

Genomic and immune heterogeneity of multiple synchronous lung adenocarcinoma at different developmental stages

Received: 12 September 2023

Accepted: 28 August 2024

Published online: 10 September 2024

 Check for updates

Yue Zhao ^{1,2,3,11} ✉, Jian Gao ^{1,2,3,4,11}, Jun Wang ^{5,11}, Fanfan Fan ^{1,2,3,11}, Chao Cheng^{6,11}, Danwen Qian ^{7,8,11}, Ran Guo^{1,2,3}, Yang Zhang^{1,2,3}, Ting Ye^{1,2,3}, Marcellus Augustine^{7,8,9}, Yicong Lin^{3,10}, Jun Shang ^{1,2,3}, Hang Li^{1,2,3}, Yunjian Pan^{1,2,3}, Qingyuan Huang^{1,2,3}, Haiqing Chen^{1,2,3}, Han Han^{1,2,3}, Zhendong Gao^{1,2,3}, Qiming Wang⁵, Shiyue Zhang⁵, Mou Zhang⁵, Fangqiu Fu^{1,2,3}, Yueren Yan^{1,2,3}, Shanila Fernandez Patel^{7,8}, Roberto Vendramin ^{7,8}, Hui Yuan^{1,2,3}, Yawei Zhang^{1,2,3}, Jiaqing Xiang^{1,2,3}, Hong Hu^{1,2,3}, Yihua Sun ^{1,2,3}, Yuan Li ^{3,10}, Kevin Litchfield ^{7,8} ✉, Zhiwei Cao ^{4,5} ✉ & Haiquan Chen ^{1,2,3} ✉

Multiple synchronous lung cancers (MSLCs) constitute a unique subtype of lung cancer. To explore the genomic and immune heterogeneity across different pathological stages of MSLCs, we analyse 16 MSLCs from 8 patients using single-cell RNA-seq, single-cell TCR sequencing, and bulk whole-exome sequencing. Our investigation indicates clonally independent tumours with convergent evolution driven by shared driver mutations. However, tumours from the same individual exhibit few shared mutations, indicating independent origins. During the transition from pre-invasive to invasive adenocarcinoma, we observe a shift in T cell phenotypes characterized by increased Treg cells and exhausted CD8⁺ T cells, accompanied by diminished cytotoxicity. Additionally, invasive adenocarcinomas exhibit greater neoantigen abundance and a more diverse TCR repertoire, indicating heightened heterogeneity. In summary, despite having a common genetic background and environmental exposure, our study emphasizes the individuality of MSLCs at different stages, highlighting their unique genomic and immune characteristics.

Lung cancer is one of the deadliest diseases worldwide, with lung adenocarcinoma (LUAD) being the most common histological subtype^{1,2}. Advances in next-generation sequencing techniques have revolutionised the field of molecular profiling, enabling personalised treatment

strategies based on specific driver mutations present in patients³⁻⁷. Additionally, immune checkpoint inhibitors (ICIs) have shown promising results in improving the survival of lung cancer patients, although reliable biomarkers for predicting treatment response to ICIs have not

¹Departments of Thoracic Surgery and State Key Laboratory of Genetic Engineering, Fudan University Shanghai Cancer Center, Shanghai, China. ²Institute of Thoracic Oncology, Fudan University, Shanghai, China. ³Department of Oncology, Shanghai Medical College, Fudan University, Shanghai, China. ⁴International Human Phenome Institutes (Shanghai), Shanghai, China. ⁵School of Life Sciences, Fudan University, Shanghai, China. ⁶Department of Thoracic Surgery, West China Hospital of Sichuan University, Chengdu, Sichuan, China. ⁷Cancer Research UK Lung Cancer Centre of Excellence, University College London Cancer Institute, London, UK. ⁸Tumour Immunogenomics and Immunosurveillance Laboratory, University College London Cancer Institute, London, UK. ⁹Division of Medicine, University College London, London, UK. ¹⁰Department of Pathology, Fudan University Shanghai Cancer Center, Shanghai, China.

¹¹These authors contributed equally: Yue Zhao, Jian Gao, Jun Wang, Fanfan Fan, Chao Cheng, Danwen Qian. ✉ e-mail: yuezhao97@outlook.com; k.litchfield@ucl.ac.uk; zwcao@fudan.edu.cn; hqchen1@yahoo.com

been fully discovered^{8–12}. Therefore, a comprehensive understanding of the molecular profile and immune microenvironment of LUAD is crucial.

Multiple synchronous lung cancers (MSLCs) represent a unique subtype of lung cancer, characterised by the simultaneous occurrence of two or more lung cancers within the same individual at the time of diagnosis¹³. The increased utilisation of CT scans has led to a higher detection rate of MSLCs in recent years^{14–16}. Differentiating between independent tumours and haematogenous spread in MSLCs is challenging but essential, as it influences treatment decisions and prognosis¹⁷. While the origins of different lesions in MSLCs and whether they share common ancestors are not yet fully understood, evidence suggests that distinct genomic profiles may exist among different lesions¹⁸. In a subset of MSLC patients, both pre-invasive and invasive stages of LUAD are present. Pre-invasive stages, namely adenocarcinoma in situ (AIS) and minimally invasive adenocarcinoma (MIA), are considered to have nearly 100% 5-year survival rates following complete surgery^{19,20}. Previous studies have identified differences in genomic and immune profiles between pre-invasive and invasive LUADs^{21–23}. However, comprehensive comparisons of the immune microenvironment and TCR repertoire between LUADs at different stages within the same patients have not yet been achieved.

In this work, we conduct bulk whole-exome sequencing (WES), single-cell RNA sequencing and single-cell TCR sequencing on 16 paired pre-invasive and invasive LUADs from 8 MSLC patients. Our aim is to investigate the genomic and immune heterogeneity of LUAD at different developmental stages within the same patients. This integrated approach allows us to examine the comprehensive landscape of molecular and immune characteristics in these tumours.

Results

Somatic mutation landscape demonstrates inter-tumour heterogeneity and different evolutionary patterns of multiple synchronous lung cancers (MSLCs)

Clinical, pathological and radiological characteristics of the samples are shown in Table 1, Fig. 1a and Supplementary Fig. 1. Bulk WES was performed on all the 16 samples and their matched normal lung tissues, to identify somatic mutations. We found that the most frequently mutated gene was *EGFR* (50%), followed by *RBM10* (25%), *PIK3CA* (12%), *KRAS* (12%) and *ERBB2* (12%, Fig. 1b, Supplementary Data 1). Median tumour mutation burden (TMB) for AIS/MIA samples was 1.08 (range:

0.40–18.03), while for LUAD samples, it was 1.78 (range: 1.13–18.60). Of note, one patient (FD4) demonstrated a much higher TMB than the others, which might be due to being a smoker and having a higher activity of smoking mutation signature (SBS4, Supplementary Fig. 2).

Interestingly, we observed only a small number of shared somatic mutations between samples from the same patients across our study cohort (Fig. 1c and Supplementary Fig. 3), indicating that the patients had multiple primary lung cancers rather than intrapulmonary metastases. To assess whether different tumours harboured the same driver mutations, we focused on the driver mutations of LUAD. Convergent evolution was found in patients FD4, FD5, FD8 and FD16, where shared driver mutations were identified in independently originated tumours (Fig. 1c). In patients FD1 and FD14, only one of the two tumours harboured a driver mutation (*EGFR* p.L858R). Notably, in patient FD9, the LUAD sample had an *EML4-ALK* fusion, while the MIA sample contained an *ERBB2* exon 20 insertion, both of which were considered driver events in LUAD (Fig. 1c). In patient FD4, the LUAD sample harboured *RBM10* p.E165Q mutation, while the MIA sample harboured *RBM10* p.A247P mutation, suggesting parallel evolution.

In addition, a combined histomolecular algorithm was used to further address this issue²⁴. When taking into account the histological manifestations, we found patients FD1, FD2, FD4, FD5, FD9 and FD14 all had 1 LUAD and 1 AIS/MIA samples, while FD8 and FD16 had 2 LUAD samples with different predominant subtypes (Table 1). These findings collectively support the notion that different tumours within the same patient originated independently rather than through intrapulmonary metastases.

A single-cell atlas of multiple synchronous lung cancers (MSLCs)

To investigate intratumoural heterogeneity and compare the immune microenvironment between paired tumour samples of different developmental stages within the same patients, we conducted single-cell RNA sequencing (scRNA-seq) on all 6 AIS/MIA and 10 LUAD samples. After quality control, a total of 92,032 cells were retained and clustered based on their expression profiles (Supplementary Data 2). Dimensional reduction using t-distributed stochastic neighbour embedding (t-SNE) was employed. By utilising canonical markers, we identified epithelial cells, stromal cells (including endothelial cells and fibroblasts) and immune cells (such as T cells, B cells, mast cells, NK cells, myeloid cells and plasma cells) (Fig. 2a–c, Supplementary Data 3).

Table 1 | Baseline clinical and pathological characteristics of patients included in this study

| Patient | Sex | Age | Smoking status | TNM stage | Tumour location | Tumour size (cm) | Radiology | Pathology | Predominant subtype |
|---------|--------|-----|----------------|-----------|-----------------|------------------|-----------|-----------|---------------------|
| FD1 | Female | 73 | Never | IIB | RUL | 1.2 | Solid | LUAD | Papillary |
| | | | | | RUL | 0.8 | Subsolid | MIA | |
| FD2 | Female | 75 | Never | IIIA | LLL | 3.2 | Solid | LUAD | Acinar |
| | | | | | LUL | 0.7 | GGO | MIA | |
| FD4 | Male | 64 | Ever | IA3 | RUL | 2.4 | Solid | LUAD | Acinar |
| | | | | | RUL | 0.9 | GGO | MIA | |
| FD5 | Male | 57 | Ever | IA2 | RUL | 1.4 | Subsolid | LUAD | Lepidic |
| | | | | | RUL | 0.5 | Subsolid | MIA | |
| FD8 | Female | 54 | Never | IA2 | RML | 1.7 | Solid | LUAD | Papillary |
| | | | | | RLL | 1.2 | Subsolid | LUAD | |
| FD9 | Female | 32 | Never | IA2 | LUL | 2.0 | Solid | LUAD | Acinar |
| | | | | | LLL | 0.4 | GGO | MIA | |
| FD14 | Female | 71 | Never | IA3 | RLL | 2.5 | Subsolid | LUAD | Papillary |
| | | | | | RUL | 0.7 | GGO | AIS | |
| FD16 | Male | 53 | Ever | IA2 | LUL | 2.0 | Solid | LUAD | Papillary |
| | | | | | LUL | 0.6 | GGO | LUAD | |

AIS adenocarcinoma in situ, GGO ground-glass opacity, LUAD lung adenocarcinoma, LLL left lower lobe, LUL left upper lobe, MIA minimally invasive adenocarcinoma, RLL right lower lobe, RML right middle lobe, RUL right upper lobe.

