nature cancer

Resource

https://doi.org/10.1038/s43018-025-01053-7

A cellular and spatial atlas of *TP53*-associated tissue remodeling defines a multicellular tumor ecosystem in lung adenocarcinoma

Received: 27 November 2023

Accepted: 26 August 2025

Published online: 7 October 2025

Check for updates

William Zhao[1,2,20], Thinh T. Nguyen[1,2,20], Atharva Bhagwat[1,2,20], Akhil Kumar[1,2,20], Bruno Giotti[1,2,20], Benjamin Kepecs[1,2], Jason L. Weirather[3], Navin R. Mahadevan[3,4,5], Asa Segerstolpe[6], Komal Dolasia[1,2], Jamshid Abdul-Ghafar[7], Naomi R. Besson[8], Stephanie M. Jones[8], Brian Y. Soong[1,2], Chendi Li[9,10], Sebastien Vigneau[11,16], Michal Slyper[6,17], Isaac Wakiro[11,18], Mei-Ju Su[11], Karla Helvie[11], Allison Frangieh[11], Judit Jane-Valbuena[6,17], Orr Ashenberg[6], Mark Awad[12], Asaf Rotem[11,19], Raphael Bueno[13], Orit Rozenblatt-Rosen[6,17], Kathleen Pfaff[8], Scott Rodig[5], Aaron N. Hata[9,10], Aviv Regev[6,14,17], Bruce E. Johnson[11] & Alexander M. Tsankov[1,2,6,15]

Tumor protein p53 (*TP53*) is the most frequently mutated gene across many cancers and is associated with shorter overall survival in lung adenocarcinoma (LUAD). Here, to define how *TP53* mutations affect the LUAD tumor microenvironment (TME), we constructed a multiomic cellular and spatial atlas of 23 treatment-naive human lung tumors. We found that *TP53*-mutant malignant cells lose alveolar identity and upregulate highly proliferative and entropic gene expression programs consistently across LUAD tumors from resectable clinical samples, genetically engineered mouse models, and cell lines harboring a wide spectrum of *TP53* mutations. We further identified a multicellular tumor niche composed of *SPP1*$^{+}$ macrophages and collagen-expressing fibroblasts that coincides with hypoxic, prometastatic expression programs in *TP53*-mutant tumors. Spatially correlated angiostatic and immune checkpoint interactions, including *CD274*–*PDCD1* and *PVR*–*TIGIT*, are also enriched in *TP53*-mutant LUAD tumors and likely engender a more favorable response to checkpoint blockade therapy. Our systematic approach can be used to investigate genotype-associated TMEs in other cancers.

Non-small-cell lung cancer (NSCLC) is a heterogeneous disease that has routinely been classified into different histological subtypes[1]. More recently, genomic alterations have been used to characterize molecular subtypes of NSCLC that can be used to select treatments for persons with specific genomic changes[2–4]. Tyrosine kinase inhibitors (TKIs) and other targeted therapies against oncogenic mutations and chromosomal rearrangements (for example, *EGFR*, *KRAS*^{G12C} and *ALK*) have led to improved survival in persons with NSCLC[5]. Treatments with immune checkpoint inhibitors (ICIs) (for example, anti-PD1/PDL1) with or without chemotherapy have also shown clinical benefit in a subset of participants[6] and are currently used as first-line treatment for select patients with both resectable[7] and nonresectable NSCLC[8]. However, because of high rates of cancer recurrence, more effective approaches toward stratification and precision therapy are urgently needed.

Tumor protein p53 (*TP53*) is the most frequently mutated gene in NSCLC (~50% of cases) and across many other cancers[9]. *TP53* mutations

A full list of affiliations appears at the end of the paper. e-mail: alexander.tsankov@mssm.edu

are more common in smokers[10] and are associated with tumor progression and metastasis, leading to shorter survival in persons with NSCLC[11–13]. Despite being the most extensively studied tumor suppressor, *TP53* has been challenging to target therapeutically because of the diversity of mutations reported and its involvement in a wide range of regulatory pathways[14–16]. *TP53* is known to have an impact on the tumor microenvironment (TME) during tumor development in mouse models[17–20] and *TP53* mutations have been associated with increased expression of PDL1 and improved response to anti-PD1 therapy in human lung adenocarcinoma (LUAD)[21–26], motivating us to investigate TME differences in persons with LUAD with *TP53*-mutant (*TP53*^{mut}) versus wild-type (*TP53*^{WT}) tumors.

Results

A multiomic atlas of NSCLC

To characterize the impact of *TP53* mutations on the molecular, cellular and spatial heterogeneity in NSCLC, we performed whole-exome sequencing (WES), single-cell RNA sequencing (scRNA-seq), spatial transcriptomics (ST; 10X Visium) and multiplex immunofluorescence (mIF) on primary tumor resections from 23 participants with NSCLC (Fig. 1a). In total, 21 of the 23 participants had a history of cigarette smoking (median: 40, range: 12–100 pack years), 52% were female and all participants were confirmed to be treatment naive before tumor resection (Supplementary Table 1). We used WES analysis to determine the total tumor mutational burden, copy-number alterations (CNAs) and mutational status of genes commonly altered in NSCLC, including *TP53*, *EGFR*, *KRAS*, *STK11*, *KEAP1*, *RBM10* and *PTPRD* in each tumor. There was high concordance between predicted CNA profiles from tumor-matched WES and those inferred from scRNA-seq data (Fig. 1b and Extended Data Fig. 1a), many of which overlapped with commonly deleted (for example, *CDKN2A*, *PTPRD* and *B2M*) and amplified (for example, *NKX2-1 TERT*, and *EGFR*) genes in LUAD[9]. After filtering, we obtained a total of 166,821 high-quality cell profiles and used CNA inference to distinguish between malignant and nonmalignant epithelial cells and then integrated[27], clustered and annotated them into 11 broad cell classes on the basis of the coherent expression of known marker genes (Fig. 1c,d, Extended Data Fig. 1b,c and Supplementary Table 2). After excluding five tumors with squamous or atypical adenocarcinoma histology (Methods), we found that different *TP53* mutations had comparable effects in downregulating the average expression of *TP53* target genes in malignant cells relative to *TP53*^{WT} LUAD tumors (Fig. 1e).

TP53^{mut} LUAD is associated with a distinct TME composition

Cell subsets in the TME of *TP53*^{mut} versus *TP53*^{WT} LUAD tumors had distinguishable cell-intrinsic expression profiles and cell compositions (Fig. 1d,f and Extended Data Fig. 1d). We observed a pronounced decrease in the proportion of endothelial cells and pericytes in *TP53*^{mut} versus *TP53*^{WT} LUAD tumors (Fig. 1f, left), which was confirmed by scoring the corresponding cell type signatures in 510 bulk RNA-seq profiles of primary LUAD tumors from The Cancer Genome Atlas (TCGA) grouped by *TP53* mutational status (Fig. 1f, right). Endothelial cell abnormalities have been linked to tumor cell progression and metastasis[28] and pericyte depletion has been linked to hypoxia-associated epithelial-to-mesenchymal transition (EMT) in breast cancer[29].

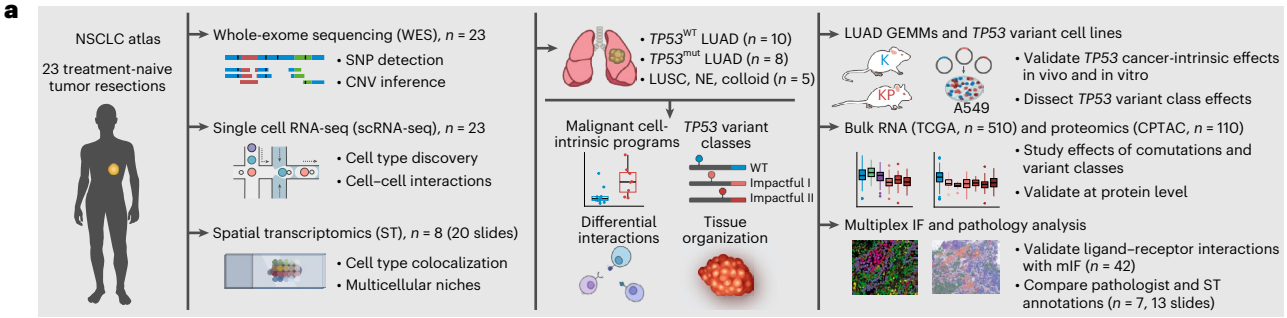
These compositional differences between *TP53*^{mut} and *TP53*^{WT} LUAD were also supported by spatial transcriptomic profiling of 14 tissue sections from six tumors in our cohort (two with *TP53*^{mut} and four with *TP53*^{WT} LUAD). All spatial slides ($n = 20$) were of similar or superior quality compared to Visium spatial data from other recently published studies[30–33] (Extended Data Fig. 1e). We excluded two cases (six slides in total) from downstream spatial analysis because of having an atypical histology (neuroendocrine; MGH1179) or insufficient number of malignant cells (BW16). Next, we used Tangram[34], a deep learning framework, to map tumor-matched scRNA-seq profiles to the corresponding ST data and generate a probabilistic measure of the cellular composition for each spatial measurement (Fig. 1g). Spatial cell composition inference was highly concordant with independent, manual annotations of the corresponding hematoxylin and eosin (H&E) stains by a pulmonary pathologist (N.R.M.; Fig. 1g, right, and Extended Data Fig. 1f–h). Moreover, the cumulative number of CNAs overlapped spatially with Tangram-predicted cancer cell abundance and pathology annotations for neoplastic regions on corresponding H&E stains, supporting the robustness of our computational methods for CNA inference and cell proportion deconvolution across ST data spots (Supplementary Figs. 1 and 2). Lastly, we also observed a decrease in the proportion of endothelial cells and pericytes across ST spots in *TP53*^{mut} versus *TP53*^{WT} tumor samples (Fig. 1h), consistent with our scRNA-seq data.

