Neoantigen Load as a Prognostic and Predictive Marker for Stage II/III Non-Small Cell Lung Cancer in Chinese Patients

Hui Li
Zhejiang cancer hospital

Ronghui He
Hangzhou YITU Healthcare Technology Co, Ltd

Yanjun Xu
Zhejiang Cancer Hospital

Taobo Luo
zhejiang cancer hospital

Kaixiu Jin
Hangzhou YITU Healthcare Technology Co, Ltd

Wuzhou Yuan
Mingma Technologies Co., Ltd

Zengguang Zheng
zhejiang cancer hospital

Lanxuan Liu
Hangzhou YITU Healthcare Technology Co, Ltd

Zebin Liang
Hangzhou YITU Healthcare Technology Co, Ltd Mingma Technologies Co., Ltd

Ao Li
Mingma Technologies Co., Ltd

Zhiguo Zheng
zhejiang cancer hospital

Yun Fan (fanyun@zjcc.org.cn)
Zhejiang cancer hospital

Research

Keywords: NSCLC, whole exome sequencing, neoantigen load, biomarker, prognosis

Posted Date: December 21st, 2020

DOI: https://doi.org/10.21203/rs.3.rs-130886/v1
Abstract

Background

Prognosis of stage II/III non–small-cell lung cancer (NSCLC) is unsatisfactory even after complete tumor resection and adjuvant chemotherapy. Tumors with high immunogenicity were defined as “hot tumors” and associated with clinical benefits from immunotherapy. Here we assessed the prognostic and predictive value of immunogenomic signatures and gene mutation characters for stage II/III non-small cell lung cancer in Chinese patients.

Methods

Ninety-one paired resected stage II/III NSCLC and normal tissues and peripheral blood samples, including 47 squamous cell lung carcinomas (SCC) and 44 lung adenocarcinomas (ADC), were collected and analyzed using the whole exome sequencing (WES) to identify gene mutation and immunogenomic signatures for association with clinicopathological variables and disease-free survival (DFS).

Results

A high number of tumor mutation burden (TMB, > 4 mutations/Mb) was associated with better DFS of NSCLC patients, although there was no such an association in SCC and ADC subgroups. Moreover, higher neoantigen load (NAL, > 2 neoantigens/Mb) exhibited better DFS and survival benefit after adjuvant chemotherapy in a low NAL subgroup of SCC patients but not in ADC subgroup. A high DNA damage repair (DDR) index (gene mutations occurred in at least three different DNA repair pathways) was associated with high NAL numbers and favorable DFS of SCC, but not in ADC patients. However, mutations of individual gene, oncogene pathways, and antigen presentation machinery genes, and human leukocyte antigen (HLA)-I number and HLA-I loss of heterozygosity (LOH) had no prognostic or predictive value for DFS of SCC or ADC patients.

Conclusion

Given the present information, NAL was a useful biomarker for lung SCC prognosis and prediction of chemotherapy responses in Chinese patients. Further study with a larger sample size from multiple institutions is needed to confirm these data.

Background

Lung cancer is still the most significant health burden in the world and in China, accounting for more than two million new cases and more than 333,000 cancer-related deaths in 2018 globally [1] and more than 733,000 new cases and more than 610,000 cancer-related deaths in 2015 in China [2]. Histologically, lung cancer can be divided into small cell lung cancer and non-small cell lung cancer (NSCLC), while approximately 85% of all lung cases are NSCLC and the latter can be further classified into lung adenocarcinoma (ADC), squamous cell carcinoma (SCC), large-cell lung carcinoma, and unclassified
carcinoma [3]. Currently, adjuvant chemotherapy of NSCLC patients after completely tumor resection is a standard treatment and improves approximately 5% five-year survival of patients [4]. However, development and identification of various biomarkers could help us to appraise clinical outcomes of different treatment options and distinguish the benefit population from adjuvant chemotherapy in management of NSCLC patients [5]. Indeed, accumulating evidence suggests that precise treatment strategy possessed a great potential to improve survival of patients with resected NSCLC using the molecular signatures; for example, both ADAURA and CTONG1104 studies demonstrated that adjuvant therapy of NSCLC with the epidermal growth factor receptor (EGFR) tyrosine kinase inhibitor (EGFR-TKI) significantly improved disease-free survival (DFS) of patients with EGFR-mutant NSCLC [6]. The recent development of immune checkpoint inhibitors, like targeting of programmed death-1 (PD-1) / programmed death ligand-1 (PD-L1), has revolutionized our control of NSCLC with an undeniable efficacy [7]. Moreover, different exogenous mutagens and distinct genomic alterations between ADC and SCC could also differ from the distinct survival outcomes of treatment selections [8]. Thus, further exploration of the genomic signatures in different pathological NSCLC types and evaluation of their impact on clinical outcome by identification of the high-risk subgroup in disease recurrence or treatment responses could lead to better therapy selections and survival benefit to our patients.

Evading immune attack has been determined as a key hallmark of cancer and immune escape has been detected in early stage of NSCLC [9]. The absence of immunogenicity was the first step of immune evasion and strongly associated with clinical outcomes of patients [10]. The tumor mutation burden was characterized as an indicator of immunogenicity given that the gene mutations can result in production of neoantigens to be specifically expressed on tumor cells rather than normal cells [11]. Several studies focused on the effect of TMB on the clinical benefits for resected early stage NSCLC. Couples of studies showed that a high TMB associated with a favorable outcome in patients with resected NSCLC [12] but other researchers reported that a high TMB was a poor prognostic factor in heterogeneous stages, histology and ethnic groups [13] or TMB was not associated with overall survival of early staged NSCLC at all [14], implying that TMB was not a perfect surrogate of immunogenicity in NSCLC. However, using the computational tools to predict the tumor neoantigens based on the WES date have been proved as a meaningful and potential method [15]. McGranahan et al. showed that neoantigen burden was associated with overall survival in ADC but not SCC cohorts of patients from The Cancer Genome Atlas (TCGA) [16], although Rosenthal et al. revealed that a higher clonal neoantigen burden was associated with better DFS in ADC and SCC of the TRACERx cohort [17]. Thus, these controversial results indicated that the function of immunogenicity-mediated immune surveillance had not been comprehensively studied in the resected NSCLC, especially in Chinese NSCLC patients, while the assessment of clinical and survival impact of these distinct mutation context and immunogenic signatures on different pathological types could be useful clinically.

In this study, we assessed gene mutation characteristics using the whole exome sequencing (WES) to identify individual, oncogenic pathways, DNA repair pathways, and antigen presentation machinery genes, human leukocyte antigen (HLA) -I numbers and loss of heterozygosity (LOH), TMB numbers and neoantigen load in stage II/III NSCLC samples and then evaluated them as prognostic and predictive
biomarkers for NSCLC patients. We expected to provide a novel insight into biomarker discovery and identification for prognosis and prediction of treatment responses NSCLC patients.