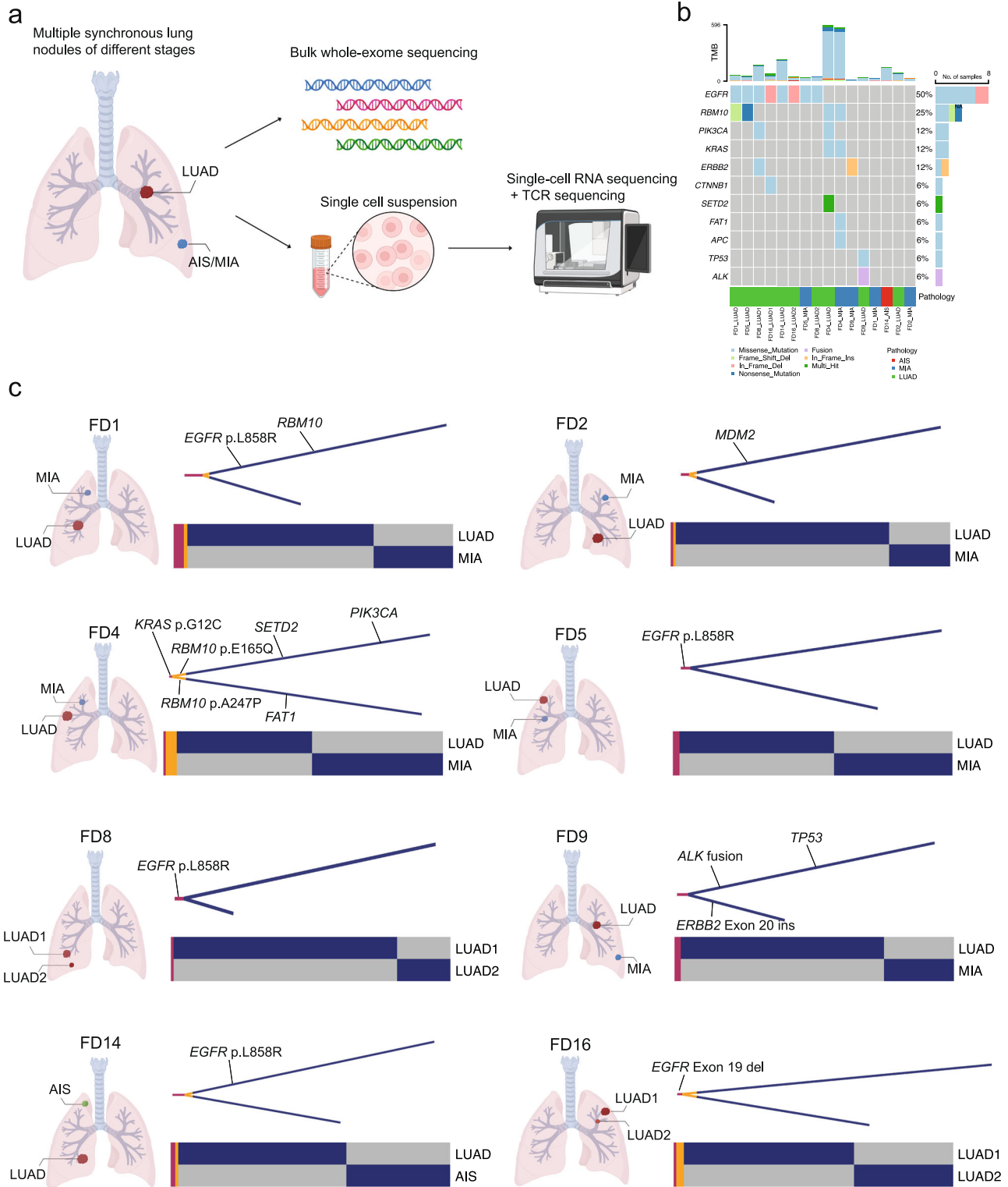


Fig. 1 | Study design and mutation landscape of 8 multiple synchronous lung cancer (MSLC) patients. a Study design. 8 multiple synchronous lung cancer (MSLC) patients were included in this study. Samples were surgically resected and were sent for bulk whole-exome sequencing (WES), single-cell RNA sequencing and single-cell TCR sequencing. **b** Mutations in major driver genes and tumour suppressor genes of sequenced tumours. **c** Intertumour heterogeneity of each MSLC patient based on mutations detected. For each individual patient, tumour location was shown on the left. A heatmap shows the mutations different tumours

harboured. Maroon, mutations that were shared by the two tumours in the same patient; yellow, parallel evolution events; blue, mutations that were private in either tumour in the same patient. A phylogenetic tree was constructed for each patient demonstrating the intertumour heterogeneity. AIS adenocarcinoma in situ, MIA minimally invasive adenocarcinoma, LUAD lung adenocarcinoma. Source data are provided as a Source Data file. Figure 1/panels a and c, created with BioRender.com, released under a Creative Commons Attribution-NonCommercial-NoDerivs 4.0 International license. <https://creativecommons.org/licenses/by-nc-nd/4.0/deed.en>.

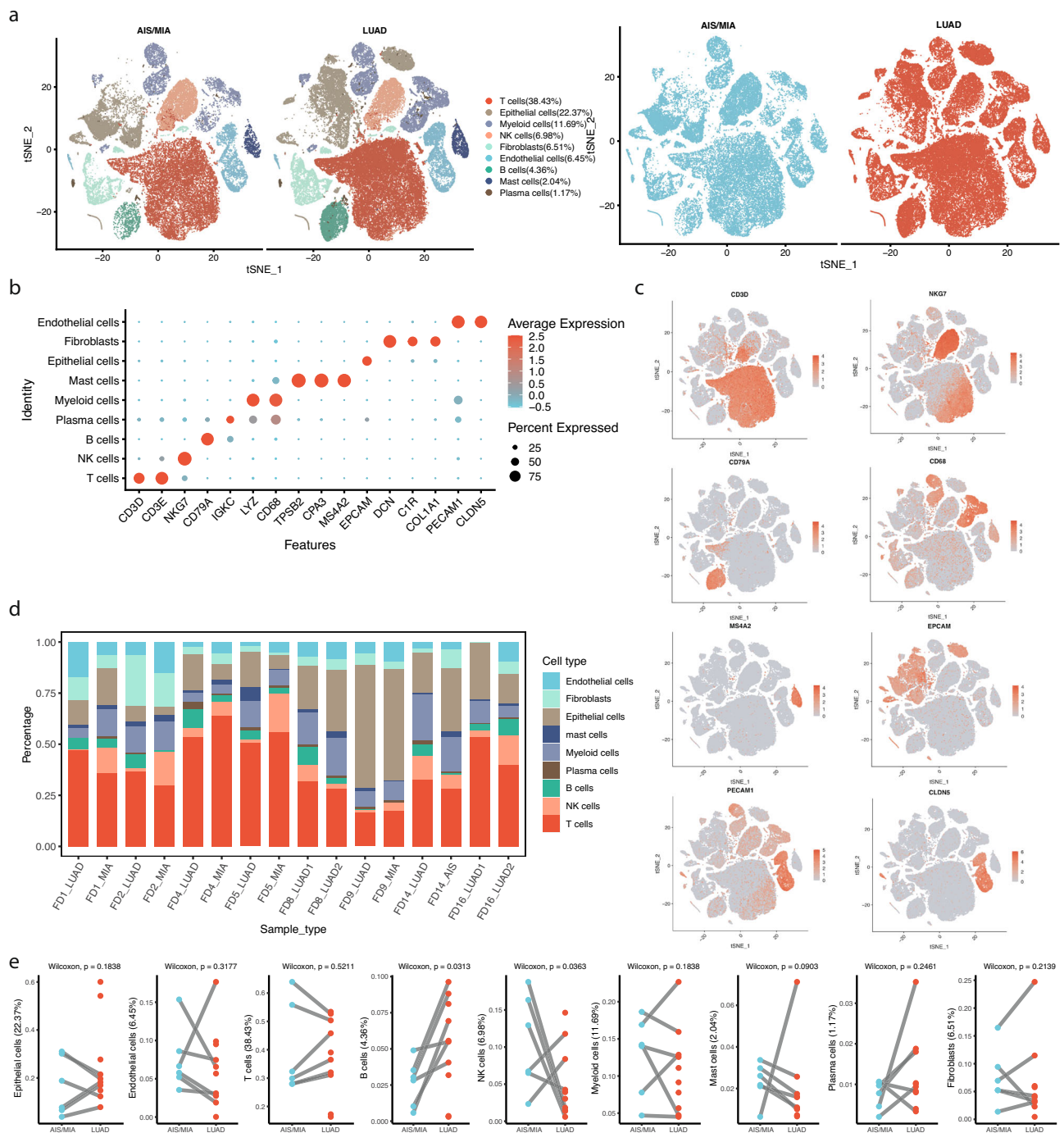


Fig. 2 | Single cell landscape of multiple synchronous lung cancer (MSLC) patients. **a** t-distributed stochastic neighbour embedding (t-SNE) plot of 92,032 sequenced cells. Grouping based on annotated cell types and pathology were shown on the left and right, respectively. **b** Expression of canonical marker genes of epithelial cells, endothelial cells, immune cells and fibroblasts; **(c)** t-SNE plots showing expression levels of canonical marker genes of identified cell types, with the gradient of colouring representing expression levels; **(d)** Bar plots showing

the percentages of each annotated cell type in each individual tumour; **(e)** Comparison of frequency of each annotated cell type between adenocarcinoma in situ (AIS)/minimally invasive adenocarcinoma (MIA) and invasive lung adenocarcinoma (LUAD). 6 AIS/MIA and 6 LUAD samples were used to derive the two-sided paired comparisons. AIS adenocarcinoma in situ, MIA minimally invasive adenocarcinoma, LUAD lung adenocarcinoma. Source data are provided as a Source Data file.

We compared the composition of these major cell types between AIS/MIA and LUAD samples. Remarkably, the number of B cells was significantly higher in LUAD ($p = 0.0313$), while the number of NK cells was significantly higher in AIS/MIA ($p = 0.0363$, Fig. 2d, e). These findings suggest that tumours at different developmental stages exhibit distinct tumour microenvironments, despite originating from the same immune system throughout the body.

