



Integrative genomic profiling reveals characteristics of lymph node metastasis in small cell lung cancer

Kangle Kong¹, Shan Hu¹, Jiaqi Yue¹, Ziheng Yang¹, Salma K. Jabbour², Yu Deng¹, Bo Zhao¹, Fan Li¹

¹Department of Thoracic Surgery, Tongji Hospital, Tongji Medical College, Huazhong University of Science and Technology, Wuhan, China;

²Department of Radiation Oncology, Rutgers Cancer Institute of New Jersey, Rutgers Robert Wood Johnson Medical School, Rutgers University, New Brunswick, NJ, USA

Contributions: (I) Conception and design: S Hu, K Kong, B Zhao, F Li; (II) Administrative support: B Zhao, F Li; (III) Provision of study materials or patients: S Hu, K Kong, J Yue, B Zhao, F Li; (IV) Collection and assembly of data: S Hu, J Yue, Z Yang, Y Deng; (V) Data analysis and interpretation: S Hu, K Kong, B Zhao, F Li; (VI) Manuscript writing: All authors; (VII) Final approval of manuscript: All authors.

Correspondence to: Fan Li; Bo Zhao. Department of Thoracic Surgery, Tongji Hospital, Tongji Medical College, Huazhong University of Science and Technology, 1095 Jiefang Avenue, Qiaokou, Wuhan 430030, China. Email: tjhtsdrli@163.com; 13006369600@163.com.

Background: Small cell lung cancer (SCLC) is the most aggressive lung cancer subtype, with more than 70% of patients having metastatic disease and a poor prognosis. However, no integrated multi-omics analysis has been performed to explore novel differentially expressed genes (DEGs) or significantly mutated genes (SMGs) associated with lymph node metastasis (LNM) in SCLC.

Methods: In this study, whole-exome sequencing (WES) and RNA-sequencing were performed on tumor specimens to investigate the association between genomic and transcriptome alterations and LNM in SCLC patients with (N+, n=15) or without (N0, n=11) LNM.

Results: The results of WES revealed that the most common mutations occurred in *TTN* (85%) and *TP53* (81%). The SMGs, including *ZNF521*, *CDH10*, *ZNF429*, *POLE*, and *FAM135B*, were associated with LNM. Cosmic signature analysis showed that mutation signatures 2, 4, and 7 were associated with LNM. Meanwhile, DEGs, including *MAGEA4*, *FOXI3*, *RXFP2*, and *TRHDE*, were found to be associated with LNM. Furthermore, we found that the messenger RNA (mRNA) levels of *RBI* (P=0.0087), *AFF3* (P=0.058), *TDG* (P=0.05), and *ANKRD28* (P=0.042) were significantly correlated with copy number variants (CNVs), and *ANKRD28* expression was consistently lower in N+ tumors than in N0 tumors. Further validation in cBioPortal revealed a significant correlation between LNM and poor prognosis in SCLC (P=0.014), although there was no significant correlation between LNM and overall survival (OS) in our cohort (P=0.75).

Conclusions: To our knowledge, this is the first integrative genomics profiling of LNM in SCLC. Our findings are particularly important for early detection and the provision of reliable therapeutic targets.

Keywords: Small cell lung cancer (SCLC); whole-exome sequencing (WES); RNA-sequencing; lymph node metastasis (LNM); multi-omics analysis

Submitted Aug 05, 2022. Accepted for publication Jan 10, 2023. Published online Feb 13, 2023.

doi: 10.21037/tlcr-22-785

View this article at: <https://dx.doi.org/10.21037/tlcr-22-785>

Introduction

Small cell lung cancer (SCLC) is a heterogeneous neuroendocrine tumor originating from the Kulchitsky cells of the bronchial mucosa epithelium. As the most aggressive subtype, SCLC accounts for about 15–20% of all lung cancers, and more than 70% of patients are already

in the extensive stage of SCLC (ES-SCLC) at the time of admission (1). However, the first-line standard treatment of etoposide combined with cisplatin or carboplatin has a poor prognosis (2,3); the median survival time is only 10 months, and the 2-year survival rate is less than 5% (4). SCLC is a recalcitrant and highly metastatic cancer (5). Patients

with lung-confined local SCLC may benefit from surgical resection to remove their primary tumor, which has been linked to a higher survival rate (6). Nearly 70% of SCLC patients, on the other hand, already have metastatic disease when they are diagnosed, most commonly in the lymph nodes (LNs), brain, liver, and bones (5,6).

LNs are the central transport hub of circulating immune cells, which exist widely in the whole body (7). Tumor cells can metastasize to LNs and quickly spread to other organs. Regional LNs are the most common metastatic site for SCLC, and lymph node metastasis (LNM) usually occurs in the early stage of SCLC. Metastasis of cancer cells from the primary tumor to regional LNs is a major factor in poor prognosis (5). Although chemotherapy and irradiation can help to reduce the development of SCLC metastases, they do not continuously improve disease-free survival and overall survival (OS) (8). This is particularly important for early detection and the provision of reliable therapeutic targets. A large number of studies have found that some novel differentially expressed genes (DEGs) are closely related to LNM in lung cancer, including SCLC (9-11). Recently, Chen *et al.* explored the characteristics of gene mutations in non-small cell lung cancer (NSCLC) with and without LNM (12). However, research into the genomic landscape or multi-omics of SCLC with various LNM statuses is limited. A high mutation rate and genomic instability almost certainly contribute to SCLC's characteristic behavior (13). The high mutational burden caused by cigarette smoking, as well as the rapid tumor growth observed in SCLC patients with metastatic

disease, indicates that primary SCLC tumors are inherently metastatic (14). The identification of markers for the various mechanisms by which premetastatic SCLC cells become metastatic will help personalize treatment for metastatic SCLC subtypes. Analysis of genomic or transcriptomic alterations is critical for understanding the molecular processes of LNM in lung cancer (15,16).

In this study, we evaluated tissue specimens by whole-exome sequencing (WES) and RNA-sequencing to investigate the association between genomic and transcriptomic alterations and the LNM in 26 patients with SCLC. We characterized and compared the key molecular features between patients with (N+) and without (N0) LNM from patients who had been operated for SCLC. Multi-omics analyses were integrated to identify significantly mutated genes (SMGs) and DEGs associated with prognosis. To our knowledge, this is the first integrative genomics analysis of the molecular characteristics of LNM in SCLC. Our findings could have a variety of implications for tumorigenesis, immunotherapy, and prognosis for SCLC patients with LNM. We present the following article in accordance with the STREGA reporting checklist (available at <https://tclcr.amegroups.com/article/view/10.21037/tclcr-22-785/rc>).

Methods

Participants

A retrospective screen was conducted on formalin-fixed, paraffin-embedded (FFPE) and fresh frozen tumor tissues from SCLC patients who underwent surgery in the Tongji Hospital of Huazhong University of Science and Technology between May 2018 and August 2019. The patient inclusion criteria were as follows: (I) pathologically diagnosed with pathologic stage T1–T4, N0–3, M0 SCLC according to the 2015 (4th) edition of the World Health Organization (WHO) guideline during the operation; (II) without underlying cardiovascular, urinary, hepatic, and blood system disease; (III) expected survival time greater than 3 months; and (IV) complete medical records available. The exclusion criteria were as follows: (I) SCLC was not the first primary cancer; (II) sample collection, storage, or test operation was not qualified; and (III) incomplete patient information. We defined OS as the time interval between surgery and death from any cause or the final follow-up visit (censored patient). The study was conducted in accordance with the Declaration of Helsinki (as revised in 2013). This

Highlight box

Key findings

- The integrative genomics profiling was used to identify SMGs and DEGs related to LNM in SCLC and to further evaluate the link between these variants and clinical characteristics and prognosis.

What is known and what is new?

- Some novel differentially expressed genes (DEGs) are closely related to LNM in lung cancer have been identified using RNA-sequencing.
- The research into the genomic landscape or multi-omics of SCLC with various LNM statuses is lacking.

What is the implication, and what should change now?

- To our knowledge, this is the first integrative genomics profiling of LNM in SCLC. Our findings are particularly important for early detection and the provision of reliable therapeutic targets.

study was approved by the Institutional Review Board of Tongji Hospital of Huazhong University of Science and Technology (No. TJ-IRB20220329) with the written informed consent of all enrolled patients.

WES and analysis

The FFPE sections were evaluated by pathologists to confirm that more than 50% of cells were tumor cells. Matching normal tissue was recovered >2 cm from the visible margin of the tumor, without having any tumor cells, according to histopathologic assessment. Using the QIAamp DNA FFPE tissue kit (Qiagen, Germantown, MD, USA), DNA was extracted from 10–15 unstained FFPE sections of 5 μ m thickness and assessed using Nanodrop and Qubit (Thermo Fisher Scientific, Waltham, MA, USA). We prepared exon-wide capture libraries using the TruePrep DNA Library Prep Kit V2 (#TD501, Vazyme, Nanjing, China) and xGen[®] Exome Research Panel (Integrated DNA Technologies, Inc., Redwood City, CA, USA). The paired-end sequencing data with an average coverage of 130 for controls and 200 for tumors were generated using an Illumina HiSeq system (Illumina, San Diego, CA, USA). GATK 4.0 (Broad Institute, Cambridge, MA, USA) was used to sort and remove polymerase chain reaction (PCR) duplication, and Burrows–Wheeler aligner (BWA) was used to align the sequences to the human reference genome (NCBI build 37).

We used Strelka2 (Illumina) to detect single nucleotide variants (SNVs), insertions, and deletions (indel) with default parameters. Ensembl Variant Effect Predictor (VEP; <https://asia.ensembl.org/info/docs/tools/vep/index.html>) was used to annotate variants and polymorphisms. The minimum reads to support somatic SNV/indel calling was 20 for a mutated region and 5 reads for a variant. In the matched normal sample, $\geq 20\times$ of coverage and <5 reads of a variant at the same site were required. Variants having a minor allele frequency (MAF) >1% were filtered out from the ExAC, gnomAD, and esp6500 databases as common germline variants. MutSigCV software on the GenePattern platform (<https://www.genepattern.org/>) was used to identify SMGs. The driver gene list was established by combining 2 sets from previous studies (17,18). The ANNOVAR annotation of mutations was conducted with cosmic89_coding, oncoKB, and other tools. Somatic copy number variants (CNVs) identified by FACETS and recurrently occurring CNVs were detected with GISTIC2.0 (19).

RNA sequencing and analysis

Total RNA extracted from the freshly frozen tissue samples. We constructed RNA-sequencing libraries with the TruSeq RNA Library Preparation Kit (Illumina) following the manufacturer's instructions. The resulting libraries qualified using Bioanalyzer 2100 (Agilent, Santa Clara, CA, USA) and quantified using real time-PCR were loaded on the Illumina NovaSeq 6000 for sequencing. The sequencing reads preprocessed using Trimmomatic were aligned to the annotation files of the human genome (UCSC hg19) using STAR with default parameters (20). The human genome sequence was downloaded from Gencode. Reads of each gene were counted using htseq-count, and cufflinks was used to calculate the fragments per kilobase of exon per million mapped fragments (FPKM) from the aligned bam files (21). The DEGs between normal and tumor samples were qualified using R package DESeq2.

Weighted gene co-expression network analysis (WGCNA)

Based on the WGCNA algorithm of the R project (Langfelder and Horvath, 2008), the screened DEGs were prepared for further scale-free network construction. Firstly, the appropriate soft threshold power β was determined by pickSoftThreshold function. Then Pearson's correlation of gene expression profiles was calculated, similarity of gene expression patterns was evaluated, and adjacency matrix was established. After the co-expression network is completed, the dynamic tree cutting algorithm were used to detect the module. Finally determine the modules that meet these criteria and generate the corresponding tree diagram.

Tumor microenvironment (TME) analysis and multi-omics integration analysis

Cell-type Identification by Estimating Relative Subsets of RNA Transcripts (CIBERSORT) is a deconvolution algorithm of R package that estimates the absolute abundance of 22 human immune cell populations using gene expression profile data (22). It was utilized to explore immune cell abundance in SCLC tissues. Multi-omics integration analysis at the gene level was performed using CNAmets (23), which is an R package that integrates copy number alteration (CNA), gene expression, and DNA methylation information. In this study, CNAmets was used to calculate a wise weight score for a gene, indicating gene

Table 1 Patient clinical characteristics

Characteristic	Total, n (%)	Lymph node metastasis, n (%)		P value
		Yes (N+, n=15)	No (N0, n=11)	
Age (years)				0.691
≥60	13 (50.0)	7 (46.7)	6 (54.5)	
<60	13 (50.0)	8 (53.3)	5 (45.5)	
Gender				0.115
Male	23 (88.5)	12 (80.0)	11 (100.0)	
Female	3 (11.5)	3 (20.0)	0 (0.0)	
Smoking				0.040
No	8 (30.8)	7 (46.7)	1 (9.1)	
Yes	18 (69.2)	8 (53.3)	10 (90.9)	
Drinking				0.067
No	17 (65.4)	12 (80.0)	5 (45.5)	
Yes	9 (34.6)	3 (20.0)	6 (54.5)	
Family history				0.364
No	23 (88.5)	14 (93.3)	9 (81.8)	
Yes	3 (11.5)	1 (6.7)	2 (18.2)	
PD-L1 expression				0.666
Negative	13 (50.0)	7 (46.7)	6 (54.5)	
Positive	12 (46.2)	7 (46.7)	5 (45.5)	
Unknown	1 (3.8)	1 (6.6)	0 (0.0)	
Tumor grade				0.234
T1/T2	18 (69.2)	9 (60.0)	9 (81.8)	
T3/T4	8 (30.8)	6 (40.0)	2 (18.2)	

PD-L1, programmed death ligand 1.

alterations due to changes in copy number levels.