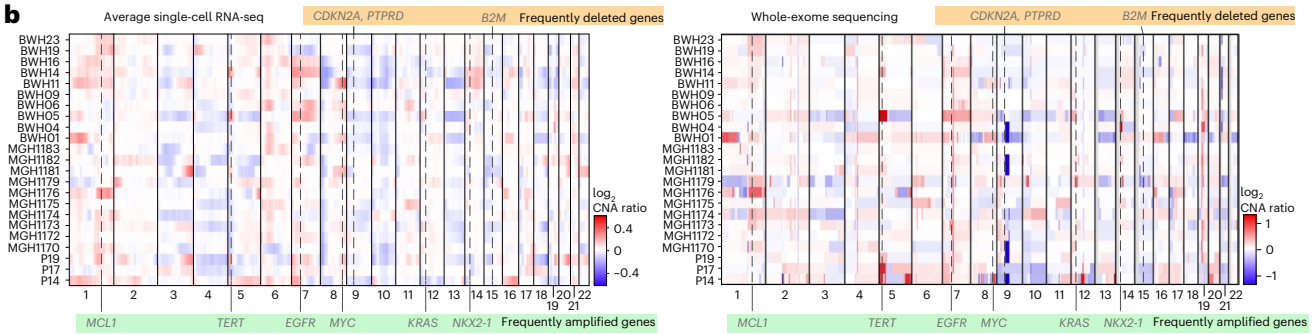
Dichotomous activity of malignant cell programs in LUAD associates with survival

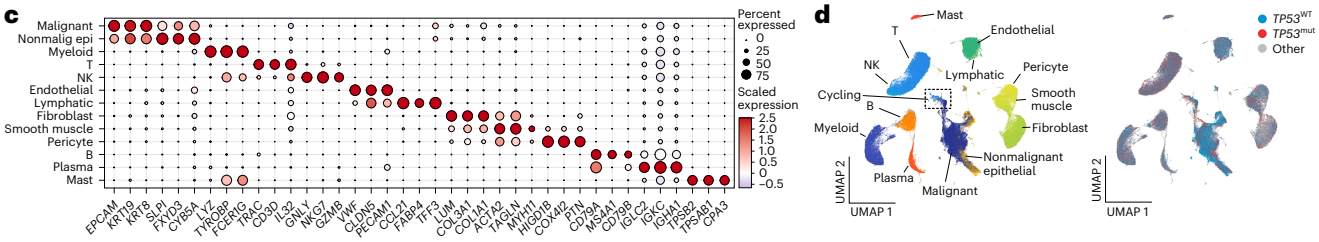
To systematically define the cancer-intrinsic expression programs shared across LUAD tumors, we used canonical correlation analysis[35] to integrate and cluster malignant cells into 18 subsets shared across tumors (Fig. 2a and Extended Data Fig. 2a) and annotated cell subsets using gene set enrichment analysis (Extended Data Fig. 2b), highlighting representative markers (Fig. 2b). These subsets were enriched for genes associated with hallmark cancer processes (for example, cell cycle, hypoxia, glycolysis, partial EMT (pEMT), interferon-γ (IFNγ) and tumor necrosis factor (TNF) signaling), general biological pathways (for example, antigen presentation (major histocompatibility complex

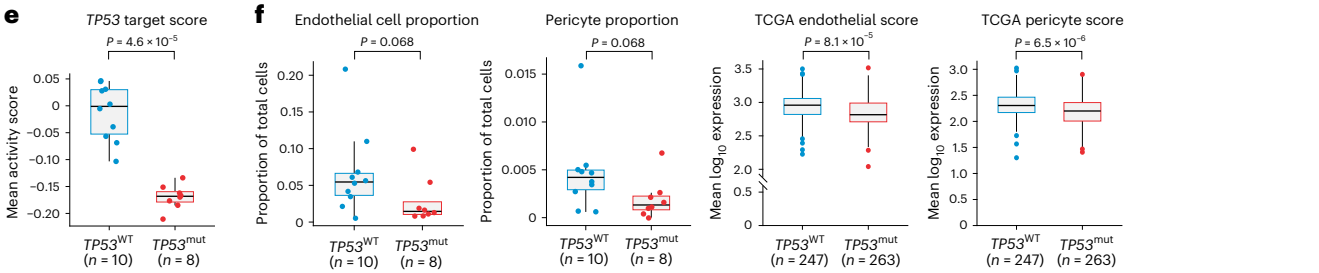
Fig. 1 | Spatially resolved, multiomic atlas of NSCLC. a, Schematic of study design, including cohort selection (left), genomic assays profiled, sample stratification, downstream analyses and validation platforms (right) ($n = 23$, where n refers to the number of NSCLC cases). **b**, Mean $\log_2$ CNA ratio inferred using aggregated scRNA-seq malignant cells (left) and bulk WES (right) from tumor-matched samples. Common focal amplifications (bottom; green) and deletions (top; orange) in LUAD are highlighted. **c**, Expression of representative markers across each annotated broad cell class. **d**, UMAP visualization of 166,821 cells, integrated to correct for batch effects and colored by annotated cell classes (left) and *TP53* mutational status (right). **e**, Average *TP53* target gene expression score in malignant cells isolated from *TP53*^{WT} ($n = 10$; blue) and *TP53*^{mut} ($n = 8$; red) LUAD primary tumor samples, where n refers to the number of participants. The P value (4.6×10^{-5}) was calculated using a two-sided Mann–Whitney–Wilcoxon test. **f**, Left: proportion of endothelial cells and pericytes from scRNA-seq of *TP53*^{WT} ($n = 10$; blue) and *TP53*^{mut} ($n = 8$; red) LUAD tumors, where n refers to the number of participants. P values (from left to right: 0.068 and 0.068) were calculated using a two-sided Mann–Whitney–Wilcoxon test. Right: mean $\log_{10}$ expression of highly specific endothelial and pericyte markers (derived from scRNA-seq) in *TP53*^{WT} ($n = 247$; blue) and *TP53*^{mut} ($n = 263$; red) bulk RNA-seq data from TCGA, where n refers to the number of participants. P values (from left to right: 8.1×10^{-5} and 6.5×10^{-6}) were calculated using a two-tailed multiple regression t-test. **g**, Left: integration of ST data with tumor-matched scRNA-seq data using Tangram allows for deconvolution of the proportion of different cell subsets. Middle: representative spatial distribution of cell classes using Tangram (top) and corresponding tissue section H&E stain before and after annotations by a pathologist (bottom). Right: violin plots comparing proportion of computationally inferred cell classes (y axis) within pathologist-annotated categories (x axis) including malignant ($n = 6{,}694$ spots), stromal ($n = 177$ spots) and immune ($n = 2{,}947$ spots). P values were calculated using a two-sided Mann–Whitney–Wilcoxon test. P values are indicated within the plot. **h**, Proportion of endothelial cells (top) and pericytes (bottom) in ST spots from *TP53*^{WT} (blue; $n = 28{,}980$ spots) and *TP53*^{mut} (red; $n = 9{,}287$ spots) LUAD sections. P values were calculated using a two-sided Mann–Whitney–Wilcoxon test. P values are indicated within the plot. The number of spots per group is depicted in parentheses on the x axis. Throughout this figure, box plot horizontal lines show the 25th, 50th (median) and 75th percentiles, with vertical whiskers extending to a maximum distance of 1.5× the interquartile range from the hinge. Data beyond the whisker ends are plotted individually.

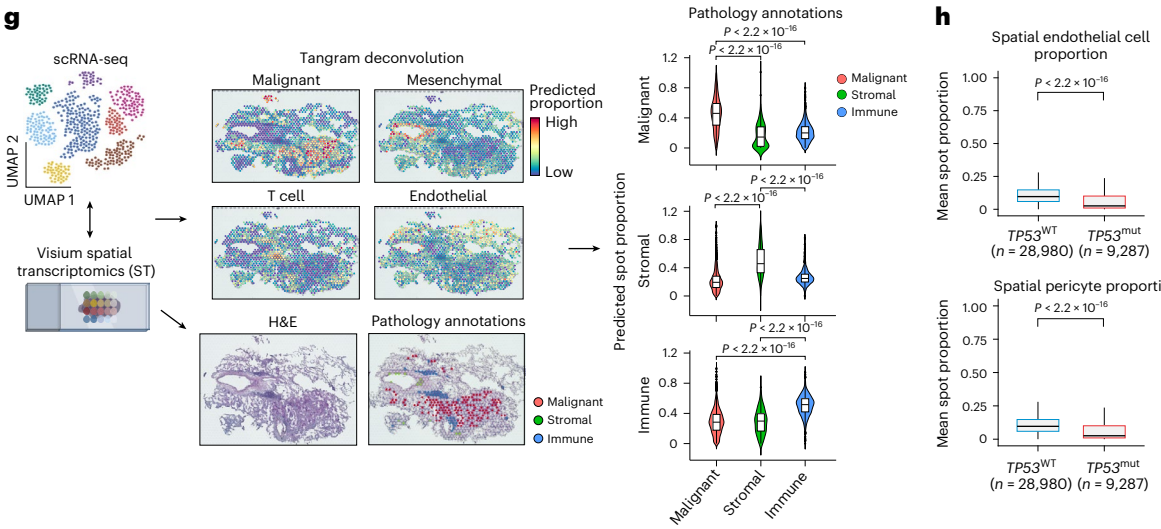
https://doi.org/10.1038/s43018-025-01053-7











