ORIGINAL ARTICLE

The average copy number variation (CNVA) of chromosome fragments is a potential surrogate for tumor mutational burden in predicting responses to immunotherapy in non-small-cell lung cancer

Yuanyuan Lei[1†], Guochao Zhang[1†], Chaoqi Zhang[1], Liyan Xue[2], Zhenlin Yang[1], Zhiliang Lu[1], Jianbing Huang[1], Ruochuan Zang[1], Yun Che[1], Shuangshuang Mao[1], Lingling Fang[1], Chengming Liu[1], Xinfeng Wang[1], Sufei Zheng[1], Nan Sun[1] & Jie He[1]

[1]Department of Thoracic Surgery, National Cancer Center/National Clinical Research Center for Cancer/Cancer Hospital, Chinese Academy of Medical Sciences and Peking Union Medical College, Beijing, China
[2]Department of Pathology, National Cancer Center/National Clinical Research Center for Cancer/Cancer Hospital, Chinese Academy of Medical Sciences and Peking Union Medical College, Beijing, China

Abstract

Objectives. The tumor mutational burden (TMB) is closely related to immunotherapy outcome. However, the cost of TMB detection is extremely high, which limits its use in clinical practice. A new indicator of genomic instability, the average copy number variation (CNVA), calculates the changes of 0.5-Mb chromosomal fragments and requires extremely low sequencing depth. **Methods.** In this study, 50 samples (23 of which were from patients who received immunotherapy) were subjected to low-depth (10X) chromosome sequencing on the MGI platform. CNVA was calculated by the formula avg (abs (copy number-2)). In addition, CNVA and TMB were compared with regard to their ability to predict immune infiltration in 509 patients from TCGA. **Results.** The high-CNVA group had higher expression levels of PD-L1, CD39 and CD19 and a higher degree of infiltration of CD8$^+$ T cells and CD3 $^+$ T cells. Among the 23 patients treated with immunotherapy, the average CNVA value of the stable disease/partial response group was higher than that of the progressive disease group ($P < 0.05$). Whole-genome sequencing data of 509 patients from TCGA and RT-PCR results of 22 frozen specimens showed that CNVA is more effective than TMB in indicating infiltration of CD8$^+$ T cells and expression of PD-L1, and CNVA also showed a specific positive correlation with TMB ($r = 0.2728$, $P < 0.0001$). **Conclusions.** Copy number variation can be a good indicator of immune infiltration and immunotherapy efficacy, and with its low cost, it is expected to become a substitute for TMB.

Keywords: CNVA, immune infiltration, immunotherapy, non-small-cell lung cancer, tumor mutational burden

INTRODUCTION

The advent of immunotherapy based on checkpoint inhibitors (ICIs) targeting the programmed cell death 1 (PD-1) axis has brought breakthrough progress to the treatment of non-small-cell lung cancer. However, due to tumor heterogeneity and the dynamics of PD-L1 expression, only a subset of patients shows a durable clinical benefit from such therapies. The precise application of ICI therapy requires reasonable predictive indicators. According to reports, patients with a PD-L1 tumor proportion score (TPS) > 50% benefit more than those with a TPS < 50%.[1,2]

In addition to the TPS, one of the most important predictors of immune efficacy is the tumor mutational burden (TMB). High TMB is associated with a poor prognosis of various tumors.[3-5] More importantly, TMB is an independent biomarker of ICI response,[6] as a high TMB is associated with high neoantigen burden and high T-cell infiltration. Initially, TMB assessment required whole-exome sequencing (WES) of the specimen. TMB reflects the number of somatic nonsynonymous mutations, including coding single-nucleotide variations (SNVs) and indels and is reported in numbers per Mb. However, the detection cost of WES is particularly high and the data are relatively difficult to analyse which is not suitable for large-scale analysis. Recently, some reports have simplified this process to estimate TMB with targeted next-generation sequencing (NGS) panels.[7] The most used assays are as follows: MSKIMPACT™[8] and DFCI OncoPanel.[9] Compared with that detected by WES, the number of genes detected by targeted NGS sequencing is small, which reduces the cost of testing to a certain extent. However, to compensate for the insufficient number of genes, the depth of NGS sequencing is far higher than that for WES, making the data difficult to analyse.

Tumor mutational burden is essentially caused by genomic instability. There are different causes of genomic instability, among which chromosomal instability (CIN), microsatellite instability (MSI) and DNA damage response (DDR) have attracted substantial attention. The relationship between microsatellite instability and cancer has been confirmed. For example, high MSI is closely related to the efficacy of immunotherapy.

Researchers found that the majority (82.1%) of MSI-high tumors were TMB-high; however, only 18.3% of TMB-high tumors were MSI-H.[10] This finding shows that TMB is not caused only by MSI. In addition, the combination of MSI and TMB detection is better at predicting the efficacy of immunotherapy than TMB alone.[10-12] Specifically, in colorectal cancer patients with high MSI, TMB is a better indicator of the efficacy of ICIs.[13,14] In small-cell lung cancer (SCLC), a large number of DDR phenomena have been discovered and the DDR pathway alterations have a positive correlation with high TMB.[15] Because the cost of DDR testing is substantially lower than that of TMB, it is expected to become a substitute for TMB.[16] Nonetheless, the direct relationship between chromosomal instability and cancer is different in different situations. CIN is mainly reflected by an increase or deletion of fragments. The sequencing depth required for detection is also far lower than that of genome TMB detection, which greatly reduces the detection cost and analysis difficulty. The deletion or amplification of the entire chromosome will lead to the generation of aneuploidy. Compared with normal diploid cells, most tumor cells are in an aneuploid state.[17,18] Different types of ploidy have different effects on the occurrence and development of tumors.[19] The overall prognosis of patients with polyploid tumors is poor,[20,21] and polyploid tumors are more malignant than haploid or diploid tumors.[22,23] However, analysis of the entire chromosome can be only used for the qualitative analysis of tumor specimens, and such analysis cannot provide specific values like TMB for the quantitative analysis of different specimens, but it can be good for predicting the therapeutic effect.

To address this problem, we broke the chromosome into small 0.5-Mb fragments, calculated the copy number changes of all fragments, averaged the absolute value (CNVA) and then evaluated the correlation between CNVA and tumor immune infiltration or immunotherapy effect. In addition, we compared the utility of CNVA with that of TMB. We found that specimens with a high CNVA had more chromosomes alterations, more obvious immune infiltration and were more sensitive to immunotherapy drugs than specimens with a low CNVA.

RESULTS

Distribution of chromosomal differences in patients treated with different immunotherapies

This study included 23 patients receiving immunotherapy and 27 non-immunotherapy patients undergoing thoracic surgery. The clinical information is shown in Table 1, and the specific immunotherapy methods are shown in Supplementary table 2. According to the method described above, we performed chromosomal variation detection with paraffin-embedded samples of these 50 specimens. Low-depth sequencing (10 X) was performed to analyse the copy numbers of these small fragments (0.5 Mb). The visualisation of the fluctuation of these small fragments of chromosomes is shown in Figure 1a–c. We observed that chromosomes 1– 23 showed different degrees of fluctuation in different specimens, and the copy number of the small fragments (0.5 Mb) of each chromosome fluctuated around the baseline. To quantify this fluctuation state, we proposed a new marker CNVA, which is the average of the copy numbers of all these small 0.5-Mb fragments and is calculated with formula avg (abs (copy number-2)). When the CNVA value was large (0.4032) (Figure 1a), the fluctuation degree of the chromosome was significantly increased, and the copy value of the 0.5 Mb fragment was randomly distributed around the baseline; however, when the CNVA value was small (0.1185) or moderate (0.2491), the fluctuation of chromosomes was low (Figure 1b, c). What's more, we performed 200X high-depth sequencing (whole-exome sequencing) on 17 specimens (with known CNVA values) that have been subjected to low-depth sequencing, among the specimens with higher CNVA (p7, p18, p6 and p17), the probability of driver gene mutation was higher (Supplementary figures 1- 3). The above results indicated that the CNVA value was a good indicator of chromosome fluctuations.

