


RESEARCH ARTICLE

Open Access



Epigenetic alterations are associated with tumor mutation burden in non-small cell lung cancer

Liangliang Cai^{1†}, Hua Bai^{1†}, Jianchun Duan¹, Zhijie Wang¹, Shugeng Gao¹, Di Wang¹, Shuhang Wang¹, Jun Jiang², Jiefei Han¹, Yanhua Tian¹, Xue Zhang¹, Hao Ye³, Minghui Li³, Bingding Huang³, Jie He¹ and Jie Wang^{1*} 

Abstract

Background: To profile genomic and epigenomic of a naïve Chinese non-small cell lung cancer (NSCLC) cohort and investigate the association between tumor mutation burden (TMB) and DNA methylation (DNAm) to explore potential alternative/complimentary biomarkers for NSCLC immunotherapies.

Methods: A total of 89 tumor tissues with matched normal tissues from Chinese NSCLC patients were collected and subjected to whole exome sequencing (WES). From comparison, each patient was evaluated for the TMB value and divided into high, medium and low TMB based on TMB tertile distribution and then relatively high and low TMB samples were selected and subjected to DNAm profiling.

Results: Patients in the low ($n = 30$), medium ($n = 29$), and high ($n = 30$) TMB tertiles had 1.1–2.5, 2.5–4.1, and 4.2–13.9 mutations/Mb, respectively. A statistical directly association between differential methylation probes (DMPs) and TMB level was observed in our cohort ($r = 0.63$, P value = 0.0003) and this was confirmed by using TCGA NSCLC dataset ($r = 0.43$, P value = 0.006). Relatively high TMB group ($n = 16$, 7.5–13.9 mutations/Mb) harbors more differential DMPs while less in relatively low TMB group ($n = 13$, 1.1–2.4 mutations/Mb). Eight hundred fifty-eight differential methylation regions (DMRs) were found in relatively high TMB group. In addition, 437 genes show DNAm aberrance status in high TMB patient group and 99 have been reported as its association with lung cancer.

Conclusion: To our knowledge, this is the first report for direct link between the methylome alterations and TMB in NSCLCs. High TMB NSCLCs had more DNAm aberrance and copy number variations (CNVs). In addition, the TMB distribution of Chinese NSCLCs population is lower than that of TCGA.

Keywords: Genetics, Epigenetics, Non-small cell lung cancer, Tumor mutation burden, Checkpoint inhibitor therapy

Background

Lung cancer is the leading cause of cancer death worldwide and highly prevalent in China [1]. Approximately 85% of lung cancer cases are non-small cell lung cancer (NSCLC) [2]. Traditional target therapies have been effective against target population but they often suffer rapid relapse [3–5], such as target therapies against EGFR mutations [6], EMLA4-ALK fusion [7], and ROS-1 rearrangement positive [8]. Recent advances in

immune checkpoint inhibitors (ICIs) [9, 10], including anti-PD-1 [11], anti-PD-L1 [12], and anti-CTLA4 [13] antibodies, may have the potential to transform cancer into chronic disease by relying on normalizing patients' own immune system in tumor microenvironment. However, so far, not all lung cancer patients give an effective clinical response to ICI therapy even the positive PD-L1 expression in tumor tissue [14]. This demands an effective biomarker for ICI responding patient stratification.

Tumor mutation burden (TMB) has been proved to be effective in differentiating responding population of ICI therapies in multiple clinical studies. In addition, PD-L1 expression, microsatellite instability and deficient mutation mismatch repair have been used as companion

* Correspondence: zhuxi@163.com

[†]Liangliang Cai and Hua Bai contributed equally to this work.

¹Department of Medical Oncology, National Cancer Center/National Clinical Research Center for Cancer/Cancer Hospital, Chinese Academy of Medical Sciences and Peking Union Medical College, Beijing, China
Full list of author information is available at the end of the article



diagnostic biomarkers for ICI therapy [15]. Tumor-infiltrating lymphocytes are another potential biomarker in tumor microenvironment [16, 17]. Among these biomarkers, TMB remains the most promising candidate up-to-date due to its relatively high positive screening rate.

Epigenetic modification, particularly DNA methylation (DNAm) has been linked to genomic instability, such as mutations in a DNA methyltransferase gene may cause chromosome instability in human and mouse [18, 19], and the LINE-1 hypomethylation has been found to associate with global loss of imprinting, which induce chromosomal instability in colorectal cancer and head and neck squamous cell carcinoma [20, 21].

However, direct correlation between DNAm status and TMB has not been addressed to date in NSCLC clinical samples. Here, we investigated DNAm profiles of a Chinese NSCLC cohort, together with whole exome sequencing (WES) data to explore their direct correlation with TMB. This can provide further insights for future novel biomarker developments for ICI therapies.

Methods

Patient cohorts

We have selected a total of 89 treatment naïve lung adenocarcinoma (LUAD) or lung squamous cell carcinoma (LUSC) patients from the Cancer Hospital, Chinese Academy of Medical Sciences & Peking Union Medical College who underwent definitive surgical resection before adjuvant therapy, including chemotherapy or radiotherapy. This study was approved by the Cancer Hospital, Chinese Academy of Medical Sciences & Peking Union Medical College, and performed in accordance with the Declaration of Helsinki Principles. All these samples were fresh frozen tissues which were under low temperature conditions (at -80°C). After obtaining informed consents, tumor tissues and their matched control were obtained for WES and DNAm profiling. To avoid the contamination of tumor tissue, all the matched normal tissues were collected at lobectomy edge. All samples had been subjected to pathology review for histological subtyping. The detailed clinical characteristics of these 89 Chinese Han population samples are summarized in Additional file 1: Table S1.

WES and data processing

Sequencing protocol: DNA libraries for tumor and their matched control samples were prepared with standard protocol using MGIEasy Exome Capture V4 Probe Set capture kit (cat. No: 1000007745, <https://en.mgitech.cn/article/detail/v4.html>) with the capture region size 36 Mb. Pair-end sequencing (2×100 bp) was performed on BGI-Seq 500 instruments. Data processing: Alignment: The raw paired end reads were mapped to the human reference genome (hg19) using bwa-mem (version 0.7.16

with $-M$ option: mark shorter split hits as secondary and the remaining setting was at default). Samtools v1.3.1 was used to sort and merge bam files from the same patient sequenced from different lanes. PCR duplicate read pairs were identified using biobambam (v.0.0.148). The quality control (all the sample QC files were available at <https://drive.google.com/open?id=1HggApA8homvpF4xD2YOI3EQ2HsY3hS4S>) was generated with FastQC (v0.11.8) and the post-alignment QC metrics information was shown in Additional file 1: Table S2. Variants calling: Variants calling was performed using a modified version of DKFZ-pipeline based on samtools mpileup and bcftools version 0.1.19 (pcawg-dkfz-workflow). Briefly, variants in the tumor sample were initially and used as query in the control sample. The raw calls were then annotated with various publicly available databases, including 1000 Genomes variants, ESP exon database, single-nucleotide polymorphism database (dbSNP), ExAC v.0.3.1 (non-TCGA variants), repeats and other elements. The functional consequence of the variants was predicted using Annovar [22] with UCSC Refseq annotations, followed by assessment of the variants in terms of their confidence and then classified into somatic or non-somatic calls. Only highly confident somatic variants with the following filtering criteria: Read Depth ≥ 10 , AF $\geq 5\%$, Number of reads indicating mutation ≥ 3 , were used for further analysis. TMB level is defined by two ways: one is as number of nonsynonymous coding somatic mutations (NOMs) per tumor, including single nucleotide variation (SNVs) and short insertion/deletion polymorphism (INDELs); the other is the number of mutations is proportion to the size of UCSC Refseq annotations (33.4 Mb). R/Bioconductor package “maftools” [23] was used for visualization and summarization of MAF files from this study. TCGA WES somatic mutations: Confident somatic mutation calls derived from the WES data of the LUAD and LUSC cohorts were directly downloaded from the TCGA GDC Data Portal (<https://portal.gdc.cancer.gov>).