Statistical analysis

The software SPSS 25.0 (IBM Corp., Armonk, NY, USA) was used for all statistical analysis. Normally distributed continuous data were tested using analysis of variance (ANOVA), and non-normally distributed continuous data were analyzed using the Wilcoxon test. Student's *t*-test was used for continuous variables with normal distributions. Fisher's exact test and the chi-square test were used for categorical variables. Kaplan-Meier survival and Cox regression were used to evaluate the OS, and the log-rank test was used to analyze the differences. Statistical

significance was considered when $P < 0.05$.

Results

Clinical characteristics

In this study, 26 qualified SCLC patients (15 N+ and 11 N0) were included in the retrospective analysis. Pathology examination showed that among the 26 resected SCLC tissues, 3 (11.5%) were diagnosed with combined SCLC (a rare subtype of SCLC; a heterogeneous tumor composed of SCLC and non-small cell carcinoma), and the remaining 23 (88.5%) patients were diagnosed with pure SCLC. Patient clinical characteristics are summarized in *Table 1*.

The mean age of the participants at diagnosis was 59 years, and 23 (88.5%) patients were male. Some 12 patients were positive and 13 patients were negative for programmed death ligand-1 (PD-L1) staining. Interestingly, we found that LNM in our cohort may be associated with smoking ($P=0.040$).

Mutational landscape of N0 and N+ patients

The results of WES revealed a total of 9,904 nonsynonymous somatic mutations in 6,294 genes, with the most common variants occurring in *TTN* (85.0%), *TP53* (81.0%), and *MUC16* (58.0%) (Figure S1). Of these genes, the N0 tumors had 3,386 mutated genes, while the N+ tumors had 4,101 mutated genes, with only 18 common mutations between the groups, indicating distinct mutation profiles related to LNM. As shown in Figure 1A, we summarized the top 30 driver genes with the highest frequency of somatic mutations. In descending order of mutation frequency, the top 5 genes were *TP53*, *APOB*, *FAM135B*, *SPTA*, and *ZNF429* in the N0 group, and *TP53*, *LRP1B*, *RB1*, *CDH10*, and *KMT2D* in the N+ group. Specifically, the mutation frequencies of *AFF3*, *COL5A1*, *GRIN2A*, *UBR5*, *ZNF521*, *CDH10*, and *FN1* were noticeably lower in the N0 cases than in the N+ cases (Wilcoxon $P<0.001$), whereas *FAM135B*, *ZNF429*, *POLE*, *ZNRF3*, *ERBB4*, and *NOTCH1* mutation frequencies were remarkably higher in the N0 cases (Wilcoxon $P=0.004$) (Figure 1B). For DNA damage repair (DDR) genes, the top 5 genes with the highest mutation frequency were *TP53*, *POLE*, *POLQ*, *HFM1*, and *LIG4* in the N0 group, and *TP53*, *POLE*, *POLQ*, *FEN1*, and *PRKDC* in the N+ group (Figure 1C). Interestingly, mutations in *FEN1*, *PRKDC*, and *TDG* occurred only in the N+ group, yet mutations in some genes, such as *ANKRD28*, *FANCD2*, and *FANCI*, occurred only in the N0 group (Figure 1D). Thus, these results suggest that the mutation rate of these genes is closely related to LNM in patients with SCLC. We also found that the mean value of tumor mutational burden (TMB) in the N0 and N+ groups was 7.871 mut/Mb and 7.474 mut/Mb, respectively, but the differences were not significant ($P=0.24$) (Figure 2A). The average weighted Genome Instability Index (wGII) score was 0.490 in the N0 group and 0.485 in the N+ group, with no significant difference ($P=1$) (Figure 2B).

To ascertain the association between mutation frequency distribution and cosmic signature in the N0 and N+ patients, we performed non-negative matrix decomposition on the frequency of 96 substitution types. The distribution results indicated a strong preference for C>A substitutions (Figure 2C).

The N+ tumors had more frequent T>A mutations than the N0 tumors (1117/8.8% vs. 659/6.8%), with no significant difference. Additionally, signatures 4, 6, and 20 were identified in both the N0 and N+ groups (Figure 2D), with signature 4 being linked to tobacco-induced mutations and signature 6/20 being linked to defective DNA mismatch repair. We also found that mutation signature 7 (associated with ultraviolet light exposure) was present only in the N+ tumors, and signature 2 (associated with activity of the Activation-induced cytidine deaminase/apolipoprotein B mRNA editing enzyme, catalytic polypeptide-like (AID/APOBEC) family of cytidine deaminases) was present only in the N0 tumors, which suggests that mutation signatures 2 and 7 are associated with LNM.

CNV profiles of N0 and N+ patients

To discover novel CNVs involved in LNM, the significantly amplified or deleted regions of the genome were identified using GISTIC2.0. We found CNVs throughout the genome, with copy number gains being more prevalent than copy number losses. Gain of chromosomes 22q11.1, 9q13, 16p11.2, 14q11.2, and 17q21.2 and loss of chromosomes 16q12.2, 6p22.1, 16p13.3, 7p21.3, and 9q21.33 were more common (Figure 3A,3B). Some chromosomes with CNVs were only identified in the N0 group, such as 10q22.2, 1p36.13, 12q12, 16q12.2, 6p21.32, and 15q21.2, and some were only identified in the N+ group, such as 2q12.2, 16q22.2, 6p21.33, 10q24.32, 17p13.2, and 8p21.3 (Table S1). Specifically, the copy number gains were found at chromosomes 2q11.2, 17q12, and 8q12.3 in the N0 tumors, but the copy number losses occurred at these chromosomes in the N+ tumors. These chromosomal loci include many genes associated with tumor metastasis. For example, the *ASPH* gene at 8q12.3 has been reported as associated with metastasis in several tumors (24), indicating that the CNVs at these chromosomal regions might play pivotal roles in LNM in SCLC. Additionally, the N0 tumors exhibited more CNV burden (36.35/case) than the N+ tumors (33.24/case), although the difference was not significant ($P=0.44$) (Figure 3C).

Transcriptomic landscape of N0 and N+ patients

To better understand LNM in SCLC, we conducted a comparative transcriptomic analysis. To analyze the DEG levels between the N+ and N0 groups, the relative gene expression levels were subdivided into upregulated and

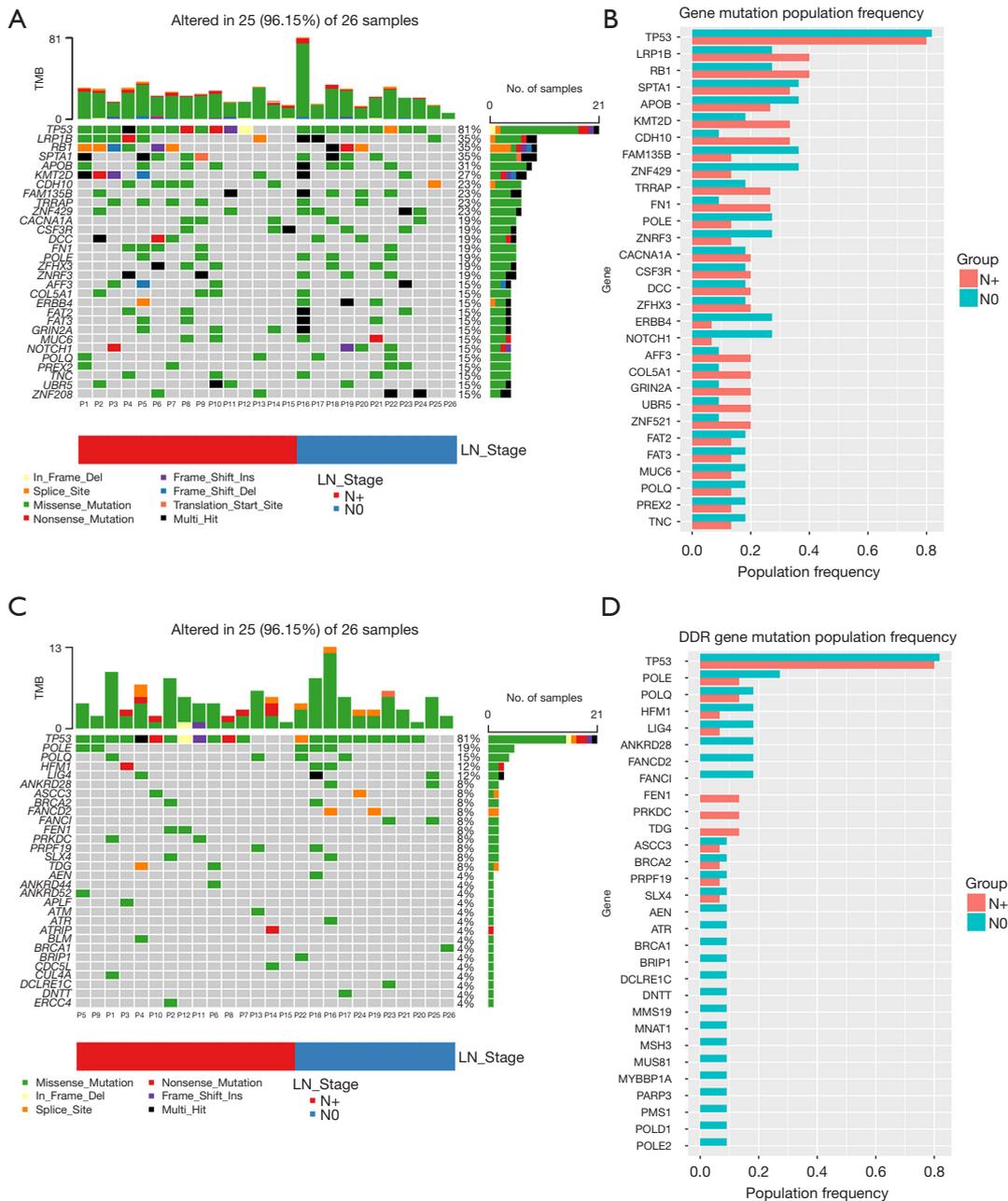


Figure 1 Comparison of mutation landscape between N0 and N+. (A) A comparison of the mutational landscapes of N0 and N+ is provided, along with the most frequently mutated driver genes. The top panel represents the TMB and the middle panel represents the matrix of frequently mutated genes. Columns represent samples, and clinicopathological characteristics of individual patients are presented below. (B) List of driver genes with significant differences between the two groups. (C) The most frequently mutated DDR genes in the N0 and N+ cohorts were compared. (D) List of DDR genes with significant differences between the two groups. TMB, tumor mutational burden; LN, lymph node; DDR, DNA damage repair.

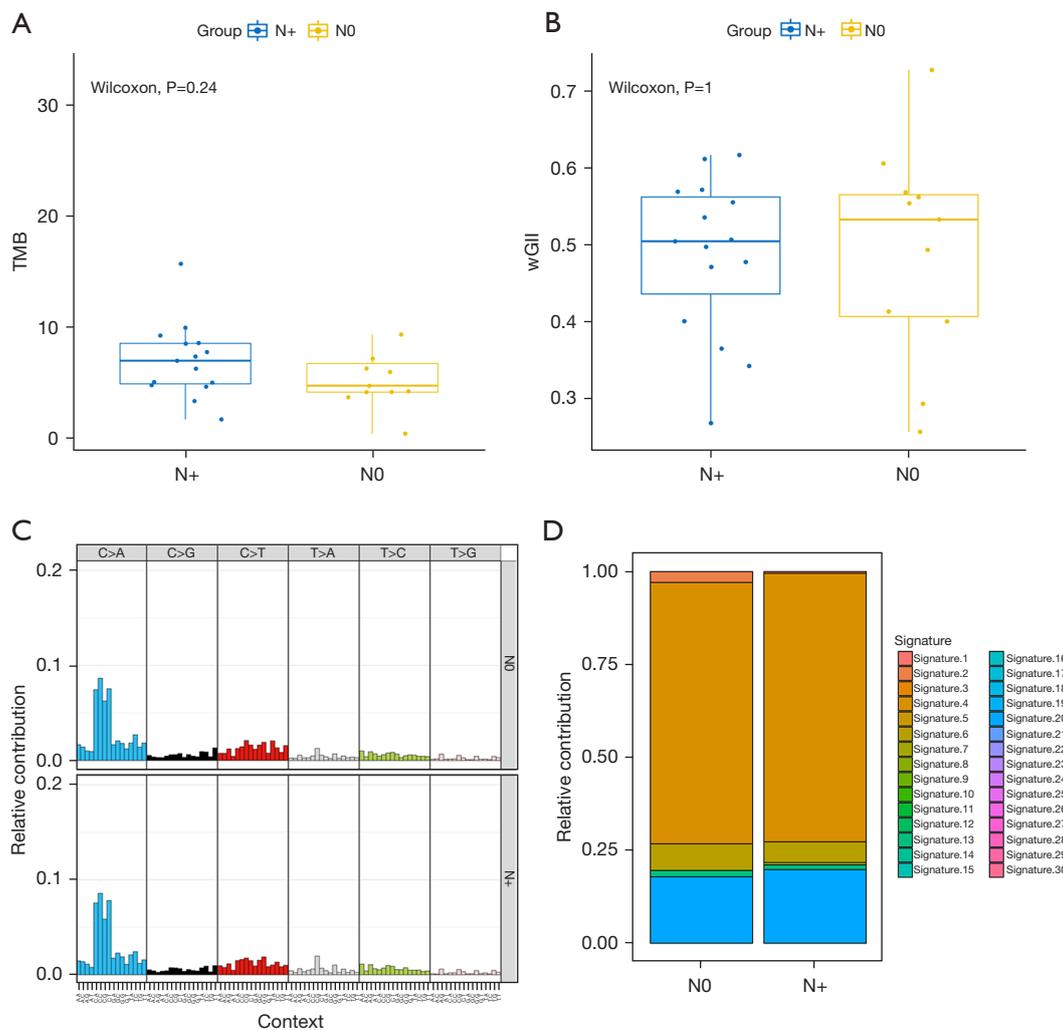


Figure 2 TMB, wGII, and variant spectrum in SCLC with LNM. (A) The TMB difference between N0 and N+. (B) Comparison of wGII between N0 and N+. (C) Relative contribution of six substitution subtypes SNV in each group. Colored bars represent different substitution subtypes. (D) Relative contributions of mutational signatures in each group. TMB, tumor mutational burden; wGII, weighted Genome Instability Index; SCLC, small cell lung cancer; LNM, lymph node metastasis; SNV, single nucleotide variation.