High CNVA value indicates high immune infiltration status

Among these 50 specimens, we found that CNVA values were mainly distributed between 0.1 and 0.5 (Figure 2a). The CNVA value was not related

Table 1. Clinical information for immunotherapy and non-immunotherapy patients

Characteristics	Patient (*n* = 50)		CNVA value		
	NO.	%	Low	High	*P*-value
All patients	50	100.0	24	26	
Age					
≤ 60	27	54.0	15	12	0.247
> 60	23	46.0	9	14	
Gender					
Male	29	58.0	12	17	0.271
Female	21	42.0	12	9	
Tumor size					
< 4 cm	36	75.0	18	18	1.000
≥ 4 cm	12	25.0	6	6	
Tumor stage					
I - II	32	65.3	14	18	0.414
III - IV	16	32.7	9	7	
LN metastasis					
Absent	28	56.0	12	16	0.412
Present	22	44.0	12	10	

to age, sex, tumor size and stage (Table 1). To further demonstrate that the CNVA value could indeed reflect the degree of chromosome fluctuations, in addition to focusing on the copy number of 0.5 Mb fragments, we used the Nipt algorithm[24] to analyse the amplification or deletion of the entire chromosome. The value returned by the algorithm was defined as the Z value, Z value > 3 indicated a triploid or higher state and Z value < -3 indicated a haploid or lower state. We divided the 50 specimens into high, moderate and low groups. Then we drew a heat map based on the CNVA and Z values. Green represents the group with a low CNVA value; orange represents the group with a high CNVA value, and blue represents the group with a moderate CNVA value. The lightest colour in the heat map, that is, the part with the least degree of chromosome aneuploidy, was basically concentrated in the group with low or moderate CNVA value. Conversely, the high-CNVA group was particularly dark in colour, and the chromosome aneuploidy state was more obvious (Figure 2b). This finding demonstrated that CNVA could indeed reflect the aneuploid or stable state of the entire chromosome through the fluctuation of 0.5 Mb fragments.

Among the 50 specimens, there were also 27 frozen tissue samples from patients who did not receive immunotherapy. To evaluate the
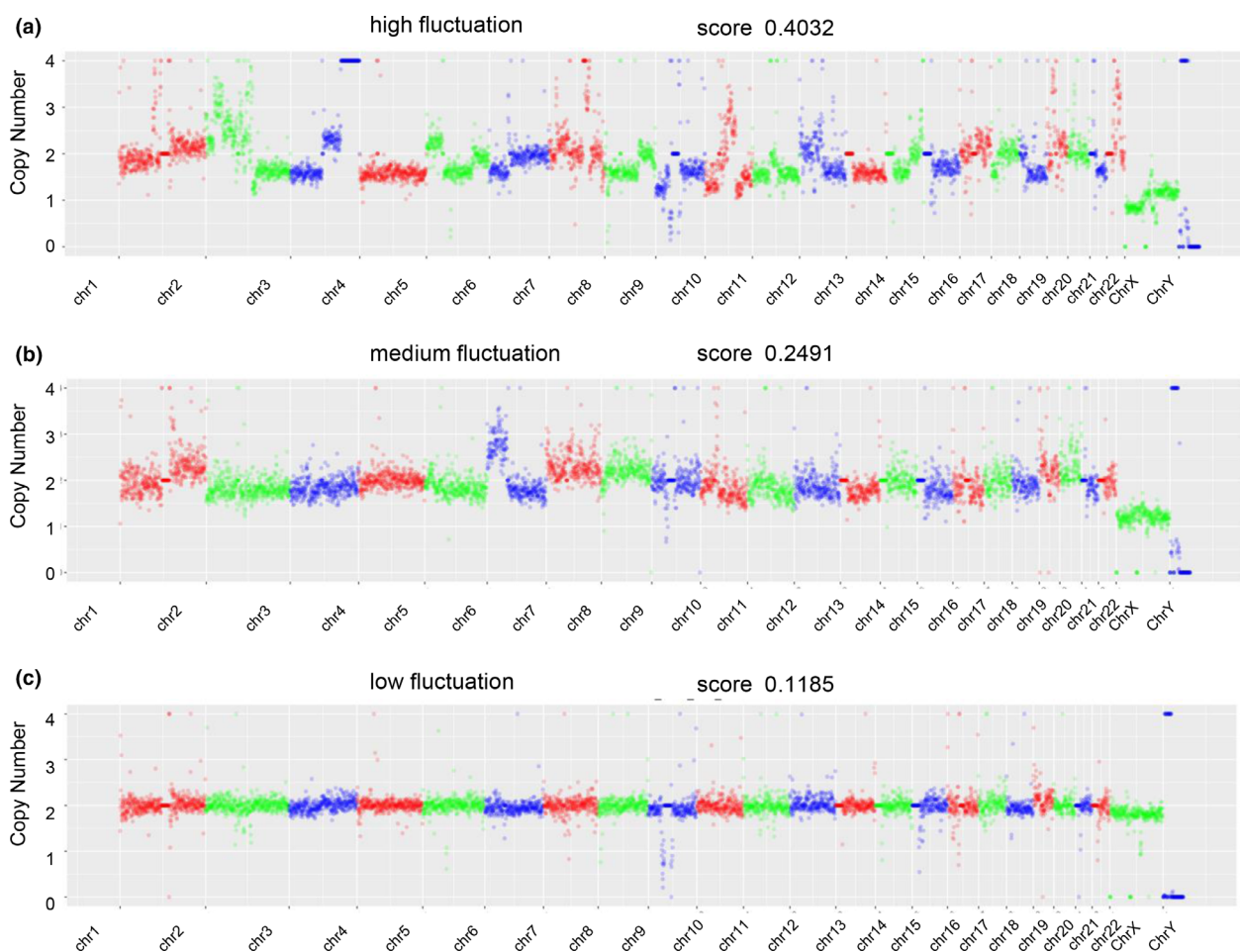
Figure 1. Fluctuation status of chromosomes in different specimens. **(a)** When the CNVA score was 0.4032, the chromosome shows a high fluctuation state. **(b)** When the CNVA score was 0.4032, the chromosome shows a moderate fluctuation state. **(c)** When the CNVA score was 0.1185, the chromosome shows a low degree of fluctuation.

relationship between CNVA and immune infiltration, we extracted RNA from these 27 frozen tissues specimens. Five of the specimens were of poor quality and were discarded. Then, we detected the expression of immune-related genes via immunohistochemistry in the remaining 22 samples by real-time quantitative PCR. The results showed that compared with the CNVA low group, the high-CNVA group had significantly upregulated levels of CD8 ($P = 0.0067$), PD-L1 ($P = 0.0477$), CD39 ($P = 0.0337$) and the B cell marker gene CD19 ($P = 0.0464$). The average expression level of immunosuppressive genes Foxp3 and CTLA-4 was downregulated in the CNVA high group but there was no significant difference (Figures 2c and 7e). In addition, we detected expression of CD3, CD8 and PD-L1 in 50 immunotherapy and

non-immunotherapy samples by immuno histochemistry. The results showed that the high-CNVA group had higher expression of CD3 (0.0370), CD8 (0.0495) and PD-L1 (Figure 2d–j) than the group with low CNVA. The above results indicate that high CNVA positively correlated with immune infiltration.