Analysis of mutational signatures

Mutational signature analysis was conducted using the deconstructSigs package v1.8.0 [24]. All the detected somatic mutations including synonymous in the cohort were imported for signature analysis. In details, the frequency of 96 possible mutation types in trinucleotide context of each patient were first calculated in somatic mutation dataset. Normalization were then processed, according to the number of times each trinucleotide context is observed in our capture region. Finally, the weights of 30 known cancer mutation signature in COSMIC (<https://cancer.sanger.ac.uk/cosmic/signatures>) were generated by linear regression based on normalized frequency of each possible mutation types. Each weight indicates that how

strongly can mutation signature influence the patient. Hierarchy cluster based on mutation signatures' weights among patients were drawn by R package 'pheatmap' [25].

Assessment of DNA methylation profiles

Five hundred nanogram of genomic DNA from each sample was bisulfite converted using the EZ DNA Methylation Kit (Zymo Research, Irvine, CA) and then analyzed on Infinium HumanMethylation 850 K EPIC BeadChip (Illumina, San Diego, CA) following the manufacturer's instructions. The array features more than 850,000 methylation sites covering 96% CpG islands and 99% gene's promoters. Raw data were analyzed using the "ChAMP" (Chip Analysis Methylation Pipeline for Illumina HumanMethylation450 and EPIC) package in R [26, 27] and all relevant parameters are default values. The differential methylated probe (DMP) of each sample was identified by the beta value of cancer and matched normal tissue with Benjamini-Hochberg (BH)-adjusted P -value < 0.05 . R/Bioconductor package "ConsensusClusterPlus" [28] was used for consensus clustering of Illumina EPIC data. Bumhunter algorithm was applied to estimate regions for which a genomic profile deviates from its baseline value. Originally implemented to detect differentially methylated genomic regions between tumors and normal controls. By default, the progress of differential methylation region (DMR) finding was done on normalized beta value. The detected DMR and estimated P value (0.05 as cutoff value) was returned.

Determination of copy number alterations (CNA) using the EPIC array and GO enrichment

The R/Bioconductor package 'conumee' [29] was employed to calculate CNAs based on the intensities generated using the EPIC array (using default settings). GISTIC [30] was then used to identify common deleted/amplified regions/genes (using default parameters). GISTIC is a tool that identifies genes targeted by somatic copy-number alterations (SCNAs) that trigger cancer growth. By classifying SCNA profiles as arm-level and focal alterations, this tool calculates the background rates of each category as well as delineates the boundaries of the SCNA regions. Aneuploidy score (AS) was calculated as is reported [31, 32], and the scores of each arm are -1 if lost, $+1$ if gained, 0 if non-aneuploid, and "NA" otherwise. For gene enrichment analysis, the function annotation tool from the DAVID website (<https://david-d.ncifcrf.gov/>) was used.

Statistics

All statistical tests were conducted in R version 3.4.1 (The R Foundation for Statistical Computing, Austria). Unpaired t test was performed to evaluate the significance of TMB value between two groups (smoking: non-

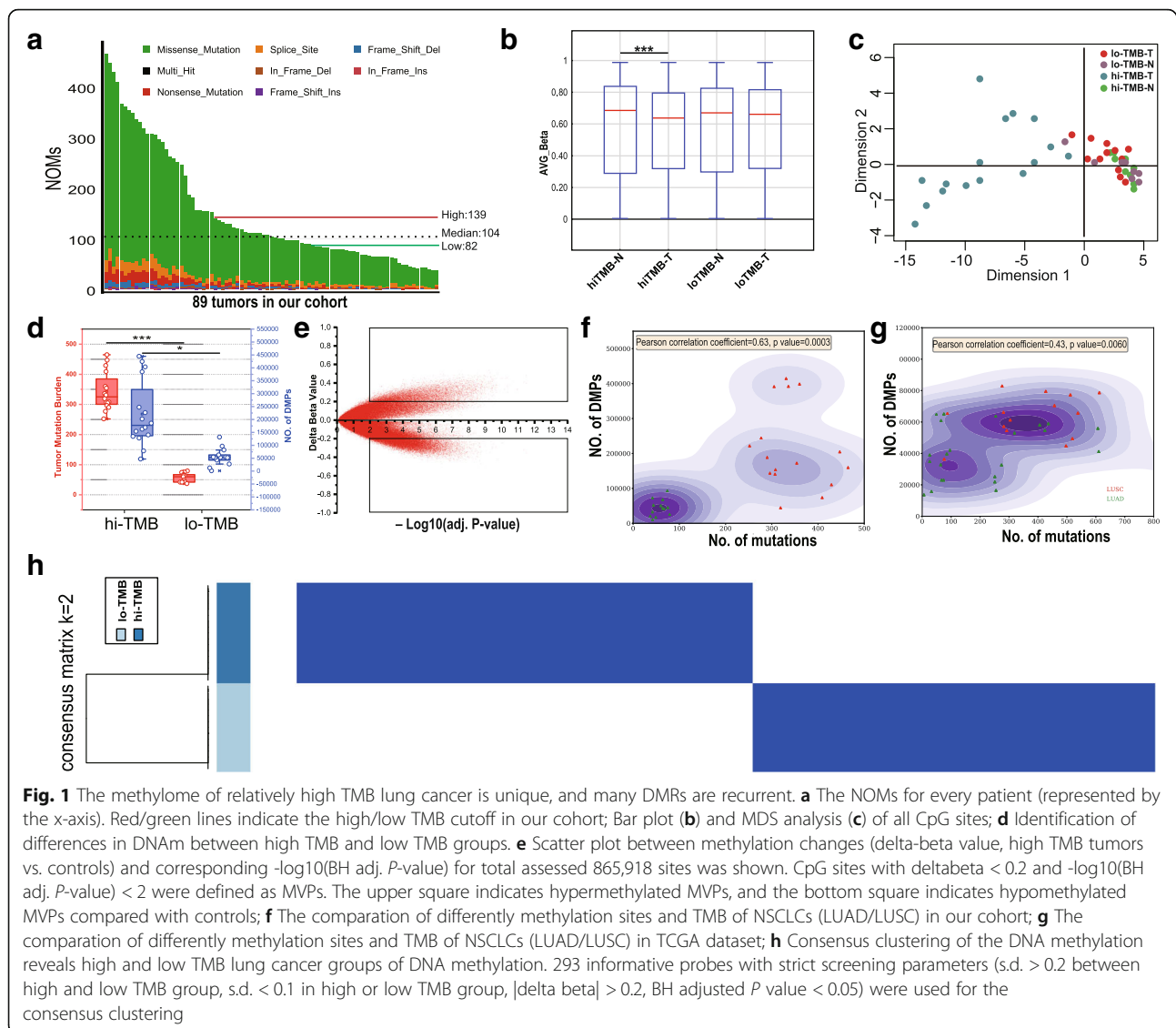
smoking, TP53+: TP53- and Chinese: TCGA LUAD/LUSC). Pearson's correlation coefficient was calculated to evaluate the strength of correlation between DNA methylation and TMB levels. * stands for P value < 0.05 . ** stands for P value < 0.01 . *** stands for P value < 0.001 .

Results

DNA methylation changes differently between high TMB and low TMB NSCLCs

These 89 patients included in this study were consist of 65 LUAD as well as 24 LUSC patients. From WES data analysis, only high confidence nonsynonymous somatic mutations (Tumor DP ≥ 10 , AF $\geq 5\%$, NO. of reads indicating mutation ≥ 3) were processed for TMB assessment. The mean coverage is achieved at $167\times$, $161\times$ in tumor samples and normal samples, respectively. More than 90% of targeted regions with coverage $> 10\times$ were found in 87/89 pair samples. TMB distribution showed a median number of 104 NOMs per tumor, ranging from 37 to 465 (Fig. 1a). Consistent with the approach of the clinical trial CheckMate 026 [33], we classified our cohort by high (139–465), medium (83–136) and low (37–82) NOMs or low (1.1–2.5), medium (2.5–4.1), and high (4.2–13.9) mutations/Mb. In order to further explore the relationship between DNAm and TMB, 13 relatively low (37–79 mutations or 1.1–2.4 mutations/Mb) and 16 relatively high (252–465 mutations or 7.5–13.9 mutations/Mb) TMB samples were selected for subsequent methylation level detection. Due to the insufficient amount of DNA after the WES experiment, these samples were not selected successively. Unless specifically mentioned, the high or low TMB group in the following text represents the relatively high TMB group and the relatively low TMB group.