https://doi.org/10.1038/s43018-025-01053-7

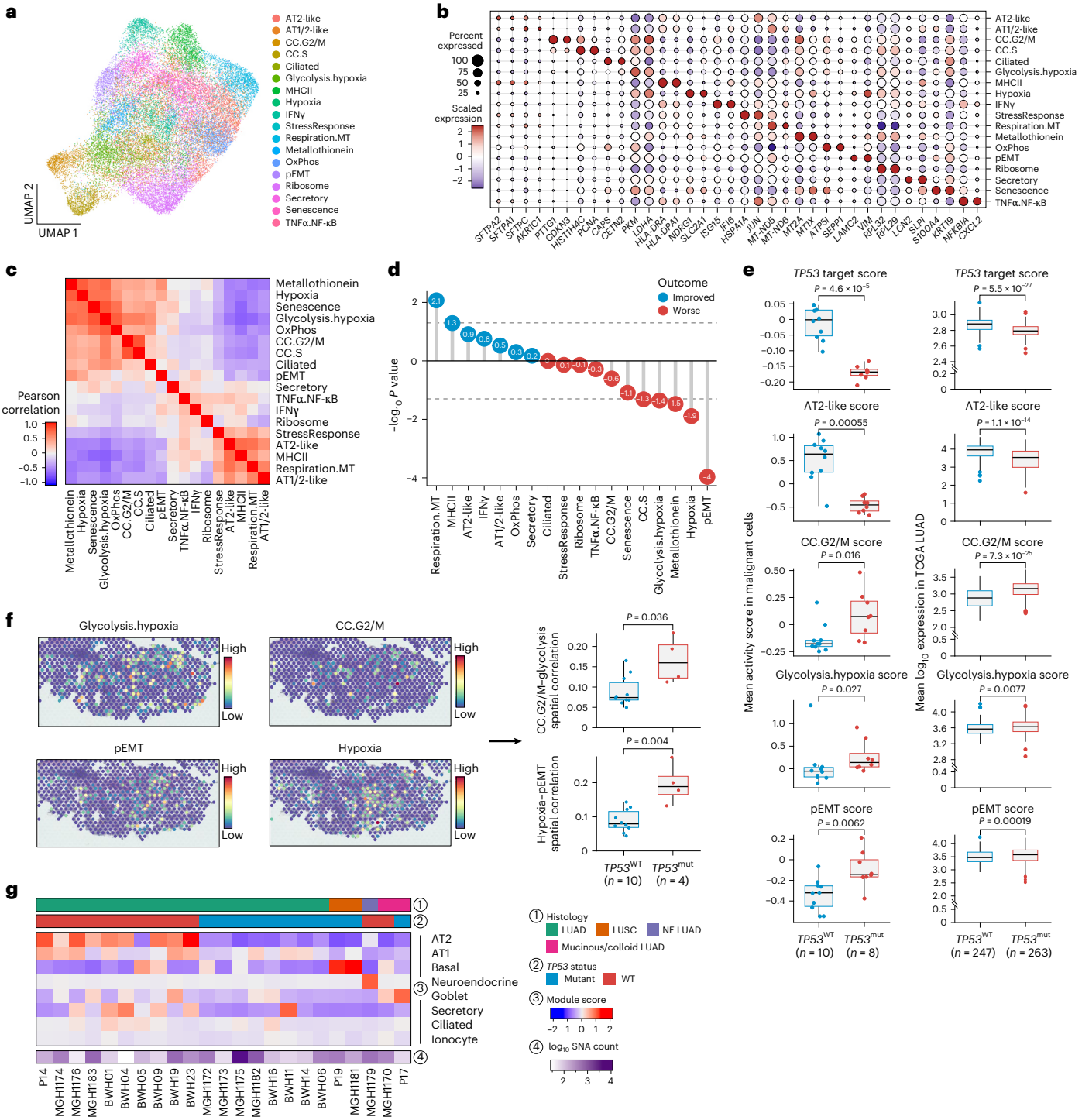


Fig. 2 | Enrichment of cell-cycle, hypoxia and pEMT programs accompanied by loss of AT2 identity in *TP53*^{mut} LUAD malignant cells. a, UMAP of 33,377 malignant cells integrated across 23 NSCLC tumors and colored by annotated malignant expression program subsets. **b**, Dot plot of two selected markers representative of each malignant program subset. **c**, Pearson correlation of malignant program scores across all malignant cells. **d**, Lollipop plot showing $-\log_{10} P$ values from Cox proportional-hazards regression analysis linking malignant programs to the disease outcome of TCGA participants, corrected for *TP53* mutational status. The sign of the *y* axis corresponds to improved (positive value) or worse (negative value) outcome. **e**, Comparison of mean malignant program expression scores between *TP53*^{WT} ($n = 10$ and $n = 247$) and *TP53*^{mut} ($n = 8$ and $n = 263$) tumors in scRNA-seq (left) and bulk RNA-seq (right) data from TCGA, where *n* refers to the number of participants. *P* values were calculated using a two-sided Mann–Whitney–Wilcoxon test (scRNA-seq, left; from top to bottom: 4.6×10^{-5}, 0.00055, 0.016, 0.027 and 0.0062) or a two-tailed multiple regression *t*-test (TCGA, right; from top to bottom: 5.5×10^{-27}, 1.1×10^{-14}, 7.3×10^{-25}, 0.0077 and 0.00019). Horizontal dashed lines indicate a *P* value of 0.05. **f**, Left: representative example of the spatial distribution of four malignant programs' expression in a participant with *TP53*^{mut} LUAD. Right: box plots comparing colocalization (spatial correlation) of pairs of malignant programs between *TP53*^{WT} ($n = 10$) and *TP53*^{mut} ($n = 4$) tumor sections, where *n* refers to the number of samples. *P* values (from top to bottom: 0.036 and 0.004) were calculated using two-sided Mann–Whitney–Wilcoxon test. **g**, Module scores for normal lung epithelial cell signatures averaged across malignant cells in each tumor, ordered by histology (1) and *TP53* mutational status (2). The $\log_{10}$ SNA count for each tumor is shown on the bottom (4). Throughout this figure, box plot horizontal lines show the 25th, 50th (median) and 75th percentiles, with vertical whiskers extending to a maximum distance of 1.5× the interquartile range from the hinge. Data beyond the whisker ends are plotted individually.

class II (MHCII)), stress response and senescence) or lung epithelial cell identity (for example, alveolar type 2 (AT2), ciliated and secretory).

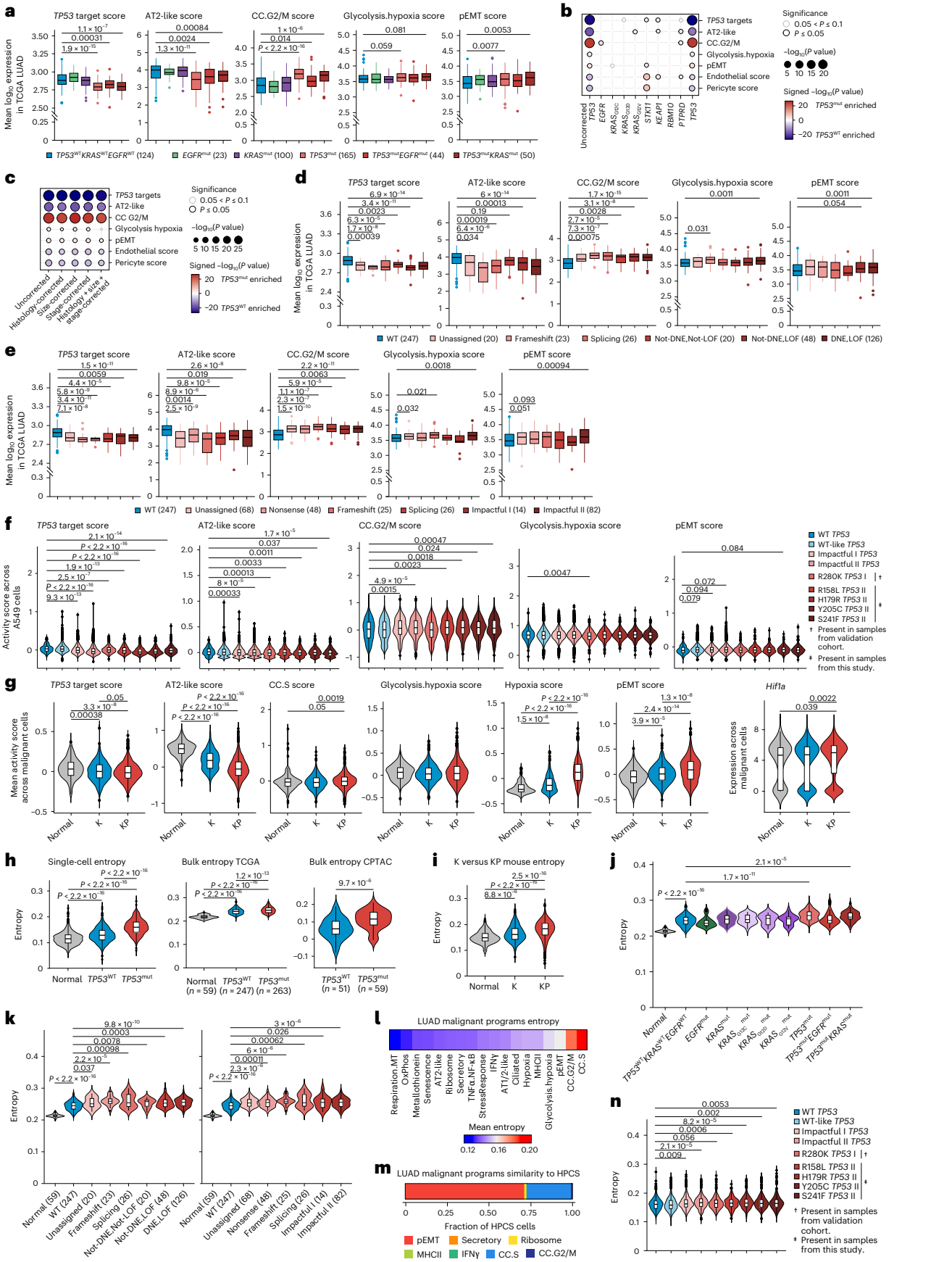
The de novo defined cancer programs clustered into two main groups after correlating their scores across all malignant cells (Fig. 2c) and tumor samples (Extended Data Fig. 2c): one consisting of cell-cycle, hypoxia, pEMT and glycolysis programs and another consisting of alveolar-like, antigen presentation and stress response programs. Malignant program scores in these two groups were strongly anticorrelated, suggesting that they were regulated by two distinct and opposing pathways (Fig. 2c and Extended Data Fig. 2c). We also used consensus non-negative matrix factorization (cNMF)[36] as a complementary method to de novo identify malignant cell coexpressed gene modules from the scRNA-seq data and found high correspondence between these modules and our clustering-derived programs (Extended Data Fig. 2d), supporting the robustness of our cancer-intrinsic program discovery methods. High expression of several malignant programs (for example, pEMT, hypoxia, cell cycle, metallothionein and glycolysis) was associated with shorter overall survival across persons with LUAD tumors in TCGA (Fig. 2d and Extended Data Fig. 2e). The association of different malignant programs with overall survival was largely independent of LUAD histology and pathological stage (Extended Data Fig. 2f). Notably, *TP53*^{mut} tumors also have shorter overall survival in LUAD TCGA (Extended Data Fig. 2g). In contrast, a high expression of the antigen presentation MHCII program was linked to prolonged overall survival (Extended Data Fig. 2f). Previously, various signatures derived from bulk RNA-seq profiling in TCGA were shown to be predictive of outcome in NSCLC[37,38] but our de novo approach allowed for the discovery of more granular programs based on malignant cell expression only, unconfounded by other cell types in the TME.

We identified several cancer-intrinsic differences associated with *TP53* mutations, including higher expression for several malignant programs linked to shorter overall survival (for example, cell cycle, glycolysis and pEMT) and a decreased expression of the AT2-like program and *TP53* targets in *TP53*^{mut} versus *TP53*^{WT} LUAD tumors (Fig. 2e, left). The malignant program differences were corroborated in bulk RNA-seq data from 263 *TP53*^{mut} and 247 *TP53*^{WT} LUAD primary tumors in TCGA (Fig. 2e, right). These findings were also replicated when analyzing a compendium of independent, publicly available scRNA-seq cohorts (Extended Data Fig. 2h and Supplementary Table 3), consisting of 22 additional treatment-naive LUAD primary tumor samples with annotated mutational status (7 *TP53*^{mut} and 15 *TP53*^{WT})[39–43]. Furthermore, spatial coexpression of cell cycle with glycolysis and of hypoxia with pEMT programs was higher in *TP53*^{mut} versus *TP53*^{WT} LUAD tumors (Fig. 2f). Notably, there was not a statistically significant difference in the number of single-nucleotide alterations (SNAs) or CNAs between *TP53*^{mut} and *TP53*^{WT} tumors in our cohort (Extended Data Fig. 2i).

