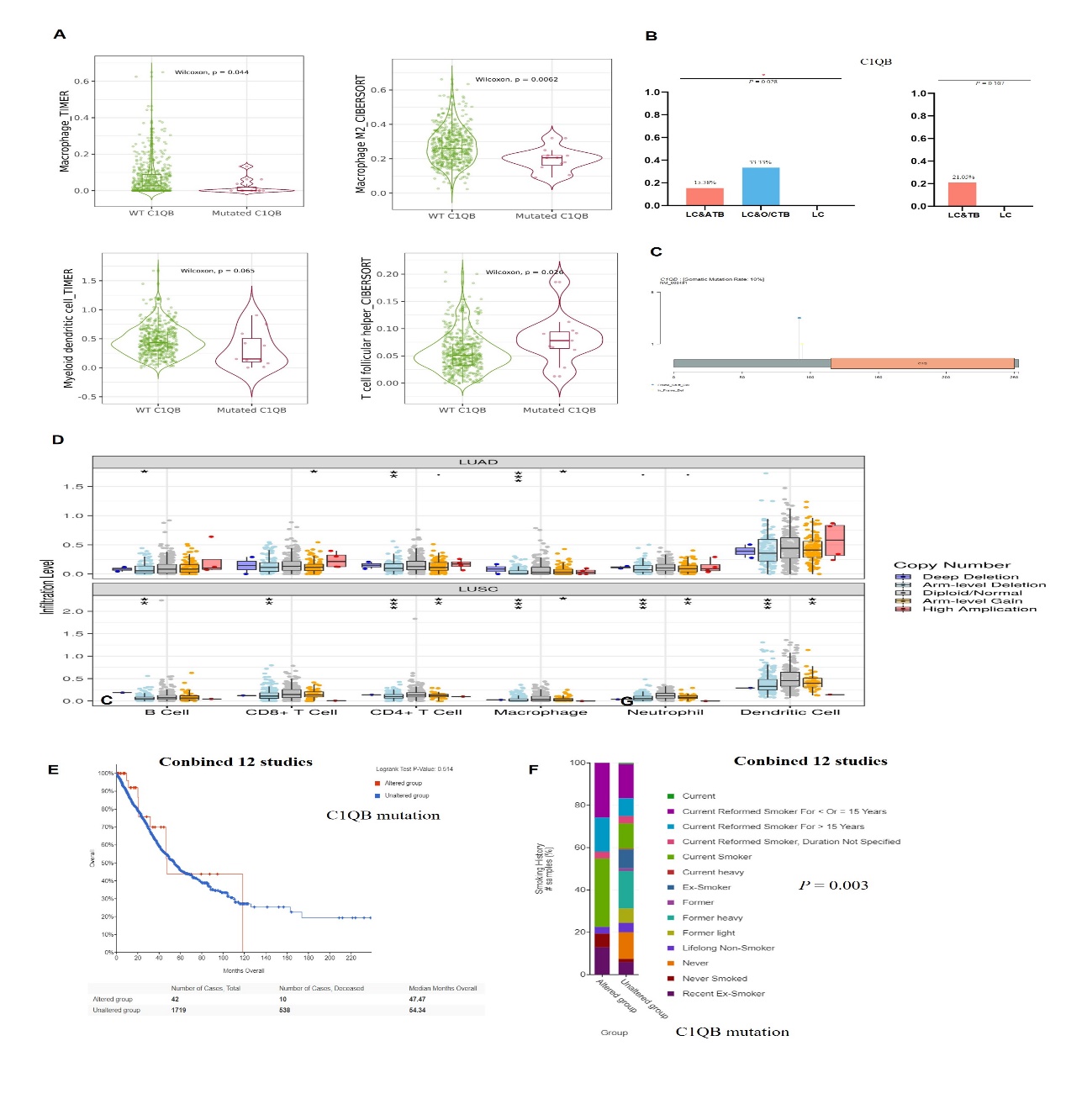
**Supplementary Figure S7. Heatmap showing the number of gains/losses (CNVs) per patient and pathway analysis.**(A) Heatmap showing the number of gains/losses (CNVs) per patient for each chromosome. (B) The results of pathway analysis showing the top 12 functional clusters for the LC&TB group (top panel), the LC&TB and LC groups (middle panel), and the LC group (bottom panel). (C) A summary of the top 50 significant gene gains and losses in the LC&TB and LC groups. (D) Bar plot showing the top 20 Gene Ontology pathways in the LC&TB and LC groups.