Correlation between the effect of immunotherapy and immune infiltration

We retrospectively collected clinical information on patients receiving immunotherapy at the Cancer Hospital of the Chinese Academy of Medical Sciences. Among them, 21 patients underwent surgery before receiving immunotherapy and had paraffin-embedded samples available. The other two patients
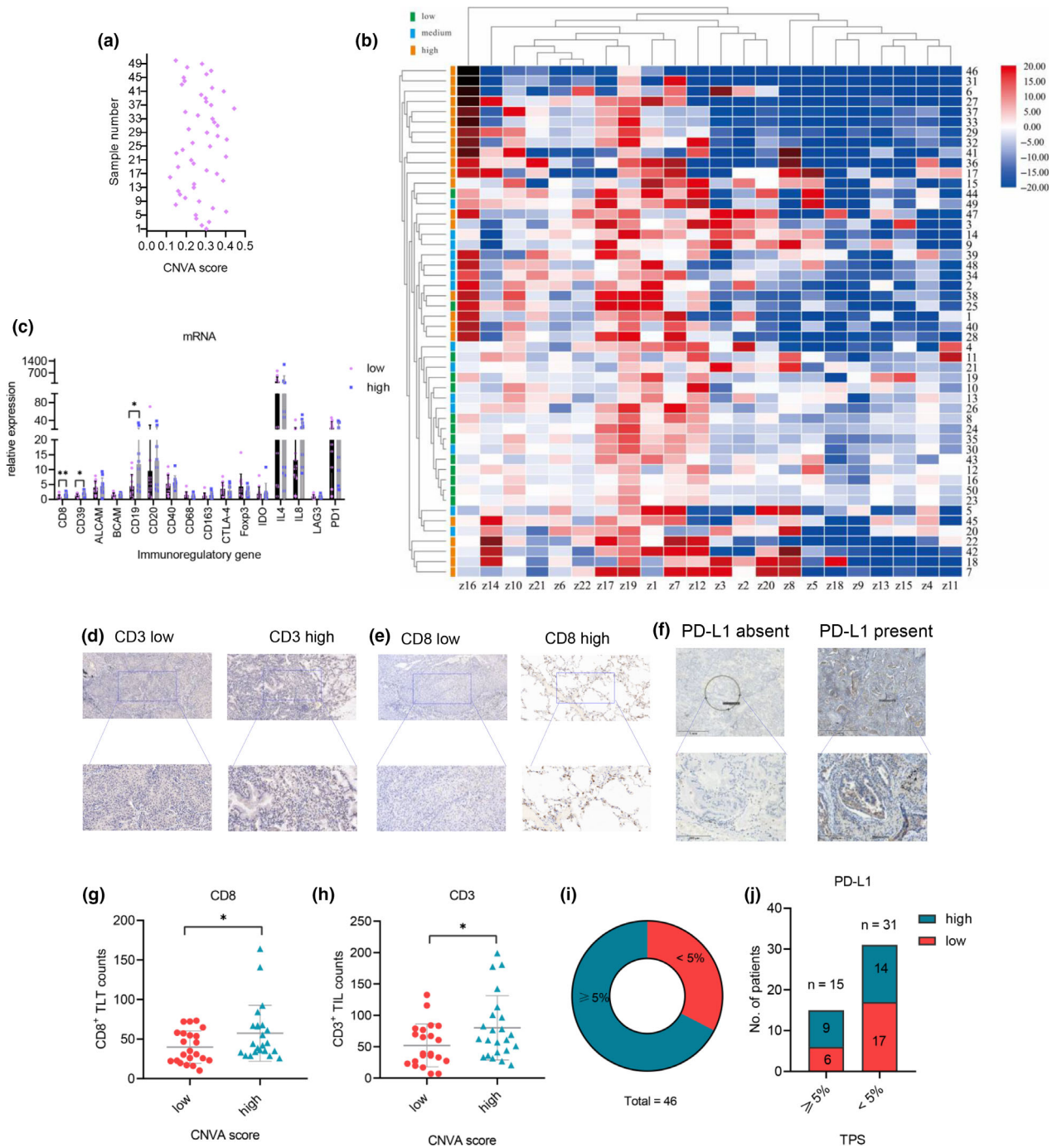
Figure 2. High CNVA value indicated high immune infiltration status **(a)**. A scatterplot depicting the distribution of CNVA score in 50 specimens. **(b)** The entire deletion or increase of chromosomes 1–23 in the three groups of CNVA score high, moderate and low. *Z* value > 3 meant chromosome is triploid or more; *Z* value < −3 meant chromosome is haploid or less. **(c)** The mRNA expression of different immunoblot-related genes in 22 frozen specimens detected by q-PCR. **(d–f)**. Representative immunohistochemical images of CD8, CD3 and PD-L1. **(g, h)**. Scatterplots show immunohistochemical scores of CD8 and CD3 in high-CNVA group and low-CNVA group. **(i)** The proportion of 46 specimens with PD-L1 TPS ≥ 5% and < 5%. **(j)**. The specific distribution of CNVA score in PD-L1 ≥ 5% and < 5% groups. The results are the average of at least 2 independent experiments with a minimum of 2 technical replicates per experiment, and the data are presented as the mean ± SD, *P < 0.05, **P < 0.01, ***P < 0.001. ns indicates no significance.

underwent biopsy before immunotherapy, and paraffin samples were also available. The clinical information is shown in Table 2. Of these 23 patients, 14 had lung adenocarcinoma and 9 had squamous cell carcinoma. All patients received second-, third-, or fourth-line immunotherapy after surgery or biopsy. Based on the evaluation of the efficacy after 12 weeks of immunotherapy, 12 of the 23 patients were judged as achieving stable disease (SD) or a (partial response) PR and 11 were judged as having (progressive disease) PD (efficacy assessment was completed by the clinician). Patients with SD or PR were classified into a single group with good efficacy, and patients with PD were classified as a group with poor efficacy. We performed immunohistochemical analysis on the paraffin samples of these 23 patients. Compared with the group with poor efficacy, the group with good efficacy showed increased expression of the classic immune infiltration marker CD3 (Figure 3a–c) ($P < 0.05$), CD8 (Figure 3d–f) (there was no significant difference which might be due to the small sample size) and PD-L1 (Figure 3g–i). This result is consistent with literature reports.[25]

High CNVA value may predict a superior immunotherapy effect

After demonstrating that CNVA had a specific correlation with immune infiltration, we next compared the relationship between immunotherapy efficacy and CNVA in the 23 patients. The CNVA values of these patients were mainly between 0.1 and 0.4 (Figure 4a). We divided these 23 patients into two groups according to efficacy, with patients with PD as one group, and patients with PR/SD as another group. The scatter plot indicates that the average CNVA value of the PD group was lower than that of the PR/SD group with a significant difference ($P = 0.0204$) (Figure 4b). However, we only included a small number of immunotherapy patients, and these patients were not strictly first-line immunotherapy patients, as many of them had undergone radiotherapy and chemotherapy or targeted therapy. These treatment methods may affect certain chromosomes fragments. Therefore, we further analysed the copy number (aneuploidy) of the whole chromosome of chromosomes 1–23 in different therapeutic groups using the Nipt method. The frequency of deletion or amplification

Table 2. Specific treatment information for immunotherapy patients

Patient ID	TNM stage	Pathology	Specimen category	Immunotherapy	Treatment effect
1	T2N2M0	Adenocarcinoma	Surgery	Third line	PD
2	T2aN0M0	Adenocarcinoma	Surgery	Second line	PR
3	T2N0M0	Adenocarcinoma	Surgery	Second line	SD
4	T2aN0M0	Adenocarcinoma	Surgery	Third line	PD
5	T2N2M0	Adenocarcinoma	Surgery	Third line	PD
6	T2aN0M0	Squamous cell carcinoma	Surgery	Second line	SD
7	T2bN0M0	Adenocarcinoma	Surgery	Second line	SD
8	T1aN1M0	Squamous cell carcinoma	Surgery	Third line	SD
9	T2N0M0	Squamous cell carcinoma	Surgery	Second line	PR-SD
10	T1aN2M0	Adenocarcinoma	Surgery	Third line	PD
11	T2N1M0	Squamous cell carcinoma	Surgery	Third line	PD
12	T3N2M0	Adenocarcinoma	Surgery	Third line	PD
13	T2N0M0	Adenocarcinoma	Surgery	Third line	PR
14	T3N0M0	Squamous cell carcinoma	Surgery	Third line	PD
15	T2aN1M0	Squamous cell carcinoma	Surgery	Third line	PR
16	T2N2M0	Adenocarcinoma	Surgery	Fourth line	PD
17	T2N2M1	Squamous cell carcinoma	Surgery	Fourth line	PD
18	T1bN0M0	Adenocarcinoma	Surgery	Fourth line	SD
19	T2N0M0	Squamous cell carcinoma	Surgery	Third line	SD
20	ypT1N2M0	Squamous cell carcinoma	Surgery	Third line	PD
21	T3N2M0	Adenocarcinoma	Surgery	Third line	PD
22	/	Adenocarcinoma	Biopsy	Fourth line	SD
23	/	Adenocarcinoma	Biopsy	Third line	SD