To further investigate the decreased expression of the AT2-like program in *TP53*^{mut} LUAD, we scored each malignant cell across tumors

Fig. 3 | Consistent cancer-intrinsic changes observed in *TP53*^{mut} LUAD across comutations, variant classes, cell lines, mouse models and bulk transcriptomic and proteomic data. a, Comparison of malignant program mean $\log_{10}$ expression between *TP53*^{WT} and different single or comutations in *EGFR*, *KRAS* and *TP53* in bulk RNA-seq data from TCGA tumor samples (from left to right: n = 124, 23, 100, 165, 44 and 50), where n refers to the number of participants. P values were calculated using a two-sided Mann–Whitney–Wilcoxon test. P values are indicated within the plot. **b**, Dot plot showing association of different mutations (columns) with mean $\log_{10}$ expression of malignant programs and endothelial and pericyte scores (rows) in TCGA LUAD. The color of the dots reflects the direction and magnitude of association between mutation status (mutant versus WT) and expression score, using a signed $-\log_{10}$ P value assessed by a two-sided Mann–Whitney–Wilcoxon (uncorrected *TP53*; left) or a two-tailed multiple regression t-test (*EGFR*, *KRAS*^{G12C}, *KRAS*^{G12D}, *KRAS*^{G12V}, *STK11*, *KEAP1*, *RBM10*, *PTPRD* and *TP53*; right). Positive values (red color) indicate enrichment in *TP53*^{mut} samples; negative values (blue color) indicate enrichment in *TP53*^{WT} samples. The size of dots corresponds to the $-\log_{10}$ P value. The black outline indicates $P \le 0.05$ and gray outline indicates $0.05 < P \le 0.1$. **c**, Dot plot showing association of *TP53* mutation status with mean $\log_{10}$ expression of malignant programs, endothelial cell and pericyte abundance scores (rows) in TCGA LUAD, with no correction (left column) and after adjusting for histology, size, stage and all covariates (right columns). The color of the dots reflects the direction and magnitude of association between mutation status (mutant versus WT) and expression score, using a signed $-\log_{10}$ P value assessed by a two-sided Mann–Whitney–Wilcoxon (uncorrected *TP53*; left) or a two-tailed multiple regression t-test (right columns). Positive values (red color) indicate enrichment in *TP53*^{mut} samples; negative values (blue color) indicate enrichment in *TP53*^{WT} samples. The size of dots corresponds to the $-\log_{10}$ P value. The black outline indicates $P \le 0.05$ and gray outline indicates $0.05 < P \le 0.1$. **d**, Comparison of malignant program mean $\log_{10}$ expression between *TP53*^{WT} and different functional *TP53*^{mut} categories in TCGA tumor samples (from left to right: n = 247, 20, 23, 26, 20, 48 and 126), where n refers to the number of participants. P values were calculated using a two-sided Mann–Whitney–Wilcoxon test. P values are indicated within the plot. **e**, Comparison of malignant program mean $\log_{10}$ expression between *TP53*^{WT} and different functional *TP53*^{mut} Impactful categories in TCGA tumor samples (from left to right: n = 247, 68, 48, 25, 26, 14 and 82), where n refers to the number of participants. P values were calculated using a two-sided Mann–Whitney–Wilcoxon test. P values are indicated within the plot. **f**, Comparison of malignant program mean expression between *TP53*^{WT} and different *TP53*^{mut} Impactful categories in A549 cells (from left to right: n = 356, 19,546, 10,077, 49,757, 971, 1,000, 517, 554 and 1,000), where n refers to the number of cells. The rightmost four violin plots show the effect of mutations also found in tumor samples from our cohort. P values were calculated using a two-sided Mann–Whitney–Wilcoxon test. P values are indicated within the plot. **g**, Comparison of malignant program scores and *Hif1a* expression across normal epithelial (n = 206; gray), K (n = 505; blue) and KP (n = 1,554; red) mouse model malignant cell scRNA-seq data, where n refers to the number of cells. P values were calculated using a two-sided Mann–Whitney–Wilcoxon test. P values are indicated within the plot. **h**, Left: distribution of entropy scores across nonmalignant epithelial cells (n = 1,796; gray), malignant *TP53*^{WT} cells (n = 976; blue) and malignant *TP53*^{mut} cells (n = 743; red) in the scRNA-seq data, where n refers to the number of cells. Middle: distribution of entropy scores across adjacent normal (n = 59; gray), *TP53*^{WT} (n = 247; blue) and *TP53*^{mut} (n = 263; red) bulk transcriptomic LUAD samples from TCGA, where n refers to the number of participants. Right: distribution of entropy scores across *TP53*^{WT} (n = 51; blue) and *TP53*^{mut} (n = 59; red) bulk proteomic LUAD samples from CPTAC, where n refers to the number of participants. P values were calculated using a two-sided Mann–Whitney–Wilcoxon test. P values are indicated within the plot. **i**, Distribution of entropy scores across normal epithelial (n = 206; gray), malignant K (n = 505; blue) and KP (n = 1,554; red) cells in K and KP mouse tumors, where n refers to the number of cells. P values were calculated using a two-sided Mann–Whitney–Wilcoxon test. P values are indicated within the plot. **j**, Distribution of entropy scores across adjacent normal, *TP53*^{WT}*KRAS*^{WT}*EGFR*^{WT}, *EGFR*^{mut} and different *KRAS*^{mut} and *TP53*^{mut} single and comutations in bulk RNA-seq LUAD tumor samples from TCGA (from left to right: n = 59, 108, 23, 69, 26, 14, 17, 144, 43 and 45), where n refers to the number of participants. Entropy for *KRAS*^{mut} variant (G12C, G12D and G12V) tumors is also shown separately. P values were calculated using a two-sided Mann–Whitney–Wilcoxon test. P values are indicated within the plot. **k**, Distribution of entropy scores across adjacent normal (gray), *TP53*^{WT} (blue) and different *TP53*^{mut} (reds) categories in bulk RNA-seq LUAD tumor samples from TCGA (left panel, from left to right: n = 59, 247, 20, 23, 26, 20, 48 and 126; right panel, from left to right: n = 59, 247, 68, 48, 25, 26, 14, 82), where n refers to the number of participants. P values were calculated using a two-sided Mann–Whitney–Wilcoxon test. P values are indicated within the plot. **l**, Heatmap of mean entropy score for malignant cell subsets in scRNA-seq data. **m**, Stacked bar plot of proportion of HPCS cells from K and KP mice, classified into different malignant subsets annotation from this study. **n**, Comparison of entropy scores between *TP53*^{WT} (n = 356 cells) and different *TP53*^{mut} (n = 517 cells subsampled within each variant category) impact categories in A549 cell lines. The rightmost four violins show the entropy distributions of mutations also found in our cohort. P values were calculated using a two-sided Mann–Whitney–Wilcoxon test. P values are indicated within the plot. Throughout this figure, box plot horizontal lines show the 25th, 50th (median) and 75th percentiles, with vertical whiskers extending to a maximum distance of 1.5× the interquartile range from the hinge. Data beyond the whisker ends are plotted individually. Nonsignificant P values are not shown.

https://doi.org/10.1038/s43018-025-01053-7



for normal lung epithelial cell markers[44,45] (Supplementary Table 4). *TP53*^{WT} LUAD malignant cells scored highly for AT2 markers (Fig. 2g), whereas cells from tumors with squamous, neuroendocrine or mucinous and colloid subtypes expressed the expected corresponding basal, neuroendocrine or goblet signatures[45]. Interestingly, AT2 scores for *TP53*^{mut} LUAD malignant cells were significantly lower than *TP53*^{WT} LUAD but were comparable to AT2 scores in the tumors from other NSCLC subtypes (Fig. 2g), suggesting a loss of alveolar cell identity, previously linked to LUAD progression[46].

Consistent cancer-intrinsic changes observed across different *TP53* variant and comutation classes

To understand the potential joint effect of *EGFR* and *KRAS* comutations on the observed *TP53*^{mut} LUAD cancer-intrinsic changes, we partitioned TCGA cohort into six subsets with distinct mutational statuses for these three genes (Fig. 3a). *TP53*^{mut} tumors with or without comutations in *EGFR* or *KRAS* showed significant decreases in *TP53* target and AT2-like scores and a significant increase in CC.G2/M score compared to *TP53*^{WT}-*KRAS*^{WT}*EGFR*^{WT} tumors, whereas *TP53*^{WT} tumors with either *EGFR* or *KRAS* mutations did not have such significant differences in these signatures. Interestingly, tumors with *TP53* mutations alone or in combination with *KRAS* but not with *EGFR* had significant increases in glycolysis. hypoxia and pEMT scores compared to *TP53*^{WT}*KRAS*^{WT}*EGFR*^{WT} tumors, suggesting that *EGFR* mutations counteracted the *TP53*-induced effects on these malignant expression programs (Fig. 3a), consistent with a previous report[47]. Multiple regression analysis also showed consistent cancer-intrinsic changes associated with *TP53* mutational status even when regressing out the effects of *EGFR*, *KRAS*, *STK11*, *KEAP1*, *RBM10* and *PTPRD* mutations, as well as tumor size, histology and pathological stage (Fig. 3b,c).

To investigate the potential impact of different *TP53* mutational variants on the expression of malignant gene programs, we partitioned TCGA tumor samples on the basis of dominant negative (DNE), loss of function (LOF) and impactful *TP53* variant classes (Supplementary Table 5) defined in previous functional studies[48,49]. Consistent with our tumor scRNA-seq analysis, decreases in *TP53* target and AT2-like scores and increases in CC.G2/M, glycolysis.hypoxia and pEMT scores were associated with TCGA LUAD tumors with different subcategories of *TP53* DNE and LOF mutations (Fig. 3d) or with *TP53* impactful mutation classes (Fig. 3e). The DNE, LOF and impactful II mutations gave rise to the most pronounced differences, where glycolysis.hypoxia and pEMT programs were significantly increased only in DNE, LOF and impactful II variant categories in TCGA (Fig. 3d,e). Moreover, many of these malignant cell-intrinsic changes were also observed in vitro in A549 LUAD cells (*TP53*^{WT}) overexpressing each of a broad spectrum of *TP53* mutational variants and profiled by Perturb-seq[49] (Fig. 3f). These trends held across all mutational categories, including all four *TP53* variants also found in our scRNA-seq cohort and one variant also found in the scRNA-seq validation compendium (Fig. 3f). As observed in TCGA data, changes in A549 LUAD cells malignant expression programs were most pronounced for impactful II TP53 mutations (Fig. 3f); however, we did not observe the same trends in A549 LUAD cells overexpressing different *KRAS* mutations (Extended Data Fig. 3a). Because we see consistent changes in malignant expression programs across a wide spectrum of *TP53* mutations, we reason that these phenotypic changes are associated with a loss of *TP53*^{WT} activity, rather than specific gain of function that arises from any specific mutation. In support of our observations, heterogeneous deletion events resulting in loss of *TP53* activity have been shown to result in homogeneous and deterministic patterns of genome evolution[50].

Lastly, we also observed consistent changes in expression associated with *TP53* mutations at the protein level. Analyzing bulk proteomic data from the Clinical Proteomic Tumor Analysis Consortium (CPTAC) showed consistent decreases in *TP53* target and AT2-like signature scores and an increase in CC.G2M signature scores (Extended Data Fig. 3b), where effects were again most pronounced for DNE, LOF and impactful II mutation categories (Extended Data Fig. 3c).