downregulated. Hierarchical clustering analysis on all DEGs of each tumor sample revealed that the N+ and N0 groups shared 6,666 DEGs. Of these, 3,790 were upregulated and 2,876 were downregulated. A total of 3,710 DEGs were only observed in the N+ groups and 1,741 private DEGs only in the N0 groups (Figure 4A). These results show that the expression levels of melanoma-associated antigen A4 (*MAGEA4*), forkhead box I3 (*FOXI3*), potassium voltage-gated channel subfamily Q member 2 (*KCNQ2*), and fibroblast growth factor 21 (*FGF21*) were only significantly increased in the N+ patients compared with the N0 patients. Relaxin family peptide receptor 2 (*RXFP2*), bactericidal/

permeability-increasing-fold-containing family B member 1 (*BPIFB1*), thyrotropin releasing hormone degrading enzyme (*TRHDE*), and ankyrin repeat domain 28 (*ANKRD28*) were reduced in the N+ patients. Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway and Gene Ontology (GO) enrichment analyses of DEGs showed differences in signaling pathways between the N+ and N0 groups (Figure 4B,4C). The shared DEGs were enriched in well-known carcinogenesis pathways including “focal adhesion”, “MicroRNAs in cancer”, and “extracellular matrix (ECM)-receptor interaction”.

In addition, the top 5,000 genes with an absolute median

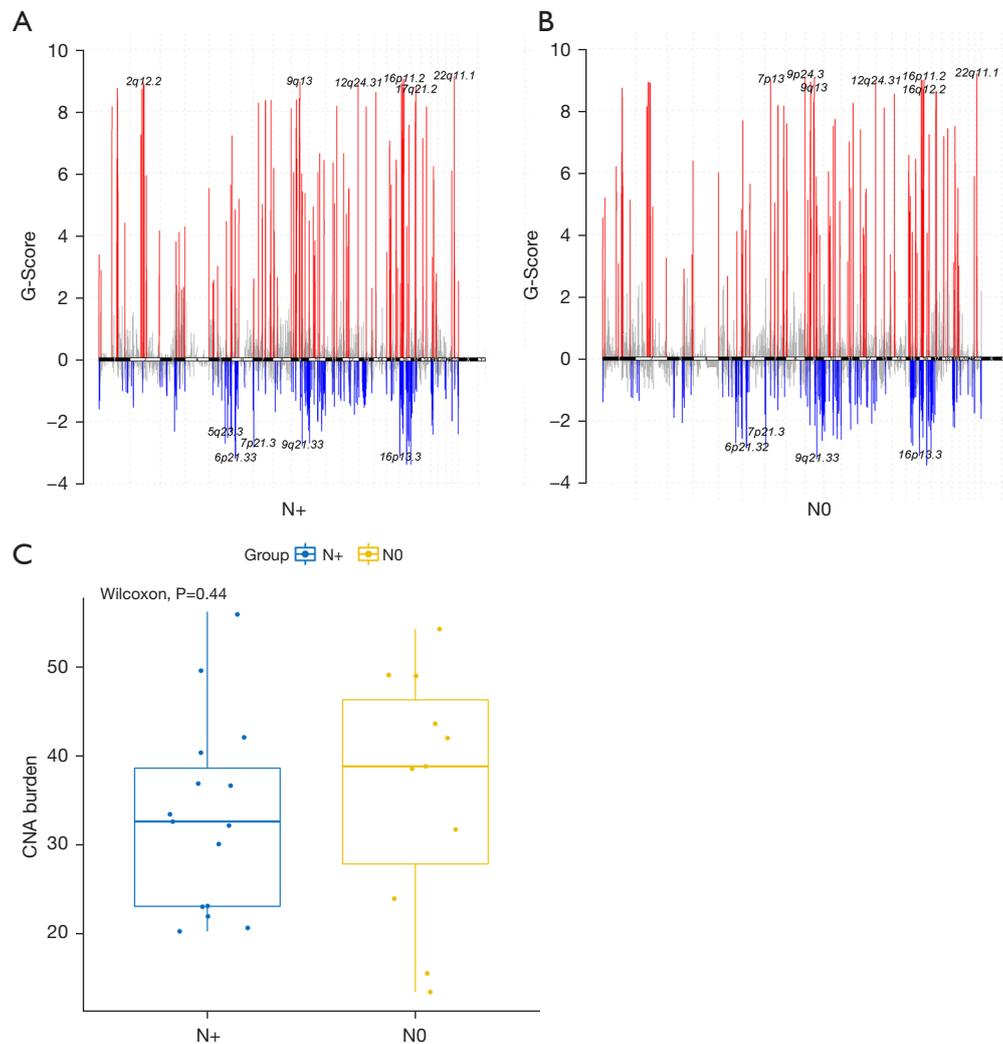


Figure 3 Distinct CNA landscape of N0 and N+. (A,B) Somatic CNAs in each group. Deletions and amplifications are represented on the y-axis by blue or red bars, respectively. Each peak region (cytoband) is displayed together with its known or potential cancer-related genes. (C) Comparison of the CNA burden between N0 and N+. CNA, copy number alteration.

difference were obtained, and a co-expression network was constructed using weighted correlation network analysis (WGCNA). The appropriate soft threshold power β was determined as 12 for subsequent adjacencies calculation (Figure S2A). Finally, five gene modules were obtained according to the results of the dynamic tree cut algorithm (Figure S2B). The relationship of modules and clinical features were evaluated and visualized in Figure S2C to explore the clinical significance of co-expressed genes. The yellow module in the figure was significantly associated with smoking and gender. Furthermore, the correlation of module membership and gene significance for LNM in grey

module is depicted in Figure S2D.

TME of N0 and N+ patients

To further study the relationship between immune cell characteristics of SCLC and LNM, TME expression profiles were clustered as N+ and N0 groups. The clusters presented different immune cell profiles including signatures of T cells, B cells, macrophages, and natural killer cells (Figure 5A). Briefly, the mean proportions of plasma cells and CD8⁺ T cells in both the N+ and N0 groups were over 10%, whereas CD4 naïve T cells and T cells gamma

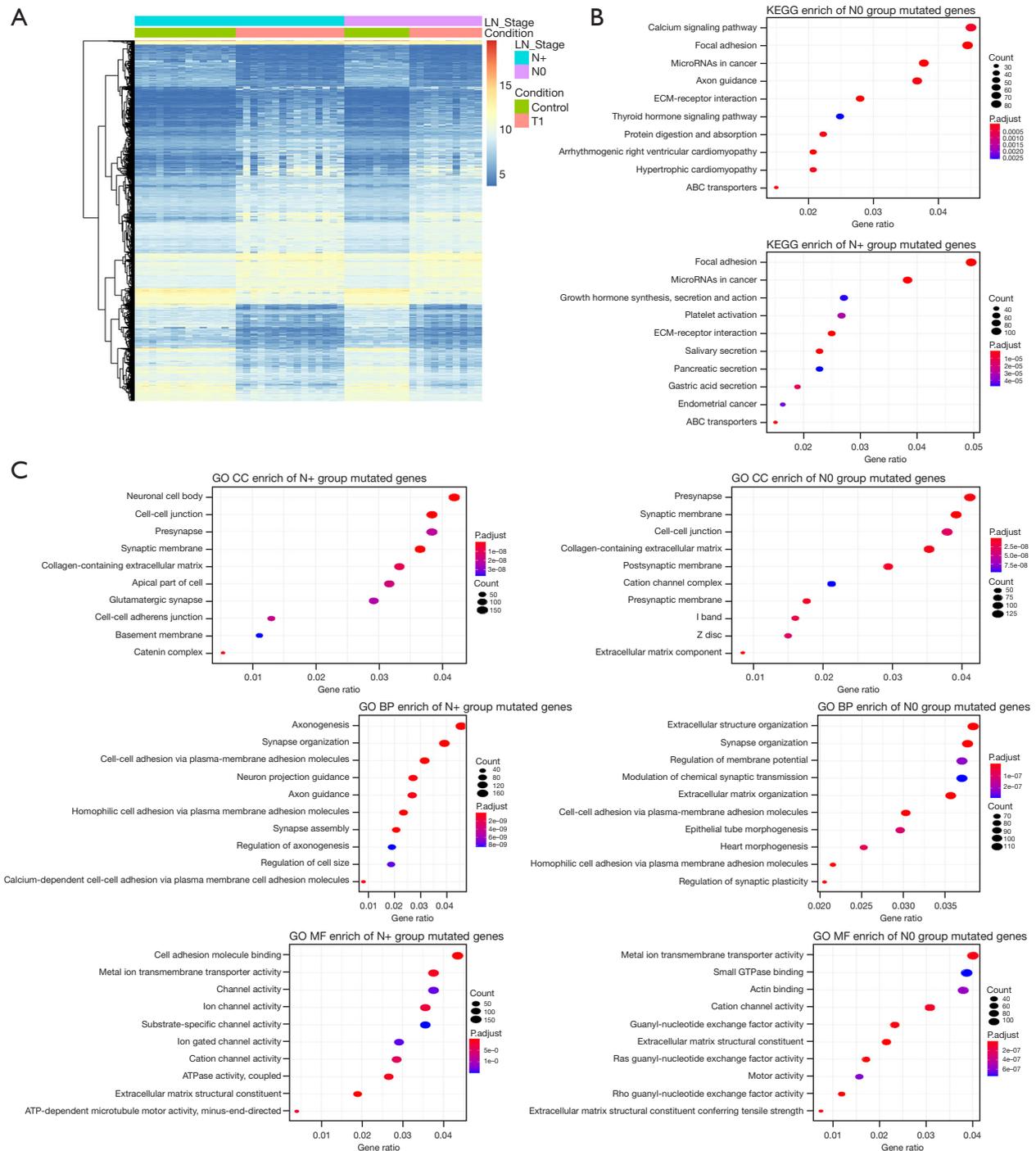


Figure 4 The transcriptomic landscape of N0 and N+ patients. (A) Heatmap of DEGs. Hierarchical clustering analysis of DEGs from each N+ and N0 tumor sample. Blue indicates the lowest amount of gene expression, whereas pink-orange indicates the highest level. (B) KEGG pathway enrichment of the DEGs. The y-axis represents the pathway name, while the x-axis represents the enriched factor in each pathway. (C) GO enrichment analysis of DEGs. Top 10 GO terms in CC, BP and MF are presented. The size of the bubbles reflects the number of genes, while the color bar indicates the corrected P value, with a greater value in red and a lower value in blue. LN, lymph node; ECM, extracellular matrix; ABC, ATP-binding cassette; ATP, adenosine triphosphate; CC, cellular component; BP, biological process; MF, molecular function; GO, Gene Ontology; KEGG, Kyoto Encyclopedia of Genes and Genomes; DEG, differentially expressed gene.