PD, progressive disease; PR, partial response; SD, stable disease.
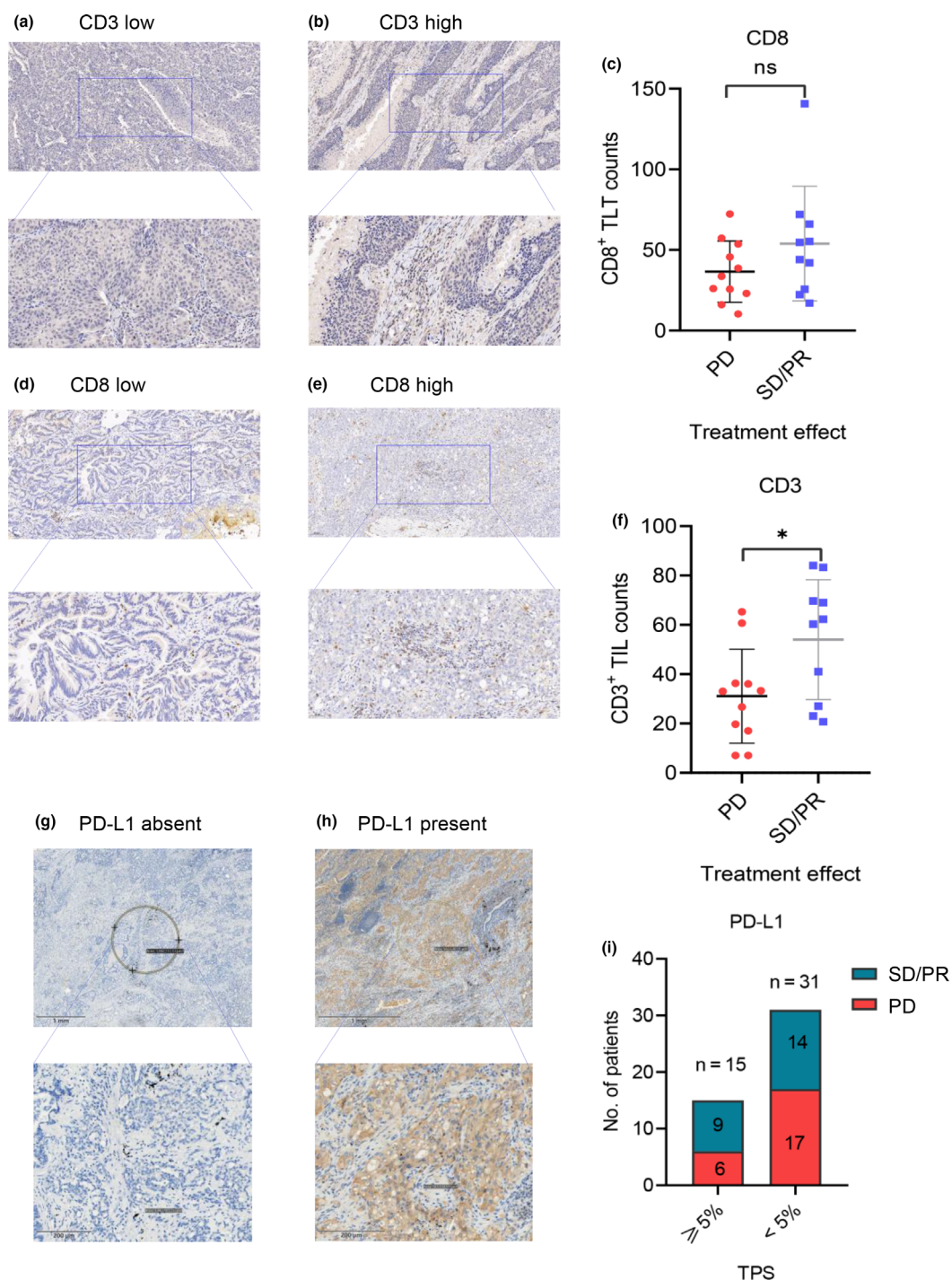
Figure 3. Relationship between immune efficacy and immune infiltration **(a)**. A scatterplot depicts the distribution of CNVA score in 50 specimens. **(b)** The entire deletion or increase of chromosomes 1–23 in the three groups of CNVA score high, moderate and low. *Z* value > 3 indicates that the chromosome is triploid or more; *Z* value < −3 meant chromosome is haploid or less. **(c)** The mRNA expression of different immunoblot-related genes in 22 frozen specimens detected by q-PCR. **(d–f)**. Representative immunohistochemical images of CD8, CD3 and PD-L1. **(g, h)**. Scatterplots show immunohistochemical scores of CD8 and CD3 in high-CNVA group and low-CNVA group. **(i)**. The proportion of 46 specimens with PD-L1 TPS ≥ 5% and < 5%. **(j)**. The specific distribution of CNVA score in PD-L1 ≥ 5% and < 5% groups. The results are the average of at least 2 independent experiments with a minimum of 2 technical replicates per experiment, and the data are presented as the mean ± SD, *P < 0.05, **P < 0.01, ***P < 0.001. ns indicates no significance.

of the entire chromosome is very likely to be lower than that of the small 0.5 Mb fragment, and it will be more stable under the influence of various treatment methods. Indeed, we found that the PR/SD group was mainly located in the lighter colour area of the heat map, while the PD group was basically in the darker colour area, that is, the area where the chromosome aneuploidy phenomenon was more obvious (Figure 4c). All the above results confirmed that the degree of chromosome fluctuation was indeed related to the therapeutic effect.

Verification of the correlation between chromosome copy number changes in terms of DNA content and immune infiltration

The most intuitive results from changes in the copy number of chromosomes are changes in the DNA content. The change in DNA content can roughly indicate whether there is chromosome aneuploidy or a major change in chromosome copy number. However, such changes will be noted only when the difference in DNA content caused by the change in the number of chromosomes can be detected. Therefore, a normal DNA content cannot exclude the existence of malignant or chromosomal abnormalities, but an abnormal DNA content always indicates chromosomal abnormalities. In other words, changes in DNA content will be more specific in reflecting chromosomal instability. Hence, we can indirectly confirm the relationship between chromosome copy number abnormalities and immune infiltration by analysing the relationship between DNA content and immune infiltration. We detected the relationship between DNA content (fluorescence value of DAPI) and PD-L1 in a series of lung cancer cell lines and lung cancer tissue samples by flow cytometry. A schematic diagram of the PD-L1 isotype control and positive specimen is shown in Figure 5a–d. We observed that in 15 lung cancer cell lines (the PD-L1 expression and DNA content of each cell line are detailed in Supplementary table 3), the fluorescence intensity of DAPI correlated negatively with PD-L1 expression (Figure 5e). However, in lung cancer tissue specimens, the optimal curve of the trend line was a polynomial of two terms (Figure 5f, g). These results showed that when a certain baseline was used as the standard (likely the DNA content of normal diploid cells), the DNA content changed towards one of two extremes, which could have either very small or very large numbers of chromosome copies (predicting chromosome fluctuations), PD-L1 expression would increase and the result was indeed similar to the CNVA effect (with an emphasis on the absolute value of copy number changes). Nevertheless, the cell line results did not exhibit two extreme, which we speculated might be related to the small number of cell lines and the relative concentration of DNA content.

The state of immune infiltration is related to the level of TMB

Tumor mutational burden is a widely studied indicator related to the efficacy of immunotherapy. We further analysed the correlation between TMB and various infiltrating immune cells in lung adenocarcinoma. We downloaded the information of 509 lung adenocarcinoma patients with both whole-genome and transcriptome data from TCGA. Whole-genome data were used to calculate the TMB value by ecTMB, and the transcriptome data were used to analyse the infiltration of various immune cells by CIBERSORT (https://cibersort.stanford.edu/). According to the TMB value, these 509 cases were divided into two groups: TMB high and TMB low. The results of the heat map clustering showed that the group with high TMB (blue) was roughly clustered together and had higher levels of CD8$^+$ T cells and M1 type macrophages and lower M2 type macrophages than the low TMB group (Figure 6a).To observe the difference in infiltrating lymphocytes between the TMB high and TMB low groups more clearly, we subdivided the 509 samples into three groups, TMB high, TMB moderate and TMB low, and plotted histograms. Compared with the group with low TMB, we found that the TMB high group, in addition to having obvious increases in CD8$^+$ T cells ($P < 0.0001$) (Figures 6b and 7c), also had significant increases in plasma cells and activated CD4$^+$ T cells ($P < 0.0001$) (Figure 6b). The ratio of M1 to M2 macrophages (which serves as an index of immune activity) in the group with high TMB was also significantly higher than that in the group with low TMB ($P < 0.0001$) (Figure 6c).

Comparison of the utility of CNVA and TMB in predicting response to immunotherapy

Both CNVA and TMB essentially reflect genomic instability. The former is concerned with the copy
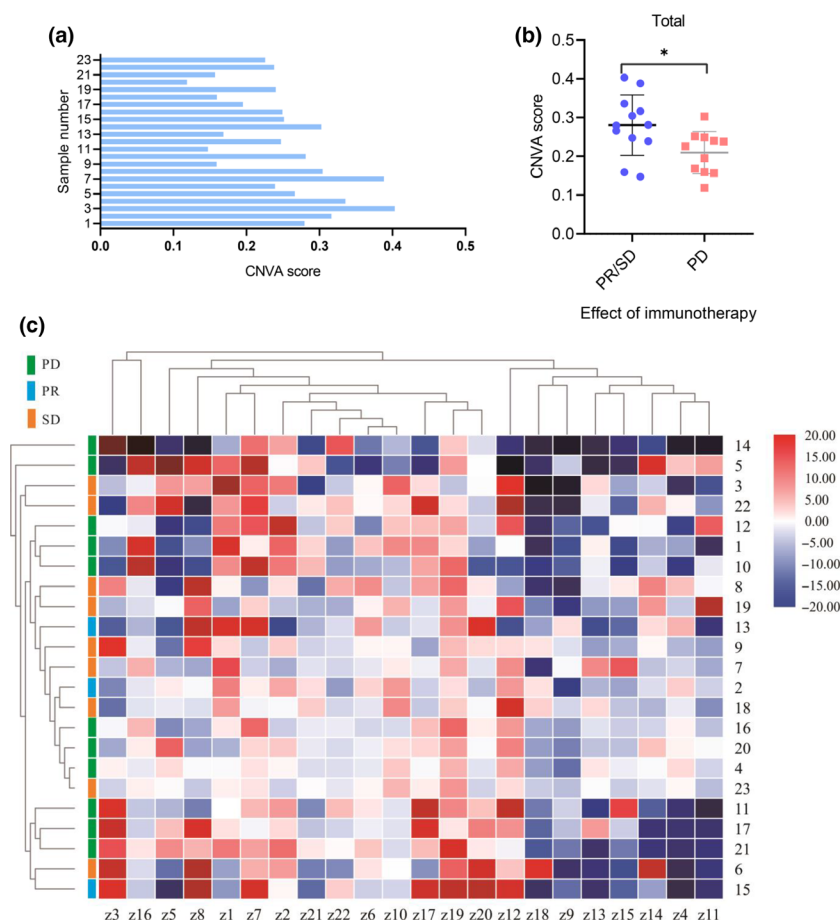
Figure 4. Correlation between CNVA score and immune efficacy. **(a)** The distribution of CNVA values in 23 immunotherapy specimens. **(b)**. The scatterplot shows the distribution of CNVA score in the PR/SD group and PD group. **(c)**. A heat map displays complete deletion or increase of chromosomes 1–23 in PR group, SD group and PD group. *Z* value > 3 indicates that the chromosome is triploid or more, *Z* value < −3 indicates that the chromosome is haploid or less. The results are the average of at least 2 independent experiments with a minimum of 2 technical replicates per experiment, and the data are presented as the mean ± SD, *$P < 0.05$, **$P < 0.01$, ***$P < 0.001$. ns indicates no significance.