DNAm profiles in tumor samples and its matched normal controls were measured using the Illumina Infinium HumanMethylation EPIC BeadChip platform (850 K), which evaluates methylation status of 865,918 CpG sites covering key features of the human whole genome. The R package 'ChAMP' for Illumina EPIC was applied for data analysis. The beta-value was chosen as a measure of the methylation level, which ranges from 0 (no methylation) to 1 (complete methylation). Based on the methylation level of 865,918 sites, differential global methylation status (unpaired t test, P value < 0.001) was seen between high TMB group (median beta-value of 0.643) and its matched controls (median beta-value of 0.629), while 0.631 and 0.629 in low TMB group (Fig. 1b). The methylation status in tumor tissue comparing normal tissue was different between high and low TMB groups, and this was further confirmed by multidimensional scaling (MDS) analysis of the CpGs (Fig. 1c). Data of tumor tissues cluster separately from the normal tissues in high



TMB patients, thereby indicating a different global methylation pattern. However, in low TMB patients, the tumor tissues cluster overlaps with their corresponding normal tissues, which indicates stable epigenomic profile between tumor and normal tissues in low TMB patients. Cluster analysis also revealed variable global methylation patterns in high TMB group comparing to low TMB group. 292121 significant DMPs with a BH-adjusted P -value below 0.05 were found while none in low TMB group. Box plot analysis further shows that high TMB group (median TMB =343) harbors significantly more differential methylation locis (31,279~391,387, with median of 188,637) with $|\text{delta beta}| > 0.2$ than low TMB group (median TMB =62; 10,479~92,932, with median of 43,340) in Fig. 1d. We obtained differentially methylated region (DMR) in high TMB group in a total number of 858 regions (Additional file 1: Table S3), while none

in low TMB group. To exclude that the observed differences in DMPs between high and low TMB samples are driven by different leukocyte enrichment, R/Bioconductor package “minfi” [34] was applied for cell type composition. The results (Additional file 2: Figure S1) showed that no significant differences were observed in CD8T, CD4T, NK, Bcell, Mono and Gran cells between high and low TMB samples.

Methylation differences between high TMB tumor and matched normal tissues were calculated as delta-beta and plotted against corresponding $-\log_{10}(\text{BH-adjusted } P\text{ value})$, as shown in Fig. 1e. With a consideration of so much DMPs and further analysis in high TMB group, we defined CpG sites with $|\text{delta-beta}| > 0.2$ and BH-adjusted P value < 0.01 as methylation variable positions (MVPs). From the over 850,000 informative probes, 61,633 MVPs were identified, representing $< 7\%$ of the total

sites surveyed and top 3000 MVPs were shown in Additional file 1: Table S4. There were more hypomethylated CpG sites (44,718 MVPs, delta-beta < -0.2, bottom square) than hypermethylated CpG sites (16,915 MVPs, delta-beta > 0.2, upper square).

After comparing each cancer and matched normal tissues of differential methylation data in high or low TMB groups, we found that high TMB patient samples contain more DMPs (Pearson correlation coefficient = 0.63, P value = 0.0003) comparing to low TMB patient samples (Fig. 1f). For further validation, we execute of analysis of tumor mutation data and DNA methylation data of 39 TCGA NSCLCs with high- (top 20% by TMB) and low-TMB (bottom 20%) as is shown in Fig. 1g. Positive correlation (Pearson correlation coefficient = 0.43, P value = 0.006) was also found between NOMs and DMPs in such independent dataset. By analyzing LUAD or LUSC samples separately, the DMPs of LUAD or LUSC was significantly correlated with TMB as is shown in Additional file 2: Figure S2. The same analysis in the TCGA NSCLCs database is consistent with this result (Top 12 high TMB vs bottom 12 low TMB: P value = 0.0026, Mann–Whitney test) as is shown in Additional file 2: Figure S3. To further identify methylation sites that distinguish high TMB and low TMB lung cancer, we utilized k-means consensus to perform cluster to these 29 primary lung cancer tissue samples from our cohort with 293 most variable methylation loci (s.d. > 0.2 between high and low TMB group, s.d. < 0.1 in high or low TMB group, $|\text{delta beta}| > 0.2$, BH adjusted P value < 0.05, Additional file 1: Table S5). We observed two distinct groups of samples (Fig. 1h), which were correlated to the high or low TMB lung cancer groups.

High TMB NSCLC patients harbor more structural variation of CNV

CpG hypomethylation status has been reported to be related to genetic instabilities, and global hypomethylation in tumor indicates more genomic instabilities [35]. We checked the copy number variation (CNV) in high TMB and low TMB group using aneuploidy score (AS) and found high TMB lung cancers possess more structural variation of CNV, while low TMB ones appeared to retain more stable genomic structural profile (Fig. 2a, Additional file 1: Table S6). The analysis results of TCGA database are consistent with our cohort study ($r = 0.18$, P value = 1×10^{-8} , Pearson correlation analysis) as is shown in Additional file 1: Table S7 and Additional file 2: Figure S4. Compared to the low TMB group, the high TMB NSCLC group exhibited more genomic deletions and amplifications (Fig. 2b), particularly a gain in chromosome arm 3q (especially 3q26) and a loss of chromosome 3p (especially 3p12). Frequent localized amplifications within chromosomal regions 8q24, 12p11,

and 15q11 loci and deletions within 8p22 and 9p23 were also detected. A total of 1237 genes (Additional file 1: Table S8) were significantly mapped to these amplified regions, whereas no genes could be significantly mapped to the deleted regions (whole chromosomal arm deletions were excluded from the analysis) (Fig. 2c). Several of these recurrent CNAs exhibited high chromosomal instability, possibly lead to TMB value increasing. The 1237 genes in CNA regions were also assessed in terms of Gene Ontology enrichment by DAVID, which revealed that the pathways of Jak-STAT signaling (hsa04630) and cytokine-cytokine receptor interaction (hsa04060) were highly represented in Fig. 2d.

437 Genes' promoter regions showed DNAm aberrance status in high TMB NSCLCs

In high TMB group, more significant DMPs was found and Fig. 3a shows the top 6 DMPs (cg16732616/DMRTA2, cg26521404/HOXA9, cg20326647/intergenic region, cg02443967/TLL2, cg09792881/DMRTA2 and cg16928066/EMX1) as representatives. We also explored the distribution of the DMPs and found hypermethylated DMPs were located closer to transcription start site (TSS), whereas the hypomethylated DMPs was shifted slightly to upstream of TSS (Fig. 3b). We focused on the MVPs with No. > 3 at the promoter region referring to TSS1,500, TSS200, 5'-UTR, and 1stExon to discover significant differential methylated gene and found 1666 genes, in which HOX family genes (26 out of 39 [36, 37]) were most effected (Additional file 2: Figure S5). In order to further exclude inappropriate genes caused by the number of samples, the same analysis was carried out in the TCGA NSCLC database, and Venn analysis (Fig. 3c, Additional file 1: Table S9) showed that there were 437 genes associated with the state of high TMB. The heatmap plot (Fig. 3d) analyzed all the 8703 probes from the 850 k chip related to these 437 genes, and the results show that they are significantly different in the high TMB group. The same analysis was performed on 4916 probes from 450 k chips in the TCGA database, and the results (Additional file 2: Figure S6) were consistent with our cohort study. To further analyze the relationship between these 437 genes and lung cancer, we used DisGeNET [38], a database of gene-disease associations, to analyze the network of these genes, and found that there were 99 genes, related to "Neoplastic Process" of lung (Additional file 1: Table S10).