## Materials And Methods

### Patients and specimens

In this retrospective study, we collected a total of 91 patients with completely resected stage II/III NSCLC (47 SCCs and 44 ADCs according to the AJCC 8th TNM staging manual) from Zhejiang Cancer Hospital (Hangzhou, China) between August 26, 2006 and September 28, 2014. No patients received neoadjuvant chemotherapy. Tissue samples from tumor and matched normal lung were snap-frozen and peripheral blood samples were taken for this study. All patients were followed up regularly in Zhejiang Cancer Hospital until recurrence or last follow-up (July 2020). The specimens comprised at least 30% of tumor cells were involved in this study assessed by histological examination for each tissue sample. Any tumor recurrence was evaluated by two different radiologists for each patient. The detailed clinicopathological features and treatment selections as well as the follow-up data were retrieved from patients’ medical records and analyzed. This study was approved by our Institutional Review Board of Zhejiang Cancer Hospital and each participant provided a written informed consent form before enrolled into this study.

### DNA extraction and WES

Genomic DNA was extracted from blood samples using the Magnetic Genomic DNA Kit (Tiangen, Beijing, China) according to the manufacturer’s protocol and from FFPE of tumor and normal samples using an internally modified magnetic extraction protocol. DNA was randomly broken into 150-200 bp using the ultrasonicator (Covaris, Woburn, MA, USA). The sheared DNA samples were subjected to the end repair step using the Agilent SureSelectXT Low Input Reagent Kit (Agilent, Santa Clara, CA, USA), followed by addition of a base A to the 3’ end to form a sticky end. After that, these DNA fragments were ligated with specific barcode adapter sequences and removed any incompletely ligated fragments using the magnetic beads. The resulted samples were subjected to a PCR amplification using universal primers that were complementary with adapter sequences to build a DNA sequencing library. The probes from the Agilent SureSelectXT All Human Exome library were utilized to capture target fragments from these samples and the library concentrations were assessed by using Qubit (Thermo Fisher, Waltham, MA, USA), while the library size was measured by using the Agilent TapeStation. The Illumina HiSeq X Ten system (San Diego, CA, USA) was used to sequence these samples according to the Illumina’s instructions for 2 × 150 paired-end sequencing by the Mingma Technologies Co., Ltd. (Shanghai, China).

### Assessment of somatic gene mutation and significantly mutated genes (SMG)

After WES of these paired samples, we determined somatic mutation variations using previously reported common algorithms. In particular, the raw reads of the WES were filtered using the skewer (v0.2.2) software according to a previous study [18], while the high-quality reads were matched to the UCSC human reference genome (hg19) with the Burrows-Wheeler Aligner (BWA v.0.7.12) accordingly with
default parameters [19]. However, all N-read bases in the reference genome were excluded for calculation of the coverage ratio. Tumor samples were sequenced with an average coverage of 387.6 x, while normal samples were with 117.2 x. We then further performed the de-duplication, indel realignment, and base quality recalibration using the Sentieon algorithm (https://www.sentieon.com/). For each sample, we assessed the somatic single nucleotide variations (SNVs) and InDels with the Sentieon TNseq according to a previous study [20]. We then filtered out genomic mutations in a low complexity region, such as the tandem repeat regions or highly homologous regions in the genome and obtained the repeat regions from the UCSC genome browser (http://genome.ucsc.edu/) and annotated all high-confident mutations with ANNOVAR (Version 2016-02-01) (http://annovar.openbioinformatics.org/) [21]. After that, we assessed the SMGs using MutSigCV (v1.4) software (https://www.genepattern.org/modules/docs/MutSigCV) [22] with the default setting. We defined a gene with the false discovery rate (FDR) < 0.05 to be significantly mutated.

**Calculation of TMB number**

To calculate TMB number, we first filtered the somatic SNVs and InDels from the WES of our paired samples and matched the variants to one of the following criteria, i.e., 1). Variant allele frequency < 5%; 2). Inclusion of the 1000 G, ExAC03 and dbSNP151 database with exceptions for the variants with a high count (≥ 4 observances) in the COSMIC database; 3). Synonymous variants; or 4). Exclusion of the coding sequence of any genes (CDS). The TMB number was calculated using the formula, i.e., \( \text{TMB} = \frac{M}{L} \) (M represents number of filtered variants, while L represents length of the covered CDS region).

**Prediction of HLA and neo-antigen**

We predicted the neo-antigens using a method modified from pVACseq according to a previous study [20], i.e., the HLA type of each paired samples was assessed using the in-house fq2HLA software analysis of the high quality reads from the normal samples. The unique HLA allele number was then counted as the HLA number, while the LOH of HLA was calculated according to a previous study [23]. Somatic SNVs and InDels were used for the HLA binding affinity prediction using the netMHCpan (v4.0) software (https://services.healthtech.dtu.dk/service.php?NetMHCpan-4.0). Peptides with an affinity < 500 nM were considered to be a neo-antigen and the number of the neoantigen load (NAL) of each paired sample was calculated using the formula, i.e., \( \text{NAL} = \frac{N}{L} \) (N represents number of neoantigen, while L represents length of the covered CDS regions).

**Statistical analysis**

We performed Pearson’s \( \chi^2 \) test to associate clinicopathological variables from patients with altered genetic features and then defined the DFS as time from surgery to tumor recurrence, death of any causes, or the last follow up of surviving patients. The Kaplan-Meier method and the log-rank test were performed to associate survival of patients with genomic alterations, like TMB number. The hazard ratio was analyzed to evaluate gene alteration as the relative hazard for patients with at least one gene mutation compared with none. The Cox regression model was used to predict the role of gene mutations in
treatment responses. After that, we performed univariate and multivariate analyses to assess gene alterations as the prognostic predictors of these NSCLC patients. A \( P \leq 0.05 \) was considered statistically significant.

**Results**

**Clinicopathological characteristics of patients**

A total of 91 patients were included in our study and the snap-frozen tumor tissue samples and matched normal peripheral blood were subjected to the Whole-exome sequencing. The tumor tissues were sequenced at a greater depth relative to paired normal tissues with the targeted sequences read an average of 387.6x and 117.2x, respectively. Here no patients received neoadjuvant chemotherapy before surgery. This study comprised of 47 SCCs and 44 ADCs and their median age was 59 years old (ranged between 30 and 71 years old) with 70 male patients (76.7%) and 46 stage II (50.5%) and 45 stage III (49.5%). Forty-three SCC patients were former or current smokers (93.6%) vs. twenty ADC patients (45.5%), while 61.4% (27/44) of patients had an EGFR mutation in ADC, but only 4.26% (2/47) had an EGFR mutation in SCC patients. In terms of treatment, 73 patients (73/91, 80.2%) received adjuvant chemotherapy after surgery for at least two cycles. The median DFS of SCC and ADC were 69 and 17.5 months, respectively. These clinicopathological data are listed in Table 1.