number changes of chromosome fragments of 0.5 Mb at the macro level, while TMB is only concerned with the number of mutations per 1 Mb of DNA. In theory, the two values should correlate. The cost of TMB testing is extremely high and the analysis is difficult. We questioned whether CNVA could replace TMB as an indicator of immunotherapy efficacy. First, we compared the changes in the expression levels of two indexes PD-L1 and CD8, which are closely related to immunotherapy efficacy under different TMB or CNVA states. The TMB value was extracted from 509 samples from TCGA. The results showed that the expression of PD-L1 and the infiltration of CD8[+] T cells in the group with high TMB were significantly increased ($P < 0.01$) (Figure 7a, c), and both had a certain correlation with TMB

(although the correlation coefficient was not large) (Figure 7b, d). The CNVA data were obtained from the abovementioned 27 non-immunotherapy specimens, 22 of which had high-quality RNA for real-time quantitative PCR. We also found high expression of PD-L1 and CD8 in samples with high CNVA ($P < 0.05$) and this correlation was stronger than the correlation between high TMB and these two parameters (Figure 7e, g). In addition, the correlation coefficient was also higher than that for TMB, although the *P*-value of PD-L1 is not as significant as TMB due to the lack of samples (Figure 7f, h).

Finally, to further verify the correlation between CNVA and TMB, we reanalysed the whole-genome data of 509 samples from TCGA. Because the whole-genome sequencing process
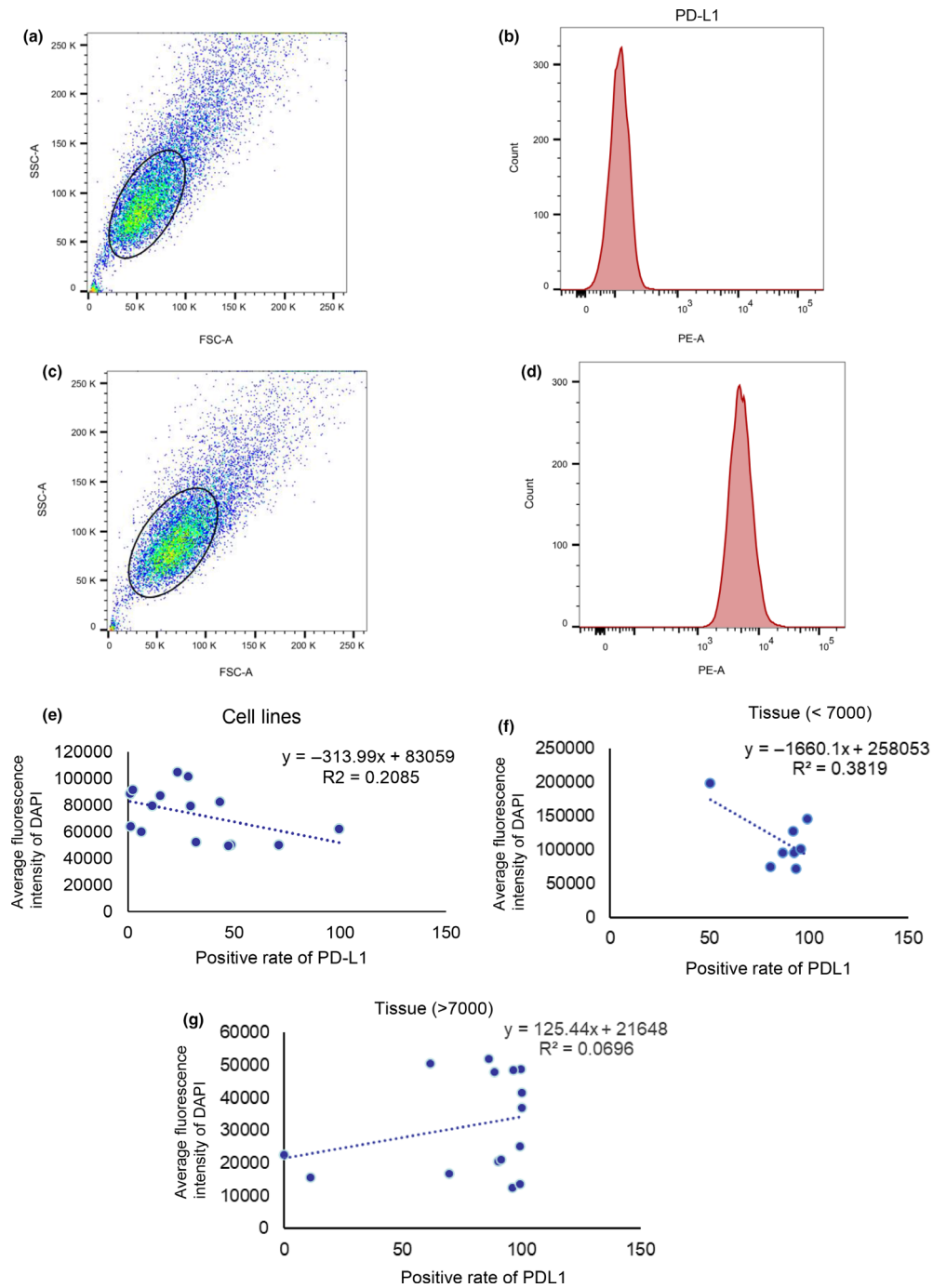
Figure 5. Verification of the correlation between chromosome copy number changes and immune infiltration at the DNA content level. **(a, b)** Representative images of the expression of PD-L1 in the isotype control by flow cytometry. **(c, d)**. Representative images of the expression of PD-L1 in the experimental group detected by flow cytometry. **(e)**. A scatterplot shows the relationship between DNA content and PD-L1 expression in 15 lung cancer cell lines (PC9, H157, H358, A549, H1299, H460, H1781, H1915, H2030, H827, SK-LU-1, H226, H322, H441, H1650). The vertical axis represents the average fluorescence value of DAPI, and the horizontal axis represented the percentage of PD-L1-positive cells. **(f, g)** A scatterplot showed the relationship between DNA content and PD-L1 expression in lung cancer tissue. The vertical axis represents the average fluorescence value of DAPI, and the horizontal axis represents the percentage of PD-L1-positive cells. The results are the average of at least 2 independent experiments with a minimum of 2 technical replicates per experiment, and the data are presented as the mean ± SD, *$P < 0.05$, **$P < 0.01$, ***$P < 0.001$, **** $P < 0.0001$ ns indicates no significance.