Chinese NSCLCs have lower NOMs than TCGA LUAD/LUSC

The TMB distribution of Chinese NSCLCs has not been well reported in the literature, therefore its description may provide insights for pharmaceutical companies or diagnostic industry to adjust their commercial strategy in China. Recent studies had demonstrated that loss of

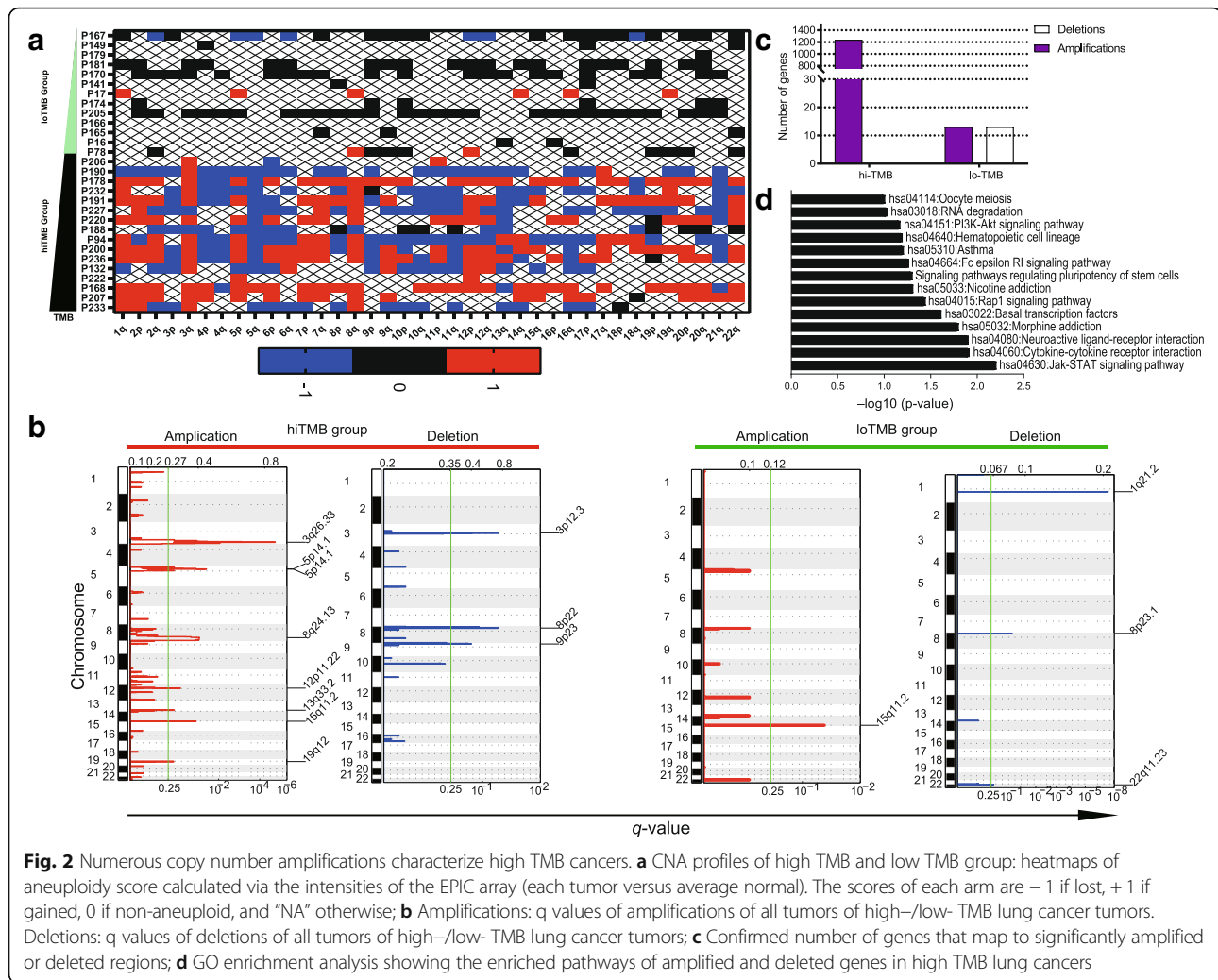


Fig. 2 Numerous copy number amplifications characterize high TMB cancers. **a** CNA profiles of high TMB and low TMB group: heatmaps of aneuploidy score calculated via the intensities of the EPIC array (each tumor versus average normal). The scores of each arm are - 1 if lost, + 1 if gained, 0 if non-aneuploid, and “NA” otherwise; **b** Amplifications: q values of amplifications of all tumors of high-/low- TMB lung cancer tumors. Deletions: q values of deletions of all tumors of high-/low- TMB lung cancer tumors; **c** Confirmed number of genes that map to significantly amplified or deleted regions; **d** GO enrichment analysis showing the enriched pathways of amplified and deleted genes in high TMB lung cancers

TP53 function increased genomic instability [39, 40]. We further investigate the mechanism of these differences between these two populations based on genetic alterations. An important driver gene of Chinese NSCLCs, EGFR mutations, which are closely related to the efficacy of molecularly targeted therapy (EGFR TKIs), have been reported to negatively correlate with TMB value [41, 42]. Heatmap plot shows that frequently mutated genes such as TP53 gene, which tends to be enriched in high TMB group (top 30 samples, range: 139–465 NOMs) in lung cancer; EGFR mutants in low TMB (bottom 30 samples, range: 37–82 NOMs), and patients with co-existence of TP53 and EGFR mutations in the intermediate TMB level (median 29 samples, range: 83–136 NOMs) (Fig. 4a). Many disease-causing genes in cancer are co-occurring or show strong exclusiveness in their mutation pattern with high TMB. In our study, gene set TP53, CSMD3, GXYLT1, PPP1R13L and TTN shows a strong co-occurrence and gene set EGFR, TTN, MUC2 and HERC2 shows a strong exclusiveness in high

TMB group (Fig. 4b). It was been confirmed in our study that the high TMB samples was mostly LUSC with smoking habit. Our study confirm that smoking was also a key factor associated with TMB (Fig. 4c). We evaluated 30 known mutational signatures for different carcinogens in COSMIC database, including UV light or tobacco by calculating the frequency of specific mutation types in trinucleotide [43]. Consistent with previous findings, we observed that high TMB patients displayed distinct mutation signatures comparing to low TMB patients as is shown in Fig. 4d. Signature 4 was the dominant mutation pattern in high TMB patients with smoking history, while high TMB patients without smoking history contained relatively stronger mutation pattern in Signature 3. Signature 3 and signature 12 occurred simultaneously in low TMB patients, regardless of smoking status. Since signature 4 is a well-known tobacco-related signature characterized by transcriptional strand bias in C > A mutations, it matches the phenotype of smoking-history among high TMB patients. The failure of DNA double-