Single-cell somatic copy number alteration (SCNA) analysis reveals a more unstable genome in LUAD than AIS/MIA samples

To investigate the differences between AIS/MIA and LUAD samples at the level of tumour cells, we focused on analysing the copy number profiles of these cells. Tumour cells were separated from epithelial cells using single-cell expression data, with 300 endothelial cells and

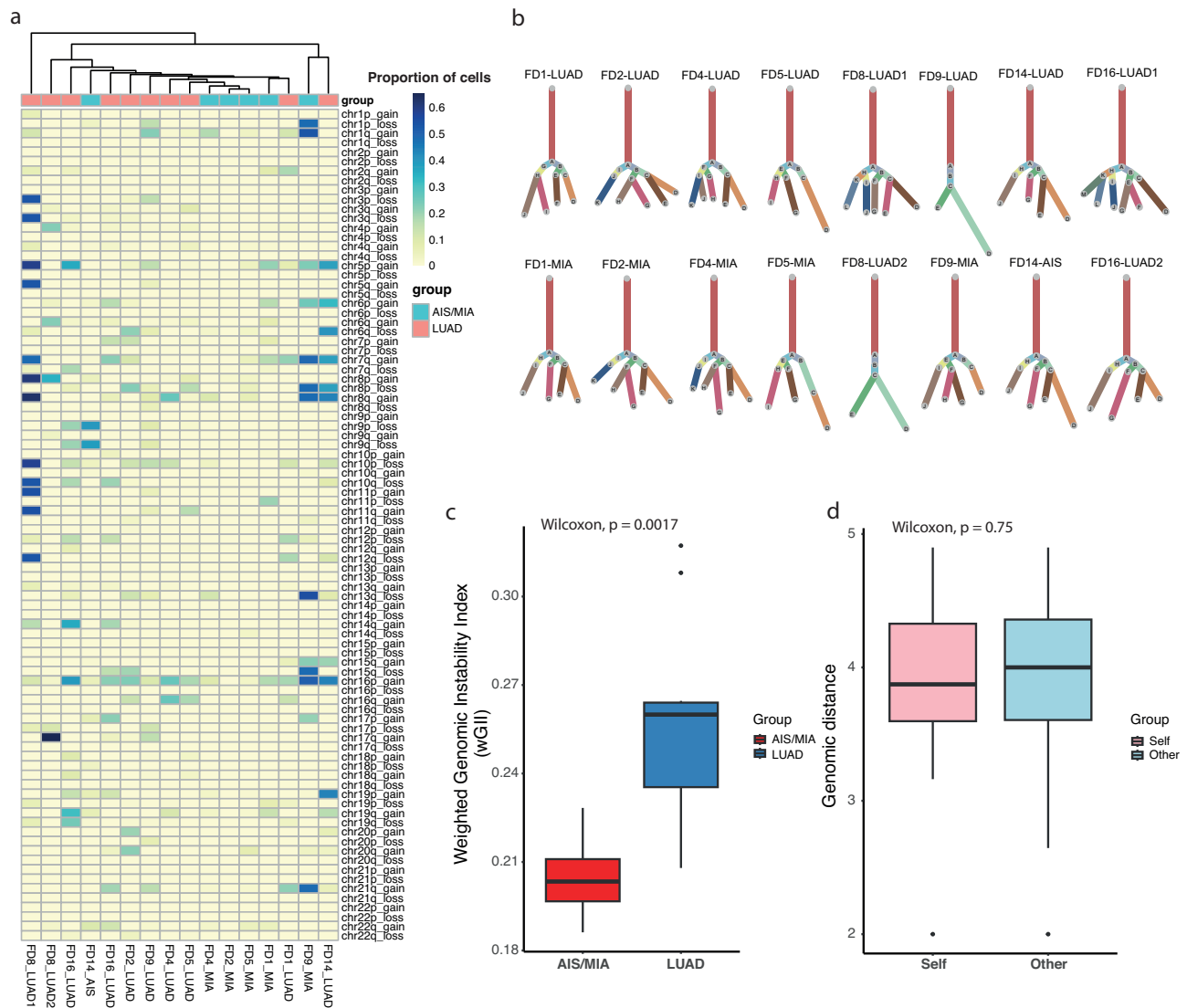


Fig. 3 | Single cell copy number variation (CNV) analysis of multiple synchronous lung cancer (MSLC) patients. **a** Heatmap showing the copy number variation profiles of MSLC patients. InferCNV was used to call CNV events using single-cell RNA-seq data. Colour bar indicates the proportion of cells having the CNV event. **b** Clonality trees were reconstructed based on CNV profiles of each tumour. Branches were scaled according to percentage of cells harbouring specific CNVs in each inferred subclone. **c** Comparison of CNV burden between AIS/MIA and LUAD samples, as measured by weighted genomic instability index (wGII). 6 AIS/MIA and 10 LUAD samples were used to derive the two-sided comparison. **d** Comparison of genomic distance between AIS/MIA and LUAD using Euclidean metrics. For each patient, intra-patient genomic distance was calculated between AIS/MIA and LUAD in the same patient (self), while inter-patient genomic distances were calculated

between AIS/MIA and each LUAD in the other patients (other). 6 AIS/MIA and 10 LUAD samples were used to derive the two-sided comparison. AIS adenocarcinoma in situ, MIA minimally invasive adenocarcinoma, LUAD lung adenocarcinoma. For boxplots in this figure, the centre of the boxes indicated the median value and upper and lower bounds of the boxes indicated the 25th and 75th percentile of data. The box covers the interquartile interval and represents the area where 50% of the data were found. Whiskers went from the minimum of data to the lower bounds of the box and the upper bounds of the box to the maximum of data. The whiskers were restricted to a maximum of 1.5 times the interquartile range (IQR) and data points outside this range were considered outliers. Source data are provided as a Source Data file.

300 fibroblasts included as spike-in controls. InferCNV was applied to infer the copy number profiles for each cell, and subsequent clustering based on these profiles was performed²⁵. Cells with quiet copy number profile similar to spike-in cells were classified as normal epithelial cells, while cells with more unstable genomes were identified as tumour cells (Supplementary Fig. 4). The frequency of arm-level somatic copy number alteration (SCNA) events across the tumours was visualised in a heatmap (Fig. 3a).

Remarkably, AIS/MIA and LUAD samples from the same patients rarely clustered together (except for FD1), further supporting the notion that different tumours within the same patient originated independently and harboured distinct genomic alterations,

despite sharing the same genetic background and environmental exposure.

Additionally, we utilised the weighted genomic integrity index (wGII) to compare the SCNA load between AIS/MIA and LUAD samples²⁶. The results demonstrated that LUAD samples had significantly higher levels of SCNA load than AIS/MIA samples ($p = 0.0017$, Fig. 3c). Finally, we calculated the genomic distance based on the SCNA profiles of the tumours using Euclidean metrics. Interestingly, the genomic distance between two samples from the same patient (self) did not differ from that between two randomly selected samples from different patients (other), further indicating that different samples within the same patient had different origins (Fig. 3d).

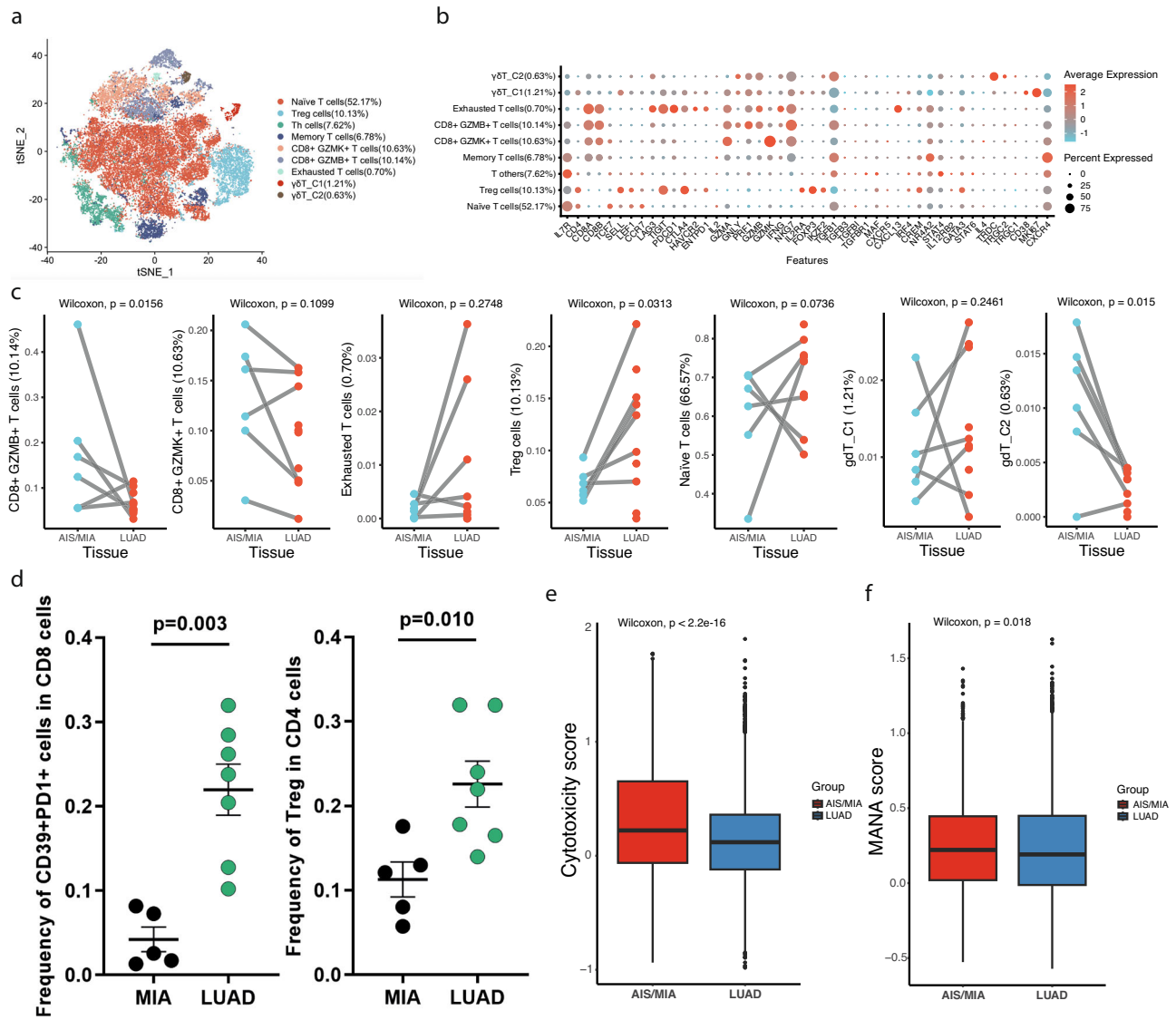


Fig. 4 | Clustering and comparison of T cells. **a** t-SNE plot showing 35,367 T cells were further clustered and annotated into 9 subgroups. **b** Expression of canonical marker genes that were used to identify different subgroups of T cells. **c** Comparison of frequency of each T cell subgroup between AIS/MIA and LUAD samples. 6 AIS/MIA and 6 LUAD samples were used to derive the two-sided paired comparisons. **d** Comparison of frequency of exhausted CD8⁺ T cells and regulatory T cells (Treg cells) using flow cytometry. 5 MIA and 7 LUAD samples were used to derive the comparison, and two-sided Student's *t* test was performed. Data were presented as mean values \pm SD. **e** Comparison of cytotoxicity score in CD8⁺ T cells using single-cell RNA-seq data between AIS/MIA and LUAD samples. 3597 cells from the AIS/MIA group and 3993 cells from the LUAD group were used to derive the two-sided comparison. **f** Comparison of MANA score in CD8⁺ T cells using single-

cell RNA-seq data between AIS/MIA and LUAD samples. 3597 cells from the AIS/MIA group and 3993 cells from the LUAD group were used to derive the two-sided comparison. AIS adenocarcinoma in situ, MIA minimally invasive adenocarcinoma, LUAD lung adenocarcinoma. For boxplots in this figure, the centre of the boxes indicated the median value and upper and lower bounds of the boxes indicated the 25th and 75th percentile of data. The box covers the interquartile interval and represents the area where 50% of the data were found. Whiskers went from the minimum of data to the lower bounds of the box and the upper bounds of the box to the maximum of data. The whiskers were restricted to a maximum of 1.5 times the interquartile range (IQR) and data points outside this range were considered outliers. Source data are provided as a Source Data file.