**Association of genomic signatures and frequently mutated pathways with ADC and SCC outcomes**

The prevalence of C>T transitions was observed in ADC and C>A transitions in SCC\(^1\) (Fig 1A). An enrichment of C>A transversions was associated with smoking status that was observed in other cancers for which smoking was a significant risk factor. In SCC, the most frequent mutations were \( TP53 (41/47, 87.2\%) \), \( TTN (38/47, 80.8\%) \), \( CSMD3 (248/47, 51.1\%) \), \( KMT2D (20/47, 42.6\%) \), \( RYR2 (19/47, 40.4\%) \) and \( CDKN2A (19/47, 40.4\%; \text{Fig 1B}) \). \( EGFR (27/44, 61.4\%) \), \( TP53 (25/44, 56.8\%) \), \( BCLAF1 (16/44, 36.4\%) \), \( RHPN2 (13/44, 30.0\%) \), \( SIGLEC10 (12/44, 27.3\%) \) and \( ANKRD38C (11/44, 25.0\%) \) were more frequent in ADC subgroup (Fig 1C). Most patients (44/47, 93.6%) were former or current smokers in SCC subgroup and almost half of patients had smoking histories in ADC subgroup, thus we detected the genomic characters within smokers and non-smokers in ADC subgroup. Frequency of C>T transition was detected in non-smokers compared to the enrichment of C>A transversion in smokers (Supplemental Fig. 1A). We also found that \( TP53 (12/20, 60.0\%) \), \( BCLAF1 (8/20, 40.0\%) \), \( EGFR (8/20, 40.0\%) \), \( RHPN2 (7/20, 35.0\%) \), and \( USH2A (7/20, 35.0\%) \) were frequent in tobacco smokers, whereas \( EGFR (19/24, 79.2\%) \), \( TP53 (13/24, 54.2\%) \), \( BCLAF1 (8/24, 33.3\%) \), \( ORTE24 (7/24, 29.2\%) \), \( ERLL11 (7/24, 29.2\%) \), and \( SIGLEC10 (7/24, 29.2\%) \) were common in patients without smoking history (Supplemental Fig. 1B). There were 61.4% of patients with an \( EGFR \) mutation (vs. 27% \( EGFR \) mutations in Caucasian NSCLC patients)\(^5\) and 4% of ADC patients with an \( K-ras \) mutation (vs. 32% of Caucasian patients\(^5\)), suggesting an extraordinary distinct genotypes in different races of patients\(^\text{24}\). Moreover, the C>T was the most frequent type of transitions in ADC patients, whereas the C>A transversion occurred more frequently in non-\( EGFR \) mutated NSCLC patients (Supplemental Fig. 1C). The most comment co-mutations in ADC patients with \( EGFR \)
were TP53 (16/27, 59.3%), BCLAF1 (12/27, 44.4%), RHPN2 (9/27, 33.3%) and ASTE1 (8/27, 29.6%), whereas the high frequent mutations in ADC without EGFR mutation were TP53 (9/17, 52.9%), SIGLEC10 (7/17, 41.1%), DST (5/17, 29.4%), LRPIB (5/17, 29.4%), MUC16 (5/17, 29.4%), NAV3 (5/17, 29.4%), OBSCN (5/17, 29.4%), TRIM48 (5/17, 29.4%), TTN (5/17, 29.4%), ZFHX3 (5/17, 29.4%), and ZFHX4 (5/17, 29.4%; Supplemental Fig. 1D).

We then analyzed the association of individual gene mutations with DFS of patients and found that mutation of NELL1, HERC2, LTN1, CYHR1, MUC5B, CUBN, OR4C15, PDE4DIR PI3KCA, NBPF10, EYS, and GPR32 was inversely associated with DFS of SCC patients, whereas USH2A mutation associated with better DFS in SCC patients (Supplemental Fig. 2A). Moreover, mutation of RELN, HMCN1, OR2L8, and NALCN was associated with better DFS in ADC patients, whereas mutated MUC5B was associated with poor DFS in ADC patients (Supplemental Fig. 2B), although there was no association of any mutated genes with DFS after correction of the false discovery rate with the multi-tests.

Furthermore, we assessed the ten oncogenic signaling pathways that were genetically altered at high frequency in ADC and SCC[25]. We found that there were more mutated genes involved in these oncogenic signaling pathways in SCC compared to those of ADC. Among them, the RTK-RAS (82% in SCC and 84% in ADC) and TP53 pathways (93% in SCC and 59% in ADC) were frequently altered. Moreover, ARHGAP35 (5/47, 10.6%), FLT3 (5/47, 10.6%), IRS2 (4/47, 8.5%), KSR2 (4/47, 8.5%), and NF1 (4/47, 8.5%) in the RTK-RAS signaling were mutated in SCC, whereas alterations of EGFR (27/44, 61.4%), IRS2 (3/44, 6.8%), ARHGAP35 (2/44, 4.5%), ROS1 (2/44, 4.5%), ERBB4 (2/44, 4.5%) and RASGRP4 (2/44, 4.5%) occurred more frequently in ADC. The HIPPO and NOTCH signaling were the third and fourth signaling pathways that were enriched with genetic variants in NSCLC in our study. In the HIPPO pathways, the most frequent mutated genes were FAT1 (12/47, 25.5%), FAT3 (9/47, 19.1%), CRB1 (8/47, 17.0%), DSCHS1 (7/47, 14.9%), and FAT4 (7/47, 14.9%) in SCC, whereas mutation of FAT3 (6/44, 13.6%), HMCN1 (6/44, 13.6%), DCHS2 (5/44, 11.4%), FAT2 (3/44, 6.8%), and DCHS1 (2/44, 4.5%) occurred in ADC (Fig. 2C). In the NOTCH pathways, predominate alterations of FBXW7 (7/47, 14.9%), SPEN (5/47, 10.6%), and CNTN6 (5/47, 10.6%) occurred in SCC, whereas NUMBL (3/44, 6.8%), MAML3 (3/44, 6.8%), and THBS2 (3/44, 6.8%) were mutated in ADC (Fig. 2E). In addition, variants of the RTK-RAS and TP53 signaling pathways were predominant in non-smoking ADC patients and smoking ADC patients, respectively (Supplemental Fig. 3A). However, a quarter of patients (5/20, 25.0%) harbored variants of the TGF-β signaling pathway genes in ADC with smoking history, but only one patient (1/24, 4.2%) had mutated genes in the TGF-β signaling pathway in non-smoking ADC subgroup (Supplemental Fig. 3A). Again, ADC patients with wild type EGFR had a higher ration with variant in the WNT signaling genes (7/17, 41.2% vs. 7/17, 25.9% in EGFR-mutated ADC) and MYC signaling genes (3/17, 17.6% vs. 2/27, 7.3% in EGFR-mutated ADC)(Supplemental Fig. 3B). After that, we analyzed the associations of each pathway gene mutations with survival of patients, but unfortunately, we did not find any association of the altered single oncogenic pathway with DFS in both SCC and ADC (Supplemental Fig. 4A and B).