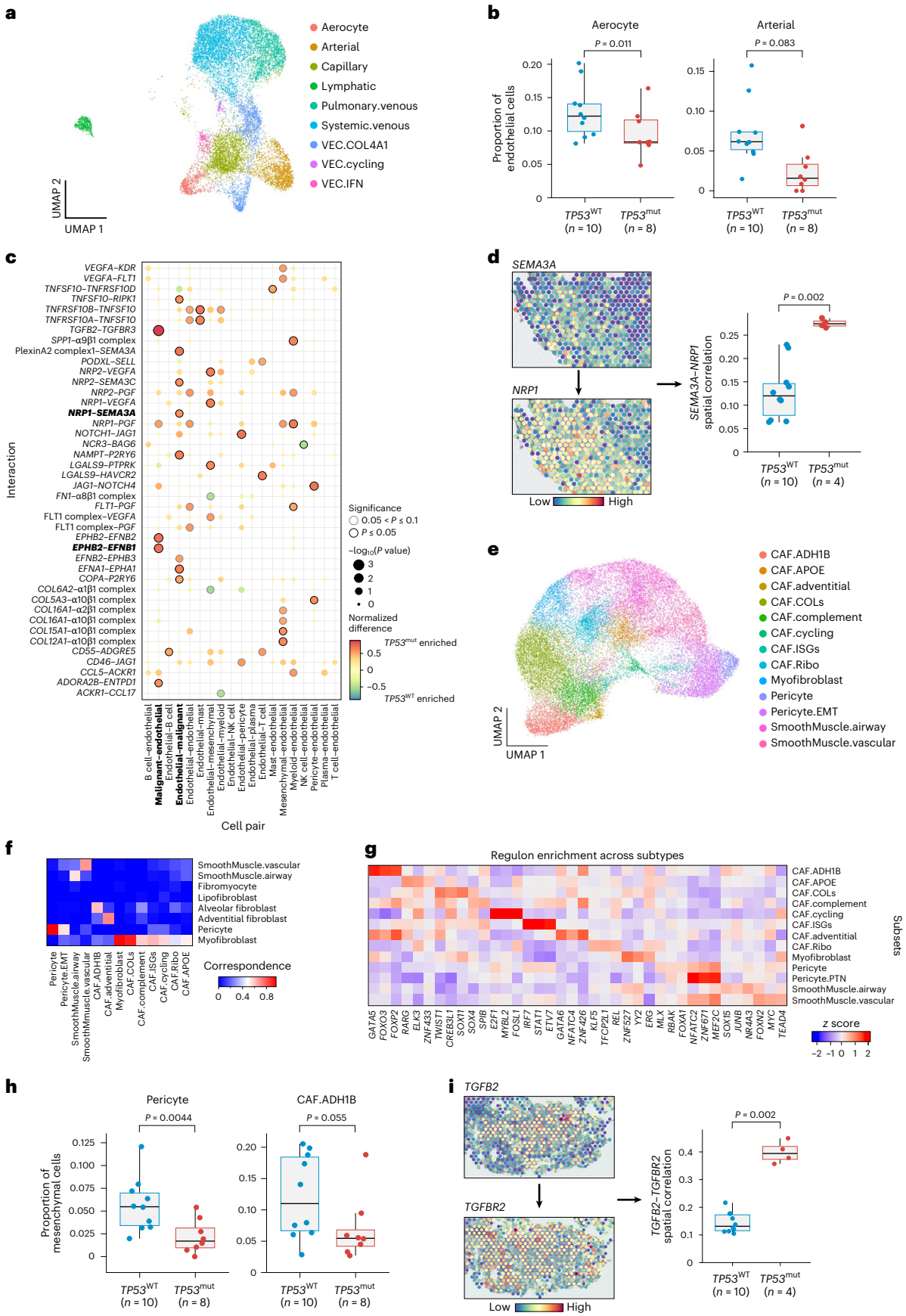
Conserved changes in signaling entropy and malignant programs in *TP53*^{mut} human tumors and mouse models

We previously reported loss of alveolar cell identity and reversion to a progenitor-like, highly plastic cell state (HPCS) during LUAD progression in mice with somatic activation of oncogenic *KRAS*^{G12D}, which was greater in tumors from mice where *p53* was deleted (KP model) than those with WT *p53* (K model)[51]. Consistent with our findings in human tumors, there was a significant decrease in AT2-like program score, and significant increases in CC.S, hypoxia and pEMT program scores when comparing scRNA-seq data from KP versus K mice malignant cells[51] (Fig. 3g). Moreover, the expression of key hypoxia transcription factor *Hif1a* was higher in KP versus K malignant cells (Fig. 3g, right). Interestingly, we did not observe significant increases in the hypoxia, glycolysis.hypoxia or pEMT programs between *TP53*^{mut} and *TP53*^{WT} A549 LUAD cells in vitro (Fig. 3f), suggesting that the TME is important for shaping these malignant programs in vivo.

The consistent loss in AT2 cell identity in *TP53*^{mut} LUAD human tumors, cell lines and mouse models prompted us to explore the consequences of this loss on signaling entropy[52], a representative measure of cellular plasticity. We found an increase in signaling entropy in *TP53*^{mut} versus *TP53*^{WT} LUAD malignant cells at the single-cell level in our cohort, which we confirmed in bulk RNA-seq and bulk proteomic profiles across a larger cohort size (Fig. 3h), as well as in the KP versus K mouse models[51] (Fig. 3i). This increase in entropy was most significant in *TP53*-only mutants and *TP53*–*KRAS* comutants but not in *TP53*–*EGFR* comutants in TCGA and most prominent in DNE, LOF and impactful II *TP53*^{mut} variant classes (Fig. 3j,k), consistent with the trends observed in the malignant program activity changes (Fig. 3a). Cell-cycle and pEMT programs

Fig. 4 | Spatially resolved interactions linked to vascular depletion and stromal remodeling in *TP53*^{mut} LUAD. a, UMAP of 15,320 endothelial cells, integrated across samples and colored by annotated subset. **b**, Comparison of aerocyte (left) and arterial (right) cell subset proportions in *TP53*^{WT} ($n = 10$; blue) and *TP53*^{mut} ($n = 8$; red) LUAD samples, where n refers to the number of participants. P values (from left to right: 0.011 and 0.083) were calculated using a two-sided Mann–Whitney–Wilcoxon test. **c**, Dot plot of differential ligand–receptor (rows) interactions between endothelial cells and other cell classes (columns). Red indicates enrichment in *TP53*^{mut} and blue indicates enrichment in *TP53*^{WT} tumor samples. The black outline indicates $P \leq 0.05$, as assessed by two-sided two-proportion Z-test. Interaction pairs (y axis) and corresponding cell pairs (x axis) discussed in the paper are in bold. **d**, Left: spatial expression of *SEMA3A* (ligand) and *NRP1* (receptor) in a representative *TP53*^{mut} ST sample. Right: box plot comparing spatial correlation of *SEMA3A* and *NRP1* expression between *TP53*^{WT} ($n = 10$) and *TP53*^{mut} ($n = 4$) tumor sections, where n refers to the number of ST samples. The P value (0.002) was calculated using a two-sided Mann–Whitney–Wilcoxon test. **e**, UMAP of 32,717 mesenchymal cells integrated across samples and colored by annotated subset. **f**, Confusion matrix of correspondence between our de novo mesenchymal subset annotations (x axis) and predicted identities from a classifier trained on annotated scRNA-seq data from normal lung mesenchymal cells (y axis). **g**, The z scores of the most specific SCENIC regulon (x axis) activity for each mesenchymal subset. **h**, Comparison of selected mesenchymal subset cell proportions *TP53*^{WT} ($n = 10$; blue) and *TP53*^{mut} ($n = 8$; red) LUAD samples, where n refers to the number of participants. P values (from left to right: 0.0044 and 0.055) were calculated using a two-sided Mann–Whitney–Wilcoxon test. **i**, Left: spatial expression of *TGFB2* (ligand) and *TGFBR2* (receptor) in a representative *TP53*^{mut} ST sample. Right: box plot comparing spatial correlation of *TGFB2* and *TGFBR2* expression between *TP53*^{WT} ($n = 10$) and *TP53*^{mut} ($n = 4$) tumor sections, where n refers to the number of ST samples. P value (0.002) was calculated using a two-sided Mann–Whitney–Wilcoxon test. Throughout this figure, box plot horizontal lines show the 25th, 50th (median) and 75th percentiles, with vertical whiskers extending to a maximum distance of 1.5× the interquartile range from the hinge. Data beyond the whisker ends are plotted individually.

https://doi.org/10.1038/s43018-025-01053-7



displayed the highest entropy among human malignant subsets and the greatest similarity to the HPCS cell state described in the KP mouse model (Fig. 3l,m). Moreover, in A549 cells overexpressing mutant *TP53*, there was significantly increased entropy in cells harboring impactful I and impactful II *TP53* mutations, including variants found in our scRNA-seq and validation cohorts (Fig. 3n). Conversely, there was no significant change in entropy in the A549 cells overexpressing different classes of *KRAS* mutants (Extended Data Fig. 3d), suggesting that the observed increase in entropy is directly linked to *TP53* mutations. Taken together, malignant cells from *TP53*^{WT} LUAD tumors were enriched for antigen presentation and AT2-like cell programs, whereas those from *TP53*^{mut} LUAD tumors were highly entropic and plastic and associated with programs predictive of poor outcome, including cell cycle, glycolysis, hypoxia and pEMT.

Stromal interactions that inhibit vascularization and promote EMT are enriched in *TP53*^{mut} LUAD

We next investigated the potential mechanisms for the observed decreased vascularization in *TP53*^{mut} LUAD tumors (Fig. 1f), which was most significant in DNE and LOF *TP53* mutations and independent of *KRAS* and *EGFR* comutations in TCGA (Extended Data Fig. 4a). We annotated nine endothelial cell subsets on the basis of the expression of established markers[44,45], including aerocytes (expressing *HPGD* and *EDNRB*), arterial (*GJA5* and *ENPP2*), capillary (*CA4* and *FCN3*), vascular endothelial cells (VECs), *COL4A1*$^{+}$ VECs (VEC.COL4A1), cycling VECs (VEC.cycling), IFN-stimulated VECs (VEC.IFN) and lymphatic, pulmonary venous and systemic venous cells (Fig. 4a and Extended Data Fig. 4b–d). The proportion of aerocytes and arterial cells was decreased in *TP53*^{mut} versus *TP53*^{WT} LUAD relative to all endothelial cells (Fig. 4b), which was further supported by bulk RNA-seq data deconvolved using scRNA-seq-derived markers (Extended Data Fig. 4e). This suggests a corresponding decrease in gas exchange in *TP53*^{mut} LUAD and a resulting decrease in oxygen diffusion and increased hypoxia at the tumor site.

Analyzing ligand–receptor pairs, several interactions between malignant and endothelial cells known to inhibit endothelial cell growth and function were enriched in *TP53*^{mut} LUAD tumors, including *SEMA3A–NRP1* and *EPHB2–EFNB1* (refs. 53–55) (Fig. 4c and Extended Data Fig. 4f). ST from matched tumor samples supports these observations, showing a significantly higher spatial correlation in the expression of *SEMA3A–NRP1* and *EPHB2–EFNB1* ligand–receptor pairs in *TP53*^{mut} versus *TP53*^{WT} LUAD sections (Fig. 4d and Extended Data Fig. 4g). Taken together, we observe a spatially correlated expression of known inhibitory interactions enriched between malignant and endothelial cells in *TP53*^{mut} LUAD, which may explain the accompanying decrease in vascularization.