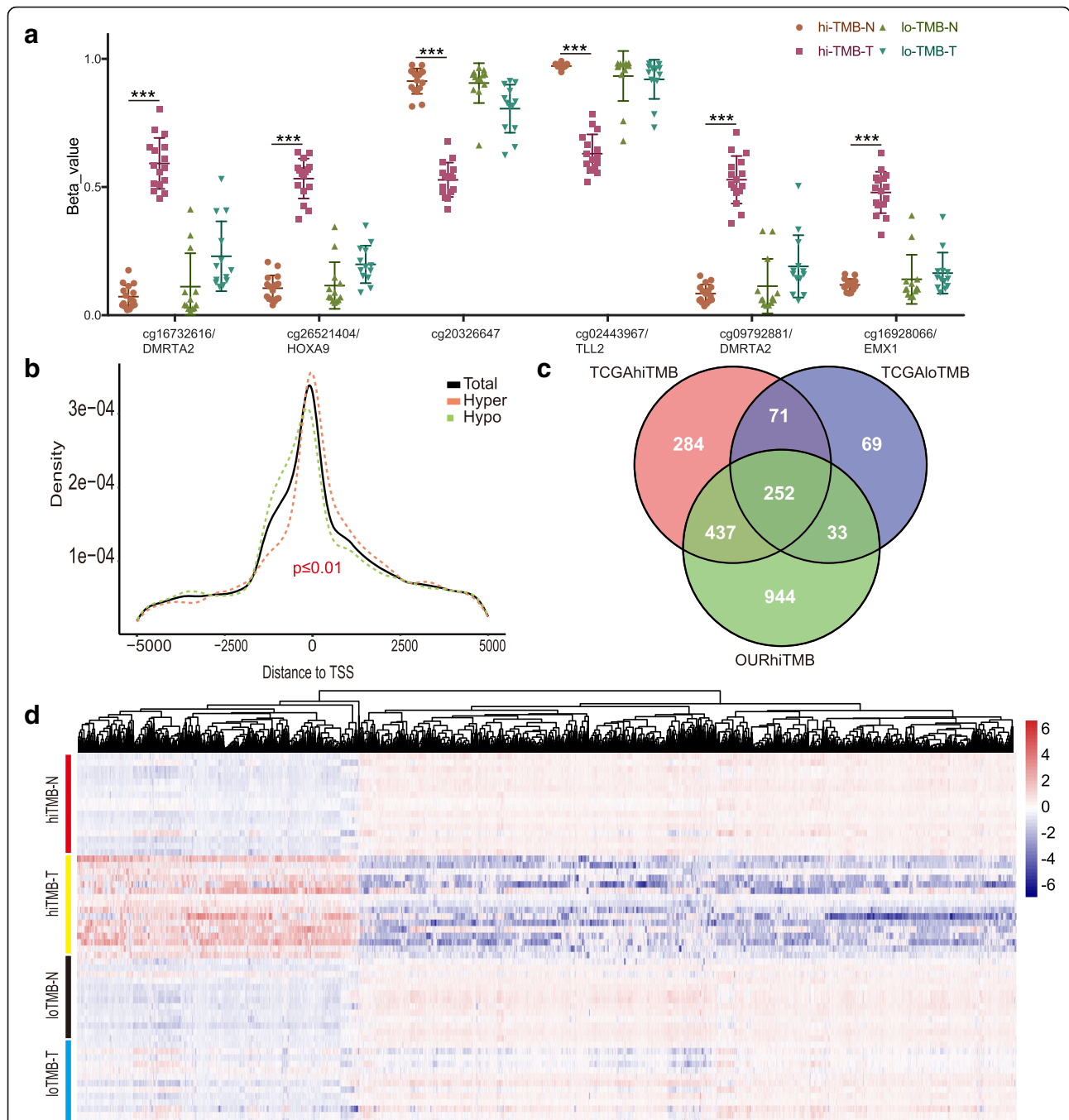


Fig. 3 The relationship of HOX gene methylation status and TMB. **a** Top 6 differential methylation sites in hi-TMB comparing with low TMB group. cg16732616/DMRTA2, cg26521404/HOXA9, cg20326647/intergenic region, cg02443967/TLL2, cg09792881/DMRTA2 and cg16928066/EMX1 were significantly methylated in high TMB tumor tissues; **b** In hi-TMB lung cancer, CpGs that are hypomethylated are more likely to be found immediately upstream of the TSS and within the 1st exon, CpGs are hypermethylated (P value < 0.01 , two-tailed Wilcoxon rank-sum test); **c** Venn diagram analysis revealed 437 genes associated with high TMB; **d** The heatmap of the all methylation probes related to 437 genes in high TMB, low TMB NSCLCs samples and the matched normal controls

strand break-repair in homologous recombination indicated by signature 3 may confer high mutation ability to patients without smoking history. Regarding to low TMB patients, efforts are needed to investigate the etiology of the strong signal in signature 12 with T > C substitutions.

Interestingly, the frequency of TP53 and EGFR mutations between our cohort and TCGA cohort was different (TP53: Chinese 46%, TCGA 70%, EGFR; Chinese 39%, TCGA 17%) (Fig. 4e). This result was consistent with previous study in Chinese lung cancer