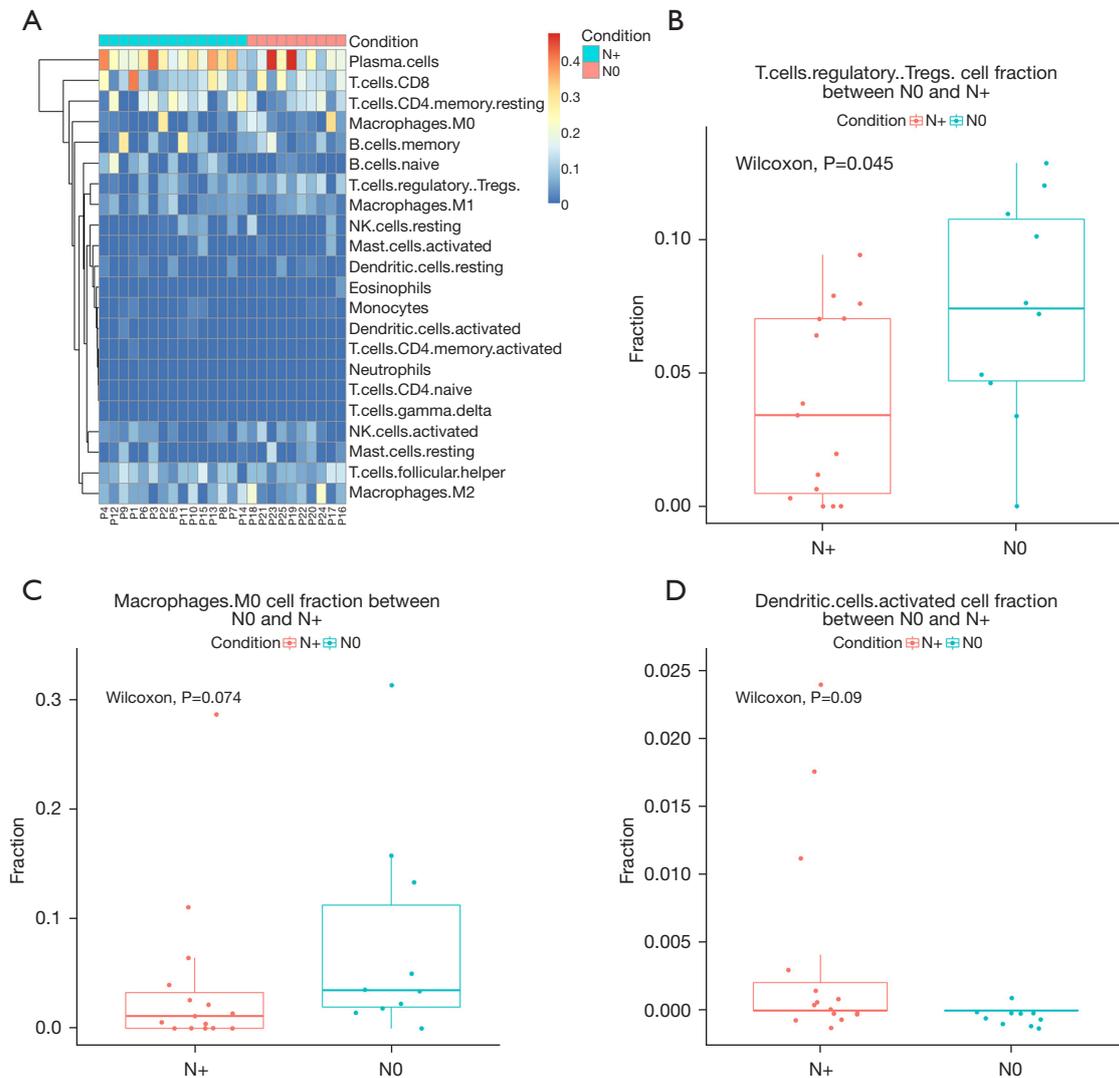


Figure 5 TME differences in N0 and N+. (A) Distinct immune cell signatures between N0 and N+. (B-D) Significantly different proportions analysis of immune cell types from each group. Fractions of Tregs (B) and M0 macrophages (C) were significantly more profound in N0 than in N+. Fraction of activated dendritic cells (D) was significantly more profound in N+ than in N0. The KEGG enriched with the germline-regulated genes. TME, tumor microenvironment; Tregs, T regulatory cells; KEGG, Kyoto Encyclopedia of Genes and Genomes.

delta were absent in all samples. There were significantly more T regulatory cells (Tregs) ($P=0.045$) (Figure 5B) and M0 macrophages cell fractions ($P=0.074$) in the N+ group (Figure 5C), yet the activated dendritic cell fraction was greater in the N0 group ($P=0.09$) (Figure 5D).

Analysis of multi-omics integration

To better understand the characteristics of LNM in SCLC, we integrated our analysis of the mutation types, copy number alterations, and mRNA expression. We found that

CNV gain of *AFF3* and *TDG* was prevalent in SCLC, and that CNV loss occurred more frequently in *ANKRD28* and *RB1*. Interestingly, *ANKRD28* expression was consistently lower in the N+ tumors than in the N0 tumors (Figure 6A). As shown in Figure 6B, mRNA levels of *RB1* ($P=0.0087$), *AFF3* ($P=0.058$), *TDG* ($P=0.05$), and *ANKRD28* ($P=0.042$) were significantly associated with the copy number alterations. Of these genes, *AFF3*, *TDG*, and *ANKRD28* mutations were closely related to LNM in our cohort. Our findings suggest that variables other than mutation status and CNV may alter gene expression.

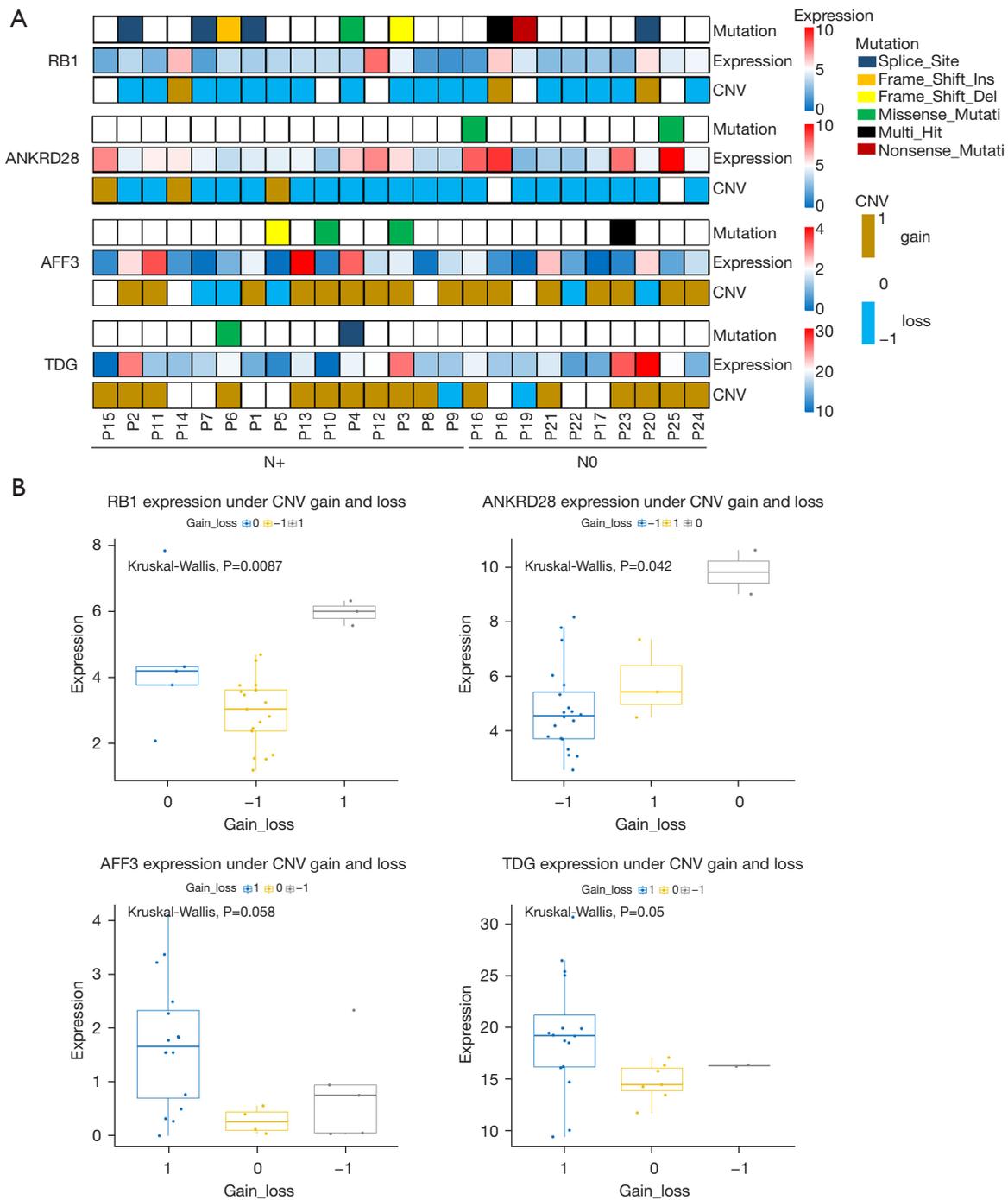


Figure 6 Correlation of mRNA expression with CNV status. (A) Distribution of *RB1*, *ANKRD28*, *AFF3*, and *TDG* expression and CNV across samples. (B) Correlation between mRNA expression and CNV of *RB1*, *ANKRD28*, *AFF3*, and *TDG*. mRNA, messenger RNA; CNV, copy number variation.

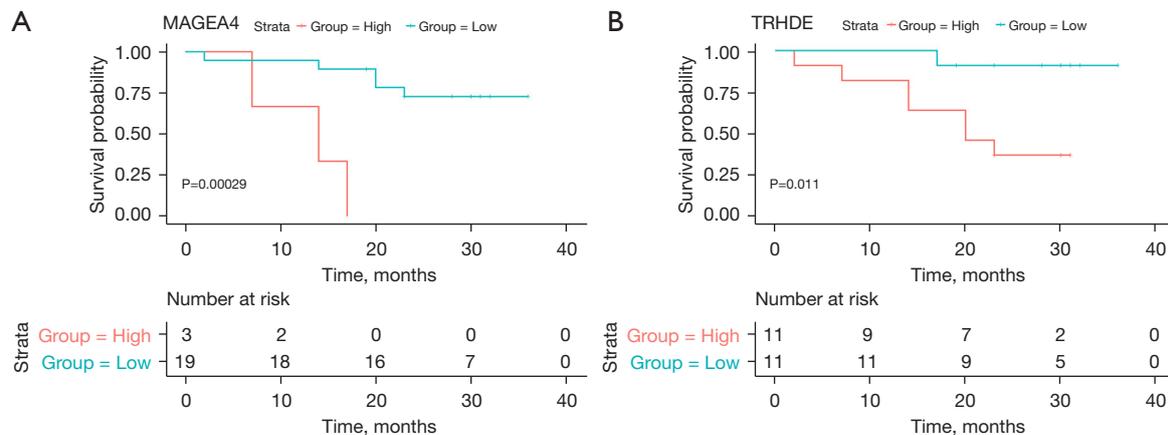


Figure 7 Association of *MAGEA4* and *TRHDE* expression status and clinical outcomes. (A) Survival analysis of *MAGEA4* expression and survival outcomes (P=0.00029); (B) survival analysis of *TRHDE* expression and survival outcomes (P=0.011).

Prognostic analysis of N0 and N+ patients

During the median 23-month post-surgery follow-up, 8/26 (30.8%) patients died. We found no significant correlation between OS and LNM in our cohort (P=0.82) (Figure S3A). Additionally, neither CNV, wGII, nor TMB had a significant impact on OS in N+ and N0 patients (Figure S3B-S3D). We further found that presence of the *ZNF521* mutation was substantially related to prognosis (OS, P<0.0001) (Figure S4A), which was significantly associated with LNM in our cohort. In N+ patients, rare *TENM4* (P=0.011), *DNAH9* (P<0.0001), *AHDC1* (P=0.038), *DNER* (P=0.045), and *ADAMTS18* (P=0.037) mutations were associated with worse prognosis (Figure S4B-S4F). In the N0 patients, OS was significantly shortened in patients with *NPHS1* (P=0.024), *PPFLA3* (P=0.019), *MUC3A* (P=0.019), *AGBL1* (P=0.015), *ATP8B3* (P=0.024), and *MYH14* (P=0.021) mutations (Figure S5). Furthermore, differential expression of *MAGEA4* and *TRHDE* between the N+ and N0 patients was substantially related to OS (log-rank test; P=0.00029 and P=0.011) (Figure 7).

cBioPortal data analysis

Genomic characteristics were further validated using cBioPortal datasets (N0 group, n=47; N+ group, n=59). The 3 most frequent somatic alterations were identified in *TP53* (86%), *RB1* (71%), and *TTN* (63%) (Figure 8A). Specifically, the mutation frequencies of *DST* (P=0.019), *SALL3* (P=0.034), *TAF1L* (P=0.034), *GPRC6A* (P=0.034), *GALNT17* (P=0.035), and *TNR* (P=0.038) were noticeably

higher in the N+ cases than in the N0 cases. Similarly, no significant difference in TMB was found between the N+ and N0 groups (Figure 8B). We found that *POLE* (P=0.005169) and *ZNF429* (P=0.001075) mutations were associated with worse prognosis (Figure 8C). Further, LNM was significantly associated with poor prognosis (P=0.014) (Figure 8D).

Discussion

Studies on LNM in SCLC have mainly focused on the transcriptome level (11,16); thus a multi-omics analysis is lacking in the literature. In this study, we performed genomic and transcriptomic profiling of SCLC patients. By integrating multi-omics data and clinical features, we provide a comprehensive view of the characteristics of LNM in SCLC.