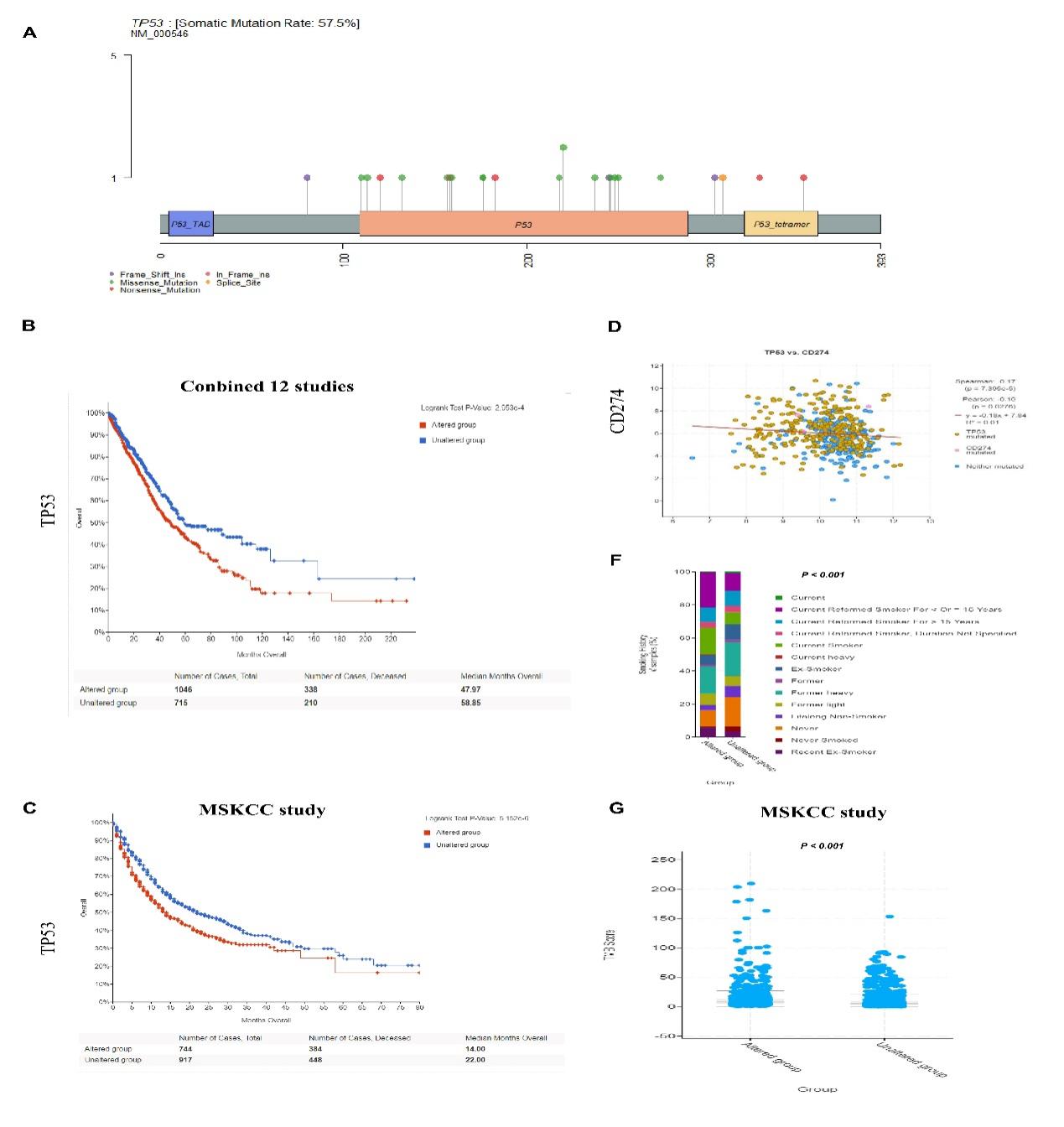
**Supplementary Figure S8. The mutation signatures and correlation analysis of genes with high mutation frequencies in the LC&TB and LC groups.** (A, B) Mutation signatures in the LC&TB and LC groups. The top panel shows the tumour mutational profile based on the fraction of mutations found in each trinucleotide context, the middle panel shows the reconstructed mutational profile created by multiplying the calculated weights by the signatures, and the bottom panel shows the sum of squares error between the tumour mutational profile and the reconstructed mutational profile. (C) Heatmaps showing the top 8 mutation signatures in each patient. (D) Mutually exclusive and cooccurring gene pairs in the LC&TB and LC groups, which are displayed as a triangular matrix. Green indicates a tendency towards cooccurrence and pink indicates a tendency towards mutual exclusivity.