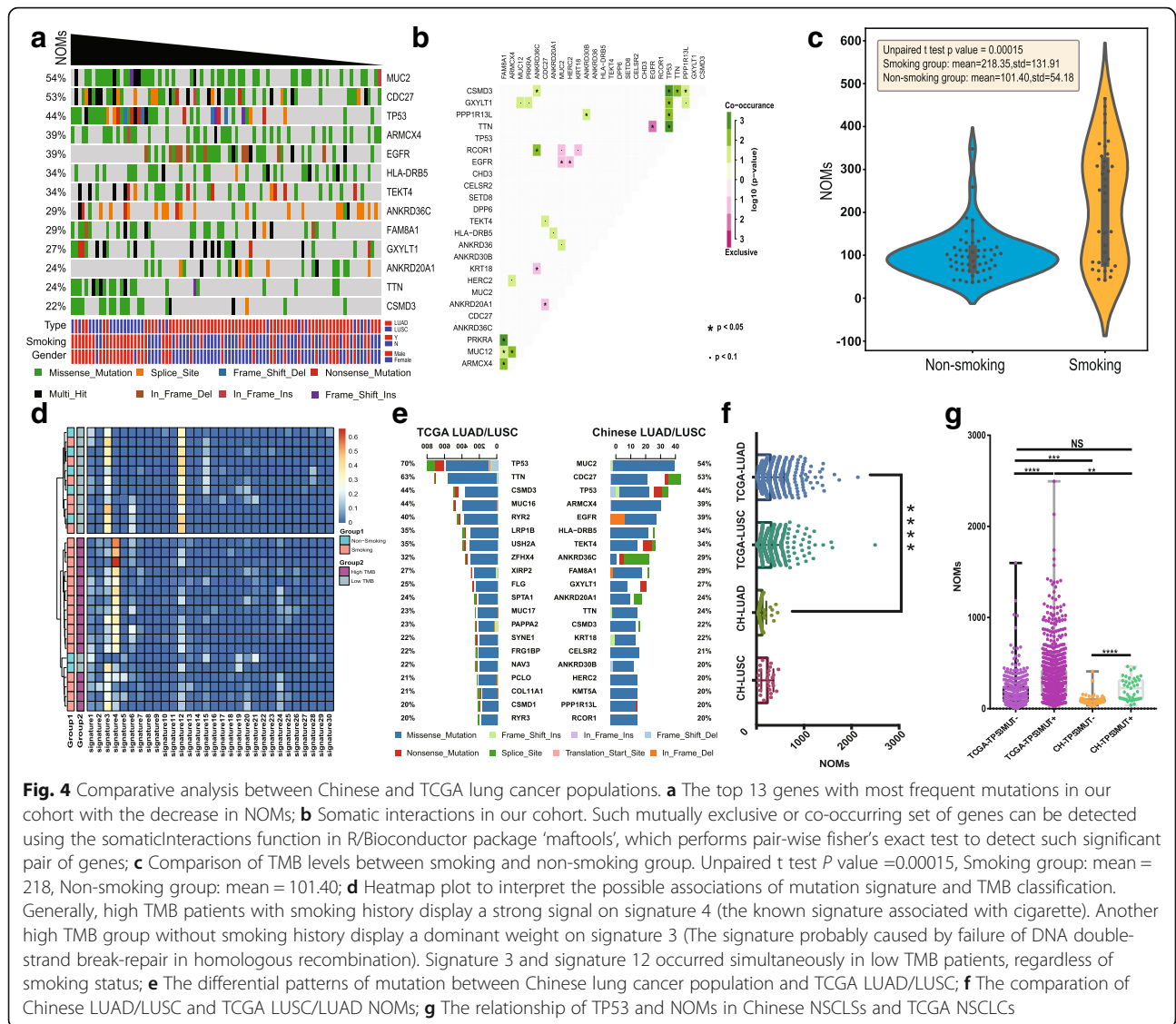


Fig. 4 Comparative analysis between Chinese and TCGA lung cancer populations. **a** The top 13 genes with most frequent mutations in our cohort with the decrease in NOMs; **b** Somatic interactions in our cohort. Such mutually exclusive or co-occurring set of genes can be detected using the somaticInteractions function in R/Bioconductor package ‘maftools’, which performs pair-wise fisher’s exact test to detect such significant pair of genes; **c** Comparison of TMB levels between smoking and non-smoking group. Unpaired t test P value =0.00015, Smoking group: mean = 218, Non-smoking group: mean = 101.40; **d** Heatmap plot to interpret the possible associations of mutation signature and TMB classification. Generally, high TMB patients with smoking history display a strong signal on signature 4 (the known signature associated with cigarette). Another high TMB group without smoking history display a dominant weight on signature 3 (The signature probably caused by failure of DNA double-strand break-repair in homologous recombination). Signature 3 and signature 12 occurred simultaneously in low TMB patients, regardless of smoking status; **e** The differential patterns of mutation between Chinese lung cancer population and TCGA LUAD/LUSC; **f** The comparison of Chinese LUAD/LUSC and TCGA LUSC/LUAD NOMs; **g** The relationship of TP53 and NOMs in Chinese NSCLCs and TCGA NSCLCs

population [44]. TMB distribution (median value = 104 NOMs per tumor) in our cohort is lower than TCGA LUAD/LUSC (median value =176 NOMs per tumor). The TMB value of LUSC was significantly (unpaired t test, P value <0.001) greater than the value of LUAD (Fig. 4f). After further analysis, it indicates that TP53 mutations significantly affect TMB level in both Chinese NSCLCs between TP53+ and TP53- mutation group (unpaired t test, P value <0.001) and TCGA LUAD/LUSC (unpaired t test, P value <0.001) (Fig. 4g). In the analysis based on Asian, Black, White populations from TCGA database, the results (Additional file 2: Figure S7 and S8) were consistent with our findings: Asian ($n = 17$, TP53: 65%, EGFR: 24%, mean NOMs: 151), Black ($n = 81$, TP53: 70%, mean NOMs: 292) and White ($n = 731$, TP53: 62%, mean NOMs: 251).

Discussion

Although linkage between methylation change and chromosomal instability has been widely reported, a direct link between differential methylation and TMB values has not been directly measured in NSCLC population. The results of the NCT02259621 trial indicated that TMB may be used as a biomarker for the pathological responses to the PD-1 blockade. Around 2 to 4 weeks after neoadjuvant nivolumab treatment, a rapid expansion of mutation-related, neoantigen-specific T-cell clones derived from a primary tumor that showed a complete response on pathological assessment, was observed in the peripheral blood of 8 out of 9 patients and a number of these clones were not identified prior to the nivolumab administration. Since then, TMB has been well reported to serve as a biomarker for stratifying patients for PD-1/PD-L1 therapies. Interestingly, a recent

study suggested that methylation pattern change may also serve as a prognosis biomarker for anti-PD-1 treatment [45]. Therefore, we conducted this study to investigate the correlation between TMB and DNAm profile. Our study revealed the significant correlation of DNAm and TMB in NSCLCs. To our knowledge, this is the first NSCLC cohort study to directly link the methylome alteration to TMB.

Methylome-wide analysis revealed widespread changes in lung cancer-associated DNAm patterns, particularly in high TMB cancer tissues. Earlier studies showed that DNA hypomethylation within the coding regions of genes is often associated with genome instability and higher mutation rate. However, those investigations were mostly performed *in vitro* in cell lines [46]. Our study results confirmed these findings with a comparison of primary NSCLC to matched normal tissues in our cohort and TCGA dataset. Based on these results, differential methylated target regions may work as a potential biomarker along with TMB or even as an alternative approach since accurate measurement of TMB demands a relatively large panel to harbor significant genomic variations which is quite expensive while methylation profiling methodology are more robust and reasonably economic. More comprehensive study on methylation regions with large scale PD-1/PD-L1 therapeutic patient samples with clinical outcome need to be conducted to lock down a panel of genes that methylation status correlate with ICI benefit.

To investigate the correlation between DNAm and TMB more extensively, we investigated 1666 genes that are significant differential methylated in our cohort. Among these genes, an important cluster of genes with hypermethylated CpGs is HOX gene family and its hypermethylation status has been reported to be associated with the low expression of HOX in lung cancer [47]. Unfortunately, such significant differences in the HOX gene family were not observed in TCGA database, so that further studies are needed to explore the function of the HOX family genes. We need to take this conclusion very cautious since the dataset is relatively small, and we only analyzed relatively high and low TMB NSCLC samples with the intermediate TMB samples to be excluded (due to samples shortage). However, our data raised a hypothesis that maybe a gene family methylation status or maybe a methylation panel can be served as a potential biomarker for ICI therapy. By integrating with the TCGA database, our study also revealed 437 potentially differentially methylated genes associated with high TMB, including 99 genes that are closely related to lung cancer disease. Since the cost for methylation panel assessment are much lower than TMB assessment, some genes' methylation status may be a potentially promising biomarker. Nevertheless, further

studies with larger size, more importantly with PD-L1 clinical outcome, are needed to further select and confirm biomarkers to improve precision management of NSCLCs ICI therapies.

EPIC 850 K arrays were employed for copy number analysis in parallel to DNAm analysis with the same DNA specimen. The 850 K array probes are as robust and sensitive as SNP arrays, resulting in CNA calls for its more wider probes coverage (> 850,000 CpGs). High TMB NSCLCs exhibited an imbalanced genome with several chromosomal gains and losses, whereas low TMB NSCLC samples showed much lower level of chromosome instability. We also confirmed that high TMB LUSC samples harbor numerous CNAs as well as aberrantly methylated sites and exhibit distinct mutational signatures.

When it was mentioned, currently, it was difficult to define an exact TMB value for its real role on ICI therapy effect prediction, although it has been explored so much. There are extensive investigations on TMB distribution on Caucasian NSCLCs, not much data has been shown on Chinese NSCLCs. Our data indicated that the TMB distribution of the Chinese NSCLCs population was significantly lower than the TMB range observed from TCGA LUAD/LUSC database. One plausible reason is that Asian, Black and Caucasian races tend to display different frequencies and patterns of tumor mutations. For example, Chinese lung cancer patients tends to harbor a much higher frequency of EGFR mutations. It has also been observed that TMB is much lower in EGFR mutated patients both in our cohort and TCGA dataset and the presence of driver alterations may provide clinically useful predictors of response to anti-PD-1/anti-PD-L1 therapies [48].

Conclusions

In our study, our results show that Chinese NSCLCs population has lower TMB level than TCGA LUAD/LUSC because of the higher mutation rate of EGFR but lower in TP53. However, the necessity to adjust the recommended TMB threshold for personalized lung cancer immunotherapy remains unclear and only clinical results can give a definitive answer. Another complication is that all patients in our study were diagnosed with NSCLC at an early stage and thus have not received any systemic treatment, including chemotherapy, targeted therapy, or ICI therapy, this may also cause different TMB distribution. Our data also confirmed the association between TP53 mutations and high TMB levels in the Chinese and TCGA LUAD/LUSC, and the association between cigarette smoking and high TMB levels. Nevertheless, our study will draw more attention on TMB cutoff adjustment on PD-1/PD-L1 therapy on Chinese NSCLCs.