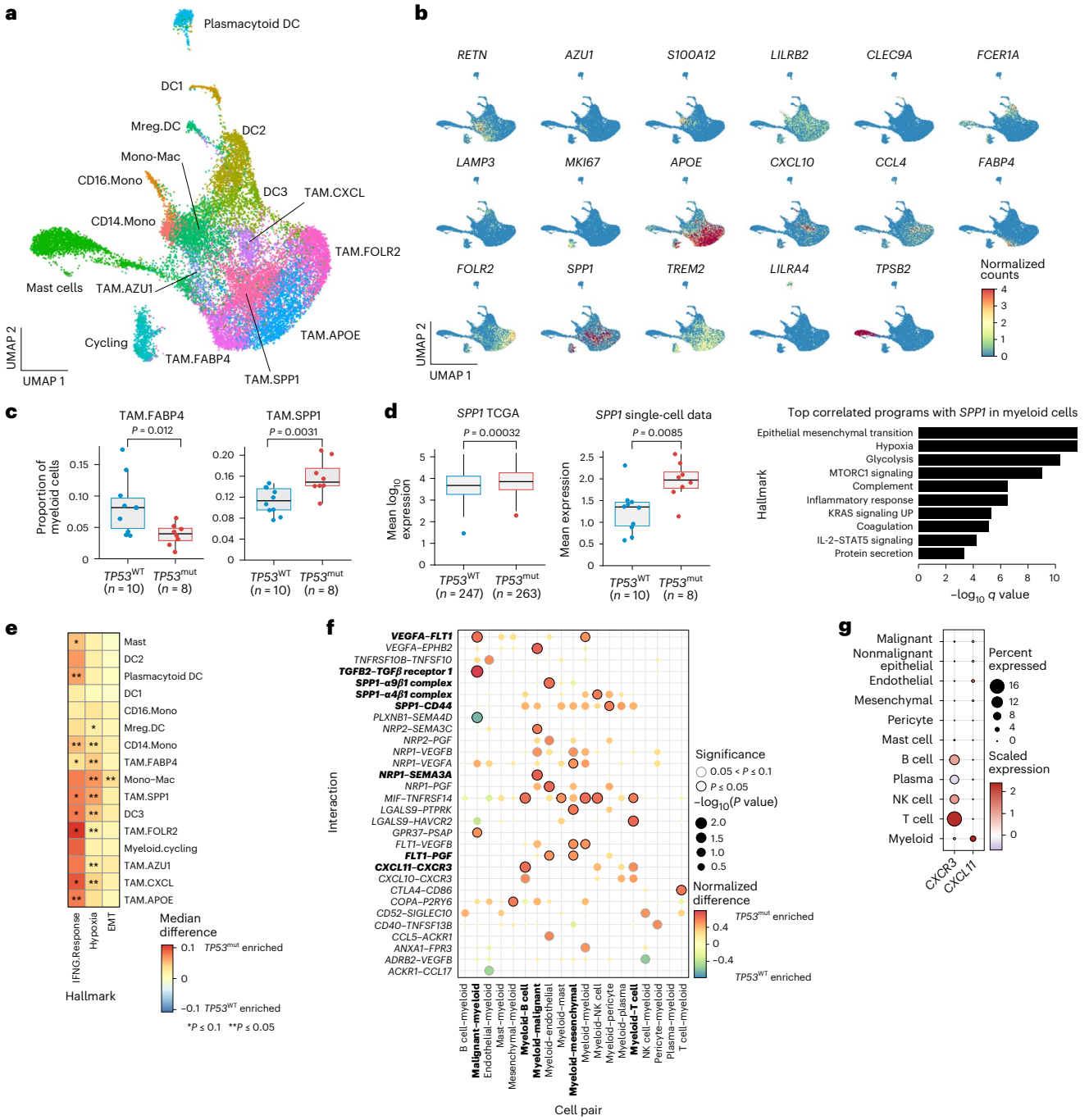
To further characterize stromal cell heterogeneity in the context of *TP53*^{mut} LUAD, we partitioned mesenchymal cells in the scRNA-seq data into 13 subsets (Fig. 4e and Extended Data Fig. 5a–c). Two subsets most closely resembled normal lung myofibroblasts[45] (Fig. 4f): cancer associated fibroblasts (CAFs) expressing high levels of collagens (*COL1A2* and *COL3A1*; CAF.COLs), which share markers with previously described *LRRC15*$^{+}$ myofibroblasts[56], and myofibroblasts expressing both fibroblast and smooth muscle cell markers. We also annotated two pericyte subsets, one expressing canonical pericyte markers (*HIGD1B* and *COX4I2*) and another expressing *COL4A2* and EMT-promoting genes, as well as airway and vascular smooth muscle subsets, alveolar (CAF.ADH1B) and adventitial (CAF.adventitial) fibroblasts, which correspond to human lung alveolar and adventitial fibroblasts[56], and fibroblast subsets expressing higher levels of lipid metabolism (CAF.APOE), ribosomal (CAF.Ribo), complement and chemotaxis (CAF.complement) and IFN-stimulated (CAF.ISGs) genes (Extended Data Fig. 5a,b). The different mesenchymal subsets were associated with different active regulons (Fig. 4g), as inferred by pySCENIC[57]. As expected, CAF.ISGs activated the *IRF7* and *STAT1* regulons, controlling IFN-responsive genes; meanwhile, CAF.COLs activated a *TWIST1* regulon, previously linked to the transdifferentiation of normal quiescent fibroblasts to CAFs, as well as EMT in cancer cells[58] (Extended Data Fig. 5d). Comparing our CAF subsets to pan-cancer CAF subsets defined previously[59] showed that CAF.COLs most closely matched the CAF_{infla} population (Extended Data Fig. 5e) involved in cytokine-mediated immune modulation and ECM remodeling[59].

According to the inferred composition of CAF subsets in ST spots across sections, pericytes spatially correlated with endothelial cells (Extended Data Fig. 5f), CAF.COLs correlated with vascular smooth muscle and CAF.ADH1B, CAF.complement correlated with CAF.adventitial, pericytes correlated with CAF.ISGs and airway smooth muscle cells correlated with CAF.APOE and CAF.Ribo. Spatially, most mesenchymal subsets were negatively correlated with malignant cells, indicating exclusion of fibroblasts from the tumor core. Comparing the relative proportions of mesenchymal cell subsets (among all mesenchymal cells) between *TP53*^{mut} versus *TP53*^{WT} LUAD, pericytes and CAF.ADH1B were decreased in *TP53*^{mut} LUAD (Fig. 4h), which was also supported by deconvolved bulk RNA-seq data from TCGA (Fig. 1f and Extended Data Fig. 5g). Similar to endothelial scores, pericyte scores were significantly decreased in *TP53*^{mut} tumors in TCGA LUAD, with or without comutations in *KRAS* and *EGFR*, and this decrease was most significant in the DNE and LOF class of *TP53* mutations (Extended Data Fig. 5h).

TGFB2–TGFBR2 had significantly higher spatial correlation in *TP53*^{mut} versus *TP53*^{WT} LUAD (Fig. 4i) and *TGFB2–TGFBR2* and *TGFBR2–TGFB3* interactions were enriched in *TP53*^{mut} versus *TP53*^{WT} LUAD

Fig. 5 | *SPP1* and other immunomodulatory factors are enriched in *TP53*^{mut} LUAD myeloid compartment. a, UMAP of 21,125 myeloid cells integrated across samples and colored by annotated subset. **b**, Feature plot showing the expression of representative myeloid subset markers. **c**, Comparison of selected myeloid subset cell proportions in *TP53*^{WT} ($n = 10$; blue) and *TP53*^{mut} ($n = 8$; red) LUAD samples, where n refers to the number of participants. P values (from left to right: 0.012 and 0.0031) were calculated using a two-sided Mann–Whitney–Wilcoxon test. **d**, Left: comparison of *SPP1* expression between *TP53*^{WT} ($n = 247$ and 10) and *TP53*^{mut} ($n = 263$ and 8) tumors in TCGA bulk RNA-seq (left) and our scRNA-seq (middle) data, where n refers to the number of participants. P values were calculated using a two-tailed multiple regression t-test (TCGA, left; $P = 0.00032$) or a two-sided Mann–Whitney–Wilcoxon test (scRNA-seq, middle; $P = 0.0085$). Right: horizontal bar plot of negative $\log_{10}$ q values from gene set enrichment analysis of the 100 most positively correlated genes with *SPP1* in myeloid cells. **e**, Differential enrichment of hallmark IFNγ response, hypoxia and EMT module score for each myeloid subset, comparing mean module scores across *TP53*^{mut} versus *TP53*^{WT} tumor samples. Red and blue indicate enrichment in *TP53*^{mut} and *TP53*^{WT} tumor samples, respectively. P values were calculated using a two-sided Mann–Whitney–Wilcoxon test. **f**, Dot plot of differential ligand–receptor (rows) interactions between myeloid cells and other cell classes (columns). Red indicates enrichment in *TP53*^{mut} and blue enrichment in *TP53*^{WT} tumor samples. The black outline indicates $P \leq 0.05$, as assessed by two-sided two-proportion Z-test. Interaction pairs (y axis) and corresponding cell pairs (x axis) discussed in the paper are in bold. **g**, Dot plot of the expression of *CXCR3* and *CXCL11* across broad cell class annotations in our scRNA-seq data. **h**, Differential expression analysis of ligand and receptor genes from the interactions displayed in **f**, comparing pseudobulk myeloid subset profiles for *TP53*^{mut} versus *TP53*^{WT} tumors in scRNA-seq data. P values were calculated using a two-sided Mann–Whitney–Wilcoxon test. Red indicates enrichment in *TP53*^{mut} and blue indicates enrichment in *TP53*^{WT} tumor samples. **i**, Left: spatial expression of *SPP1* (ligand) and *CD44* (receptor) in a representative *TP53*^{mut} ST sample. Right: box plot comparing spatial correlation of *SPP1* and *CD44* expression between *TP53*^{WT} ($n = 10$) and *TP53*^{mut} ($n = 4$) tumor sections, where n refers to the number of ST samples. The P value (0.002) was calculated using a two-sided Mann–Whitney–Wilcoxon test. Throughout this figure, box plot horizontal lines show the 25th, 50th (median) and 75th percentiles, with vertical whiskers extending to a maximum distance of 1.5× the interquartile range from the hinge. Data beyond the whisker ends are plotted individually.

https://doi.org/10.1038/s43018-025-01053-7



between malignant and mesenchymal cells (Extended Data Fig. 5i), although the enrichment was not significant. Transforming growth factor-β signaling is a well-known inducer of EMT in cancer cells[60], which is known to be mediated by CAFs found in the TME[61]. Moreover, *VEGFA* expression in myofibroblasts is higher in *TP53*^{mut} versus *TP53*^{WT} LUAD (Extended Data Fig. 5j, bottom), consistent with previous reports that myofibroblasts can regulate angiogenesis in response to hypoxia[62]. Overall, enriched stromal cell compositions, interactions and spatial colocalization suggest distinct TME mechanisms by which *TP53*^{mut} LUAD tumors can induce hypoxia and EMT, thereby contributing to a more aggressive tumor phenotype and poorer clinical outcomes.