**Supplementary Figure S9. *TP53* mutations in patients with LC&O/CTB, LC&ATB, or LC, as well as their roles in lung cancer immune infiltration according to public databases.**(A) Bar plots showing the frequency of *TP53* somatic mutations in the LC&O/CTB, LC&ATB, and LC groups, as well as in the LC&TB and LC groups. (B) The signalling pathways that are most related to *TP53* mutations based on cBioportal data. (C) Comparing tumour infiltration levels according to the different somatic copy number alterations of *TP53*, which were defined according to the [GISTIC 2.0](http://portals.broadinstitute.org/cgi-bin/cancer/publications/view/216) module as: deep deletion (–2), arm-level deletion (–1), diploid/normal (0), arm-level gain (1), and high amplification (2). Boxplots show the distributions of each immune subset at each copy number status in LUAD and LUSC. The infiltration level for each category was compared to that of the normal group using the two-sided Wilcoxon rank-sum test. (D) Comparing immune infiltration between tumours with and without the *TP53* gene based on the two-sided Wilcoxon rank-sum test (\*\*\* *P* < 0.001, \*\* *P* < 0.01, \* *P* < 0.05).

**Supplementary Figure S10. *TP53* mutations in patients with LC&TB or LC patients, as well as their associations with clinicopathological features and survival of LC patients according to public databases.**

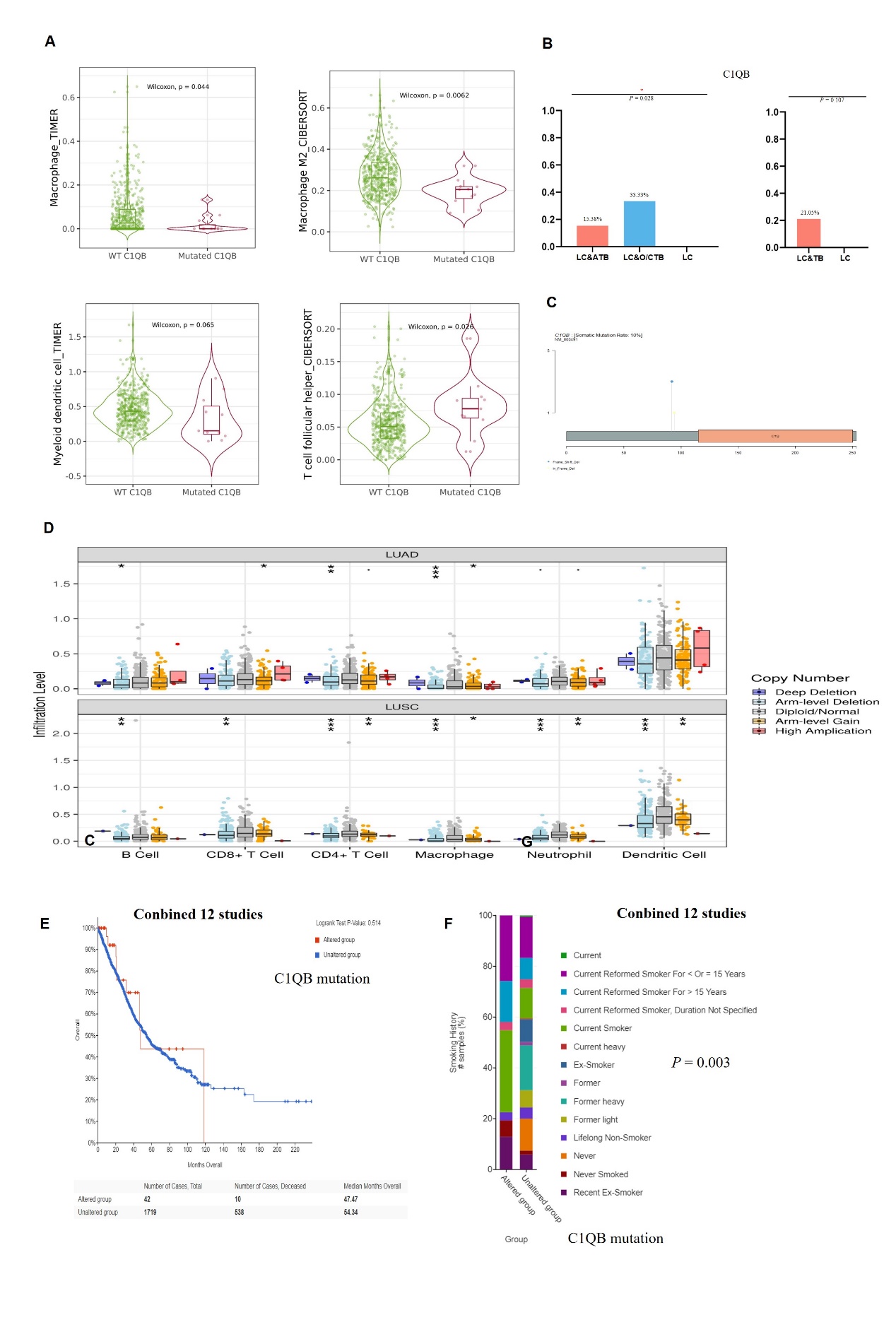
(A) Lollipop plot showing the *TP53* mutation distributions and protein domains in 40 patients. The mutation hotspots are labelled and the somatic mutation frequencies and transcript names are indicated. The effects of *TP53* mutations on overall survival are shown based on data from 12 studies (B) and the MSKCC study (C) from the cBioPortal database. (D) Correlations between CD274 and *TP53* mutation status (Spearman: –0.17, *P* < 0.001; Pearson: –0.1, *P* = 0.028). (F) Correlation between smoking history and *TP53* mutation status (*P* < 0.001) among LC patients from 12 studies. (G) Correlation between TMB and *TP53* mutation status in the pan-cancer MSKCC study (*P* < 0.001).