Additional files

Additional file 1 : Table S1. Clinical characteristics of 89 NSCLCs in this study. **Table S2.** The post-alignment QC metrics information of WES. The post-alignment QC metrics information of WES including expected depth, the mean target coverage tumor and matched normal controls and the percentage of duplicates read, off targets and not mapped reads, GC content were provided. **Table S3.** 858 DMR in high TMB group comparing to matched normal controls. **Table S4.** BA list of the top 3000 MVPs. **Table S5.** 293 methylation sites that can distinguish between high or low TMB. **Table S6.** ASs and NOMs for each sample of our cohort. **Table S7.** ASs and NOMs for each sample of TCGA. **Table S8.** The 1237 gene list for the GO analysis with location info. **Table S9.** The gene list with the No. of MVP > 3 in our cohort and TCGA database for the Venn plot analysis. **Table S10.** Ninety-nine genes involved in "Neoplastic Process" of lung within DisGeNET database from 437 gene list. (RAR 1336 kb)

Additional file 2 : Figure S1. Cell type composition by sample type. **Figure S2.** The comparison of DMPs in high TMB or low TMB based on lung cancer type LUAD or LUSC in our cohort. **Figure S3.** The comparison of DMPs in high TMB or low TMB in TCGA NSCLCs. **Figure S4.** The correlation of ASs with NOMs in TCGA database. **Figure S5.** 4 HOX family gene cluster's methylation status in our cohort. **Figure S6.** Heatmap plot of all the 437 genes' methylation probes in TCGA database. **Figure S7.** Subgroup analysis of NOMs based on ethnicity (ASIAN, BLACK and WHITE) in TCGA database. **Figure S8.** The first 20 high frequency mutated genes of ASIAN, BLACK or WHITE population in TCGA database. (RAR 16686 kb)

Abbreviations

BH: Benjamini-Hochberg; CD4T: Cluster of differentiation 4 positive T cell; CD8T: Cluster of differentiation 8 positive T cell; CNV: Copy number variation; CTLA4: Cytotoxic T-lymphocyte-associated antigen 4; DMP: Differential methylation probe; DMR: Differential methylation region; DNAm: DNA methylation; EGFR: Epidermal growth factor receptor; FDR: False discovery rate; GO: Gene ontology; HOX: Homeobox; INDEL: Insertion/deletion polymorphism; Jak-STAT: Janus kinase/signal transducers and activators of transcription; LUAD: Lung Adenocarcinoma; LUSC: Lung Squamous Cell Carcinoma; MVP: Methylation variable position; NK: Natural killer cell; NOM: Number of mutation; NSCLC: None small cell lung cancer; PCR: Polymerase chain reaction; PD-1: Programmed cell death protein 1; PD-L1: Programmed death-ligand 1; SCNA: somatic copy-number alteration; SNV: Single nucleotide variation; TMB: tumor mutation burden; TP53: Tumor protein p53; TSS: Transcription start site; WES: Whole exome sequencing

Acknowledgements

We would like to thank the patients for their contribution to this research.

Authors' contributions

Conceptualization, LLC and JW; methodology, HB; software, HY; validation, BDH, MHL and DW; formal analysis, JCD and JFH; investigation, ZJW and SGG; resources, JH; data curation, JJ; writing—original draft preparation, LLC; writing—review and editing, JW; visualization, YHT and XZ; supervision, JW; project administration, JH; funding acquisition, JW. All authors read and approved the final manuscript.

Funding

This research was funded by National Natural Sciences Foundation Key Program, 81630071 and 81330062; National key research and development project Precision Medicine Special Research, 2016YFC0902300; National High Technology Research and Development Program 863, SS2015AA020403; CAMS Innovation Fund for Medical Sciences, CIFMS 2016-12M-3-008; Aiyu foundation, KY201701.

Availability of data and materials

The datasets generated and/or analyzed during the current study are available from the corresponding author on reasonable request.

Ethics approval and consent to participate

Patients undergoing removal a primary tumor were consented in accordance with IRB-approved institutional protocols.

Consent for publication

Not Applicable.

Competing interests

The authors declare that they have no competing interests.

Author details

¹Department of Medical Oncology, National Cancer Center/National Clinical Research Center for Cancer/Cancer Hospital, Chinese Academy of Medical Sciences and Peking Union Medical College, Beijing, China. ²Department of Oncology, Affiliated Hospital of Qinghai University, Xining, China. ³Sinotech Genomics Ltd, Shanghai, China.

Received: 12 February 2019 Accepted: 28 June 2019

Published online: 26 July 2019

References

- Chen W, Zheng R, Baade PD, Zhang S, Zeng H, Bray F, et al. Cancer statistics in China, 2015. *CA Cancer J Clin.* 2016;66(2):115–32.
- Cetta F, Renieri A, Frullanti E. Germline mutations in lung cancer and personalized medicine. *Familial Cancer.* 2017;17(3):429–430.
- Pogrebniak KL, Curtis C. Harnessing tumor evolution to circumvent resistance. *Trends Genet.* 2018;34(8):639–51.
- Holohan C, Van Schaeybroeck S, Longley DB, Johnston PG. Cancer drug resistance: an evolving paradigm. *Nat Rev Cancer.* 2013;13(10):714–26.
- Rotow J, Bivona TG. Understanding and targeting resistance mechanisms in NSCLC. *Nat Rev Cancer.* 2017;17(11):637–58.
- Cross DA, Ashton SE, Ghorghiu S, Eberlein C, Nebhan CA, Spitzler PJ, et al. AZD9291, an irreversible EGFR TKI, overcomes T790M-mediated resistance to EGFR inhibitors in lung cancer. *Cancer Discov.* 2014;4(9):1046–61.
- Bang Y, Kwak E, Shaw A, Camidge D, Iafate A, Maki R, et al. Clinical activity of the oral ALK inhibitor PF-02341066 in ALK-positive patients with non-small cell lung cancer (NSCLC). *J Clin Oncol.* 2010;28(18_suppl):3.
- Arnaoutakis K. Crizotinib in ROS1-rearranged non-small-cell lung cancer. *N Engl J Med.* 2015;372(7):683.
- Emens LA, Ascierto PA, Darcy PK, Demaria S, Amm E, Redmond WL, et al. Cancer immunotherapy: opportunities and challenges in the rapidly evolving clinical landscape. *Eur J Cancer.* 2017;81:116–29.
- Wakelee H, Kelly K, Edelman MJ. 50 Years of progress in the systemic therapy of non-small cell lung cancer. *Am Soc Clin Oncol.* 2014;34(34):177.
- Brahmer J, Reckamp KL, Baas P, Crinò L, Eberhardt WE, Poddubska E, et al. Nivolumab versus docetaxel in advanced squamous-cell non-small-cell lung cancer. *N Engl J Med.* 2015;373(2):123–35.
- Herbst RS, Soria JC, Kowanetz M, Fine GD, Hamid O, Gordon MS, et al. Predictive correlates of response to the anti-PD-L1 antibody MPDL3280A in cancer patients. *Nature.* 2014;515(7528):563–7.
- Hellmann MD, Rizvi NA, Goldman JW, Gettinger SN, Borghaei H, Brahmer JR, et al. Nivolumab plus ipilimumab as first-line treatment for advanced non-small-cell lung cancer (CheckMate 012): results of an open-label, phase 1, multicohort study. *Lancet Oncol.* 2017;18(1):31–41.
- Alexander WJPT. The checkpoint immunotherapy revolution: what started as a trickle has become a flood, despite some daunting adverse effects; new drugs, indications, and combinations continue to emerge. *Pharmacy and Therapeutics.* 2016;41(3):185–91.
- Howitt BE, Shukla SA, Sholl LM, Ritterhouse LL, Watkins JC, Rodig S, et al. Association of Polymerase e-mutated and microsatellite-unstable endometrial cancers with Neoantigen load, number of tumor-infiltrating lymphocytes, and expression of PD-1 and PD-L1. *sJAMA Oncol.* 2015;1(9):1319–23.
- Zacharakis N, Chinnasamy H, Black M, Xu H, Lu YC, Zheng Z, et al. Immune recognition of somatic mutations leading to complete durable regression in metastatic breast cancer. *Nat Med.* 2018;24(6):724–30.
- Bonanno L, Pavan A, Dieci M, Di Liso E, Schiavon M, Comacchio G, et al. The role of immune microenvironment in small-cell lung cancer: distribution of PD-L1 expression and prognostic role of FOXP3-positive tumour infiltrating lymphocytes. *Eur J Cancer.* 2018;101:191–200.
- Xu GL, Bestor TH, Bourchis D, Hsieh CL, Tommerup N, Bugge M, et al. Chromosome instability and immunodeficiency syndrome caused by mutations in a DNA methyltransferase gene. *Nature.* 1999;402(6758):187–91.
- Gaudet F, Hodgson JG, Eden A, Jackson-Grusby L, Dausman J, Gray JW, et al. Induction of tumors in mice by genomic hypomethylation. *Science.* 2003;300(5618):489–92.