Except for *TP53* and *RB1*, there is little consensus regarding shared gene mutations among SCLC patients. In our study, the most common mutations were found in *TTN* and *TP53* genes, and we found a high frequency (69.2%) of a *TTN/TP53* double mutation. The results were further corroborated using cBioPortal data, which showed a high frequency (55.7%) of *TTN/TP53* co-mutation. Mutation of the *TTN* gene has been significantly associated with lung squamous cell carcinoma (LUSC) (25); LUSC patients have favorable OS benefits from *TTN* mutation or *TTN/TP53* double mutation. However, the identification of *TTN* as an important cancer-related gene, especially in SCLC, remains controversial since mutations have only been demonstrated through bioinformatics analysis. The focus of this

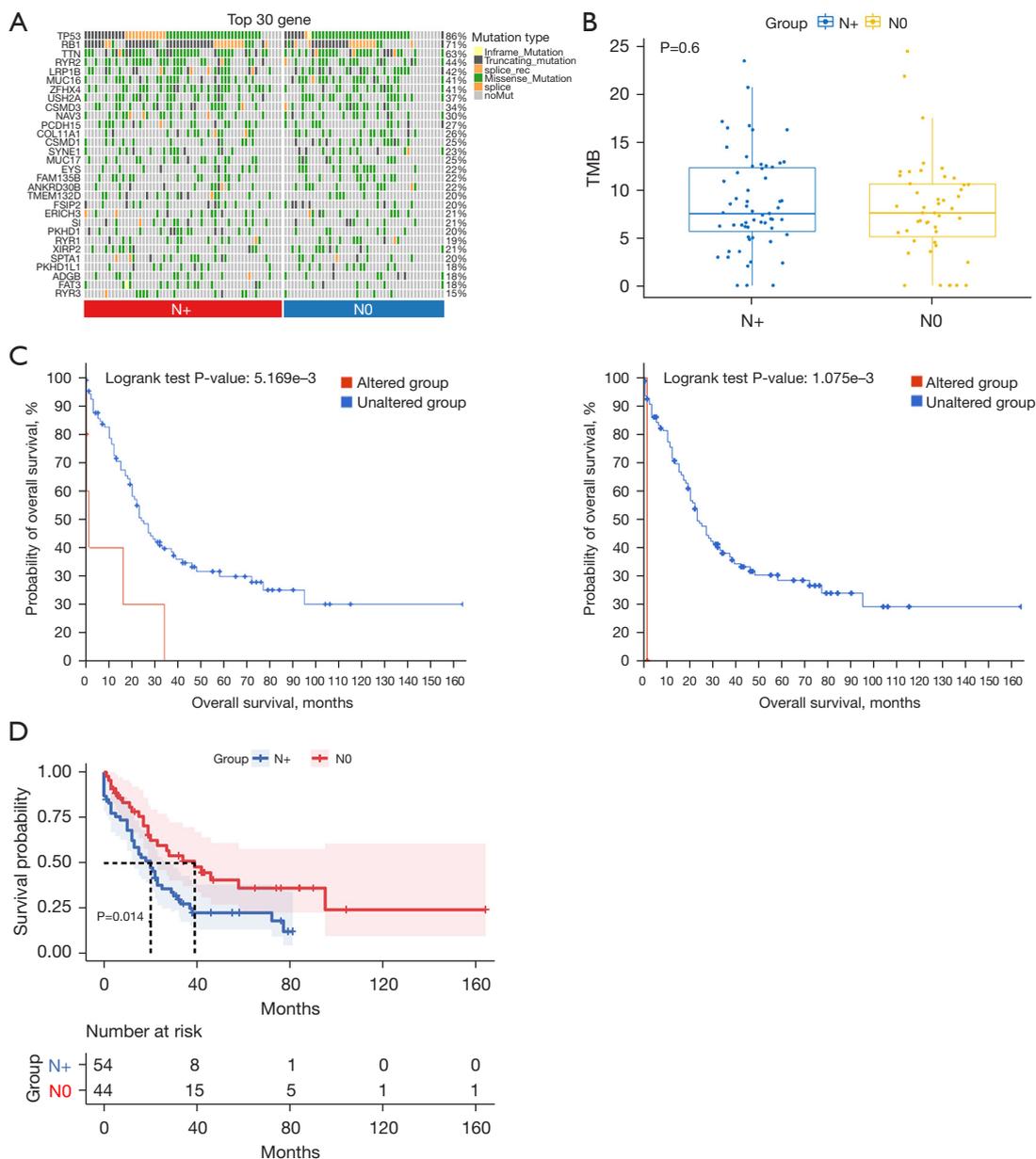


Figure 8 Genomic characteristics analysis of SCLC with LNM in the cBioPortal data set. (A) Landscape of somatic mutations in each group. (B) Compared TMB difference between N0 and N+. (C) Association of SMG alterations and worse prognosis in the cBioPortal cohort. The left shows POLE mutations (P=0.005169). The right shows ZNF429 mutations (P=0.001075). (D) Survival analysis of LNM status and clinical outcomes (P=0.014). TMB, tumor mutational burden; SCLC, small cell lung cancer; LNM, lymph node metastasis; SMG, significantly mutated genes.

controversy lies in its large complex structure and the false positive results caused by heterogeneity in the mutation process (26,27). By combining our data with previous SCLC studies, we were able to identify other prevalent mutant driver genes with a prevalence of more than 20%,

including *LRP1B*, *KMT2D*, and *CDH10*. Furthermore, the significantly mutated driver genes were identified in LNM and non-metastasis groups, some of which were specifically associated with LNM, including *ZNF521*, *CDH10*, *ZNF429*, and *FAM135B*. Recurrent *CDH10* mutations have been

reported in non-smokers who are negative for *EGFR/KRAS/ALK* mutations in lung adenocarcinoma (28) and as a prognostic signature in colorectal cancer (29). Jiang *et al.* showed that, as the most common and significant mutant gene, *CDH10* is also related to poor survival in SCLC (30). Both *ZNF521* and *FAM135B* have also been reported to be associated with LNM in other cancers (31,32). Further studies on the relationship between these differentially mutated genes and prognosis showed that *ZNF521* was significantly associated with prognosis of SCLC patients with LNM. In addition, we found that rare mutations in *TENM4*, *DNAH9*, *AHDC1*, *DNER*, and *ADAMTS18* were associated with poorer prognosis in patients with lymph node metastasis, while patients without metastasis with mutations in *NPHS1*, *PPF1A3*, *MUC3A*, *AGBL1*, *ATP8B3*, and *MYH14* had significantly shorter OS. Although this conclusion should be confirmed with additional research and larger sample size.

Cosmic signature analysis revealed widespread defective DNA mismatch repair and DDR gene alterations in our cohort. The initial high response rate of SCLC to platinum (cisplatin and carboplatin) and topoisomerase I/II inhibitors may be explained by the high-proliferative capacity of the tumor cells and the evidence indicating that DDR pathway defects are more susceptible to DNA damage-inducing treatments, such as chemotherapy and radiotherapy (etoposide and irinotecan). The Fanconi anemia (FA) pathway is critical for repairing DNA damage caused by inter-strand crosslinkers such as cisplatin. It has also been shown that mutations in FA pathway genes may lead to initial hypersensitivity to platinum-based therapy (cisplatin) in SCLC (30,33). In our cohort, most patients received conventional chemoradiotherapy after surgery, but there was no significant association between chemoradiotherapy and prognosis. Since various chemotherapeutics require a functional *TP53* protein to effectively induce apoptosis, the functional loss caused by high frequency mutations of *TP53* may enhance resistance to cytotoxic drugs in SCLC (34,35). In addition, the combination of *TP53* inactivation and MMR defects or other DNA repair mutations has also been observed to lead to early sensitivity and potential resistance of SCLC to cisplatin (30,36). This resistance requires urgent attention to improve clinical diagnosis and treatment of SCLC.

Over 95% of patients with SCLC have a history of smoking, with a 5-year survival rate of less than 2% (37). In our cohort, a history of smoking in 74% of patients was significantly associated with LNM, further verified

by cosmic profiling. Tobacco is known to contain various addictive compounds, including nicotine, which binds to the nicotinic acetylcholine receptor (nAChR). Studies have shown that nicotine induces the expression of achaete-scute family bHLH transcription factor 1 (*ASCL1*), which in turn induces the expression of the *nAChR* gene cluster (38,39). In particular, $\alpha 3$, $\alpha 5$, and $\beta 4$ *nAChR* subunits (e.g., *CHRNA5/A3/B4*) encoded by the *nAChR* gene cluster are differentially expressed in SCLC (40). There is even an emerging consensus to classify SCLC subtypes based on the differential expression of four transcription factors: *ASCL1* (SCLC-A), *NEUROD1* (SCLC-N), *POU2F3* (SCLC-P), and *YAP1* (SCLC-Y) (41). Similarly, *ASCL1* and *CHRNA5/A3/B4* were found to be over-expressed in our cohort with a greater than 4-fold change in expression, and their differential expression changes were greater in LNM patients than in those without metastasis. Chan *et al.* found that SCLC-A subtype showed less T-cell dysfunction than SCLC-N subtype using single-cell transcriptome sequencing. SCLC subtypes have the potential of plasticity and mutual transformation, especially SCLC-A and -N subtypes. In addition, SCLC showed greater immune isolation and less immune invasion than lung adenocarcinoma (42). Transcriptome data were also used to analyze and compare DEGs in the N0 patients and N+ patients. Several DEGs that have been reported to be associated with tumor metastasis were differentially expressed between the N+ and N0 groups, including *MAGEA4*, *FOXI3*, and *TRHDE*. As a cancer/testis antigen (CTA), *MAGEA4* is overexpressed in a variety of cancers. Previous studies have found that *MAGEA4* overexpression is associated with LNM of various tumors (43-45). Yoshida *et al.* found that *MAGEA4* expression was negatively correlated with survival in NSCLC patients (46). In our study, the presence of differentially expressed *MAGEA4* and *TRHDE* were also significantly associated with OS. Previous reviews indicated that the expression levels of *FOXI3* were higher in bone and lung metastases and LNM (47,48). The *FOXI3* target gene is also associated with breast cancer metastasis (49). In addition, Mukherjee *et al.* analyzed the expression data of the time from occurrence of primary breast tumors to distant metastasis and found that tumors with the highest expression of *FOXI3* characteristic genes metastasized earlier than those with the lowest expression (47). Similarly, *FOXI3* as a DEG was higher in the N+ patients but not differentially expressed in the N0 patients, supporting previous reports. Previous study has indicated that in patients with stage IA lung adenocarcinoma, especially when

CEA level is over 5 ng/dL and SUVmax is over 5, the risk of lymph node metastasis is higher, indicating the predictive factors of lymph node metastasis in clinical stage IA lung adenocarcinoma (50). Here we identify factors associated with SCLC lymph node metastasis at the molecular level by integrating genomic and transcriptomic data, which is the first report to our knowledge.

Our study had several limitations that should be noted. First, our sample size was small. Thus, it is necessary to verify the results of this study in the future with more patients and a longer follow-up period, as well as functional studies of key genes. Second, because this was a retrospective study, there was a possibility of bias in the case selection and limited medical information about the patients. Third, our study lacked proteomics and epigenetics analysis, which may provide more molecular information related to SCLC with LNM. In addition, the mechanism of early sensitivity and potential resistance to radiotherapy or chemotherapy for SCLC is still unclear, which needs to be further explored with a larger sample size in the future. Despite these limitations, this is the first integrative genomics profiling of LNM in SCLC and our results provide a comprehensive view of the genomics characteristics of LNM in SCLC. Our findings are particularly important for early detection and the provision of reliable therapeutic targets, although these targets still require extensive studies to be validated in the future.

Conclusions

We used WES and RNA sequencing to identify SMGs and DEGs related to LNM in SCLC and to further evaluate the link between these variants and clinical characteristics and prognosis. Our findings are particularly important for early detection of metastases in SCLC and the provision of reliable therapeutic targets.

Acknowledgments

We thank Shanghai Tongshu Biotechnology Co., Ltd. for technical support. The authors also appreciate the academic support from the AME Lung Cancer Collaborative Group.

Funding: None.

Footnote

Reporting Checklist: The authors have completed the STREGA reporting checklist. Available at <https://tlcr.amegroups.com/article/view/10.21037/tlcr-22-785/rc>

[amegroups.com/article/view/10.21037/tlcr-22-785/rc](https://tlcr.amegroups.com/article/view/10.21037/tlcr-22-785/rc)

Data Sharing Statement: Available at <https://tlcr.amegroups.com/article/view/10.21037/tlcr-22-785/dss>

Conflicts of Interest: All authors have completed the ICMJE uniform disclosure form (available at <https://tlcr.amegroups.com/article/view/10.21037/tlcr-22-785/coif>). All authors report technical support from Shanghai Tongshu Biotechnology Co., Ltd. SKJ receives consulting fees from Merck & Co, Inc., Novocure, Radiologica, and grants from Beigene and Merck & Co, Inc., and is a reviewer for IMX Medical, on the adjudication committee for Syntactx and participates in a DMSB for Advarra. The authors have no other conflicts of interest to declare.

Ethical Statement: The authors are accountable for all aspects of the work in ensuring that questions related to the accuracy or integrity of any part of the work are appropriately investigated and resolved. The study was conducted in accordance with the Declaration of Helsinki (as revised in 2013). The studies was reviewed and approved by the Institutional Review Board of Tongji Hospital of Huazhong University of Science and Technology (No. TJ-IRB20220329). The patients provided their written informed consent to participate in this study.

Open Access Statement: This is an Open Access article distributed in accordance with the Creative Commons Attribution-NonCommercial-NoDerivs 4.0 International License (CC BY-NC-ND 4.0), which permits the non-commercial replication and distribution of the article with the strict proviso that no changes or edits are made and the original work is properly cited (including links to both the formal publication through the relevant DOI and the license). See: <https://creativecommons.org/licenses/by-nc-nd/4.0/>.