Increased *SPP1* expression and immunomodulatory interactions in *TP53*^{mut} LUAD myeloid cells

To investigate the impact of *TP53* mutations on tumor-associated macrophages (TAMs), we de novo identified 16 myeloid cell subsets, largely consistent with previous NSCLC atlases[24] (Fig. 5a,b and Extended Data Fig. 6a). We found a significant decrease in the proportion of alveolar macrophage-like *FABP4*-expressing TAMs (TAM.FABP4) and an increase in *SPP1*-expressing TAMs (TAM.SPP1) in *TP53*^{mut} LUAD tumors (Fig. 5c). In agreement, *SPP1* expression was significantly increased in TCGA bulk RNA-seq data and more prominently in myeloid pseudobulk comparisons between *TP53*^{mut} and *TP53*^{WT} LUAD tumors (Fig. 5d, left). The proportion of *CXCL9/10/11*-expressing TAMs (TAM.CXCLs) was also higher in *TP53*^{mut} LUAD tumors both in the scRNA-seq data and in deconvolution analysis of TCGA bulk RNA-seq data (Extended Data Fig. 6b). Consistently, expression of cytokines *CXCL9/10/11* was also increased in *TP53*^{mut} bulk TCGA LUAD tumors (Extended Data Fig. 6c). Notably, increases in *SPP1* and *CXCL9/10/11* were most significant in *TP53*^{mut} tumors without comutations in *EGFR* or *KRAS* and in DNE, LOF and impactful II classes of *TP53*^{mut} tumors in TCGA (Extended Data Fig. 6d,e).

Myeloid cell subsets in *TP53*^{mut} and *TP53*^{WT} LUAD were also associated with distinct gene expression programs, where IFNγ response, hypoxia and EMT hallmark pathways were differentially upregulated between myeloid subsets from *TP53*^{mut} versus *TP53*^{WT} tumors, especially in TAM.CXCL and TAM.SPP1 cells (Fig. 5e). *SPP1* expression has been previously associated with EMT and early lymph node metastasis in LUAD[63], as well as with shorter overall survival in TCGA LUAD independent of *TP53* mutational status (Extended Data Fig. 6f). Notably, the genes most correlated with *SPP1* expression across single myeloid cells were enriched for EMT, hypoxia and glycolysis functions (Fig. 5d, right) and included VEGF receptors *NRP1* and *FLT1* (Extended Data Fig. 6g). In summary, we find increased expression of *SPP1* in *TP53*^{mut} LUAD myeloid cells that is tightly linked with induction of genes involved in EMT and hypoxia programs.

To identify ligand–receptor interactions that may contribute to a distinct myeloid compartment in *TP53*^{mut} LUAD, we again conducted differential cell–cell interaction analysis. Putative interactions enriched in *TP53*^{mut} LUAD include *VEGFA–FLT1* and *NRP1–SEMA3A* (Fig. 5f), involved in recruitment of myeloid cells to tumor sites[64], and *TGFB2–TGFBR1* interactions between malignant and myeloid cells, which are known to promote monocyte differentiation into TAMs[65] and may help explain the relative increase in the monocyte-derived macrophage population in *TP53*^{mut} LUAD. In addition, *CXCL11–CXCR3* interactions between myeloid and B or T cells were enriched in *TP53*^{mut} LUAD (Fig. 5f) with *CXCR3* specifically expressed by natural killer (NK), B and T cells (Fig. 5g). Moreover, *CXCL11* was upregulated in multiple myeloid subsets in *TP53*^{mut} LUAD (Fig. 5h) and expressed most highly in TAM.CXCLs myeloid cells (Extended Data Fig. 6h). Lastly, there was a significant increase in putative *SPP1*-mediated interactions between myeloid and endothelial cells or pericytes in *TP53*^{mut} LUAD (Fig. 5f), also supported by increased spatial correlation of *SPP1* and *CD44* expression in *TP53*^{mut} tumors (Fig. 5i). Taken together, *TP53*^{mut} LUAD has increased expression of putative interactions involved in monocyte and lymphocyte recruitment, TAM differentiation, hypoxia, EMT and other downstream effector functions through *SPP1*, *CXCL11* and other TAM-associated immunomodulators.

Enrichment of T cell exhaustion and immune checkpoint interactions in *TP53*^{mut} LUAD

Lymphoid cells in the scRNA-seq data partitioned into 13 subsets of T and NK cells (Fig. 6a,b and Extended Data Fig. 7a) and five subsets of B cells, including cycling, follicular and marginal zone B cells and IgA and IgG plasma cells (Extended Data Fig. 7b). Among T and NK cells, *TP53*^{mut} LUAD had a decreased relative proportion of *CD4*$^{+}$ tissue-resident memory T cells (CD4.TRM) and an increased proportion of pre-exhausted-like[66], *GZMK*-expressing *CD8*$^{+}$ (CD8.GZMK) and exhausted-like (T.Exhausted) T cells (Fig. 6c), which was supported by our scRNA-seq validation cohort (Extended Data Fig. 7c and Supplementary Table 3) and by deconvolved bulk RNA-seq data from TCGA primary tumors (Extended Data Fig. 7d). Higher abundance of exhausted T cells may be mediated by increased chemotaxis through *CXCL11–CXCR3* interactions (Fig. 5f–h), where *CXCR3* was upregulated in *TP53*^{mut} exhausted T cells (Extended Data Fig. 7e, right). We also observed a higher proportion of T follicular helper (TFH) cells relative to all T and NK cells in *TP53*^{mut} versus *TP53*^{WT} tumors in both our discovery and validation scRNA-seq cohorts (Extended Data Fig. 7c), which was statistically significant when combining all samples. Expression of *CXCL13*, a top predictor for response to checkpoint inhibition therapy[67], was most highly expressed in exhausted-like and TFH subsets and significantly increased in *TP53*^{mut} tumors (Extended Data Fig. 7e). Furthermore, expression of immune checkpoint molecules *PDCD1*, *CTLA4*, *HAVCR2* and *TIGIT* were significantly higher across multiple T cell subsets in *TP53*^{mut} versus *TP53*^{WT} tumors, whereas expression of *KLRB1*, a marker of favorable prognosis in cancer[68], was downregulated (Extended Data Fig. 7e, right). Taken together, increased expression of *CXCL13* and immune checkpoint molecules (*PDCD1*, *CTLA4*, *HAVCR2* and *TIGIT*) and proportions of TFH, *GZMK*-expressing *CD8*$^{+}$ and exhausted-like T cells all suggest a heightened immunogenic potential in the *TP53*^{mut} LUAD TME that may contribute to a more favorable ICI immunotherapy response[67,69–71].

Ligand–receptor analysis of putative immune checkpoint interactions[72] among T, myeloid and malignant cells showed an enrichment in *PDCD1–CD274* and *TIGIT–PVR* putative interactions between T and malignant cells in *TP53*^{mut} versus *TP53*^{WT} LUAD (Fig. 6d), as well as *CTLA4–CD86* and *HAVCR2–LGALS9* putative interactions between T and myeloid cells. In agreement, we found an enrichment in TP53^{mut} versus TP53^{WT} LUAD tumors of *CD274*, *CD86*, *PVR*, *PDCD1*, *CTLA4* and *TIGIT* expression in bulk RNA-seq data from TCGA (Extended Data Fig. 7f), of PVR and CD274 protein levels in bulk proteomic data from CPTAC (Extended Data Fig. 7g) and of *PDCD1*, *CTLA4* and *TIGIT* RNA expression in multiple T cell subsets in our scRNA-seq data (Extended Data Fig. 7e, right), where *TIGIT* was the most frequently expressed across T cells compared to *CTLA4* and *PDCD1* (Fig. 6e). *PVR* expression was also higher in malignant cells from *TP53*^{mut} versus *TP53*^{WT} tumor scRNA-seq data and highly correlated with the expression of the cell-cycle program in malignant cells (Fig. 6f), suggesting that *PVR* may be activated during cancer cell proliferation. Furthermore, there was significantly higher spatial colocalization of *TIGIT* and *PVR* expression in *TP53*^{mut} versus *TP53*^{WT} LUAD tumor samples (Fig. 6g). In agreement, there was also a higher proportion of *PVR*$^{+}$ malignant cells in *TP53*^{mut} versus *TP53*^{WT} tumors and even more prominently in cycling malignant cells (Fig. 6h).

To investigate the *TP53*^{mut} enrichment in interactions between PDL1 (*CD274*) and PD1 (*PDCD1*) at a single-cell resolution, we applied mIF to participant-matched *TP53*^{mut} and *TP53*^{WT} formalin-fixed paraffin-embedded (FFPE) slides from our cohort (Fig. 7a and Extended Data Fig. 8a–e). There was a significant increase in log count