Compared with AIS/MIA, LUAD has a more inhibitory tumour microenvironment (TME)

To gain insights into the composition and functional status of T cells within the tumour microenvironment (TME), we applied clustering analysis to single-cell expression profiles of 35,367 T cells (13,137 for the AIS/MIA group, and 22,130 for the LUAD group, Fig. 4a). By employing t-SNE for dimensional reduction, distinct T cell subsets were identified (Fig. 4a). These included naïve T cells (52.17%), Treg cells (10.13%), memory T cells (6.78%), CD8 + GZMK + T cells (10.63%), CD8 + GZMB + T cells (10.14%), exhausted T cells (0.70%), two clusters of γ δT cells (1.84%), and other T cells (7.62%), as characterised by specific markers (Fig. 4b, Supplementary Data 3). To compare the

frequencies of these T cell subsets between AIS/MIA and LUAD samples, we conducted a paired analysis. The results revealed that the frequency of Treg cells was significantly higher in LUAD compared to AIS/MIA ($p = 0.0313$, Fig. 4c). Although not statistically significant, there was a trend suggesting a higher frequency of exhausted CD8⁺ T cells in LUAD compared to AIS/MIA ($p = 0.2748$, Fig. 4c). To further validate these findings, we performed flow cytometry on an independent cohort comprising 5 MIA and 7 LUAD samples. Consistently, the frequency of both exhausted T cells and Treg cells was significantly higher in LUAD (Fig. 4d), indicating the presence of a more immunosuppressive TME in LUAD compared to AIS/MIA. We further compared exhausted CD8⁺ T cells and Treg cells in tumours with or without

mutations in specific driver genes (*EGFR*, *TP53*, *RBM10*, *PIK3CA* and *KRAS*), and found tumours with *EGFR* mutations tended to have a higher level of Treg cells, while no significant difference was observed for exhausted CD8⁺ T cells (Supplementary Fig. 5). Moreover, PD-L1 (*CD274*) expression between AIS/MIA and LUAD samples was compared, and no significant difference was observed between the two groups using either scRNA-seq or immunofluorescent staining (Supplementary Fig. 6).

Anti-tumour T cell response is higher in AIS/MIA than in LUAD

Furthermore, we observed a significantly higher frequency of CD8⁺ GZMB⁺ T cells in AIS/MIA samples ($p=0.0156$), indicating a greater cytotoxic function in AIS/MIA compared to LUAD (Fig. 4c). To assess the overall cytotoxicity between AIS/MIA and LUAD samples, we defined a cytotoxic score based on established gene signatures²⁷, which demonstrated a significantly higher score in AIS/MIA (Fig. 4e). To further investigate the T cell activation and cytotoxicity markers in pre-invasive and invasive LUAD samples, we performed flow cytometry on an additional cohort comprising 5 MIA and 10 LUAD samples. Consistently, the frequency of GZMA⁺ and GNLY⁺ cells was significantly higher in MIA samples ($p=0.003$ and 0.015 , respectively), indicating a higher level of cytotoxicity in T cells within the TME of pre-invasive samples (Supplementary Fig. 7). Ki67, a marker of proliferation, was also found to be significantly higher in the pre-invasive group ($p=0.014$, Supplementary Fig. 7). It is reported that the majority of T cells do not recognise tumour antigens and are considered bystander T cells, whereas T cells recognising mutation-associated neoantigens (MANA) play a crucial role in anti-tumour immune responses^{28,29}. Therefore, we defined a MANA score based on genes enriched in MANA-specific T cells from a previous study²⁸. The analysis revealed a significantly higher MANA score in AIS/MIA compared to LUAD, indicating a lower proportion of bystander T cells in AIS/MIA samples (Fig. 4f). Collectively, these findings suggest that AIS/MIA samples exhibit a heightened anti-tumour T cell response compared to LUAD.

TCR repertoire and neoantigen prediction analysis

To explore the T cell receptor (TCR) repertoire in the different stages of MSLC tumours, single-cell TCR sequencing was conducted on the 16 MSLC samples. Firstly, we examined the relationship between TCR diversity and tumour heterogeneity, observing that tumours with high TCR diversity displayed greater level of heterogeneity (Fig. 5a). Subsequently, we compared the number of TCR clonotypes and TCR diversity between the two groups and found that LUAD samples had a higher number of TCR clonotypes and greater diversity measured by Shannon's Diversity Index ($p=0.042$ and $p=0.031$, respectively, Fig. 5b, c). However, despite having a higher number of TCR clonotypes, LUAD exhibited a lower frequency of large and hyperexpanded TCR clonotypes ($p=0.056$, Fig. 5d, e and Supplementary Fig. 8), as illustrated in Supplementary Fig. 8 which depicts changes in the top TCR clonotypes.

We next used netMHCpan (version 4.0) to predict neoantigens and used Panpep to predict their affinity to TCRs. TCRs with an affinity score >0.5 were considered neoantigen-committed TCRs³⁰. Significantly more neoantigens were detected in LUAD samples ($p=0.016$, Fig. 5f). Notably, CD8⁺ T cells containing neoantigen-committed TCRs exhibited a higher MANA score compared to those without such TCRs, indicating a reduced likelihood of being bystander T cells (Fig. 5g). This trend remained consistent when analysing AIS/MIA and LUAD samples separately (Supplementary Fig. 9a, b), suggesting that CD8⁺ T cells containing predicted neoantigen-committed TCRs were less likely to be bystander T cells. No significant difference was observed in MANA scores for CD8⁺ T cells containing neoantigen-committed TCRs between AIS/MIA and LUAD samples, indicating comparable efficacy in non-bystander CD8⁺ T cells in both groups (Supplementary Fig. 9c). Exhaustion scores, as calculated using genes

associated with T cell exhaustion²⁷, were found to be higher in CD8⁺ T cells containing neoantigen-committed TCRs (Fig. 5h). This trend was observed in both AIS/MIA and LUAD samples (Supplementary Fig. 10a, b), with the exhaustion score significantly higher in LUAD compared to AIS/MIA for CD8⁺ T cells containing neoantigen-committed TCRs (Supplementary Fig. 10c). Moreover, immune checkpoint scores, based on known immune checkpoints, were significantly higher in CD8⁺ T cells containing neoantigen-committed TCRs. While this trend was not observed in AIS/MIA samples, it was evident in LUAD samples (Supplementary Fig. 11a, b). CD8⁺ T cells containing neoantigen-committed TCRs in LUAD exhibited a higher immune checkpoint score than those in AIS/MIA (Supplementary Fig. 11c). Finally, we found genes with non-silent mutations had significantly lower expression than genes with silent mutations, indicating that expression of neoantigens was more suppressed in more invasive stages (Fig. 5j, Supplementary Fig. 12). Collectively, these findings highlight the significance of neoantigens and their corresponding TCRs in LUAD progression and provide additional insights into the immune escape mechanisms within the LUAD progression.

Comparison of Myeloid cells and natural killer (NK) cells between AIS/MIA and LUAD

We next performed clustering analysis on myeloid cells and NK cells to investigate their differences between AIS/MIA and LUAD. This analysis revealed the presence of 10 distinct myeloid cell types and two NK cell types (Fig. 6a, c). Among the myeloid cells, we observed significantly higher frequencies of anti-inflammatory macrophages and proliferating macrophages in LUAD samples, while conventional dendritic cells (type 1) and non-classical CD16⁺ monocytes were significantly more abundant in AIS/MIA samples (Fig. 6b). Regarding NK cells, although the difference did not reach statistical significance ($p=0.0903$), a lower frequency of CD56dimCD16⁺ NK cells was observed in LUAD compared to AIS/MIA (Fig. 6d). NoTable, this trend was validated in a flow cytometry experiment comprising 5 MIA and 7 LUAD samples, where a significant decrease in the frequency of CD56dimCD16⁺ NK cells was observed in LUAD samples (Fig. 6e).

Discussion

MSLCs present a unique subtype of lung cancer that poses challenges in determining whether they are independent tumours or intrapulmonary metastases^{13,14,16}. In this study, we conducted WES on 16 samples obtained from 8 patients. The results revealed that shared driver mutations were only identified in 4 out of the 8 patients (FD4, FD5, FD8 and FD16, as shown in Fig. 1). Additionally, we found few shared passenger mutations, indicating that MSLCs within the same patient originated independently. A combined histomolecular analysis further validated their independency. Notably, despite being clonally independent lesions, tumours in patients FD4, FD5, FD8 and FD16 exhibited patterns of convergent evolution. Convergent evolution has been reported in various cancer types and is associated with treatment resistance^{31–34}. An intriguing case (FD9) was observed, where one tumour contained the *EML4-ALK* fusion while the other tumour harboured *ERBB2* exon 20 insertion. Both are driver alterations and therapeutic targets in LUAD^{5,35,36}. Cases with parallel evolution in driver genes were also identified. Taken together, these findings suggest that MSLCs arise independently and demonstrate multiple evolutionary subtypes.