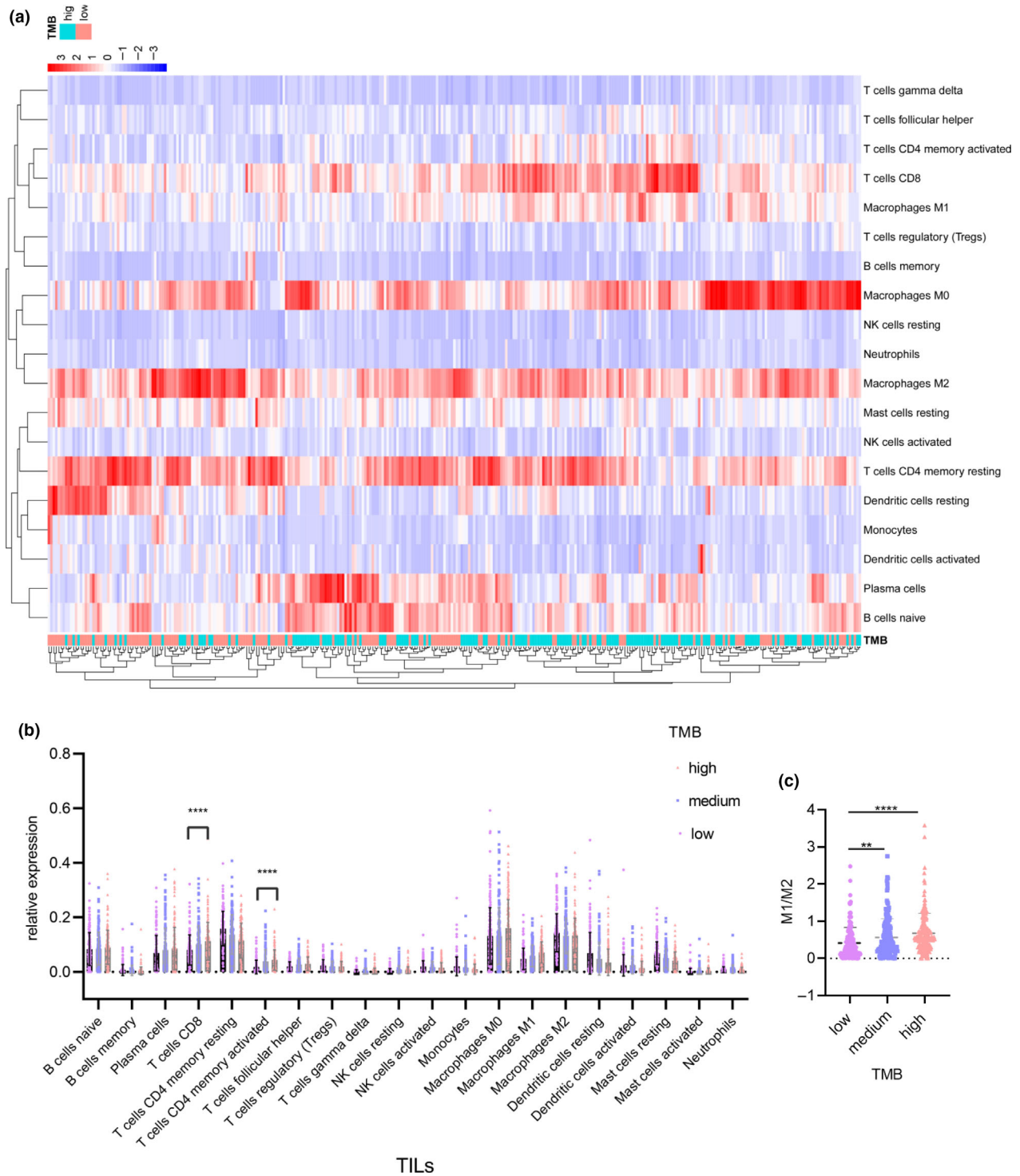
Figure 6. The relationship between TMB and immune infiltration. **(a)** A heat map shows the infiltration of 19 immune cell subclasses in the 509 TCGA specimens in the group with high TMB and low TMB. **(b)** The infiltration of 19 immune cells in r TMB high, moderate and low groups. **(c)** A scatterplot shows the ratio of M1 and M2 macrophages at different TMB levels in the 509 specimens. The results are the average of at least 2 independent experiments with a minimum of 2 technical replicates per experiment, and the data are presented as the mean $\pm$ SD, *$P < 0.05$, **$P < 0.01$, ***$P < 0.001$, ****$P < 0.0001$. ns indicates no significance.
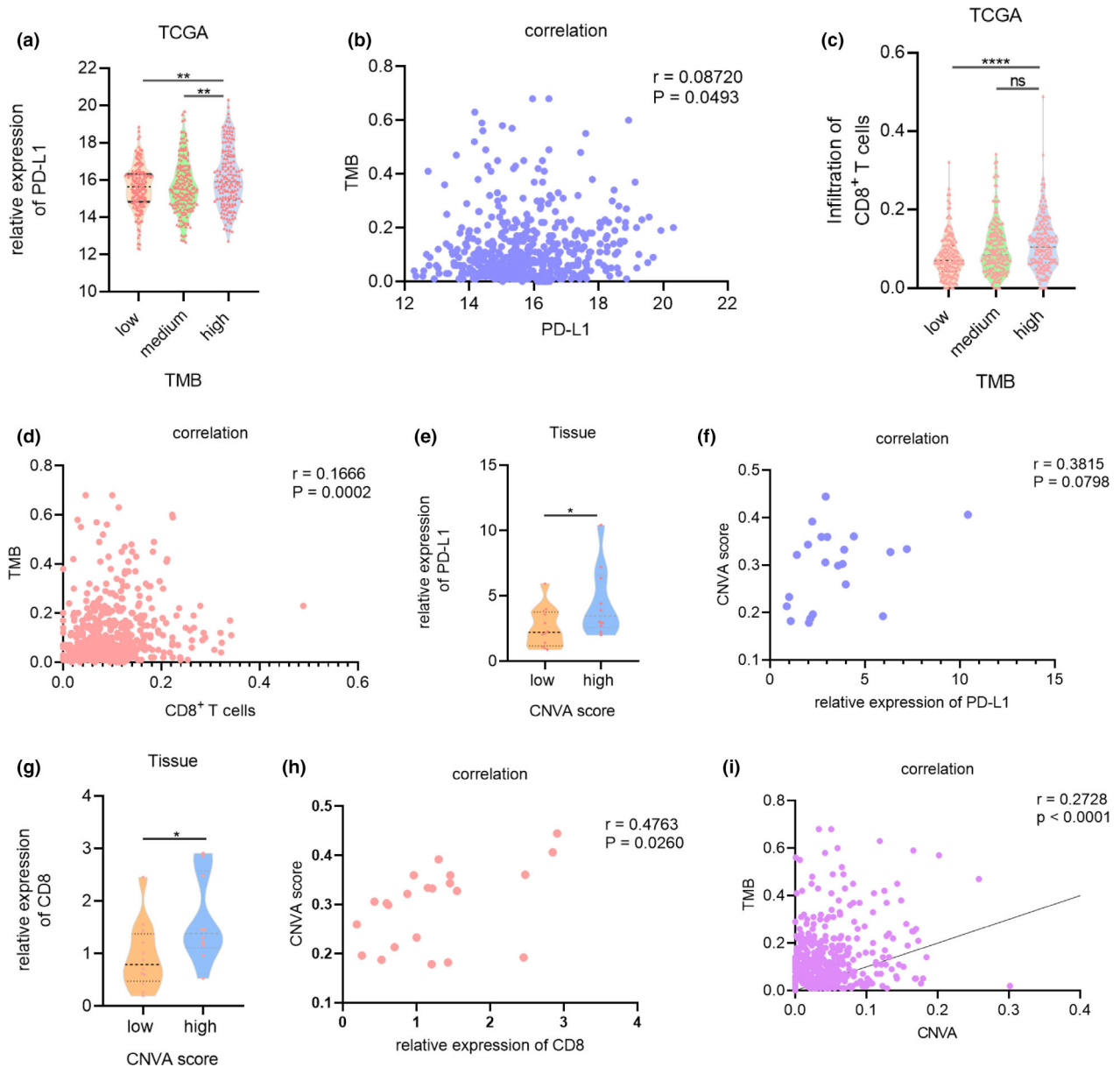
Figure 7. Comparison of the role of CNVA and TMB in the indication of immune efficacy and the relationship between the two. **(a)** A violin diagram showing the expression of PD-L1 in groups of 509 TCGA specimens with different levels of CNVA. **(b)** A scatterplot analysis of the correlation between PD-L1 and TMB. **(c)** A violin diagram showed the expression of CD8 in groups of 509 TCGA specimens with different levels of CNVA. **(d)** A scatterplot showing the correlation between CD8 and TMB. **(e)** The expression level of PD-L1 in 22 frozen tissue samples with different CNVA levels. **(f)** A scatterplot showing the relationship between PD-L1 and CNVA values in 22 frozen specimens. **(g)** The expression level of CD8 mRNA in 22 frozen tissue samples with different CNVA levels. **(h)** A scatterplot showing the relationship between CD8 and CNVA values in 22 frozen specimens. **(i)** Correlation between the average copy number change of all genes (including introns and exons) and TMB in 507 TCGA specimens. The results are the average of at least 2 independent experiments with a minimum of 2 technical replicates per experiment, and the data are presented as the mean ± SD, *$P < 0.05$, **$P < 0.01$, ***$P < 0.001$, ****$P < 0.0001$. ns indicates no significance.

does not include breaking the chromosome into small fragments of 0.5 Mb, and the data present only changes in mutations or copy number throughout the genome, we analysed the changes

in the copy number of all exons and introns, and the average value was calculated and used to replace the CNVA value. A Pearson correlation showed that there was indeed a clear correlation

between TMB and CNVA values ($r = 0.2728$, $P < 0.0001$) (Figure 7i). These results indicated that CNVA has the potential to replace TMB.