References

1. Semenova EA, Nagel R, Berns A. Origins, genetic landscape, and emerging therapies of small cell lung cancer. *Genes Dev* 2015;29:1447-62.
2. Sundström S, Bremnes RM, Kaasa S, et al. Cisplatin and etoposide regimen is superior to cyclophosphamide, epirubicin, and vincristine regimen in small-cell lung cancer: results from a randomized phase III trial with 5 years' follow-up. *J Clin Oncol* 2002;20:4665-72.
3. Rossi A, Di Maio M, Chiodini P, et al. Carboplatin- or

- cisplatin-based chemotherapy in first-line treatment of small-cell lung cancer: the COCIS meta-analysis of individual patient data. *J Clin Oncol* 2012;30:1692-8.
4. Kalemkerian GP, Schneider BJ. Advances in Small Cell Lung Cancer. *Hematol Oncol Clin North Am* 2017;31:143-56.
 5. Gazdar AF, Bunn PA, Minna JD. Small-cell lung cancer: what we know, what we need to know and the path forward. *Nat Rev Cancer* 2017;17:725-37.
 6. Wang Y, Zheng Q, Jia B, et al. Effects of Surgery on Survival of Early-Stage Patients With SCLC: Propensity Score Analysis and Nomogram Construction in SEER Database. *Front Oncol* 2020;10:626.
 7. Girard JP, Moussion C, Förster R. HEVs, lymphatics and homeostatic immune cell trafficking in lymph nodes. *Nat Rev Immunol* 2012;12:762-73.
 8. Takahashi T, Yamanaka T, Seto T, et al. Prophylactic cranial irradiation versus observation in patients with extensive-disease small-cell lung cancer: a multicentre, randomised, open-label, phase 3 trial. *Lancet Oncol* 2017;18:663-71.
 9. Bi MM, Shang B, Wang Z, et al. Expression of CXCR4 and VEGF-C is correlated with lymph node metastasis in non-small cell lung cancer. *Thorac Cancer* 2017;8:634-41.
 10. Takanami I. Overexpression of CCR7 mRNA in nonsmall cell lung cancer: correlation with lymph node metastasis. *Int J Cancer* 2003;105:186-9.
 11. Wang Z, Lu B, Sun L, et al. Identification of candidate genes or microRNAs associated with the lymph node metastasis of SCLC. *Cancer Cell Int* 2018;18:161.
 12. Chen B, Li R, Zhang J, et al. Genomic Landscape of Metastatic Lymph Nodes and Primary Tumors in Non-Small-Cell Lung Cancer. *Pathol Oncol Res* 2022;28:1610020.
 13. Kim KB, Dunn CT, Park KS. Recent progress in mapping the emerging landscape of the small-cell lung cancer genome. *Exp Mol Med* 2019;51:1-13.
 14. Yang D, Denny SK, Greenside PG, et al. Intertumoral Heterogeneity in SCLC Is Influenced by the Cell Type of Origin. *Cancer Discov* 2018;8:1316-31.
 15. Zhang J, Fujimoto J, Zhang J, et al. Intratumor heterogeneity in localized lung adenocarcinomas delineated by multiregion sequencing. *Science* 2014;346:256-9.
 16. Feng ZX, Zhao LJ, Guan Y, et al. Identification of risk factors and characteristics of supraclavicular lymph node metastasis in patients with small cell lung cancer. *Med Oncol* 2013;30:493.
 17. Martínez-Jiménez F, Muiños F, Sentís I, et al. A compendium of mutational cancer driver genes. *Nat Rev Cancer* 2020;20:555-72.
 18. Bailey MH, Tokheim C, Porta-Pardo E, et al. Comprehensive Characterization of Cancer Driver Genes and Mutations. *Cell* 2018;174:1034-5.
 19. Mermel CH, Schumacher SE, Hill B, et al. GISTIC2.0 facilitates sensitive and confident localization of the targets of focal somatic copy-number alteration in human cancers. *Genome Biol* 2011;12:R41.
 20. Dobin A, Davis CA, Schlesinger F, et al. STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* 2013;29:15-21.
 21. Anders S, Pyl PT, Huber W. HTSeq—a Python framework to work with high-throughput sequencing data. *Bioinformatics* 2015;31:166-9.
 22. Newman AM, Liu CL, Green MR, et al. Robust enumeration of cell subsets from tissue expression profiles. *Nat Methods* 2015;12:453-7.
 23. Louhimo R, Hautaniemi S. CNAmets: an R package for integrating copy number, methylation and expression data. *Bioinformatics* 2011;27:887-8.
 24. Kanwal M, Smahel M, Olsen M et al. Aspartate beta-hydroxylase as a target for cancer therapy. *J Exp Clin Cancer Res* 2020;39:163.
 25. Cheng X, Yin H, Fu J, et al. Aggregate analysis based on TCGA: TTN missense mutation correlates with favorable prognosis in lung squamous cell carcinoma. *J Cancer Res Clin Oncol* 2019;145:1027-35.
 26. Kim YA, Madan S, Przytycka TM. WeSME: uncovering mutual exclusivity of cancer drivers and beyond. *Bioinformatics* 2017;33:814-21.
 27. Hofree M, Shen JP, Carter H, et al. Network-based stratification of tumor mutations. *Nat Methods* 2013;10:1108-15.
 28. Ahn JW, Kim HS, Yoon JK, et al. Identification of somatic mutations in EGFR/KRAS/ALK-negative lung adenocarcinoma in never-smokers. *Genome Med* 2014;6:18.
 29. Yu J, Wu WK, Li X, et al. Novel recurrently mutated genes and a prognostic mutation signature in colorectal cancer. *Gut* 2015;64:636-45.
 30. Jiang L, Huang J, Higgs BW, et al. Genomic Landscape Survey Identifies SRSF1 as a Key Oncodriver in Small Cell Lung Cancer. *PLoS Genet* 2016;12:e1005895.
 31. Huan C, Xiaoxu C, Xifang R. Zinc Finger Protein 521, Negatively Regulated by MicroRNA-204-5p, Promotes Proliferation, Motility and Invasion of Gastric Cancer Cells. *Technol Cancer Res Treat* 2019;18:1533033819874783.

32. Wang T, Lv X, Jiang S, et al. Expression of ADAM29 and FAM135B in the pathological evolution from normal esophageal epithelium to esophageal cancer: Their differences and clinical significance. *Oncol Lett* 2020;19:1727-34.
33. Auerbach AD. A test for Fanconi's anemia. *Blood* 1988;72:366-7.
34. Lowe SW, Ruley HE, Jacks T, et al. p53-dependent apoptosis modulates the cytotoxicity of anticancer agents. *Cell* 1993;74:957-67.
35. Vasey PA, Jones NA, Jenkins S, et al. Cisplatin, camptothecin, and taxol sensitivities of cells with p53-associated multidrug resistance. *Mol Pharmacol* 1996;50:1536-40.
36. Lin X, Howell SB. DNA mismatch repair and p53 function are major determinants of the rate of development of cisplatin resistance. *Mol Cancer Ther* 2006;5:1239-47.
37. Jackman DM, Johnson BE. Small-cell lung cancer. *Lancet* 2005;366:1385-96.
38. Improgo MR, Schlichting NA, Cortes RY, et al. ASCL1 regulates the expression of the CHRNA5/A3/B4 lung cancer susceptibility locus. *Mol Cancer Res* 2010;8:194-203.
39. Improgo MR, Scofield MD, Tapper AR, et al. The nicotinic acetylcholine receptor CHRNA5/A3/B4 gene cluster: dual role in nicotine addiction and lung cancer. *Prog Neurobiol* 2010;92:212-26.
40. Improgo MR, Scofield MD, Tapper AR, et al. From smoking to lung cancer: the CHRNA5/A3/B4 connection. *Oncogene* 2010;29:4874-84.
41. Rudin CM, Poirier JT, Byers LA, et al. Molecular subtypes of small cell lung cancer: a synthesis of human and mouse model data. *Nat Rev Cancer* 2019;19:289-97.
42. Chan JM, Quintanal-Villalonga Á, Gao VR, et al. Signatures of plasticity, metastasis, and immunosuppression in an atlas of human small cell lung cancer. *Cancer Cell* 2021;39:1479-1496.e18.
43. Laban S, Giebel G, Klümper N, et al. MAGE expression in head and neck squamous cell carcinoma primary tumors, lymph node metastases and respective recurrences-implications for immunotherapy. *Oncotarget* 2017;8:14719-35.
44. Sani SA, Forghanifard MM, Sharifi N, et al. Investigation of melanoma-associated antigen A4 cancer/testis antigen clinical relevance in esophageal squamous cell carcinoma. *J Cancer Res Ther* 2018;14:1059-64.
45. Forghanifard MM, Gholamin M, Farshchian M, et al. Cancer-testis gene expression profiling in esophageal squamous cell carcinoma: identification of specific tumor marker and potential targets for immunotherapy. *Cancer Biol Ther* 2011;12:191-7.
46. Yoshida N, Abe H, Ohkuri T, et al. Expression of the MAGE-A4 and NY-ESO-1 cancer-testis antigens and T cell infiltration in non-small cell lung carcinoma and their prognostic significance. *Int J Oncol* 2006;28:1089-98.
47. Mukherjee A, Hollern DP, Williams OG, et al. A Review of FOXP3 Regulation of Development and Possible Roles in Cancer Progression and Metastasis. *Front Cell Dev Biol* 2018;6:69.
48. Haider M, Zhang X, Coleman I, et al. Epithelial mesenchymal-like transition occurs in a subset of cells in castration resistant prostate cancer bone metastases. *Clin Exp Metastasis* 2016;33:239-48.
49. Györfy B, Lanczky A, Eklund AC, et al. An online survival analysis tool to rapidly assess the effect of 22,277 genes on breast cancer prognosis using microarray data of 1,809 patients. *Breast Cancer Res Treat* 2010;123:725-31.
50. Ye B, Cheng M, Li W, et al. Predictive factors for lymph node metastasis in clinical stage IA lung adenocarcinoma. *Ann Thorac Surg* 2014;98:217-23.

(English Language Editor: J. Jones)

Cite this article as: Kong K, Hu S, Yue J, Yang Z, Jabbour SK, Deng Y, Zhao B, Li F. Integrative genomic profiling reveals characteristics of lymph node metastasis in small cell lung cancer. *Transl Lung Cancer Res* 2023;12(2):295-311. doi: 10.21037/tlcr-22-785

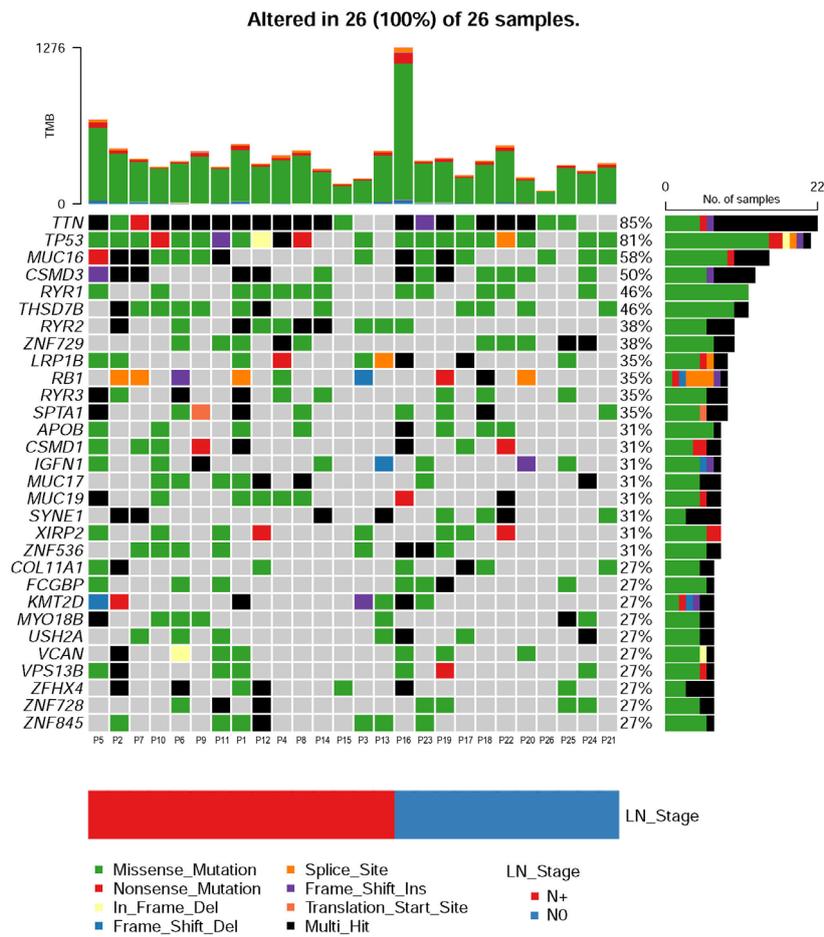


Figure S1 Landscape of somatic mutations in each group. The most frequently mutated genes in this cohort were shown. The top panel represents the TMB in each sample. The middle panel represents the matrix of alterations in a selection of frequently mutated genes. Columns represent samples. Clinicopathological characteristics of the LN stage are presented below. TMB, tumor mutational burden; LN, lymph node.