**Association of a comprehensive immunogenomic profiling with DFS of ADC and SCC patients**
After that, we performed a comprehensive analysis for the immunogenic profiling using the WES data for both ADC and SCC patients, including HLA-I number, HLA LOH, TMB, DNA repair pathway, and antigen presentation machinery and the calculation to predict the neoantigen to determine the HLA-binding affinity (< 500 nM of peptides derived from somatic SNVs and Indels). We found a similar HLA-I distribution of allele frequency net of China Jiangsu Han (HLA-A/B, n=3238) & China South Han pop 2 (HLA-C, n=1098; Fig. 3A). Most patients had six HLA-I loci in both SCC (32/47, 68.1%) and ADC (34/44, 77.3%), while near a half of SCC patients had HLA LOH, but only a third ADC patients had HLA LOH (Fig. 3B). Moreover, there was no significant difference in HLA numbers and HLA LOH from *EGFR* mutation or tobacco smoking, respectively (Supplemental Fig. 5A-B). Our survival analysis revealed that neither HLA number nor HLA LOH status was associated with DFS of SCC or ADC patients (Supplemental Fig. 5C-D and Table 2).

The DNA damage repair (DDR) signaling includes eight pathways, i.e., the check point factors (CPF); homologous recombination repair (HRR); Fanconi anemia (FA); nucleotide excision repair (NER); mismatch repair (MMR); base excision repair (BER); nonhomologous end-joining (NHEJ); and DNA translation synthesis (TLS)[26]. Genomic variants in CPF (93% in SCC and 61% in ADC) and the HRR pathway (43% in SCC and 25% in ADC) were more obvious in both SCC and ADC. Specifically, only two SCC patients had no mutation in the DDR pathway, whereas one SCC patient had mutations in seven pathways of the DDR pathway. Overall, there was as high as 42.6% of SCC patients harboring genomic alterations in at least three pathways of the DDR. In contrast, ten ADC patients (10/44, 22.7) had no DDR related mutations and a quarter of ADC patients had deficiency in at least one DDR pathway (Fig. 4C), and mutations in at least two DDR pathways were observed in 45.5% ADC patients (20/44; Fig. 4D). Indeed, SCC patients showed a higher prevalence of gene alterations in the DNA repair pathway and the antigen presentation machines compared to those of ADC patients (Fig. 4C). Nevertheless, the frequency of the mutated genes participated in the DNA repair pathway and antigen presentation machinery were equally distributed in ADC patients with or without *EGFR* mutations (Supplemental Fig. 6A). However, interestingly, patients with tobacco smoking history had a distinct distribution of gene mutations involved in the DDR signaling, and serious deficiency of the DDR signaling was associated with TMB and NAL (Supplemental Fig. 6B).

The deficiency in the antigen presentation plays a key role in impairment of tumor neoantigen production and contributes to immune escape in lung cancer [23]. In this study, we explored the mutations of the antigen presentation machinery (APM)-related genes for association with prognosis of these Chinese lung cancer patients, but there was no positive or useful finding (Supplemental Fig. 7C-D). When tumors were stratified into dichotomy (low DDR index with gene mutations in < 3 pathways vs. high DDR index with gene mutations in ≥ 3 pathways), the low DDR index was associated with poor DFS of SCC patients (Fig. 4A and Table 2), although the DDR index did not have a predictable value for chemotherapy of SCC patients (Fig. 4A). However, interestingly, the DDR index was significantly associated with prediction of the neoantigen load, although there was no correlation observed between the DDR index and TMB in SCC patients, suggesting a key role of the DDR index in neoantigen production in SCC (Supplemental Fig. 8A). However, the HLA number and LOH were no differences between the low and high DDR index groups.
Supplemental Fig. 8B). Furthermore, we observed a significant difference in TMB and NAL between low and high DDR index groups in ADC patients (Supplemental Fig. 8C), although there was no association of the DDR index in prediction of DFS in ADC (Supplemental Fig. 8D). There was also no significant difference in HLA number and LOH between low and high DDR indexes in ADC (Supplemental Fig. 8E).

Next, we found an average of 4.48 somatic mutations per Mb and 2.44 predicted neo-antigens identified in SCC samples, whereas 2.28 somatic mutations and 0.99 predicted neoantigens per Mb in ADC. Specifically, SCC had a higher level of TMB and NAL than those of ADC (Fig. 3C). However, there was no significant difference in TMB and NAL between mutated and wild type EGFR in ADCs (TMB, 0.46 mutations per Mb to 32.3 mutations per Mb; the median=1.66 mutations per Mb; NAL, 0.26 neoantigens per Mb to 21.82 neoantigens per Mb with the median of 0.98 neoantigens per Mb), but a narrow range of TMB and NAL in EGFR-mutated ADC (TMB, 1.11 mutations per Mb to 5.27 mutations per Mb with the median of 2.31 mutations per Mb; NAL, 0.52 neoantigens per Mb to 1.64 neo-antigens per Mb with a median of 1.45 neoantigens per Mb) (Supplemental Fig. 6B and Supplemental Fig. 9B). Similarly, there was no significant difference in TMB and NAL observed between former or current smokers and non-smokers (Supplemental Fig. 6A and Supplemental Fig. 9A).

A high TMB number per Mb (i.e., 4 mutations per Mb) was associated with better DFS of these 91 NSCLC patients (Supplemental Fig. 9C). However, there was no association observed between TMB number and DFS of ADC (Supplemental Fig. 9D) or SCC patients (Supplemental Fig. 9E). We then used NAL number (> 2 neoantigens per Mb as a cut-off point) and further analyzed these data and still not find any statistical significance in ADC (Supplemental Fig. 9F). However, there was an association of low NAL number with poor DFS (months, Hazard Ratio(HR)=2.56, 95% CI: 1.15-5.68, p=0.021) in SCC patients (Fig 4B and Table 2). Our multivariate analysis also showed that NAL number was an independent prognostic predictor for SCC patients, while the DDR index was not included in the multivariate analysis considering the strong association between the DDR index and NAL (see supplemental Table 1). Furthermore, we also found the benefit of the adjuvant chemotherapy in improvement of DFS in SCCs with a lower NAL number (Fig 4B).