**Supplementary Figure S11. *C1QB* mutations in patients with LC&O/CTB, LC&ATB, or LC, as well as their associations with clinicopathological features, immune tumour infiltration, and patient survival according to public databases.**

(A) Correlations between *C1QB* mutations and macrophage/dendritic cell infiltration of lung cancer based on the TIMER 2.0 database. The violin plots show differences between LUAD tumours with mutant or wild-type *C1QB* in terms of macrophage infiltration (TIMER-estimated), myeloid dendritic cell infiltration (TIMER-estimated), M2 macrophage infiltration (CIBERSORT-estimated), and T follicular helper cell infiltration (CIBERSORT-estimated). (B) Differences in *C1QB* mutation frequencies between the LC&O/CTB, LC&TB, and LC groups. (C) Lollipop plot showing the *C1QB* mutations and protein domains in 40 patients. The mutation hotspots are labelled and the somatic mutation frequencies and transcript names are indicated. (D) Comparing tumour infiltration levels according to different somatic copy number alterations of *C1QB*, which were defined according to the [GISTIC 2.0](http://portals.broadinstitute.org/cgi-bin/cancer/publications/view/216) module as: deep deletion (–2), arm-level deletion (–1), diploid/normal (0), arm-level gain (1), and high amplification (2). (E) Effect of *C1QB* mutation status on overall survival in 12 studies. (F) Correlation between smoking history and *C1QB* mutation status (*P* = 0.003) in LC patients from 12 studies.

\*\*\* *P* < 0.001, \*\* *P* < 0.01, \* *P* < 0.05.



**Supplementary Figure S12. *C1QB* expression in LUAD and LUSC tumours and its correlation with tumour-infiltrating lymphocyte levels estimated using the TIMER 1.0 database.**

(A) Boxplots showing *C1QB* expression in tumour and adjacent normal tissues across all TCGA tumours. (B) Scatter plots showing the correlations of *C1QB* expression in LUAD and LUSC with tumour purity and the infiltration of B-cells, CD8+ T-cells, CD4+ T-cells, macrophages, neutrophils, and dendritic cells as estimated using the TIMER 1.0 database. (C) Scatter plots showing the correlations of *TP53* expression in LUAD and LUSC with tumour purity and the infiltration of B-cells, CD8+ T-cells, CD4+ T-cells, macrophages, neutrophils, and dendritic cells as estimated using the TIMER 1.0 database.

\*\*\* *P* < 0.001, \*\* *P* < 0.01, \* *P* < 0.05.