By sequencing MSLCs from the same patients, we were able to analyse intra-patient heterogeneity and directly compare the immune microenvironment between pre-invasive and invasive LUAD (LUAD), while controlling patients' genetic background and environmental exposure. Mutations in canonical driver genes were observed in both AIS/MIA and LUAD samples (Fig. 1b, c). In comparison to pre-invasive LUADs, invasive LUADs exhibited a higher level of TMB and genomic instability, which could possibly be caused by a higher level of somatic

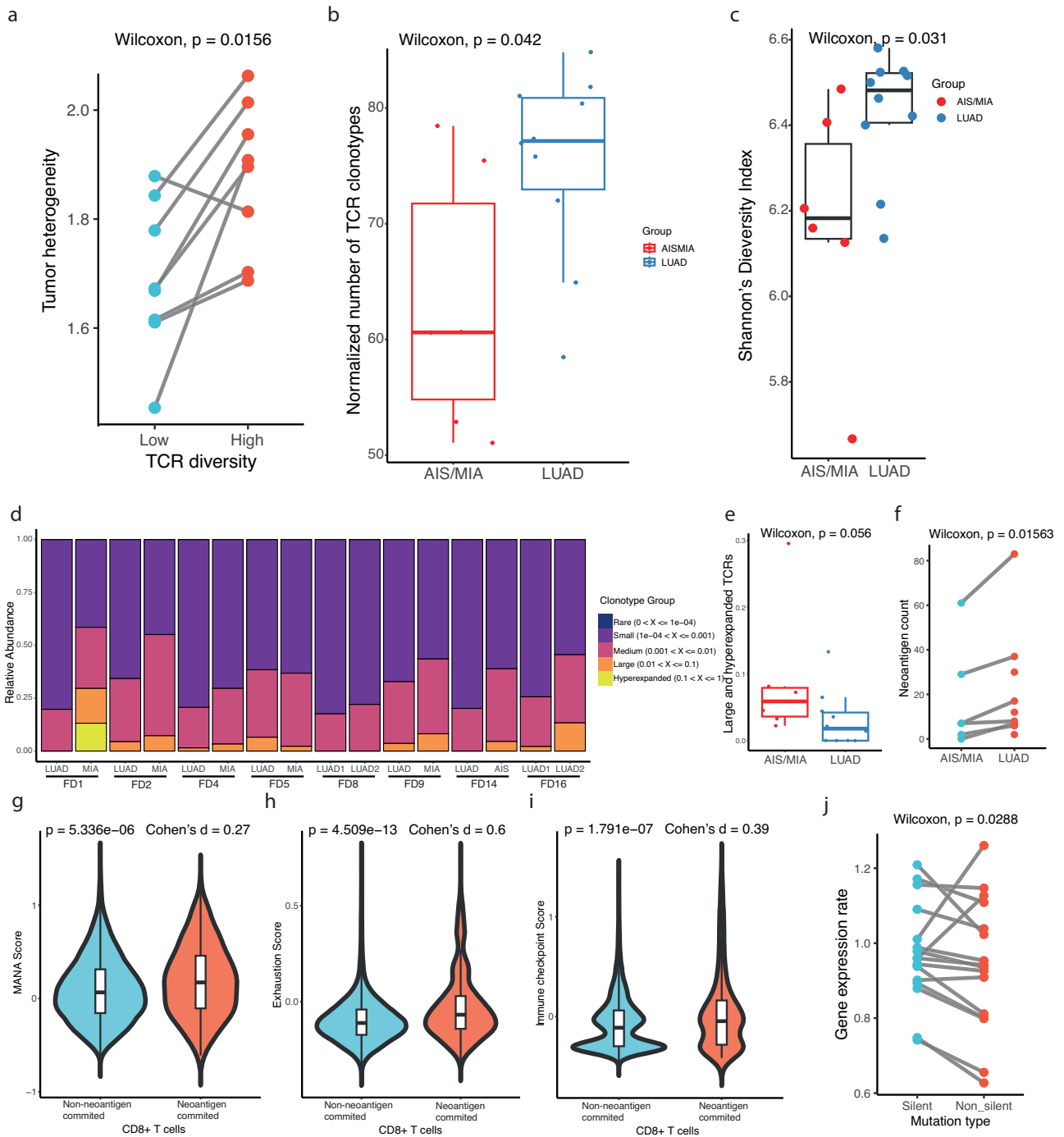


Fig. 5 | T cell receptor (TCR) repertoire and neoantigen prediction of multiple synchronous lung cancer (MSLC) patients. **a** Comparison of tumour heterogeneity between patients with different levels of TCR diversity. 8 samples with lower TCR diversity and 8 samples with higher TCR diversity were used to derive the two-sided comparison. **b** Comparison of number of TCR clonotypes between AIS/MIA and LUAD samples. 6 AIS/MIA and 10 LUAD samples were used to derive the two-sided comparison. **c** Comparison of TCR diversity between AIS/MIA and LUAD sample, as measured by Shannon's Diversity Index. 6 AIS/MIA and 10 LUAD samples were used to derive the two-sided comparison. **d** Proportion of occupied clonal space specific clonotypes across all samples. **e** Comparison of number of neoantigen specific TCRs between AIS/MIA and LUAD samples. 6 AIS/MIA and 10 LUAD samples were used to derive the two-sided comparison. **f** Comparison of number of predicted neoantigens between AIS/MIA and LUAD samples. 5 AIS/MIA and 5 LUAD samples were used to derive the paired two-sided comparison. **g** Comparison of MANA score between CD8+ T cells containing non-neoantigen committed and neoantigen committed TCRs. Two-sided Wilcoxon's rank-sum test

was used. **h** Comparison of exhaustion score between CD8+ T cells containing non-neoantigen committed and neoantigen committed TCRs. Two-sided Wilcoxon's rank-sum test was used. **i** Comparison of immune checkpoint score between CD8+ T cells containing non-neoantigen committed and neoantigen committed TCRs. Two-sided Wilcoxon's rank-sum test was used. **j** Comparison of the expression of genes with silent and non-silent mutations. For boxplots in this figure, the centre of the boxes indicated the median value and upper and lower bounds of the boxes indicated the 25th and 75th percentile of data. The box covers the interquartile interval and represents the area where 50% of the data were found. Whiskers went from the minimum of data to the lower bounds of the box and the upper bounds of the box to the maximum of data. The whiskers were restricted to a maximum of 1.5 times the interquartile range (IQR) and data points outside this range were considered outliers. AIS, adenocarcinoma in situ; MIA, minimally invasive adenocarcinoma; LUAD, lung adenocarcinoma; TCR, T cell receptor; MANA, mutation-associated neoantigens. Source data are provided as a Source Data file.

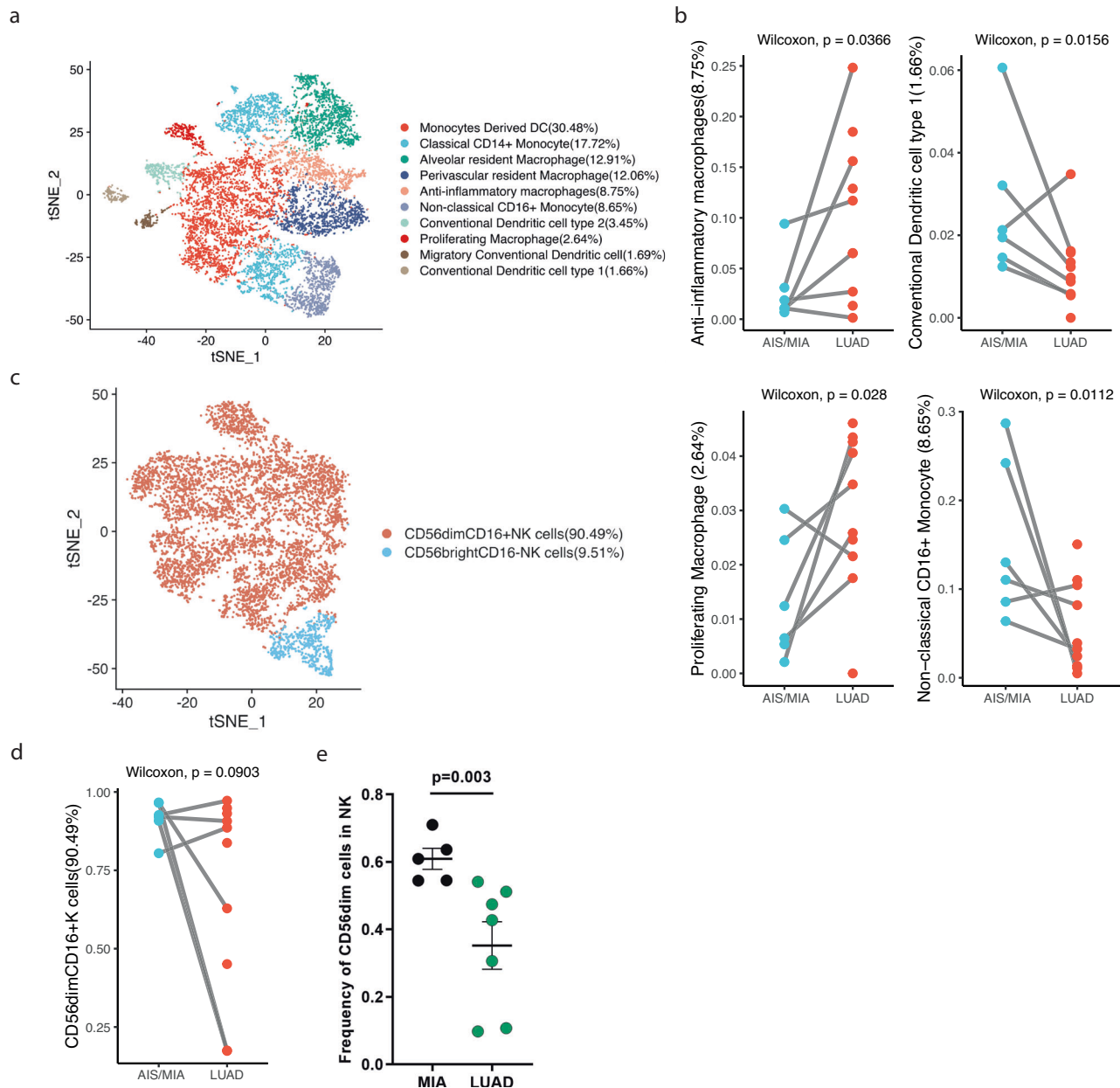


Fig. 6 | Clustering and comparison of myeloid cells and natural killer (NK) cells.

a t-SNE plot showing 10,763 myeloid cells were further clustered and annotated into 10 subgroups. **b** Comparison of frequency of anti-inflammatory macrophages (top-left), proliferating macrophages (bottom-left), conventional dendritic cell (top-right) and non-classical CD16+ monocytes (bottom-right) between AIS/MIA and LUAD samples. 6 AIS/MIA and 6 LUAD samples were used to derive the paired two-sided comparisons. **c** t-SNE plot showing 6590 NK cells were further clustered into

2 subgroups. **d** Comparison of frequency of CD56dimCD16+ NK cells between AIS/MIA and LUAD samples. 6 AIS/MIA and 6 LUAD samples were used to derive the paired two-sided comparison. **e** Comparison of frequency of CD56dim NK cells using flow cytometry on 5 MIA and 7 LUAD samples, and two-sided Student's *t* test was performed. Data are presented as mean values \pm SD. AIS adenocarcinoma in situ, MIA minimally invasive adenocarcinoma, LUAD lung adenocarcinoma, NK natural killer. Source data are provided as a Source Data file.

copy number alterations, structural variations, impaired DNA mismatch repair functions and higher activities of APOBEC signatures. Previous studies have reported a higher frequency of mutations in tumour suppressor genes, especially *TP53*, in LUAD compared to AIS/MIA²¹. However, in this study, we found only one LUAD sample (FD9) harboured a *TP53* mutation (Fig. 1b, c). The low frequency of *TP53* mutation could be due to the small sample size and relatively early stage of the tumours examined.

Regarding the evolutionary patterns of tumours, we observed convergent evolution in the driver mutation space, where the same driver alterations were identified in clonally independent lesions from

the same patient (Fig. 1c). However, despite a significant increase in SCNAs from AIS/MIA to LUAD, no evidence of convergent selection in copy number space was observed, and there was no difference in genetic similarity between matched lesions (self) and randomly sampled unmatched lesions (other, Fig. 3d). Intra-tumour heterogeneity was inferred based on single-cell level SCNAs and was observed in both pre-invasive and invasive LUAD samples (Fig. 3b).