These results allowed us to explore the detailed association of DDR pathway and NAL numbers in SCC. As shown in Fig 4C, patients with high NAL numbers were enriched in high DDR index group and CPF, FA, and HRR pathways were the most frequently mutated among the high DDR index and NAL groups in SCC, indicating that the DNA damage repair pathway contributed to the neoantigen productions and neoantigen-directed immune surveillance favored SCC patients. However, the immune escape may be adapted by other immune escape mechanisms in ADC patients (Fig. 4C). In addition, we analyzed the different effects of TMB and NAL number in NSCLC and found that one half of oncogenic mutations did not create neoantigen, and indels variants created 1.75 folds neoantigens compared to SNV in our study among ADC and SCC, suggesting TMB number was not a good surrogate marker of the immunogenic neoantigen (Fig. 4D).

In addition, we assessed the different genomic features stratified by NAL numbers in SCC. Frequency of C>A transition was detected in high NAL compared to the enrichment of C>T transversion in low NAL number of SCC patients (Supplemental Fig. 10A). Our data showed that the frequent mutations enriched
in the high NAL SCC subgroup were TP53 (26/28, 92.9%), TTN (25/28, 89.3%), CSMD3 (20/28, 71.4%), USH2A (15/28, 53.6%), RYR2 (14/28, 50.5%), and CDKN2A (13/28, 46.4%; Supplemental Fig. 10B), whereas mutations of TP53 (17/19, 89.5%), TTN (13/19, 68.4%), KMT2D (7/19, 36.8%), BCLAF1 (6/19, 31.6%), FLG (6/19, 31.6%), and LRP1B (6/19, 31.6%) were more frequent in the low NAL SCC subgroup (Supplemental Fig. 10B). Among the different signaling pathways, the RTK-RAS and TP53 pathways exhibited frequent alterations in both high and low NAL SCC cases. Interestingly, the higher frequent alteration of these ten signal pathways was enriched in the high NAL SCC group compared to that of the low NAL SCC group (Supplemental Fig. 10C and Supplemental Table 2). However, there was no significant difference in HLA number and LOH between high and low NAL SCC groups (Supplemental Fig. 10D and Supplemental Table 2). In line with previous results, patients with high NAL presented more frequent mutations involved in DNA damage repair pathway, and positively associated with DDR index, suggesting the deficiency in the DDR signaling contributed to production of neoantigens (Supplemental Fig. 10E and Supplemental Table 2). However, mutations involved in antigen presentation machine were equally in high and low neoantigen burden subgroup, indicating the deficiency of processing of antigens may be achieved by other mechanisms rather than gene mutations (supplement Fig. 10F and Supplemental Table 2).

In summary, predicted neoantigen load acted as a more useful indicator of immunogenicity than TMB, and provided effective stratification variable in prognosticating disease outcome and benefits from adjuvant chemotherapy for patients with lung squamous cell carcinoma.

**Discussion**

Lung cancer is still the significantly global healthy problem and adjuvant chemotherapy a standard therapy strategy for NSCLC, only improved the 5-year survival by 5% generally [4]. Recent utilization of the immune checkpoint blockage treatment promoted the overall survival of advanced NSCLC patients [7]. However, further research will help us to successfully conquer this deadly disease. In one hand, we need novel agents and treatment options to effectively control lung cancer progression, while in other hand, we will discovery, identify, and evaluate novel biomarkers to predict prognosis of patients and treatment responses, which will selectively treat the responded patients continuously and other treatment options for those that don't respond well to a given therapy. Towards this end, identification of comprehensively genomic signatures and gene alterations could be useful biomarkers to predict NSCLC survival and treatment responses. Thus, our current study performed the WES analysis to identify and evaluate different genomic alterations in order to provide biomarkers for NSCLC. The results from this study identified a larger numbers of gene mutations and alterations in these NSCLC sample. However, the high TMB number was able to predict better DFS of these NSCLC patients, but not for each of SCC and ADC subgroups of patients, which may be due to decrease in sample size, depending further study. Moreover, a higher NAL number was able to better predict favorable DFS and successful adjuvant chemotherapy in SCC patients (but not for ADC patients). In addition, the high DNA damage repair (DDR) index also predicted a high NAL number and favorable DFS of SCC patients (also not for ADC patients). However, we failed to find any predicting values of mutations of individual, oncogene pathway, and antigen
presentation machinery genes, and HLA-I number and LOH for DFS of SCC or ADC patients or treatment responses. In conclusion, our current data revealed analysis of NAL number as a biomarker for lung SCC prognosis and prediction of chemotherapy responses. Future prospective study with larger samples from multiple institutions is needed to confirm our current data.

In our current study, we identified the gene mutational landscape in Chinese NSCLC cases. Consistent with other previous studies [27], TP53 was shown to be the most frequently mutated gene in lung SCC and secondary in lung ADC, while EGFR was the most commonly mutated gene in lung ADC. As we know, lung cancer development is involved in accumulation of large gene mutations and epigenetic alterations to activate oncogenes and inhibit tumor suppressor genes [28]. Our current study indeed showed different gene mutations, e.g., TTN, CSMD3, KMT2D, RYR2, and CDKN2A in SCC, while PLLAF1, RHPN2, SIGLEC10 and ANKRD38C in ADC samples. Our pathway analysis identified alterations of multiple gene pathways in these NSCLC samples; for example, the RTK-RAS-RAF signaling pathway was the most aberrant signaling in ADC, while the TP53 pathway alteration occurred in SCC. Moreover, the DNA repair pathway and antigen presentation machinery was shown a higher frequency of mutations in SCC than in ADC, whereas the HLA-I number was similar between ADC and SCC, although SCC samples had a higher trend of HLA-I LOH than in ADC, all of which were not shown any associations with survival of ADC or SCC patients. Previous studies also revealed the alterations of similar genes in NSCLC [28] and the frequency of EGFR mutation in Caucasians and East Asians was quite different. Approximately 7% overall, 13% among patients with adenocarcinoma and 35% among never smokers with EGFR mutation among Caucasians, whereas the average mutation prevalence was approximately 30% overall, 47% among lung ADC patients and 56% in never smokers [24]. K-ras has also been demonstrated different characteristics between NSCLC patients among Asian and Caucasian populations. Approximately 26% of Caucasian ADC patients had K-ras mutations [29], but the K-ras mutation prevalence was usually less than 10% in Eastern Asian NSCLC patients [30]. Interestingly, ADC patients with EGFR mutation have significantly lower levels of TMB compared with that of EGFR wild-type patients, while the same trend was not observed in NSCLS patients with K-ras mutation [31]. Thus, the prevalence of critical driver genes such as EGFR and K-ras contributes to the different tumor burden distribution and survival outcomes in Asian and Caucasian populations, which lead to the complicated correlation between TMB and clinical survival. Second, genomics studies have showed distinct genomic alterations of ADC and SCC. K-ras, STK11, EGFR, PTPTD, and STK11 were frequent mutated in lung ADC patients, while high prevalence of CDKN2A, KMT2D, NFE2L2, LRRK2, and PTEN mutations were frequently observed in lung SCC patients [32]. These distinct genomic mutations between ADC and SCC may be due to different risk factors or environmental exposure factors, like tobacco smoke[33] as predictors for survival of patients [32].