DISCUSSION

Current immunotherapy drugs are mainly inhibitors targeting PD-L1 / PD1 targets. By weakening the activity of the PD-L1/ PD-1 pathway, the patient's own immune system can re-recognise and kill tumor cells to overcome immune suppression. However, PD-L1 inhibitors are not effective for all cancer patients. For patients with lung cancer, the effective rate is only approximately 30%, while the effective rate of patients with positive PD-L1 expression may reach approximately 50%. Therefore, whether PD-L1 is highly expressed is currently the most important indicator to judge the efficacy of these drugs. Regardless, there are some obvious limitations regarding the detection of PD-L1. For example, some patients with negative PD-L1 will have a good response to immunotherapy; the detection of PD-L1 mainly depends on immunohistochemistry and is more affected by the type of antibody. In addition, the scoring standards of immunohistochemistry are different and subjective, bringing some difficulties to the accurate evaluation of the efficacy of immunotherapy. Therefore, a new evaluation index of immunotherapy efficacy is needed.

Genomic instability is a common characteristic of tumor cells. Gene mutations caused by this instability will further promote the production of a series of new antigens and change the immune microenvironment of the tumor tissue. Therefore, tumor mutation load (TMB) is a good indicator of immunotherapy. However, the detection of TMB requires whole-exon sequencing, and the cost is a large obstacle to its routine clinical use. In this study, we propose a new marker, CNVA, which does not consider the number of specific mutant genes, but detects the change in the average copy number of small chromosomes fragments (0.5 Mb) at the macro level. Similar to TMB, it can reflect the instability of the genome through the fluctuation of chromosomes to predict the immune infiltration state and immunotherapy efficacy.

A change in chromosome copy number can be reflected in a change in DNA content. Conversely, a change in DNA content can also be used to speculate whether there is an increase or decrease

in chromosome copy number. In fact, this concept was utilised very early in the DNA index (DI), that is, the ratio of the actual DNA content of the target sample to the DNA content of normal diploid cells. If the ratio is outside the diploid and tetraploid stages, the sample is defined as aneuploid,[22,26] such as haploid, triploid and polyploid. Normal cells have a relatively constant DNA content, but structural and/or chromosomal abnormalities are common during cell carcinogenesis. Compared with diploid tumors, these aneuploid tumors are generally associated with a poor prognosis and are insensitive to treatment,[18-20,23] although the mechanism has not been studied. We found that the DNA content and PD-L1 level have a certain degree of correlation by flow cytometry. Thus, we speculated that chromosomal abnormalities might cause gene instability and lead to increased mutation load (TMB), ultimately resulting in the dysregulation of the immune microenvironment. This was the original intention behind creating the CNVA marker. CNVA obtained by low-power sequencing may be a good substitute for TMB.

In this study, we confirmed that CNVA has a good correlation with immune infiltration. Among the 23 immunotherapy specimens, the average value of CNVA in the SD/PR group was significantly higher than that in the PD group. Nonetheless, the sample size was small, and further studies with larger sample sizes are needed to verify this conclusion. In addition, some patients treated with third- or fourth-line immunotherapy had received prior chemotherapy or radiotherapy, and it was unclear whether these treatments had any effect on chromosome fragments. The stability of the copy number of 0.5 Mb fragments was not as stable and fluctuated with the amplification or deletion of the entire chromosome, and the occurrence of chromosomal deletions requires greater external interference. We recommend that for low-depth sequencing samples, in addition to determining CNVA, the Nipt algorithm which analyses copy number of the entire chromosome of chromosomes 1–23 should be used.

To conclude, low-depth whole-genome sequencing data can reflect the bigger picture and be used to overcome inaccurate baseline data associated with high-precision sequencing and accurately reflect the fluctuation of chromosomes. The average copy number change (CNVA) of small chromosome fragments (0.5 Mb) is an important

indicator of genomic instability and can well predict the degree of immune infiltration and the effect of immunotherapy. The use of this marker may also greatly reduce testing costs and analysis difficulty, offering a convenient method for clinical testing.

METHODS

Patient samples

A total of 23 patients with non-small-cell lung cancer (NSCLC) treated with a PD-1 or PD-L1 inhibitor after surgery or biopsy were retrospectively identified from the Cancer Hospital of the Chinese Academy of Medical Sciences from April 2017 to June 2019. The fresh and frozen specimens used for flow sorting and real-time quantitative PCR came from another 27 surgical patients from the Cancer Hospital of the Chinese Academy of Medical Sciences from March 2019 to February 2020. The clinical characteristics, tumor genomics, and outcome data of patients were collected. Responses to PD-1/PD-L1 targeted therapy were assessed using RECIST 1.1 guidelines and confirmed by the corresponding institute investigator. Ethics approval was granted by the Committee for the Ethics Review of Research Involving Human Subjects of the Cancer Hospital of the Chinese Academy of Medical Sciences.

CNVA analysis

First, the collected paraffin-embedded samples were dissolved, DNA was extracted, and the DNA library was constructed with the MGI platform (CWBIO, Beijing, China). High-quality genomic DNA with A260/A280 = 1.8 ~ 2.0 was selected for fragmentation. After the fragmentation, the range of DNA distribution was wide, and we used magnetic bead to select specific fragments to control the final library fragment concentration. Those samples with concentrated DNA and high purity after physical fragmentation could be directly subject to end repair. If the DNA concentration range of the sample after physical fragmentation was wide, the fragments were sorted by length. Next, the ligated products were purified and PCR amplified. After constructing the library, we performed subsequent chromosome low-depth (10X) sequencing. During data analysis, the entire chromosome was cut into t fragments of 0.5 Mb, and each chromosome was cut into more than six thousand fragments. Raw reads were aligned with Burrows-Wheeler Aligner software to determine the chromosomal origin of each sequenced DNA fragment. To calculate the copy number of each position (0.5 Mb), we subtracted 2 from the copy number, then took the absolute value and calculated the average number of all positions. CNVA was calculated with the formula avg (abs (copy number-2)).

Noninvasive prenatal testing (Nipt) analysis

The sequencing platform and database building method for Nipt analysis are the same as those for CNVA analysis. Unlike CNVA, which only focuses on copy number changes of 0.5 Mb chromosome fragments, Nipt focuses on changes in the entire chromosome, that is, macroscopic increases or deletions affecting the copy number of the entire chromosome. The specific data analysis method is described in the literature.[24]

Analysis of TMB using TCGA

TCGA data were downloaded from the Genomic Data Commons Data Portal (GDC Data Portal, https://portal.gdc.cancer.gov/). In total, 509 lung adenocarcinoma (LUAD) patients from TCGA were analysed. Somatic mutations were identified using the ecTMB algorithm. The correlation analysis between CNVA of selected samples and whole-exome mutation burden was performed using Pearson correlation coefficient analysis.

Immunohistochemistry

The paraffin sections were gradually dewaxed and hydrated with xylene and alcohol (Beijing Chemical Works, China). Antigens retrieval was performed in a microwave. After washing 3 times with PBS buffer, goat serum (ZSGB-BIO, Beijing, China) was applied to tissue slides to cover the whole section. The slides were incubated for 30 min at room temperature, taking care to prevent the surrounding tissue from drying out. After the goat serum had been used to block the slides, the excess serum was shaken off and diluted primary antibodies against PD-L1 (CST 13683S, 1:200), CD8 (Abcam ab93278, 1:500) and CD3 (PTG 17617-1-AP, 1:100) were added to the specimen and incubated at 4 °C overnight. The primary antibodies were washed off the next day, the secondary antibody (Servicebio, GB23303, 1:200) was added, and the sample was incubated and finally developed using the BCA (Servicebio, G5001) method.

Flow cytometry

Flow cytometric analysis was undertaken on representative fresh samples of lung cancer obtained at surgical resection. After mechanical disaggregation to obtain monodispersed cell suspensions, the tissue samples were treated with collagenase (Sigma-Aldrich, 35285124, 1 mg mL^{-1}) and DNase I (0.33 U mL^{-1}, Roche) (diluted with DMEM medium) for 1 h in 37 °C and prepared for the staining of flow cytometry. The concentration of the single cell suspension was adjusted to 1×10^6 mL^{-1}; then, 1 mL of the suspension was added to a 1.5-mL EP tube, and an appropriate amount of PBS was added to wash the cells, the samples were centrifuged at 100 *g* for 5 min, and the supernatant was discarded. Next, 5 μL of blocking agent was added to each tube, and PBS was added to reach a volume of 100 μL. The samples were incubated on ice for 10 min in the dark. Then, 5 μL of anti-PD-L1 antibody (Biolegend 329706) and anti-EpCAM antibody (Biolegend 369814) were added to each tube, and the samples were incubated on ice for 30 min in the dark. Finally, the cells were washed twice with PBS, the nuclear dye DAPI (KeyGEN BioTECH, KGA215-50) was added, and the samples were incubated at room

temperature in the dark for 10 min. The stained cells were washed twice with PBS and then evaluated by flow cytometry.