We hereby report that LUAD demonstrate a more inhibitory tumour immune microenvironment (TME) than AIS/MIA, as evidenced by an increase in exhausted CD8⁺ T cells and Treg cells and a concomitant drop in cytotoxicity (as depicted in Fig. 4c–e). Our findings

align with a previous study that integrated single-cell RNA sequencing (scRNA-seq) and spatial transcriptomics on 6 AIS/MIA and 3 LUAD samples, which also reported an increase in Treg cells from AIS/MIA to LUAD, albeit with samples obtained from different patients³⁷. Another study which performed scRNA-seq on tumours and adjacent normal lung tissues of 5 MSLC patients found Treg cells to be more enriched in tumour tissue compared to normal tissue, while no significant difference was observed between pre-invasive and invasive LUADs³⁸. The shift towards a more inhibitory immune microenvironment during the progression from pre-invasive to invasive cancer has also been reported in pancreatic ductal adenocarcinoma (PDAC)³⁹, highlighting the critical role played by the local immune microenvironment in tumour evolution. Furthermore, we observed fewer bystander T cells in AIS/MIA, as indicated by a higher MANA score in AIS/MIA than in LUAD (as shown in Fig. 4f), which further confirms the greater inhibition of the local immune response as tumours progress. The frequencies of other immune cells, such as gamma-delta T cells, macrophages, dendritic cells and monocytes also demonstrate significant differences between AIS/MIA and LUAD samples, suggesting dynamic changes as tumours progress (as illustrated in Figs. 4c and 6b, d, e). Taken together, these results provide valuable insights into the importance of the local immune response in the progression of LUAD, as systemic immune responses are controlled for.

Tumours express neoantigens that can be recognised by TCRs and induce immune response⁴⁰. In the current study, we performed single-cell TCR sequencing on MSLC samples of different stages to investigate how the TCR repertoire changes as tumours progress. TCR diversity was found to positively correlate with tumour heterogeneity, suggesting that differences in the TCR repertoire may arise from distinct tumour subclones. We observed a higher number of TCR clonotypes and increased TCR diversity in LUAD, while hyperexpanded TCRs were more frequently observed in AIS/MIA. These findings indicate a higher level of intra-tumour heterogeneity in LUAD compared to AIS/MIA. To gain further insights, we predicted neoantigens and assessed their affinity for TCRs. We discovered that both the number of neoantigen counts and the number of neoantigen-committed TCR counts were higher in LUAD, suggesting that distinct neoantigens potentially drive the differences in the T cell repertoire between the two stages. By predicting the affinity between neoantigens and TCRs, we were able to analyse the interplay between neoantigens and TCRs, leading to the identification of CD8⁺ T cells specific to neoantigens. Notably, CD8⁺ T cells containing neoantigen-committed TCRs were less likely to be bystander T cells, as evidenced by higher MANA scores (as shown in Fig. 5g). This difference in MANA scores remained significant when analysing AIS/MIA and LUAD individually (as shown in Supplementary Fig. 9a, b). However, these T cells exhibited a higher level of exhaustion and expressed higher levels of immune checkpoints (Fig. 5h, i), indicating the presence of immune escape mechanisms during the progression of LUAD. Additionally, we found that the expression of genes with non-silent mutations was lower than that of genes with silent mutations (as shown in Fig. 5j), suggesting reduced immunogenicity by tumours as they progressed.

The current TNM staging system considers multiple nodules in the same lobe to be T3 disease, and multiple nodules in a different ipsilateral lobe to be T4 disease⁴¹. However, this classification assumes multiple nodules are intrapulmonary metastases. Although currently there is no definitive consensus for this specific type of patients, patients with multiple primary lung nodules might benefit more from surgical resections⁴². Our study demonstrates the different genomic and immune profiles of independent MSLCs, providing evidence for the feasibility for surgical resection of the lesions.

One limitation of our study is that it remains challenging to accurately predict the affinity between TCRs and neoantigens from computational perspective. More experimentally validated data in the future will be crucial for improving the accuracy of such predictions.

We hope that additional data will emerge enhance our understanding of the TCR repertoire and its association with pre-invasive cancers.

In summary, our study integrated bulk WES, single-cell RNA sequencing (scRNA-seq) and single-cell TCR sequencing (scTCR-seq) on MSLCs, providing evidence of the independent origins of each lesion, the evolutionary pattern of MSLCs and the interaction between tumours and the immune microenvironment of MSLCs. This research sheds light on the unique characteristics of this particular subtype of lung cancer.

Methods

Patient cohort and sample collection

This study complies with all relevant ethical regulations and was approved by the Committee for Ethical Review of Research (Fudan University Shanghai Cancer Center Institutional Review Board No. 090977-1). From December 2021 to September 2022, 8 multiple synchronous lung cancer (MSLC) patients who underwent surgery at the Department of Thoracic Surgery, Fudan University Shanghai Cancer Center were prospectively enrolled in this study. Preoperative tests, including contrast-enhanced chest computed tomography (CT) scanning, were performed to determine the clinical stage of the disease. CT-guided hookwire localisation was performed preoperatively when nodules were too small or too deep to localise. Resected tumours were sent for intraoperative frozen section pathology immediately after they were resected, and intraoperative pathological diagnosis was made. After surgery, tumours were sent to at least 2 pathologists to determine the histological classification and pathological stage. Usually, 3–5 sections of each specimen were used to reach the final pathological stage of the tumour. Tumours were classified into AIS, MIA and invasive adenocarcinoma according to the standards of the International Association for the Study of Lung Cancer, American Thoracic Society and European Respiratory Society⁴³. For invasive adenocarcinomas, the occupancy of each one of these several patterns, namely, lepidic, acinar, papillary, micropapillary, solid and invasive mucinous adenocarcinoma, was recorded in a 5% increment. Informed consents of all patients were obtained.

Sample preparation, single-cell RNA sequencing and single-cell TCR sequencing

Fresh tumour tissue samples were collected and immediately stored in MACS Tissue Storage Solution (Miltenyi Biotec) at 2–8 °C. Within 12 h after surgery, tumour tissues were dissociated and digested. For the quality check and counting of single cell suspension, the cell viability was above 80%. The cells that have passed the test were washed and resuspended to prepare a suitable cell concentration of 700–1200 cells/ μ L. Cells were then loaded onto the 10X Chromium Single Cell Platform (10X Genomics). Each cell suspension was subjected to 5' single-cell RNA sequencing using Chromium Next GEM Single Cell 5' Kit v2 (PN-1000263). Generation of gel beads in emulsion (GEMs), barcoding, GEM-RT clean-up, cDNA amplification and library construction were performed following manufacturer's instructions. Single-cell TCR sequencing libraries were simultaneously constructed using Chromium single cell V(D)J Enrichment Kit, Human T cell (PN-1000252). Libraries were sequenced on the Illumina Nova6000 platform to generate 150-bp paired-end reads, and mapped to GRCh38 human genome using the Cell Ranger toolkit (v2.1.0). Average reads per cell for scRNA-seq data was 69,353, and average reads per cell for scTCR-seq data was 24,350.

Filtering and normalisation of scRNA-seq data

Scrublet (version 0.2.3) was used to identify and remove doublet cells with expected doublet rate 0.06⁴⁴. Downstream data processing of single-cell RNA sequencing (scRNA-seq) data was conducted using R (version 4.1.3), and the foundational analysis pipeline was established using the Seurat package (version 4.3.0)⁴⁵. Cells with either lower than

200 or higher than 6000 expressed genes were excluded from the analysis. Next, cells with a percentage of RNA encoding proteins less than 80% were also discarded. Because of the heterogeneity among different samples, a dynamic adjustment strategy was used for additional filtering parameters. Specifically, cells with a total number of UMIs that was greater or smaller than median $\pm 5 \times$ MAD (median absolute deviation) or a percentage of molecules mapped to mitochondrial genes greater than median $+ 6 \times$ MAD were then excluded from the downstream analysis. RNA expression was then log-normalised using the scale factor of 10,000.

Identification of epithelial cells, stromal cells and immune cells

Cells were at first clustered and annotated into 3 types, namely epithelial cells (*EPCAM*), stromal cells (*MME*, *PECAM*) and immune cells (*PTPRC*) based on their gene expression profiles.

Stromal cells were further clustered and annotated as fibroblasts (*DCN*, *CIR*, *COL1A1*) and endothelial cells (*PECAMI*, *RAMP2*, *CLDN5*), while immune cells were further clustered and annotated as T cells (*CD3D*, *CD3E*), NK cells (*NGK7*), B cells (*CD79A*), plasma cells (*IGKC*), myeloid cells (*LYZ*, *CD68*) and mast cells (*TPSB2*, *CPA3*, *MS4A2*).

Regarding immune cells, T cells were first clustered and annotated into CD4+ T cells, CD8+ T cells and CD4- CD8- T cells. CD4+ T cells were mainly divided into Naïve T cells (*TCF7*, *SELL*, *LEFI*, *CCR7*), Treg cells (*IL2RA*, *FOXP3*, *IKZF2*, *TGFB1*, *TGFB3*, *TGFB1*, *TGFBRI*), and CD8+ T cells were defined different subclusters according to the expression of *GZMK* and *GZMB*. In addition, the expression of exhaustion marker genes (*LAG3*, *TIGIT*, *PDCDI*, *CTLA4*, *HAVCR2*, *ENTPDI*) helped us define a subcluster of exhausted killing T cells. The identification process of myeloid and NK cells followed a similar approach, and the specific markers for these cell types can be found in Supplementary Data 3.

Epithelial cells were further clustered into normal epithelial cells and tumour cells based on the copy number alteration profiles of each single cell (see method inferCNV).

InferCNV and tumour cell identification

In order to isolate malignant tumour cells from all epithelial cells, copy number alterations were inferred from the patterns of chromosomal gene expression using inferCNV (version = 1.10.1)⁴⁶. The expression profiles of randomly selected 300 fibroblasts and 300 endothelial cells were used as reference. We used the following parameters: denoise = TRUE, default hidden markov model settings, and a value of 0.1 for 'cutoff'. Arm level copy number alterations were determined based on GRCh38 cytoband location information, and Each CNV event was annotated as gain or loss. After the copy number landscape for all the cells was generated, cells were clustered based on their copy number profiles. Cells with more quiet genomes similar to reference cells were considered normal epithelial cells, while the other cells with more unstable genomes were considered tumour cells. Finally, phylogenetic trees were constructed based on the copy number information, and Uphyloplot2 was used to visualise the intra-tumour evolutionary trees for each sample⁴⁷.

Genomic distance calculation

Genomic distance was calculated by taking the Euclidean distance of the detected genomic SCNAs. Briefly, SCNAs of all samples were represented as a matrix, with 1 defined as an existing event and 0 as a non-existing event. Pairwise Euclidean distances were then calculated either between the MSLCs from the same patients or randomly selected samples from different patients.

MANA score calculation

T cells recognising mutation-associated neoantigens (MANA) play a crucial role in anti-tumour immune responses. By using the AddmoduleScore function embedded in the Seurat package, we calculated MANA score was calculated based on a set of genes that were reported

to be enriched in MANA-specific T cells²⁸. The full list of genes used for calculating MANA score was shown in Supplementary Data 4. Wilcoxon one-sided test was employed to calculate the significance level, where the Cohen's d value was used to measure the effect size.