In our current study, we found that a higher TMB number was associated with better DFS in NSCLC patients but not in ADC or SCC subgroup of NSCLC patients. Our current study developed a DDR index according to the numbers of mutated DDR pathway genes to associate with DFS of NSCLC patients, especially in ADC patients. In addition, our current study focused on NAL for association with DFS of NSCLC patients and found positive data on SCC patients. Obviously, mutations of the DNA repair pathway genes, such as high microsatellite instability (MSI-H), led to defects in mismatch repair (MMR-D)
or co-mutations in two DNA damage pathways and increase in overall genomic instability, which was shown to associate with improved sensitivity to the immune checkpoint blockade (ICB)-based immunotherapy of different human cancers[34]. The mismatch repair deficiency, which is associated with high TMB, was a positive indicator for prognosis and negative biomarker for benefit from adjuvant chemotherapy[35]. Nivolumab plus Ipilimumab treatment showed much better progression-free survival of NSCLC patients with the high TMB numbers[36]. TMB number was approved by US Food and Drug Administration (FDA) according to the KEYNOTE-158 study to predict a successful immunotherapy as a biomarker[37]. Several other previous studies focused on TMB for prediction of the clinical benefits of treatment of resected early stage NSCLC, e.g., Siddhartha et al. showed that a high TMB (> 8 mutations per Mb) was associated with a favorable outcome [overall survival (OS), disease-free survival (DFS), and lung cancer-specific survival] of patients with resected NSCLC after adjuvant chemotherapy[16]. However, controversially, Owada-Ozaki et al. reported that a high TMB number (> 62 mutations per Mb) was associated with poor NSCLC prognosis in Japanese patients[9]. Chun et al. showed that lung adenocarcinoma patients with low TMB numbers had better DFS than those with high TMB in Korean patients[38]. The TMB number was an independent predictor for favorable prognosis of lung ADC patients from Europe[39]. Low TMB number was associated with a longer DFS of Chinese lung adenocarcinoma patients[40], but not with the OS of Chinese lung SCC patients[41] or of early-stage lung SCC of Caucasian patients[42]. Our current data also confirmed that use of TMB isn't sufficient to predict prognosis of NSCLC.

Indeed, other recent studies using the WES and the computational analysis were able to utilize the NAL as a biomarker to predict and assess cancer immunity and immunotherapy in human cancers[15]. Specifically, tumor neoantigens have been defined as peptides that are derived from somatic gene mutations and express in tumor cells but not in normal cells that can be recognized by tumor-infiltrating lymphocytes (TILs)[43]; thus, NAL is an ideal surrogate of immunogenicity. Indeed, a high NAL number was able to predict response of cancer patients to the immune checkpoint blockade (ICB)-immunotherapy[44], while detection of the neoantigen-MHC complex rather than mutated genes could help to initiate the anti-tumor immune response[45] and frequent recognition of neoantigens by CD4+ T cells occurred in human melanoma[46]. Moreover, various mutation types contributed to different neo-antigen productions; for example, the probability of the indels that altered a given gene open-reading frame could generate a neo-antigen was three folds higher than that of non-synonymous SNV[47]; thus, use of NAL number to predict DFS could be much better than TMB numbers. Furthermore, in our study, we also found that neoantigen burden was a positive indicator of DFS in SCC but not ADC patients, suggesting that neoantigen-directed immune surveillance could depend on the genetic landscapes in a given tumor, such as tumor heterogeneity, ethnic groups, and environmental exposure factors. Our study further showed that patients with a low NAL SCC had a benefit from adjuvant chemotherapy. Indeed, neoantigen load acted as a useful indicator of immunogenicity than TMB and could provide a stratification variable for prognosis or outcome of patients with lung SCC cancer after adjuvant chemotherapy. In addition, tobacco smoker was a single most important risk factor for lung cancer[48]. Our current study revealed that the C>A transversion was associated with tobacco smoking of NSCLS
patients, while mutations of TP53, BCLAF1, EGFR, RHPN2, and USH2A also frequently occurred in tobacco smokers, although EGFR and TP53 mutation were also shown in NSCLC patients without tobacco use, for a reason that is not clear.

In addition, our current data showed that neither TMB or NAL have prognostic and predictive value in DFS of ADC patients, suggesting the complication of biological process of this subgroup. As high as 40-60% of ADC patients from East Asian countries had EGFR mutations and ADC with EGFR mutation exhibited different distribution of TMB compared to wild type EGFR NSCLC[31]. In our current study, we showed that there was no significant difference in TMB and NAL between EGFR wild type and mutated ADC. As we know, EGFR mutation was a negative predictor for the immune checkpoint inhibitors[49]. Thus, other genetic variants but not immunogenicity calculated by TMB or NAL responsible for prognosis should be explored in ADC patients with EGFR mutations. Our current data showed that there was no significant difference in TMB or NAL between EGFR-wild type ADC and SCC, suggesting that other immunogenomic variants may play a role in resected lung ADC without EGFR mutations.

However, our current study does have some limitations; for example, this study is just retrospective with a relatively small sample size. Considering the possibility for clinical decision-making, future validation is needed in an independent larger cohort of patients. Moreover, NAL used in our current study was just based on the HLA-I binding prediction; however, other processes involved in neo-antigen productions, like the processing and presentation of antigens, the stability of the MHC-peptide complex and immune recognition have not been calculated in our current prediction algorithms. The level of the mutation combined with HLA I binding could be better optimization or prediction of neo-antigen, but we didn't involve them in our current study. Further, the HLA binding was also restricted to HLA-I, but the HLA-II binding prediction was not included in our current algorithms. In addition, tissue samples used in the current study were freshly frozen tissues; however, use of formalin-fixed and paraffin-embedded tissues may be more useful in most of hospitals and be more clinical utility for broad applications in personalizing cancer treatment care. Last, but not least, our current study didn’t evaluate the overall survival (OS).