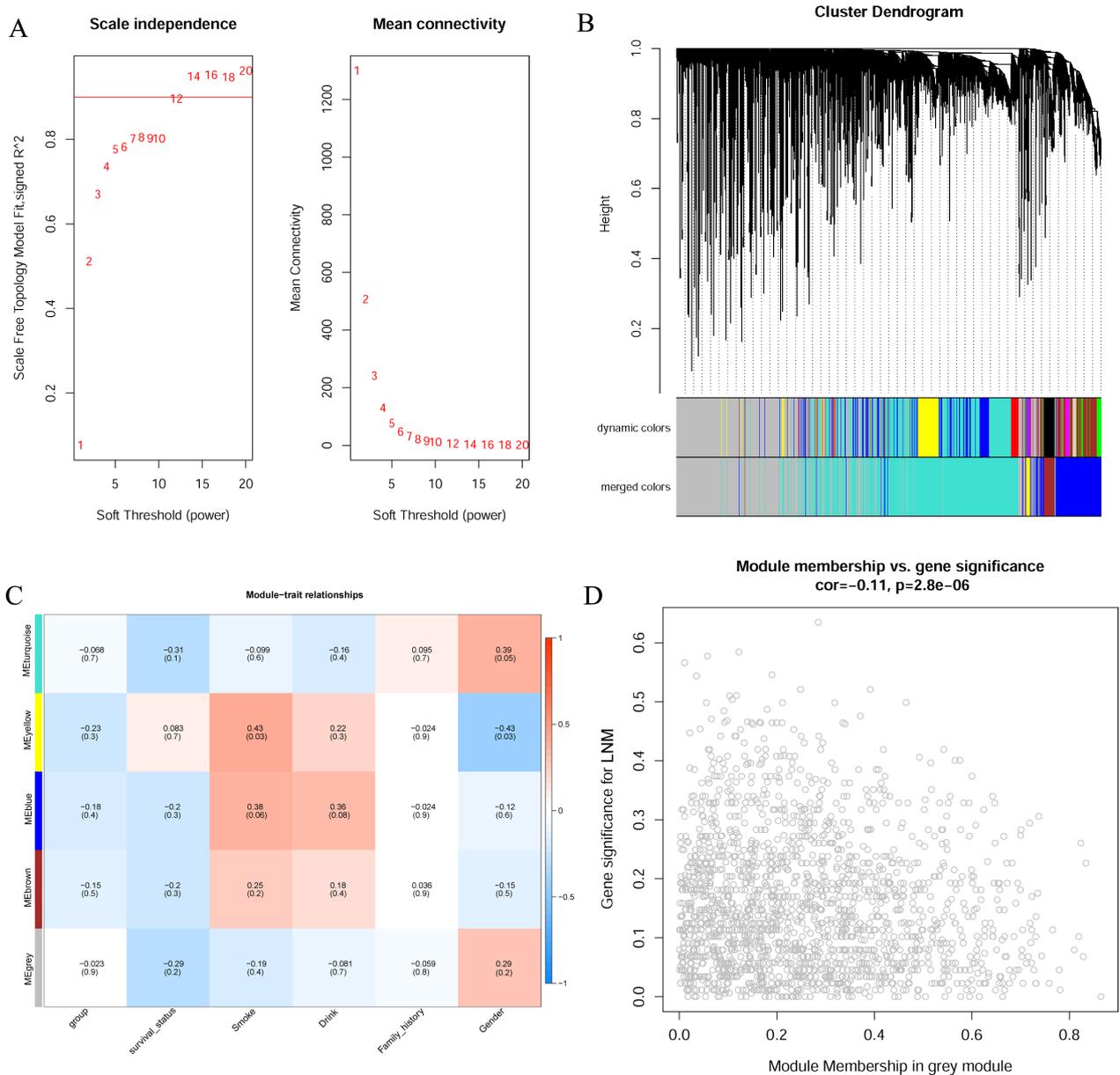


Figure S2 Construction of co-expression network through WGCNA. (A) Network topology for different soft-thresholding powers. (B) A cluster diagram of gene cluster of SCLC. (C) Heatmap of the correlation between module eigengenes and the clinical features. (D) The correlation of LNM-related module membership and gene significance. WGCNA, weighted gene co-expression network analysis. LNM, lymph node metastasis.

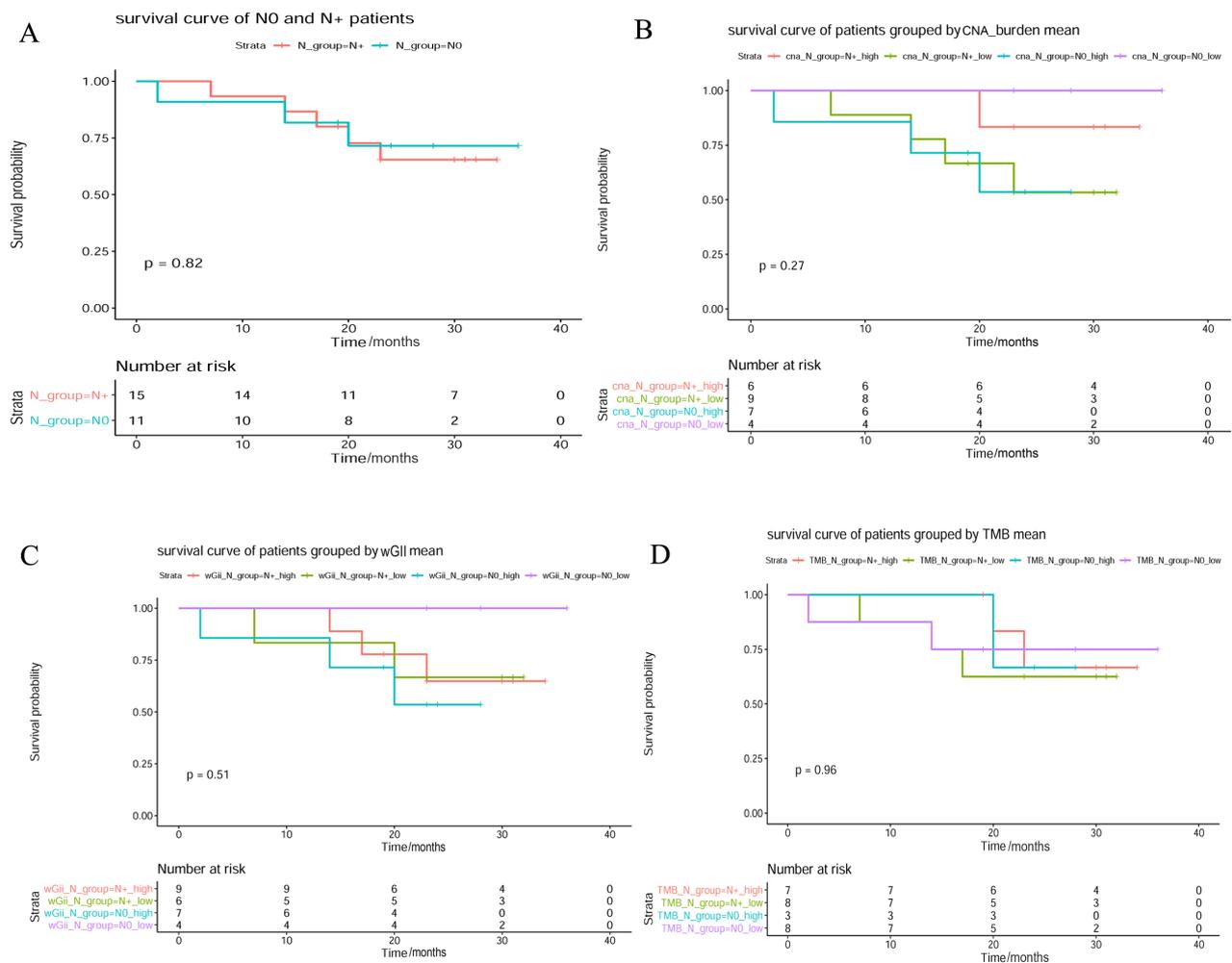


Figure S3 Overall survival analysis. (A) Survival analysis of N0 and N+; (B) survival analysis of CNA burden in N0 and N+ patients; (C) survival analysis of wGII in N0 and N+ patients; (D) survival analysis of TMB in N0 and N+ patients. CNA, copy number alteration; wGII, weighted Genome Instability Index; TMB, tumor mutational burden.

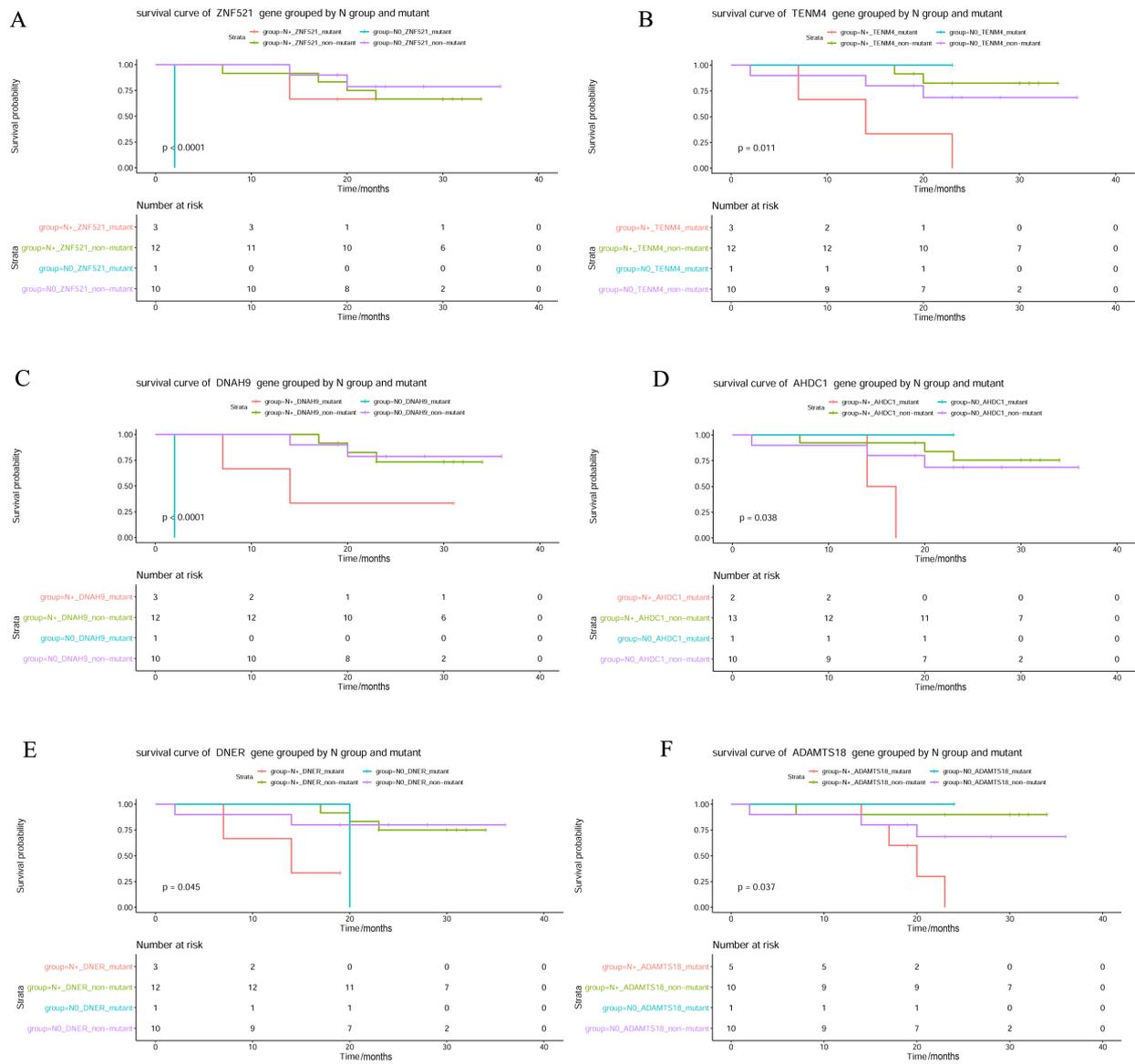


Figure S4 Overall survival analysis. (A) Survival analysis of ZNF521 mutation in N0 and N+ patients; Survival analysis of TENM4, DNAH9, AHDC1, DNER, and ADAMTS18 mutation (B-F) in N0 and N+ patients.

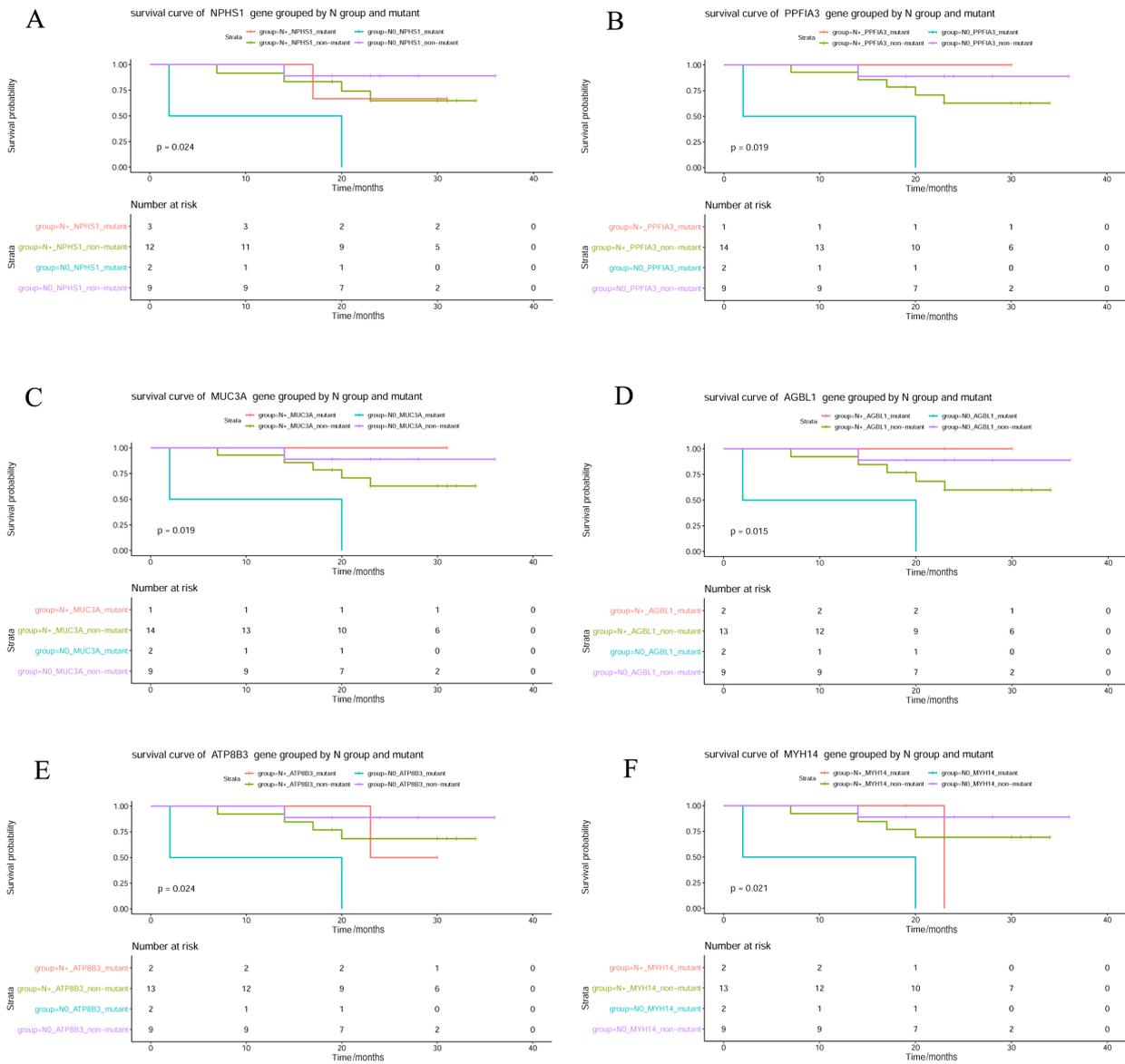


Figure S5 Overall survival analysis. Survival analysis of *NPHS1*, *PPFIA3*, *MUC3A*, *AGBL1*, *ATP8B3*, and *MYH14* mutation (A-F) in N0 and N+ patients.