Calculation of exhaustion score, cytotoxic score and immune checkpoint score

To quantify the level of exhaustion and cytotoxicity of CD8+ T cells in the TME, we utilised the AddmoduleScore function embedded in the Seurat package to calculate exhaustion score and cytotoxicity score based on the expression of specific sets of genes (Supplementary Data 4). To explore the expression level of immune checkpoints, we defined immune checkpoint score using a list of genes that were regarded as immune checkpoints (*CTLA4*, *PDCDI*, *LAG3*, *HAVCR2*, *TIGIT* and *ENTPDI*). AddmoduleScore function embedded in the Seurat package was used to calculate the immune checkpoint score. Wilcoxon one-sided test was employed to calculate the significance level, where the Cohen's d value was used to measure the effect size.

Flow cytometry

Fresh pathologically confirmed LUAD or MIA tumour tissues were transported in ice-cold H1640 (Gibco, Life Technologies) immediately after surgical resection and then were dissected, mechanically minced, treated with RPMI 1640 supplemented plus collagenase IV (1 mg/ml, Gbicol) and DNase I (100 µg/ml, Sigma) for 1 h at 37 °C and filtered to single-cell suspension (70 µm, Corning). The single-cell suspensions were washed in phosphate-buffered saline twice, enriched by Lymphoprep density gradient centrifugation, washed and resuspended in phosphate-buffered saline to gain peripheral mononuclear cells. Live/dead staining was performed using the Zombie Yellow Fixable Viability Kit (BioLegend #329920). For extracellular staining, antibodies were incubated with cells for 20 min in FACS buffer. For intracellular staining, cells were fixed with the Foxp3/Transcription Factor Staining Buffer Set (eBiosciences) and then stained with intracellular antibodies, as per the manufacturer's instructions. After staining, data was collected using a LSR Fortessa (BD Biosciences) and analysis was performed using FlowJo software (v.10.5.3, TreeStar). A full list of antibodies used in flow cytometry is shown in Supplementary Data 5. Gating strategies are shown in Supplementary Fig. 13.

Multiplex immunofluorescence staining

We used the PD-L1 rabbit anti-human antibody (clone: SP142, Cat: ab228462, Abcam) and the PanCK mouse anti-human antibody (clone: IHC-M067, Cat: AE1/AE3, Novus) to perform three-colour immunofluorescence staining. Briefly, lung sections were deparaffinized and rehydrated successively, followed by antigen retrieval and incubation in 3% H₂O₂ and normal goat serum. Then, sections were serially stained with primary antibodies, secondary antibodies and DAPI. Slides were scanned using the PerkinElmer Vectra3 platform and analysed by inform (v2.3, PerkinElmer) using machine learning algorithm with a visual cutoff followed by single-cell-based mean pixel fluorescence intensity to achieve accuracy and calculate⁴⁸. A full list of antibodies used in flow cytometry is shown in Supplementary Data 5.

Whole-exome sequencing (WES)

Genomic DNA the 16 tumour samples obtained from 8 patients and their matched normal tissues were extracted using the QIAamp DNA Mini Kit (Qiagen #51304) following the manufacturer's instructions. Extracted DNA was fragmented into an average size of ~200 bp before they were amplified and purified. DNA was captured to generate indexed, target-enriched library amplicons. Twist Human Core Exome EF Multiplex Complete Kit, 96 Samples was used for library preparation (Twist Bioscience #PN-100803). Constructed libraries were then sequenced on the Illumina Novaseq6000 and 150 bp paired-end reads

were generated. Sequence depth was 200x for tumour samples and 100x for matched normal lung tissues.

Qualification control, alignment and mutation calling of WES data

To ensure the quality of raw sequencing data, fastp (version 0.23.2) was employed with the relevant quality threshold set to 25 (default 20)⁴⁹. Sequencing data were aligned against reference human genome (hg38) using BWA-MEM⁵⁰. The Genome Analysis Toolkit (GATK, version 4.2.3.0) was used to perform base quality recalibration and realignments⁵¹. Somatic single-nucleotide variants (SNVs) and indels were called using Mutect2⁵², and annotated using ANNOVAR⁵³. TMB was calculated as the total number of nonsynonymous SNVs and indels divided by 30, given coverage of ~30 Mb.

Detection of *ALK*, *ROS1* and *RET* rearrangement events

A capture-based targeted sequencing panel was used to analyse selected introns in *ALK*, *RET* and *ROS1* for the detection of rearrangement events. Libraries were prepared by amplifying the targeted regions using multiplex polymerase chain reaction, followed by adapter DNA ligation. Multiplexed sequencing was then performed using the Illumina HiSeq 2500 platform. Rearrangements were identified using Illumina suite software.

Neoantigen prediction

We first used scHLACount to calculate the HLA molecular subtypes of each sample⁵⁴. Based on the mutation sites called from WES data and HLA subtypes calculated, we utilised netMHCpan (version 4.0) to calculate the affinity score between each mutant peptide and HLA molecule⁵⁵. Binding activity was divided into three groups: strong binding, weak binding and non-binding. We focused on mutations with strong and weak binding activities for predicting the affinity between neoantigens and TCRs.

TCR clonotype calling and prediction of affinity between neoantigens and TCRs

TCR clonotypes were called using ScRepertoire⁵⁶. TCR clones with a copy number greater than 5 in each sample were used as candidate libraries. By utilising the neoantigens calculated by netMHCpan, the affinity between neoantigens and the TCRs of the sample was calculated using Panpep³⁰. We specifically sought combinations where the affinity between the neoantigen and the corresponding TCR was strong (affinity score ≥ 0.5), and defined those TCRs as neoantigen-committed TCRs. T cells containing neoantigen-committed TCRs were matched using shared barcodes. There were occasions where multiple strong binding TCRs were matched to one neoantigen. In this situation, we chose the TCR with the highest affinity score as the neoantigen-committed TCR.

TCR diversity calculation

We utilised the widely employed diversity metric, the Shannon's Diversity Index (SDI), to compute TCR diversity and population heterogeneity. The formula is as follows:

$$\text{Shannon's Diversity Index} = - \sum_{i=1}^n (P_i \ln(P_i))$$

where n represents the total number of types of elements in the overall composition, and P_i denotes the proportion of an individual element in the overall composition.

When calculating TCR diversity, the frequency of all TCRs in a single sample was input. Two tumour samples from the same patient were categorised into two groups based on the median value of SDI. Tumours with a TCR SDI less than the median value was considered of

low TCR diversity and tumours with a TCR SDI more than the median value was considered of high diversity.

Reporting summary

Further information on research design is available in the Nature Portfolio Reporting Summary linked to this article.

Data availability

The raw data generated in this study have been deposited in the GSA-human data repository with the accession numbers HRA005685 for whole-exome sequencing data (<https://ngdc.cncb.ac.cn/gsa-human/browse/HRA005685>), HRA005684 for single-cell RNA-seq data (<https://ngdc.cncb.ac.cn/gsa-human/browse/HRA005684>) and HRA005697 for single-cell TCR-seq data (<https://ngdc.cncb.ac.cn/gsa-human/browse/HRA005697>). The data used in the figures in this study are provided in the Source Data file. All other data are available in the article and its Supplementary files or from the corresponding author upon request. Source data are provided with this paper.

Code availability

Custom codes for data analysis were stored in the GitHub repository: https://github.com/yuezha097/FUSCC_MSLC-paper-codes⁵⁷.

References

- Sung, H. et al. Global cancer statistics 2020: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA Cancer J. Clin.* **71**, 209–249 (2021).
- Chen, W. et al. Cancer statistics in China, 2015. *CA Cancer J. Clin.* **66**, 115–132 (2016).
- Mok, T. S. et al. Gefitinib or carboplatin-paclitaxel in pulmonary adenocarcinoma. *N. Engl. J. Med.* **361**, 947–957 (2009).
- Mok, T. S. et al. Osimertinib or platinum-pemetrexed in EGFR T790M-positive lung cancer. *N. Engl. J. Med.* **376**, 629–640 (2017).
- Solomon, B. J. et al. First-line crizotinib versus chemotherapy in ALK-positive lung cancer. *N. Engl. J. Med.* **371**, 2167–2177 (2014).
- Shaw, A. T. et al. Lorlatinib in advanced ROS1-positive non-small-cell lung cancer: a multicentre, open-label, single-arm, phase 1-2 trial. *Lancet Oncol.* **20**, 1691–1701 (2019).
- Planchard, D. et al. Dabrafenib plus trametinib in patients with previously untreated BRAF(V600E)-mutant metastatic non-small-cell lung cancer: an open-label, phase 2 trial. *Lancet Oncol.* **18**, 1307–1316 (2017).
- Mok, T. S. K. et al. Pembrolizumab versus chemotherapy for previously untreated, PD-L1-expressing, locally advanced or metastatic non-small-cell lung cancer (KEYNOTE-042): a randomised, open-label, controlled, phase 3 trial. *Lancet* **393**, 1819–1830 (2019).
- Herbst, R. S. et al. Pembrolizumab versus docetaxel for previously treated, PD-L1-positive, advanced non-small-cell lung cancer (KEYNOTE-010): a randomised controlled trial. *Lancet* **387**, 1540–1550 (2016).
- Borghaei, H. et al. Nivolumab versus docetaxel in advanced non-squamous non-small-cell lung cancer. *N. Engl. J. Med.* **373**, 1627–1639 (2015).
- Doroshov, D. B. et al. PD-L1 as a biomarker of response to immune-checkpoint inhibitors. *Nat. Rev. Clin. Oncol.* **18**, 345–362 (2021).
- Shum, B., Larkin, J. & Turajlic, S. Predictive biomarkers for response to immune checkpoint inhibition. *Semin. Cancer Biol.* **79**, 4–17 (2022).
- Martini, N. & Melamed, M. R. Multiple primary lung cancers. *J. Thorac. Cardiovasc Surg.* **70**, 606–612 (1975).
- Huo, J. W. et al. Radiological classification, gene-mutation status, and surgical prognosis of synchronous multiple primary lung cancer. *Eur. Radio.* **32**, 4264–4274 (2022).