20. Ogino S, Kawasaki T, Nosho K, Ohnishi M, Suemoto Y, Kirkner GJ, et al. LINE-1 hypomethylation is inversely associated with microsatellite instability and CpG island methylator phenotype in colorectal cancer. *Int J Cancer*. 2008; 122(12):2767–73.
21. Martínez JG, Perez-Escuredo J, Castro-Santos P, Marcos CA, Pendas JL, Fraga MF, et al. Hypomethylation of LINE-1, and not centromeric SAT- α , is associated with centromeric instability in head and neck squamous cell carcinoma. *Cell Oncol (Dordr)*. 2012;35(4):259–67.
22. Wang K, Li M, Hakonarson H. ANNOVAR: functional annotation of genetic variants from high-throughput sequencing data. *Nucleic Acids Res*. 2010; 38(16):e164.
23. Mayakonda A, Lin D-C, Assenov Y, Plass C, Koeffler HP. Maftools: efficient and comprehensive analysis of somatic variants in cancer. *Genome research*. 2018;28(11):1747–56.
24. Rosenthal R, McGranahan N, Herrero J, Taylor BS, Swanton CJGB. deconstructSigs: delineating mutational processes in single tumors distinguishes DNA repair deficiencies and patterns of carcinoma evolution. *Genome Biol*. 2016;17(1):31.
25. Kolde RJ. Pheatmap: pretty heatmaps; 2012. p. 61.
26. Morris TJ, Butcher LM, Feber A, Teschendorff AE, Chakravarthy AR, Wojdacz TK, et al. ChAMP: 450k Chip analysis methylation pipeline. *Bioinformatics*. 2014;30(3):428–30.
27. Butcher LM, Beck S. Probe lasso: a novel method to rope in differentially methylated regions with 450k DNA methylation data. *Methods*. 2015;72:21–8.
28. Seiler M, Huang CC, Szalma S, GJOAJolB B. ConsensusCluster: a software tool for unsupervised cluster discovery in numerical data. *OMICS A J Integr Biol*. 2010;14(1):109–13.
29. Hovestadt V, Zapatka M. conumee: enhanced copy-number variation analysis using Illumina 450k methylation arrays. R package version 099; 2015. p. 4.
30. Mermel CH, Schumacher SE, Hill B, Meyerson ML, Beroukhir R, Getz G. GISTIC2.0 facilitates sensitive and confident localization of the targets of focal somatic copy-number alteration in human cancers. *Genome Biol*. 2011;12(4):R41.
31. Taylor AM, Shih J, Ha G, Gao GF, Zhang X, Berger AC, et al. Genomic and functional approaches to understanding cancer aneuploidy. *Cancer Cell*. 2018;33(4):676–89. e3.
32. Auslander N, Heselmeyer-Haddad K, Patkar S, Hirsch D, Camps J, Brown M, et al. Cancer-type specific aneuploidies hard-wire chromosome-wide gene expression patterns of their tissue of origin. *BioRxiv*. 2019:563858. <https://doi.org/10.1101/563858>.
33. Peters S, Creelan B, Hellmann MD, Socinski MA, Reck M, Bhagavatheeswaran P, et al. Abstract CT082: impact of tumor mutation burden on the efficacy of first-line nivolumab in stage iv or recurrent non-small cell lung cancer: an exploratory analysis of CheckMate 026. *Cancer Res*. 2017;77(13 Supplement): CT082.
34. Aryee MJ, Jaffe AE, Corrada-Bravo H, Ladd-Acosta C, Feinberg AP, Hansen KD, et al. Minfi: a flexible and comprehensive Bioconductor package for the analysis of Infinium DNA methylation microarrays. *Bioinformatics*. 2014; 30(10):1363–9.
35. Eden A, Gaudet F, Waghmare A, Jaenisch R. Chromosomal instability and tumors promoted by DNA hypomethylation. *Science*. 2003;300(5618):455.
36. Lappin TR, Grier DG, Thompson A, Halliday HL. HOX genes: seductive science, mysterious mechanisms. *The Ulster medical journal*. 2006;75(1):23.
37. Scott MPJC. Vertebrate homeobox gene nomenclature. *Cell*. 1992;71(4): 551–3.
38. Piñero J, Bravo À, Queralt-Rosinach N, Gutiérrez-Sacristán A, Deu-Pons J, Centeno E, et al. DisGeNET: a comprehensive platform integrating information on human disease-associated genes and variants. *Nucleic Acids Res*. 2016;45(D1):D833–D9.
39. Cortez MA, Ivan C, Valdecanas D, Wang X, Peltier HJ, Ye Y, et al. PDL1 regulation by p53 via miR-34. *J Natl Cancer Inst*. 2016;108(1):djv303.
40. Mei J, Liu Y, Li Q, Li X, Ning Z, Zhao W, et al. PD-1/PD-L1 expression in non-small-cell lung cancer and its correlation with EGFR/KRAS mutations. *Cancer Biol Ther*. 2016;17(4):407–13.
41. Zhang XT, Li LY, Mu XL, Cui QC, Chang XY, Song W, et al. The EGFR mutation and its correlation with response of gefitinib in previously treated Chinese patients with advanced non-small-cell lung cancer. *Ann Oncol*. 2005;16(8):1334.
42. Mu XL, Li LY, Zhang XT, Wang MZ, Feng RE, Cui QC, et al. Gefitinib-sensitive mutations of the epidermal growth factor receptor tyrosine kinase domain in chinese patients with non-small cell lung cancer. *Clin Cancer Res*. 2005; 11(12):4289.
43. Alexandrov LB, Nikzainal S, Wedge DC, Aparicio SAJR, Behjati S, Biankin AV, et al. Signatures of mutational processes in human cancer. *Nature*. 2013; 500(7463):415–21.
44. Wu K, Zhang X, Li F, Xiao D, Hou Y, Zhu S, et al. Frequent alterations in cytoskeleton remodelling genes in primary and metastatic lung adenocarcinomas. *Nat Commun*. 2015;6:10131.
45. Duruisseaux M, Martínez-Cardús A, Calleja-Cervantes ME, Moran S, de Moura MC, Davalos V, et al. Epigenetic prediction of response to anti-PD-1 treatment in non-small-cell lung cancer: a multicentre, retrospective analysis. *The Lancet Respiratory Medicine*. 2018;6(10):771–81.
46. Rauch T, Wang Z, Zhang X, Zhong X, Wu X, Lau SK, et al. Homeobox gene methylation in lung cancer studied by genome-wide analysis with a microarray-based methylated CpG island recovery assay. *Proc Natl Acad Sci U S A*. 2007;104(13):5527–32.
47. Hwang J-A, Lee BB, Kim Y, Park S-E, Heo K, Hong S-H, et al. HOXA11 hypermethylation is associated with progression of non-small cell lung cancer. *Oncotarget*. 2013;4(12):2317.
48. Spigel DR, Schrock AB, Fabrizio D, Frampton GM, Sun J, He J, et al. Total mutation burden (TMB) in lung cancer (LC) and relationship with response to PD-1/PD-L1 targeted therapies. *Journal of Clinical Oncology*. 2016;34(15_ suppl):9017

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Ready to submit your research? Choose BMC and benefit from:

- fast, convenient online submission
- thorough peer review by experienced researchers in your field
- rapid publication on acceptance
- support for research data, including large and complex data types
- gold Open Access which fosters wider collaboration and increased citations
- maximum visibility for your research: over 100M website views per year

At BMC, research is always in progress.

Learn more biomedcentral.com/submissions