Table S1 Comparison of CNV in N0 and N+ SCLC

	N0			N+				
	Cytoband	q value	Residual q value	Wide peak boundaries	Cytoband	q value	Residual q value	Wide peak boundaries
	9p24.3	1.15E-18	1.15E-18	chr9:107381-179012	16p11.2	3.19E-25	3.19E-25	chr16:31917855-46467006
	22q11.1	1.15E-18	1.15E-18	chr22:1-17048481	22q11.1	3.19E-25	3.19E-25	chr22:1-17048418
	9q13	1.15E-18	3.71E-11	chr9:47307204-70928318	9q13	4.35E-25	5.98E-18	chr9:47307203-70928318
	7p13	1.19E-18	1.19E-18	chr7:43982540-44095178	2q12.2	4.68E-25	4.68E-25	chr2:106800892-107094587
	16p11.2	1.19E-18	1.19E-18	chr16:31927566-46467006	12q24.31	7.71E-25	7.71E-25	chr12:125296404-125408863
	12q24.31	1.64E-18	2.35E-15	chr12:125297491-125407110	17q21.2	1.95E-24	1.95E-24	chr17:39183147-39427187
	2q11.2	1.79E-18	1.79E-18	chr2:97771742-98215296	1q21.2	3.69E-24	2.56E-16	chr1:147456277-149709685
	1q21.2	7.00E-18	9.37E-14	chr1:147994869-149859509	14q11.2	1.40E-23	1.40E-23	chr14:1-20181853
	17q12	1.67E-17	1.67E-17	chr17:36275041-36463073	7q22.1	1.89E-22	1.89E-22	chr7:100635464-100656953
	14q11.2	3.41E-17	3.41E-17	chr14:1-20181853	7p13	4.71E-22	4.71E-22	chr7:43982540-44095178
	11q14.3	3.49E-16	3.49E-16	chr11:89522931-89712272	11q14.3	1.33E-21	1.33E-21	chr11:89443434-89711422
	7q22.1	5.62E-16	5.62E-16	chr7:101952394-101996816	1p21.1	1.41E-21	1.78E-21	chr1:104083769-104761379
	2p11.2	7.46E-16	7.46E-16	chr2:88043238-88337138	13q21.1	1.52E-21	1.52E-21	chr13:57704909-57903580
	13q21.1	9.39E-16	9.39E-16	chr13:57705133-57903414	18q21.1	1.52E-21	1.52E-21	chr18:44517288-44569285
	13q21.1	9.39E-16	1	chr13:1-115169878	9p24.3	2.58E-21	2.58E-21	chr9:107381-179012
	10q23.2	1.47E-14	3.46E-07	chr10:88760075-88782161	16q23.1	4.57E-19	4.57E-19	chr16:74419231-74455225
	6p25.2	1.91E-14	1.91E-14	chr6:3152707-3236338	2p11.2	8.97E-18	8.97E-18	chr2:88043238-88337271
	8p23.1	4.06E-14	4.06E-14	chr8:6784636-7282345	6p25.2	1.12E-17	1.12E-17	chr6:3152707-3236275
	19p12	7.07E-14	7.07E-14	chr19:22145150-22166723	18p11.21	2.48E-17	2.48E-17	chr18:14132216-14758294
Gain	10q22.2	7.07E-14	0.000222	chr10:75405781-75515680	15q11.2	4.58E-17	4.58E-17	chr15:1-22378561
	18q21.1	1.18E-13	1.18E-13	chr18:44534810-44569285	10q23.2	1.60E-15	1.60E-15	chr10:88760075-88791677
	12p13.2	1.46E-13	1.46E-13	chr12:11420126-11557260	12p13.2	1.60E-15	1.60E-15	chr12:11496049-11685746
	16q23.1	3.64E-13	3.64E-13	chr16:74415497-74455225	10q23.2	1.60E-15	1	chr10:1-135534747
	18p11.21	5.40E-13	5.40E-13	chr18:14502027-14758294	1q21.1	6.29E-15	0.011336	chr1:121474967-144611963
	1q21.1	1.74E-11	0.11796	chr1:121474974-144614416	19p12	4.82E-14	4.82E-14	chr19:22145150-22166061
	15q11.2	2.53E-11	2.53E-11	chr15:20737152-22378437	8p23.1	7.24E-14	7.24E-14	chr8:6784636-7282306
	3q29	8.27E-11	8.27E-11	chr3:195296404-195456623	21q22.3	1.43E-13	1.43E-13	chr21:45944022-46140892
	1p21.1	2.27E-10	0.0001026	chr1:104083752-104162410	9q21.32	2.95E-13	2.45E-09	chr9:84413337-84608030
	10p11.21	6.24E-10	6.24E-10	chr10:37426483-37500112	5q35.2	4.84E-12	4.84E-12	chr5:175394899-175540875
	5p15.33	7.66E-10	7.66E-10	chr5:741667-850724	5p15.33	1.11E-11	1.11E-11	chr5:741667-850724
	21q22.3	1.62E-09	1.62E-09	chr21:45943807-46076172	10p11.21	8.34E-10	1.19E-08	chr10:37426483-37486430
	9q21.32	1.73E-09	4.70E-05	chr9:84413337-84571467	10p11.21	8.34E-10	1	chr10:1-135534747
	19q13.31	1.18E-08	1.18E-08	chr19:43218317-43717356	3q29	7.24E-08	7.24E-08	chr3:195296864-195456623
	1p36.13	6.64E-08	0.0013874	chr1:16775662-17215903	19q13.31	0.00061854	0.00061854	chr19:43218317-43702466
5q35.2	4.93E-07	4.93E-07	chr5:175394899-175541064	11p11.12	0.01835	0.01835	chr11:48501609-49178190	
5q35.2	4.93E-07	1	chr5:1-180915260	4p16.1	0.051361	0.051361	chr4:9153195-9793628	
12q12	2.78E-05	0.057356	chr12:33550349-39074501	8q24.21	0.14024	0.14024	chr8:127846011-130761793	
11p11.12	0.0012849	0.0012849	chr11:48501415-49168521					
8q12.3	0.057356	0.057356	chr8:62550476-63171905					
4p16.1	0.21674	0.21674	chr4:9154130-9410691					

Table S1 (continued)

Table S1 (continued)

	N0				N+			
	Cytoband	q value	Residual q value	Wide peak boundaries	Cytoband	q value	Residual q value	Wide peak boundaries
	16q12.2	1.07E-13	1.07E-13	chr16:56226267-56294594	16q22.2	1.05E-20	1.05E-20	chr16:56196234-88872136
	9q21.33	3.42E-13	3.42E-13	chr9:89772196-90341960	6p21.33	2.11E-20	2.11E-20	chr6:28851652-33382945
	16p13.3	1.51E-12	1.51E-12	chr16:2153702-2169209	16p13.3	2.11E-20	2.11E-20	chr16:2142906-2168486
	6p21.32	1.03E-11	1.03E-11	chr6:32066515-33382945	7p21.3	4.10E-17	4.10E-17	chr7:7308189-7681446
	7p21.3	1.10E-11	1.06E-11	chr7:7313638-7592328	9q21.33	1.00E-16	1.09E-16	chr9:89771576-90341960
	15q21.2	1.40E-11	1.40E-11	chr15:51519055-51535657	10p14	6.01E-16	6.01E-16	chr10:6883574-7601373
	5q23.3	4.04E-11	4.04E-11	chr5:129090788-130496432	19p13.11	2.00E-13	1.98E-13	chr19:18122653-18228056
	10q26.3	1.87E-10	1.81E-10	chr10:135028776-135053367	22q13.33	3.05E-13	1.13E-12	chr22:50687350-50750508
	10p14	7.17E-09	7.17E-09	chr10:6883574-7601895	3q13.31	1.38E-12	1.38E-12	chr3:113528789-113677406
	11q13.4	1.46E-08	1.46E-08	chr11:72315314-72342211	10q24.32	1.64E-12	1.64E-12	chr10:103922569-104011301
	19p13.11	3.29E-08	3.16E-08	chr19:18121090-18228345	21q22.3	7.11E-10	7.11E-10	chr21:46646003-47610594
	12q23.1	1.91E-07	1.91E-07	chr12:96340707-96404735	17p13.2	4.46E-09	4.46E-09	chr17:3553132-3714472
	3q13.31	2.43E-07	2.43E-07	chr3:113597195-113677359	12q23.1	2.76E-08	2.76E-08	chr12:96311085-96394765
	22q13.33	1.09E-06	1.07E-06	chr22:50687350-50719421	8p21.3	1.16E-07	1.16E-07	chr8:21166235-21766771
	3p21.31	1.75E-06	1.75E-06	chr3:50248198-50316035	1p36.32	4.62E-07	4.62E-07	chr1:2341792-2461229
	21q22.3	9.66E-06	9.67E-06	chr21:46646003-47610504	5q32	2.70E-14	6.57E-07	chr5:149465024-149676755
	7q22.1	3.32E-05	3.32E-05	chr7:100808810-100883075	13q14.11	1.16E-06	1.16E-06	chr13:40174636-40765798
	12p12.3	0.0001999	0.0002052	chr12:14719948-14923682	2p23.3	1.26E-06	1.26E-06	chr2:26357529-26531400
	17q11.2	5.99E-06	0.0002516	chr17:27879812-27920755	11q13.4	1.67E-06	1.68E-06	chr11:72145557-72396776
	17p11.2	2.33E-07	0.0002977	chr17:18061962-18164521	12p12.3	2.11E-05	2.08E-05	chr12:14706357-14923682
	1p36.32	0.0004121	0.0004356	chr1:2341911-2461229	15q14	2.11E-05	2.08E-05	chr15:34490977-34651296
Loss	19q13.12	0.0004401	0.0004356	chr19:35422473-35725649	9p13.2	2.57E-05	2.56E-05	chr9:37592485-37771784
	2p23.3	0.0005916	0.0005916	chr2:26404841-26539970	20q13.33	5.68E-05	5.68E-05	chr20:60892413-60908102
	1q42.13	0.0015564	0.0014458	chr1:228375245-228581878	5q23.3	5.20E-16	0.00014408	chr5:129090958-130502899
	8q23.3	0.0015398	0.0014458	chr8:113645048-113674856	3p21.1	0.00022158	0.00021336	chr3:52414398-52530528
	5p13.3	0.0018373	0.0018151	chr5:31918938-31948832	5p13.3	0.00029203	0.00027496	chr5:31926019-31955247
	9p13.2	1.02E-07	0.0024337	chr9:37745977-37790444	14q32.33	0.00039547	0.00038518	chr14:105753235-105895234
	20p13	0.0033853	0.0032572	chr20:3215366-3458645	7q22.1	0.00044295	0.00044295	chr7:100808810-101014359
	20q13.33	0.0038791	0.0039992	chr20:62180738-62279698	1q42.2	0.00063422	0.00063511	chr1:231357049-231471514
	13q21.32	0.0043252	0.0042037	chr13:67370724-67517500	2q11.2	0.00073236	0.00071167	chr2:101767032-102603229
	2q21.3	0.017652	0.017391	chr2:136631169-136872405	6q16.3	0.0049843	0.0049064	chr6:101328965-105389199
	6q16.3	0.022785	0.022552	chr6:101327299-105388592	19q13.42	0.0050347	0.0049064	chr19:55284532-55587667
	9p13.2	1.02E-07	0.022552	chr9:37582748-37790444	20p13	0.027174	0.027174	chr20:3213061-3451889
	8p12	0.027744	0.028298	chr8:30231258-30601690	17q11.2	1.48E-05	0.034344	chr17:27869849-27953970
	14q32.33	0.030971	0.030971	chr14:105842533-105895202	17q12	0.0021862	0.034344	chr17:38030908-38105010
	4p16.3	0.044324	0.044095	chr4:476415-686055	18q12.2	0.060671	0.060734	chr18:34388917-36793563
	11p15.5	0.055869	0.055869	chr11:1-3242809	8q12.3	0.00045153	0.088859	chr8:62623338-63973276
	17p13.1	9.67E-06	0.11086	chr17:9684951-10204333	4p14	0.14042	0.1356	chr4:39780579-40046173
	4q35.1	0.18979	0.1905	chr4:184236428-190949996	4q13.2	0.14859	0.14777	chr4:69207411-69513266
	18p11.21	0.22781	0.22359	chr18:11649717-11983108	22q11.21	0.002507	0.1654	chr22:19707589-19868126
	17q11.2	5.99E-06	0.71063	chr17:1-81195210	8q12.3	0.00045153	0.18993	chr8:62623338-63973276
	2q21.3	0.017652	1	chr2:1-243199373				
	16q12.2	1.07E-13	1	chr16:1-90354753				

"Red" represents private chromosomes with CNVs in N0 or N+ group. "Bold" indicates shared chromosomes with different CNV in N0 and N+ group. CNV, copy number variation; SCLC, small cell lung cancer.