Real-time quantitative PCR

Total RNA was isolated with the standard TRIzol-based protocol (Invitrogen). RT-PCR was performed on an ABI 7900HT Real-Time PCR thermocycler (Life Technologies). The kit for cDNA synthesis and amplification both were from Thermo Fisher (K1622, 752). After the PCR was completed, the base and threshold of the amplification curve were manually set. The data were analysed using the $2^{-\Delta\Delta CT}$ method. For the primer sequences used in the article, please refer to the Supplementary table 1.

Statistical analysis

Data analysis was performed using GraphPad Prism 6 (GraphPad Software, Inc.) or SPSS 24 (IBM, USA). Comparisons between two groups were performed using the Student's *t*-test (two-tailed). Pearson correlations were employed to evaluate the significance of the association between TMB and immune infiltration or between TMB and CNVA. *P*-values < 0.05 were considered significant.

CONFLICTS OF INTEREST

The authors declare that they have no competing interests.

AUTHOR CONTRIBUTION

Yuan Lei: Conceptualization; Data curation; Formal analysis; Methodology; Project administration; Resources; Software; Supervision; Validation; Visualization; Writing-original draft; Writing-review & editing. **Guo Zhang:** Conceptualization; Data curation; Project administration; Resources; Supervision; Validation. **Chaoqi Zhang:** Investigation; Supervision; Validation. **Li Xue:** Resources; Supervision. **Zhen Yang:** Resources; Validation. **Zhi Lu:** Investigation; Writing-review & editing. **Jian Huang:** Software; Supervision; Writing-review & editing. **Ruo Zang:** Investigation; Resources. **Yun Che:** Supervision; Validation. **Shuang Mao:** Software; Supervision. **Ling Fang:** Investigation; Methodology. **Cheng Liu:** Resources; Supervision. **Xin Wang:** Project administration. **Su Zheng:** Project administration. **Nan Sun:** Conceptualization; Data curation; Funding acquisition; Project administration; Writing-review & editing. **Jie He:** Conceptualization; Formal analysis; Funding acquisition; Project administration; Supervision; Writing-review & editing.

REFERENCES

1. Gainor JF, Rizvi H, Jimenez Aguilar E *et al*. Clinical activity of programmed cell death 1 (PD-1) blockade in never, light, and heavy smokers with non-small-cell lung cancer and PD-L1 expression >/=50. *Ann Oncol* 2020; **31**: 404–411.
2. Reck M, Rodriguez-Abreu D, Robinson AG *et al*. Pembrolizumab versus chemotherapy for PD-L1-positive non-small-cell lung cancer. *N Engl J Med* 2016; **375**: 1823–1833.
3. Samstein RM, Lee CH, Shoushtari AN *et al*. Tumor mutational load predicts survival after immunotherapy across multiple cancer types. *Nat Genet* 2019; **51**: 202–206.
4. Lee DW, Han SW, Bae JM *et al*. Tumor Mutation Burden and Prognosis in Patients with Colorectal Cancer Treated with Adjuvant Fluoropyrimidine and Oxaliplatin. *Clin Cancer Res* 2019; **25**: 6141–6147.
5. Wang F, Wei XL, Wang FH *et al*. Safety, efficacy and tumor mutational burden as a biomarker of overall survival benefit in chemo-refractory gastric cancer treated with toripalimab, a PD-1 antibody in phase Ib/II clinical trial NCT02915432. *Ann Oncol* 2019; **30**: 1479–1486.
6. Ready N, Hellmann MD, Awad MM *et al*. First-line nivolumab plus ipilimumab in advanced non-small-cell lung cancer (CheckMate 568): outcomes by programmed death ligand 1 and tumor mutational burden as biomarkers. *J Clin Oncol* 2019; **37**: 992–1000.
7. Budczies J, Allgauer M, Litchfield K *et al*. Optimizing panel-based tumor mutational burden (TMB) measurement. *Ann Oncol* 2019; **30**: 1496–1506.
8. Rizvi H, Sanchez-Vega F, La K *et al*. Molecular Determinants of Response to Anti-Programmed Cell Death (PD)-1 and Anti-Programmed Death-Ligand 1 (PD-L1) Blockade in Patients With Non-Small-Cell Lung Cancer Profiled With Targeted Next-Generation Sequencing. *J Clin Oncol* 2018; **36**: 633–641.
9. Ricciuti B, Kravets S, Dahlberg SE *et al*. Use of targeted next generation sequencing to characterize tumor mutational burden and efficacy of immune checkpoint inhibition in small cell lung cancer. *J Immunother Cancer* 2019; **7**: 87.
10. Goodman AM, Sokol ES, Frampton GM, Lippman SM, Kurzrock R. Microsatellite-Stable Tumors with High Mutational Burden Benefit from Immunotherapy. *Cancer Immunol Res* 2019; **7**: 1570–1573.
11. Georgiadis A, Durham JN, Keefer LA *et al*. Noninvasive Detection of Microsatellite Instability and High Tumor Mutation Burden in Cancer Patients Treated with PD-1 Blockade. *Clin Cancer Res* 2019; **25**: 7024–7034.
12. Luchini C, Bibeau F, Ligtenberg MJL *et al*. ESMO recommendations on microsatellite instability testing for immunotherapy in cancer, and its relationship with PD-1/PD-L1 expression and tumour mutational burden: a systematic review-based approach. *Ann Oncol* 2019; **30**: 1232–1243.

13. Schrock AB, Ouyang C, Sandhu J *et al*. Tumor mutational burden is predictive of response to immune checkpoint inhibitors in MSI-high metastatic colorectal cancer. *Ann Oncol* 2019; **30**: 1096–1103.

14. Innocenti F, Ou FS, Qu X *et al*. Mutational Analysis of Patients With Colorectal Cancer in CALGB/SWOG 80405 Identifies New Roles of Microsatellite Instability and Tumor Mutational Burden for Patient Outcome. *J Clin Oncol* 2019; **37**: 1217–1227.

15. Park S, Lee H, Lee B *et al*. DNA Damage Response and Repair Pathway Alteration and Its Association With Tumor Mutation Burden and Platinum-Based Chemotherapy in SCLC. *J thorac Oncol* 2019; **14**: 1640–1650.

16. Arulananda S, Mitchell P, John T. DDR Alterations as a Surrogate Marker for TMB in SCLC - Use it or Lose it? *J thorac Oncol* 2019; **14**: 1498–1500.

17. Pinto AE, Roque L, Rodrigues R, Andre S, Soares J. Frequent 7q gains in flow cytometric multiploid/ hypertetraploid breast carcinomas: a study of chromosome imbalances by comparative genomic hybridisation. *J Clin Pathol* 2006; **59**: 367–372.

18. Sugai T, Habano W, Jiao YF, Suzuki M, Takagane A, Nakamura S. Analysis of genetic alterations associated with DNA diploidy, aneuploidy and multiploidy in gastric cancers. *Oncology* 2005; **68**: 548–557.

19. Sugai T, Takahashi H, Habano W *et al*. Analysis of genetic alterations, classified according to their DNA ploidy pattern, in the progression of colorectal adenomas and early colorectal carcinomas. *J Pathol* 2003; **200**: 168–176.

20. Michels JJ, Duigou F, Marnay J, Denoux Y, Delozier T, Chasle J. Flow cytometry in primary breast carcinomas: prognostic impact of multiploidy and hypoploidy. *Cytometry B Clin Cytom* 2003; **55**: 37–45.

21. Michels JJ, Marnay J, Plancoulaine B, Chasle J. Flow cytometry in primary breast carcinomas: prognostic impact of S-phase fraction according to different analysis patterns. *Cytometry B Clin Cytom* 2004; **59**: 32–39.

22. Carloni S, Gallerani G, Tesei A *et al*. DNA ploidy and S-phase fraction analysis in peritoneal carcinomatosis from ovarian cancer: correlation with clinical pathological factors and response to chemotherapy. *Onco Targets Ther* 2017; **10**: 4657–4664.

23. Dayal JH, Sales MJ, Corver WE *et al*. Multiparameter DNA content analysis identifies distinct groups in primary breast cancer. *Br J Cancer* 2013; **108**: 873–880.

24. Chiu RW, Chan KC, Gao Y *et al*. Noninvasive prenatal diagnosis of fetal chromosomal aneuploidy by massively parallel genomic sequencing of DNA in maternal plasma. *Proc Natl Acad Sci USA* 2008; **105**: 20458–20463.

25. Jing CY, Fu YP, Yi Y *et al*. HHLA2 in intrahepatic cholangiocarcinoma: an immune checkpoint with prognostic significance and wider expression compared with PD-L1. *J Immunother Cancer* 2019; **7**: 77.

26. Monteiro LS, Palmeira C, Bento MJ, Lopes C. DNA content in malignant salivary gland tumours. *Oral Dis* 2009; **15**: 295–301.

Supporting Information

Additional supporting information may be found online in the Supporting Information section at the end of the article.