In conclusion, our current study demonstrated use of NAL number to predict DFS and treatment responses of SCC patients, which could be a more useful indicative marker for immunogenicity than TMB to provide an effective stratification as a prognostic marker for SCC outcome and treatment response. Further study with a larger size from multiple institutions is needed to confirm our current data.

**Conclusion**

In summary, our study illustrated that higher neoantigen load (NAL) exhibited better DFS for squamous cell lung carcinomas (SCC) but not lung adenocarcinomas (ADC) in Chinese patients. The benefit of the adjuvant chemotherapy in improvement of DFS were observed in SCC with a lower NAL but not ADC. Our current data revealed NAL as a biomarker for lung SCC prognosis and prediction of chemotherapy
responses. Further study with a larger sample size from multiple institutions is needed to confirm these data.

**Abbreviations**

**NSCLC**: non–small-cell lung cancer  
**SCC**: squamous cell lung carcinomas  
**ADC**: adenocarcinomas  
**WES**: whole exome sequencing  
**DFS**: disease-free survival  
**TMB**: tumor mutation burden  
**NAL**: neoantigen load  
**DDR**: DNA damage repair  
**HLA**: human leukocyte antigen  
**LOH**: loss of heterozygosity  
**EGFR**: epidermal growth factor receptor  
**EGFR-TKI**: EGFR tyrosine kinase inhibitor  
**PD-1**: programmed death-1  
**PD-L1**: programmed death ligand-1  
**TCGA**: The Cancer Genome Atlas  
**SMG**: significantly mutated genes  
**CDS**: coding sequence  
**CPF**: check point factors  
**HRR**: homologous recombination repair  
**NER**: nucleotide excision repair  
**FA**: Fanconi anemia
MMR: mismatch repair
BER: base excision repair
NHEJ: nonhomologous end-joining
TLS: DNA translation synthesis
APM: antigen presentation machinery
HR: Hazard Ratio
FDA: Food and Drug Administration
TILs: tumor-infiltrating lymphocytes
ICB: immune checkpoint blockade
OS: overall survival

Declarations

Acknowledgments

Not applicable.

Authors’ contributions

HL performed the research and wrote the manuscript. YJX, TBL and ZZG performed the research. RHH, KXJ, WZY, LXL, ZBL and AL performed the statistical analysis. ZGZ and YF designed and supervised the research project. The authors read and approved the final manuscript.

Funding

This work was supported by the National Natural Science Foundation of China (Grant No. 81972718), the Natural Science Foundation of Zhejiang Province (Grant No. LQ21H160004), and Zhejiang Medical technology program (Grant No. 202143913).

Availability of data and materials

All data generated or analyzed during this study are included in this published article.

Ethics approval and consent to participate

This study has been approved by the Ethics Committee and written informed consent was obtained from all participants.
Consent for publication

The authors declare that they consent for publication.

Competing interests

The authors declare that they have no competing interests.

Author details

1 Department of Medical Oncology, the Cancer Hospital of the University of Chinese Academy of Sciences (Zhejiang Cancer Hospital) Institute of Basic Medicine and Cancer (IBMC), Chinese Academy of Sciences, Hangzhou, Zhejiang 310022, China.

2 Hangzhou YITU Healthcare Technology Co, Ltd, Hangzhou, Zhejiang, 310000, China.

3 Department of Thoracic Surgery, the Cancer Hospital of the University of Chinese Academy of Sciences (Zhejiang Cancer Hospital) Institute of Basic Medicine and Cancer (IBMC), Chinese Academy of Sciences, Hangzhou, Zhejiang 310022, China.

4 Mingma Technologies Co., Ltd, Shanghai, 200131, China.

5 Department of Pathology, the Cancer Hospital of the University of Chinese Academy of Sciences (Zhejiang Cancer Hospital) Institute of Basic Medicine and Cancer (IBMC), Chinese Academy of Sciences, Hangzhou, Zhejiang 310022, China.

6 Zhejiang Cancer Research Institute, the Cancer Hospital of the University of Chinese Academy of Sciences (Zhejiang Cancer Hospital) Institute of Basic Medicine and Cancer (IBMC), Chinese Academy of Sciences, Hangzhou, Zhejiang 310022, China.

References


Tables
<table>
<thead>
<tr>
<th>Variables</th>
<th>SCC (n=47)</th>
<th>ADC (n=44)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>n (%)</td>
<td>n (%)</td>
</tr>
<tr>
<td><strong>Age (yrs.)</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>&gt; 65</td>
<td>5 (10.6)</td>
<td>7 (16.0)</td>
</tr>
<tr>
<td>≤ 65</td>
<td>42 (89.4)</td>
<td>37 (84.1)</td>
</tr>
<tr>
<td><strong>Gender</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Male</td>
<td>45 (95.7)</td>
<td>25 (56.8)</td>
</tr>
<tr>
<td>Female</td>
<td>2 (4.30)</td>
<td>19 (43.2)</td>
</tr>
<tr>
<td><strong>Tobacco smoking status</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Never smoker</td>
<td>3 (6.40)</td>
<td>24 (54.5)</td>
</tr>
<tr>
<td>Former/current smoker</td>
<td>44 (93.6)</td>
<td>20 (45.5)</td>
</tr>
<tr>
<td><strong>T stage</strong>*</td>
<td></td>
<td></td>
</tr>
<tr>
<td>T1-T2</td>
<td>27 (57.4)</td>
<td>32 (72.7)</td>
</tr>
<tr>
<td>T3-T4</td>
<td>20 (42.6)</td>
<td>12 (27.3)</td>
</tr>
<tr>
<td><strong>N stage</strong>*</td>
<td></td>
<td></td>
</tr>
<tr>
<td>N0-N1</td>
<td>40 (85.1)</td>
<td>21 (47.7)</td>
</tr>
<tr>
<td>N2</td>
<td>7 (14.9)</td>
<td>23 (52.3)</td>
</tr>
<tr>
<td><strong>Stage</strong>*</td>
<td></td>
<td></td>
</tr>
<tr>
<td>II</td>
<td>31 (66.0)</td>
<td>15 (34.1)</td>
</tr>
<tr>
<td>III</td>
<td>16 (34.0)</td>
<td>29 (65.9)</td>
</tr>
<tr>
<td><strong>EGFR status</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Wild type</td>
<td>45 (95.7)</td>
<td>17 (38.6)</td>
</tr>
<tr>
<td>Mutation</td>
<td>2 (4.26)</td>
<td>27 (61.4)</td>
</tr>
<tr>
<td><strong>Chemotherapy</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>None</td>
<td>12 (25.5)</td>
<td>6 (13.6)</td>
</tr>
<tr>
<td>Adjuvant chemotherapy</td>
<td>35 (74.5)</td>
<td>38 (86.4)</td>
</tr>
<tr>
<td><strong>HLA number</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>-------</td>
<td>-------</td>
<td>-------</td>
</tr>
<tr>
<td>6</td>
<td>32 (68.1)</td>
<td>34 (77.3)</td>
</tr>
<tr>
<td>≤ 5</td>
<td>15 (31.9)</td>
<td>10 (22.7)</td>
</tr>
<tr>
<td>HLA LOH</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Yes</td>
<td>24 (51.1)</td>
<td>14 (31.8)</td>
</tr>
<tr>
<td>No</td>
<td>23 (48.9)</td>
<td>30 (68.2)</td>
</tr>
<tr>
<td>TMB</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Median TMB (range)</td>
<td>4.8 (1.0-29.0)</td>
<td>2.3 (0.5-32.3)</td>
</tr>
<tr>
<td>NAL</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Median NAL (range)</td>
<td>2.4 (0.6-7.4)</td>
<td>1.0 (0.3-21.8)</td>
</tr>
</tbody>
</table>