15. Warth, A. et al. Clonality of multifocal nonsmall cell lung cancer: implications for staging and therapy. *Eur. Respir. J.* **39**, 1437–1442 (2012).
16. Yu, Y. C. et al. Surgical results of synchronous multiple primary lung cancers: similar to the stage-matched solitary primary lung cancers? *Ann. Thorac. Surg.* **96**, 1966–1974 (2013).
17. Gazdar, A. F. & Minna, J. D. Multifocal lung cancers—clonality vs field cancerization and does it matter? *J. Natl Cancer Inst.* **101**, 541–543 (2009).
18. Liu, Y. et al. Genomic heterogeneity of multiple synchronous lung cancer. *Nat. Commun.* **7**, 13200 (2016).
19. Maeshima, A. M. et al. Histological scoring for small lung adenocarcinomas 2 cm or less in diameter: a reliable prognostic indicator. *J. Thorac. Oncol.* **5**, 333–339 (2010).
20. Borczuk, A. C. et al. Invasive size is an independent predictor of survival in pulmonary adenocarcinoma. *Am. J. Surg. Pathol.* **33**, 462–469 (2009).
21. Chen, H. et al. Genomic and immune profiling of pre-invasive lung adenocarcinoma. *Nat. Commun.* **10**, 5472 (2019).
22. Dejima, H. et al. Immune evolution from preneoplasia to invasive lung adenocarcinomas and underlying molecular features. *Nat. Commun.* **12**, 2722 (2021).
23. Hu, X. et al. Multi-region exome sequencing reveals genomic evolution from preneoplasia to lung adenocarcinoma. *Nat. Commun.* **10**, 2978 (2019).
24. Mansuet-Lupo, A. et al. Proposal for a combined histomolecular algorithm to distinguish multiple primary adenocarcinomas from intrapulmonary metastasis in patients with multiple lung tumors. *J. Thorac. Oncol.* **14**, 844–856 (2019).
25. Puram, S. V. et al. Single-cell transcriptomic analysis of primary and metastatic tumor ecosystems in head and neck cancer. *Cell* **171**, 1611–1624.e24 (2017).
26. Endesfelder, D. et al. Chromosomal instability selects gene copy-number variants encoding core regulators of proliferation in ER+ breast cancer. *Cancer Res.* **74**, 4853–4863 (2014).
27. Guo, X. et al. Global characterization of T cells in non-small-cell lung cancer by single-cell sequencing. *Nat. Med.* **24**, 978–985 (2018).
28. Caushi, J. X. et al. Transcriptional programs of neoantigen-specific TIL in anti-PD-1-treated lung cancers. *Nature* **596**, 126–132 (2021).
29. Simoni, Y. et al. Bystander CD8(+) T cells are abundant and phenotypically distinct in human tumour infiltrates. *Nature* **557**, 575–579 (2018).
30. Gao, Y. et al. Pan-Peptide Meta Learning for T-cell receptor–antigen binding recognition. *Nat. Mach. Intell.* **5**, 236–249 (2023).
31. Turajlic, S. et al. Deterministic evolutionary trajectories influence primary tumor growth: TRACERx renal. *Cell* **173**, 595–610.e11 (2018).
32. Thierry, A. R. et al. Circulating DNA demonstrates convergent evolution and common resistance mechanisms during treatment of colorectal cancer. *Clin. Cancer Res.* **23**, 4578–4591 (2017).
33. Misale, S., Di Nicolantonio, F., Sartore-Bianchi, A., Siena, S. & Bardelli, A. Resistance to anti-EGFR therapy in colorectal cancer: from heterogeneity to convergent evolution. *Cancer Discov.* **4**, 1269–1280 (2014).
34. Griffiths, J. I. et al. Serial single-cell genomics reveals convergent subclonal evolution of resistance as early-stage breast cancer patients progress on endocrine plus CDK4/6 therapy. *Nat. Cancer* **2**, 658–671 (2021).
35. Friedlaender, A. et al. EGFR and HER2 exon 20 insertions in solid tumours: from biology to treatment. *Nat. Rev. Clin. Oncol.* **19**, 51–69 (2022).
36. Han, H. et al. Targeting HER2 Exon 20 Insertion-Mutant Lung Adenocarcinoma with a Novel Tyrosine Kinase Inhibitor Mobocertinib. *Cancer Res.* **81**, 5311–5324 (2021).
37. Zhu, J. et al. Delineating the dynamic evolution from preneoplasia to invasive lung adenocarcinoma by integrating single-cell RNA sequencing and spatial transcriptomics. *Exp. Mol. Med* **54**, 2060–2076 (2022).
38. He, Y. et al. Mechanisms of progression and heterogeneity in multiple nodules of lung adenocarcinoma. *Small Methods* **5**, e2100082 (2021).
39. Chen, K. et al. Single-cell RNA-seq reveals dynamic change in tumor microenvironment during pancreatic ductal adenocarcinoma malignant progression. *EBioMedicine* **66**, 103315 (2021).
40. Morazan-Fernandez, D., Mora, J. & Molina-Mora, J. A. In Silico pipeline to identify tumor-specific antigens for cancer immunotherapy using exome sequencing data. *Phenomics* **3**, 130–137 (2023).
41. Goldstraw, P. et al. The IASLC lung cancer staging project: proposals for revision of the TNM stage groupings in the forthcoming (Eighth) Edition of the TNM classification for lung cancer. *J. Thorac. Oncol.* **11**, 39–51 (2016).
42. Arai, J. et al. Clinical and molecular analysis of synchronous double lung cancers. *Lung Cancer* **77**, 281–287 (2021).
43. Travis, W. D. et al. International association for the study of lung cancer/american thoracic society/European respiratory society international multidisciplinary classification of lung adenocarcinoma. *J. Thorac. Oncol.* **6**, 244–285 (2011).
44. Wolock, S. L., Lopez, R., Klein, A. M. & Scrublet Computational identification of cell doublets in single-cell transcriptomic data. *Cell Syst.* **8**, 281–291.e9 (2019).
45. Hao, Y. et al. Integrated analysis of multimodal single-cell data. *Cell* **184**, 3573–3587.e29 (2021).
46. Tickle, T. I., Georgescu, C., Brown, M. & Haas, B. inferCNV of the Trinity CTAT Project. (Klarman Cell Observatory, Broad Institute of MIT and Harvard, Cambridge, MA, USA, 2019) <https://github.com/broadinstitute/inferCNV>.
47. Kurtenbach, S., Cruz, A. M., Rodriguez, D. A., Durante, M. A. & Harbour, J. W. UPhyloplot2: visualizing phylogenetic trees from single-cell RNA-seq data. *BMC Genomics* **22**, 419 (2021).
48. Fan, F. et al. Elevated mast cell abundance is associated with enrichment of CCR2+ cytotoxic T cells and favorable prognosis in lung adenocarcinoma. *Cancer Res.* **83**, 2690–2703 (2023).
49. Chen, S., Zhou, Y., Chen, Y. & Gu, J. fastp: an ultra-fast all-in-one FASTQ preprocessor. *Bioinformatics* **34**, i884–i890 (2018).
50. Li, H. Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. arXiv:1303.3997v2 [q-bio.GN] (2013).
51. DePristo, M. A. et al. A framework for variation discovery and genotyping using next-generation DNA sequencing data. *Nat. Genet* **43**, 491–498 (2011).
52. Cibulskis, K. et al. Sensitive detection of somatic point mutations in impure and heterogeneous cancer samples. *Nat. Biotechnol.* **31**, 213–219 (2013).
53. Wang, K., Li, M. & Hakonarson, H. ANNOVAR: functional annotation of genetic variants from high-throughput sequencing data. *Nucleic Acids Res.* **38**, e164 (2010).
54. Darby, C. A., Stubbington, M. J. T., Marks, P. J., Martinez Barrio, A. & Fiddes, I. T. scHLAcount: allele-specific HLA expression from single-cell gene expression data. *Bioinformatics* **36**, 3905–3906 (2020).
55. Jurtz, V. et al. NetMHCpan-4.0: improved peptide-MHC class I interaction predictions integrating eluted ligand and peptide binding affinity data. *J. Immunol.* **199**, 3360–3368 (2017).
56. Borcherding, N., Bormann, N. L. & Kraus, G. scRepertoire: an R-based toolkit for single-cell immune receptor analysis. *F1000Res.* **9**, 47 (2020).

57. Zhao, Y. & Wang, J. Genomic and immune heterogeneity of multiple synchronous lung adenocarcinoma at different developmental stages (Paper). Zenodo. <https://doi.org/10.5281/zenodo.12878282> (2024).

Acknowledgements

This study is supported by the National Natural Science Foundation of People's Republic of China (81930073 for Haiquan Chen and 82203037 for Y. Zhao), the Shanghai Technology Innovation Action Project (20JC1417200, Haiquan Chen), Shanghai Municipal Key Clinical Specialty Project (SHSLCZDZK02104, Haiquan Chen), Shanghai Sailing Program (22YF1408900, Y. Zhao) and the Cooperation Project of Conquering Major Diseases in Xuhui District (XHLHGG202101, Haiquan Chen). The computations in this research were performed using the CFFF platform of Fudan University. Figure 1/Panels a and c, and Supplementary Fig. 1, created with BioRender.com, released under a Creative Commons Attribution-NonCommercial-NoDerivs 4.0 International license. We thank Genergy Bio-Technology (Shanghai) Co., especially Mr. Xiang Wei, for providing sequencing service and technical support.

Author contributions

Y. Zhao, K.L., Z.C. and Haiquan Chen conceived the concept. Y. Zhao, J.G., J.W., J.S., Q.W. and M.Z. performed WES data analyses. Y. Zhao, J.G., J.W., D.Q. and S.Z. performed scRNA-seq analyses. Y. Zhao, J.G. and J.W. performed scTCR-seq analyses. Y. Zhao, J.G., J.W., C.C., D.Q., M.A., S.F.P., R.V., K.L. and Z.C. contributed analysis tools and provided technical support. F. Fan performed flow cytometry and immunofluorescent staining experiments. R.G., Yang Zhang, T.Y., Y. Lin, H.L., Y.P., Q.H., Haiqing Chen, H. Han, Z.G., F. Fu, Y.Y., H.Y., Yawei Zhang, J.X., H. Hu, Y.S., Y. Li and Haiquan Chen helped with surgical sample collection. Y. Lin and Y. Li helped with pathological diagnosis. All authors contributed data and to the interpretation. Y. Zhao, J.G., J.W., F. Fan, C.C., D.Q., K.L., Z.C. and Haiquan Chen wrote the paper.

Competing interests

The authors declare the following competing interests: K.L. has the following disclosures (all outside of the submitted work): patent on indel burden and CPI response pending, patent on ctDNA minimal residual disease calling methods; speaker fees from Roche tissue diagnostics and Ellipses pharma; research funding from CRUK TDL/Ono/LifeArc alliance and Genesis Therapeutics; and consulting roles with

Monopteros Therapeutics, Kynos Therapeutics and Tempus Labs, Inc. M.A. is supported by the City of London Centre Clinical Academic Training Programme (Year 3, SEBSTF-2021\100007). All the other authors declare no competing interests.

Additional information

Supplementary information The online version contains supplementary material available at <https://doi.org/10.1038/s41467-024-52139-2>.

Correspondence and requests for materials should be addressed to Yue Zhao, Kevin Litchfield, Zhiwei Cao or Haiquan Chen.

Peer review information *Nature Communications* thanks the anonymous reviewer(s) for their contribution to the peer review of this work. A peer review file is available.

Reprints and permissions information is available at <http://www.nature.com/reprints>

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

Open Access This article is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License, which permits any non-commercial use, sharing, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if you modified the licensed material. You do not have permission under this licence to share adapted material derived from this article or parts of it. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by-nc-nd/4.0/>.

© The Author(s) 2024