NSCLC, non-small-cell lung cancer; SCC, squamous cell carcinoma; ADC, adenocarcinoma.

TMB, tumor mutation burden; NAL, tumor neo-antigen burden; LOH, loss of heterozygosity.

*Using the 8th TNM staging classification.
<table>
<thead>
<tr>
<th>Variables</th>
<th>SCC (n=47)</th>
<th>ADC (n=44)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>HR</td>
<td>95% CI</td>
</tr>
<tr>
<td>Age (yrs.)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>&gt; 65 vs. ≤ 65</td>
<td>1.56</td>
<td>0.53-4.57</td>
</tr>
<tr>
<td>Gender</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Male vs. Female</td>
<td>0.07</td>
<td>0.01-0.37</td>
</tr>
<tr>
<td>Smoking status</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Former/current smoker vs. never smoker</td>
<td>0.38</td>
<td>0.09-1.65</td>
</tr>
<tr>
<td>T stage*</td>
<td></td>
<td></td>
</tr>
<tr>
<td>T1-T2 vs. T3-T4</td>
<td>0.67</td>
<td>0.31-1.48</td>
</tr>
<tr>
<td>N stage*</td>
<td></td>
<td></td>
</tr>
<tr>
<td>N2 vs. N0-N1</td>
<td>2.12</td>
<td>0.78-5.81</td>
</tr>
<tr>
<td>Stage*</td>
<td></td>
<td></td>
</tr>
<tr>
<td>II vs. III</td>
<td>0.50</td>
<td>0.22-1.1</td>
</tr>
<tr>
<td>HLA number</td>
<td></td>
<td></td>
</tr>
<tr>
<td>6 vs. ≤ 5</td>
<td>0.95</td>
<td>0.42-2.14</td>
</tr>
<tr>
<td>HLA LOH</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Yes vs. No</td>
<td>0.52</td>
<td>0.23-1.17</td>
</tr>
<tr>
<td>Treatment</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Adjuvant Chemotherapy vs. none</td>
<td>0.44</td>
<td>0.19-1.00</td>
</tr>
<tr>
<td>TMB</td>
<td></td>
<td></td>
</tr>
<tr>
<td>≤ 4 vs. &gt; 4</td>
<td>1.28</td>
<td>0.52-3.20</td>
</tr>
<tr>
<td>NAL</td>
<td></td>
<td></td>
</tr>
<tr>
<td>≤ 2 vs. &gt; 2</td>
<td>2.56</td>
<td>1.15-5.68</td>
</tr>
<tr>
<td>DDR index</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>----------------------</td>
<td>----</td>
<td>------</td>
</tr>
<tr>
<td><strong>Low vs. high</strong></td>
<td>2.79</td>
<td>1.15-6.78</td>
</tr>
<tr>
<td><strong>APM status</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Mutation vs. wild type</strong></td>
<td>0.85</td>
<td>0.38-1.90</td>
</tr>
</tbody>
</table>

NSCLC, non-small-cell lung cancer; SCC, squamous cell carcinoma; ADC, adenocarcinoma.

TMB, tumor mutation burden; NAL, Neoantigen load; LOH, loss of heterozygosity.

*Using the 8th TNM staging classification.

**Figures**
Figure 1

Mutation spectra and significantly mutated genes in ADC and SCC. A, Gene mutation spectra in ADC and SCC. *P ≤ 0.05 and ****P ≤ 0.0001 using Student's t-test. B, The top 30 significantly mutated genes in SCC. The samples were aligned according to their somatic non-synonymous mutation burden (in the top panel) and genes were ranked by mutation frequencies. C, The top 30 significantly mutated genes in ADC.
The samples were made in order based on their somatic non-synonymous mutation burden (in the top panel) and genes were ranked by mutation frequencies.

**Figure 2**

Frequent gene mutations enriched by the oncogenic pathways in ACC and ADC. 

A. The mutation status of the top ten oncogenic pathways in SCC. 

B. The mutation status of the top ten oncogenic pathways in ADC. 

C. Gene mutation distribution and types in the RTK-RAS, HIPPO, and NOTCH signal pathways.
Immunogenomic profiling of altered genes in SCC and ADC. A. Comparison of HLA-I type in our cohort of patients vs. the online database. Our data are consistent with AlleleFrequency.net of China Jiangsu Han (HLA-A/B, n=3238) & China South Han pop 2 (HLA-C, n=1098). B, Comparison of HLA-I number and LOH between our ADC and SCC samples. C, TMB and NAL. Our data showed that TMB and NAL were higher in SCC than in ADC. D, Illustration of immunogenomic features between ADC and SCC. The HLA number,
LOH, MS status, DNA damage repair pathway and antigen presentation machinery pathway were profiled stratified by NSCLC histological types.

Figure 4

NAL as a prognostic and predictive indicator for lung SCC. A, Kaplan-Meier curves and log rank test stratified by the DDR index. Prognostic and predictive effect of the low vs. high DDR index on DFS of SCC patients. B, Kaplan-Meier curves, and log rank test stratified by the NAL. Prognostic and predictive effect
of NAL on DFS of SCC patients. C, Comparison of NAL with the DDR index in SCC patients. D, Comparison of the different gene mutation types and characters between NAL and TMB in ADC and SCC.

**Supplementary Files**

This is a list of supplementary files associated with this preprint. Click to download.

- Supp.Figs.docx
- supplementTable.docx