https://doi.org/10.1038/s43018-025-01053-7

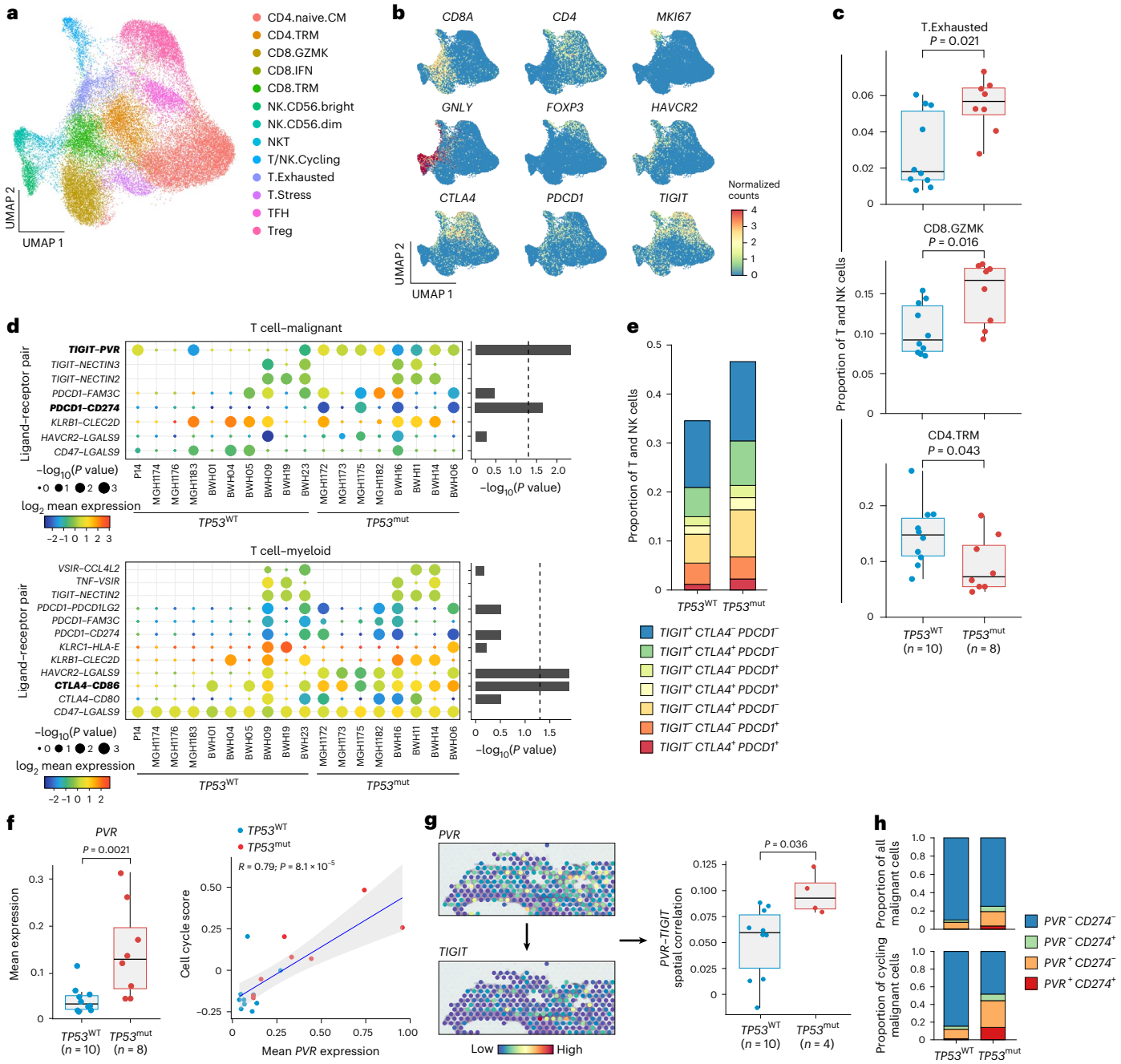


Fig. 6 | Immune checkpoint interactions and exhausted-like lymphoid cells are enriched in *TP53*^{mut} LUAD. a, UMAP of 40,293 T and NK cells integrated across samples and colored by annotated subset. **b**, Feature plot showing the expression of nine representative T and NK cell subset markers. **c**, Comparison of T.Exhausted (top), CD8.GZMK (middle) and CD4.TRM (bottom) T cell subset proportions relative to all T and NK cells in *TP53*^{WT} ($n = 10$; blue) and *TP53*^{mut} ($n = 8$; red) LUAD samples, where n refers to the number of participants. P values (from top to bottom: 0.021, 0.016 and 0.043) were calculated using a two-sided Mann–Whitney–Wilcoxon test. **d**, Left: the $\log_2$ mean gene expression and predicted significance of ICI-targetable ligand–receptor interactions between T cells and malignant cells (top) or T cells and myeloid (bottom) for each tumor. Interaction pairs (y axis) discussed in the paper are in bold. Right: horizontal bar plots showing the negative $\log_{10}$ P value (Fisher's exact test) of the enrichment in interactions between *TP53*^{mut} and *TP53*^{WT} tumors. Vertical dashed lines indicate a P value of 0.05. **e**, Stacked bar plots depicting proportion of T and NK cells expressing different combinations of immune checkpoint molecules *TIGIT*, *CTLA4* and *PDCD1* in *TP53*^{WT} (left) and *TP53*^{mut} (right) tumors. **f**, Left: box plot of mean *PVR* expression in *TP53*^{WT} ($n = 10$; blue) and *TP53*^{mut} ($n = 8$; red) sample malignant cells, where n refers to the number of participants. The P value (0.0021) was calculated using a two-sided Mann–Whitney–Wilcoxon test. Right: correlation of cell-cycle score versus mean *PVR* expression in malignant cells for each participant, where blue represents *TP53*^{WT} and red represents *TP53*^{mut} tumors. The gray band indicates 95% confidence intervals around the linear model fit (blue line). **g**, Left: spatial expression of *PVR* (ligand) and *TIGIT* (receptor) in a representative *TP53*^{mut} ST sample. Right: box plot comparing spatial correlation of *PVR* and *TIGIT* expression across ST spots at the tumor periphery (spots containing ≤50% malignant cells) between *TP53*^{WT} ($n = 10$) and *TP53*^{mut} ($n = 4$) tumor sections, where n refers to the number of ST samples. The P value (0.036) was calculated using a two-sided Mann–Whitney–Wilcoxon test. **h**, Stacked bar plots depicting the proportion of malignant cells expressing different combination of *PVR* and *CD274* among all malignant cells (top) and among all cycling malignant cells (bottom) in *TP53*^{WT} (left) and *TP53*^{mut} (right) tumors in scRNA-seq data. Among all malignant cells (top), $n = 22{,}373$ came from *TP53*^{WT} tumors and $n = 7{,}560$ came from *TP53*^{mut} tumors, whereas, among all cycling malignant cells (bottom), $n = 1{,}454$ came from *TP53*^{WT} tumors and $n = 713$ came from *TP53*^{mut} tumors, where n refers to the number of cells. Throughout this figure, box plot horizontal lines show the 25th, 50th (median) and 75th percentiles, with vertical whiskers extending to a maximum distance of 1.5× the interquartile range from the hinge. Data beyond the whisker ends are plotted individually.

https://doi.org/10.1038/s43018-025-01053-7

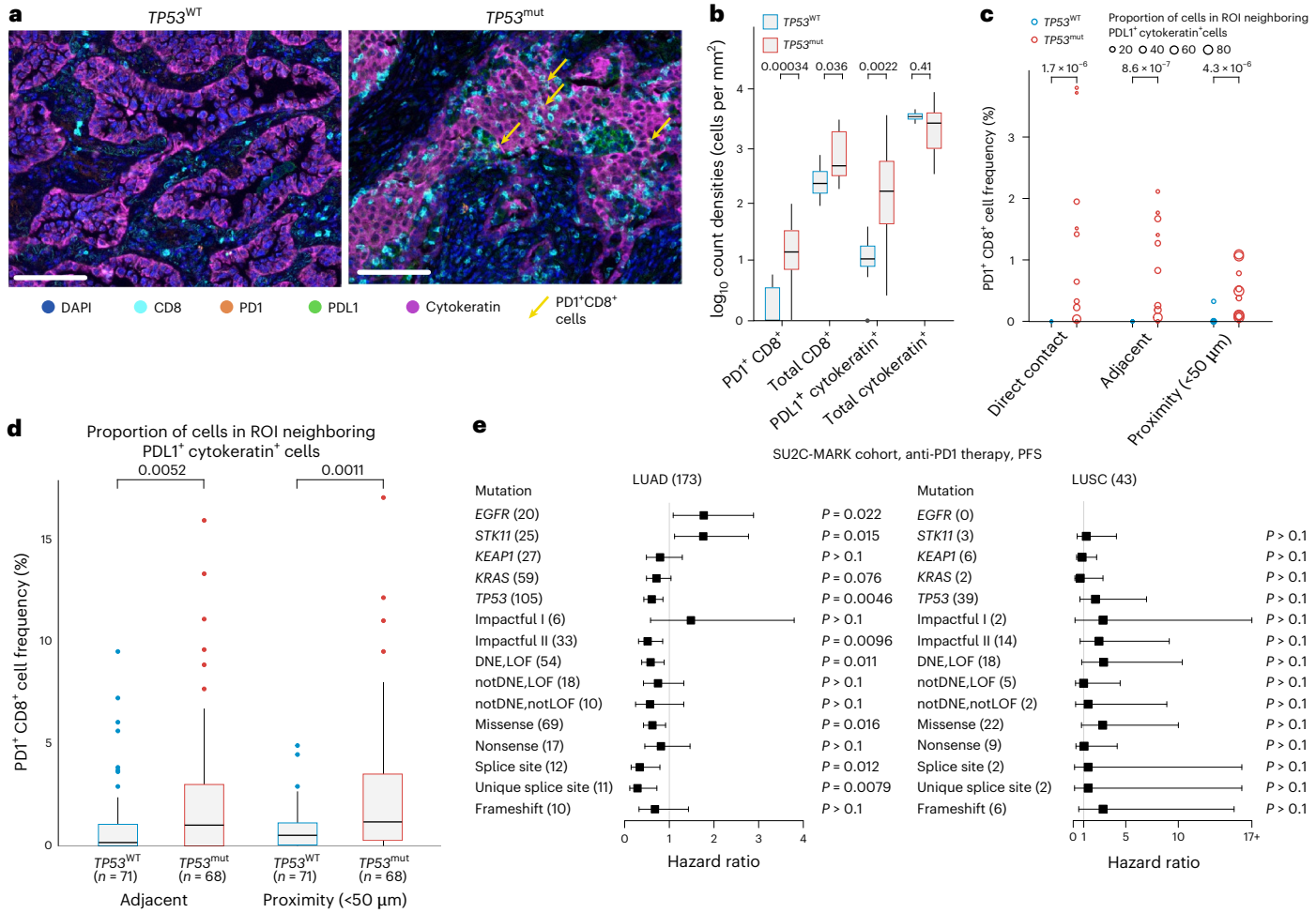


Fig. 7 | Increased colocalization of PD1$^+$CD8$^+$ cells with PDL1$^+$cytokeratin$^+$ cells and improved response to anti-PD1 therapy in *TP53*^{mut} LUAD tumors. **a**, Representative mIF images of *TP53*^{WT} (left; BWH01) and *TP53*^{mut} (right; BWH06) tumors. Yellow arrows correspond to PD1$^+$CD8$^+$ cells. Scale bar, 100 µm. **b**, Box plots showing the $\log_{10}$ count density (cells per mm^2) of PD1$^+$CD8$^+$, total CD8$^+$, PDL1$^+$cytokeratin$^+$ and total cytokeratin$^+$ cells across ROIs in *TP53*^{WT} (BWH01 and BWH04, $n = 11$; blue) and *TP53*^{mut} (BWH06 and BWH11, $n = 12$; red) tumor samples used for mIF, where n refers to the number of ROIs. P values (from left to right: 0.00034, 0.036, 0.0022 and 0.41) were calculated using a two-sided Mann–Whitney–Wilcoxon test. **c**, Grouped scatter plot showing the proportion of PD1$^+$CD8$^+$ cells among all cells in proximity to PDL1$^+$cytokeratin$^+$. Each circle represents an ROI from *TP53*^{WT} (BWH01 and BWH04; blue) versus *TP53*^{mut} (BWH06 and BWH11; red) mIF tumor stains, where the size of the circle represents the proportion of PDL1$^+$cytokeratin$^+$-neighboring cells, among all cells within the ROI. For *TP53*^{WT} mIF tumor stains, $n = 16$, 19 and 21 for direct contact, adjacent and proximity analyses, respectively, whereas, for *TP53*^{mut} mIF tumor stains, $n = 22$, 22 and 24 direct contact, adjacent and proximity analyses, respectively, where n refers to the number of ROIs. P values (from left to right: 1.7×10^{-6}, 8.6×10^{-7} and 4.3×10^{-6}) were calculated using a two-sided Mann–Whitney–Wilcoxon test. **d**, Box plots showing the proportion of PD1$^+$CD8$^+$ cells among all cells in proximity to PDL1$^+$cytokeratin$^+$, between *TP53*^{WT} ($n = 71$) versus *TP53*^{mut} ($n = 68$) LUAD tumor mIF data from the external cohort, where n refers to the number of ROIs. P values (from left to right: 0.0052 and 0.0011) were calculated using a two-sided Mann–Whitney–Wilcoxon test. **e**, Association with PFS for participants with different mutations frequently encountered in LUAD (*EGFR*, *STK11*, *KEAP1*, *KRAS* and *TP53*, also split by classes of mutations) following anti-PD1 therapy in the SU2C-MARK cohort. The analysis was performed separately for participants with LUAD (left) and LUSC (right). Hazard ratios (squares, where lines indicate 95% confidence intervals) and P values were determined using univariate Cox regression models comparing each mutation category and its respective WT category. The numbers in brackets indicate the number of participants within a mutation category. Throughout this figure, box plot horizontal lines show the 25th, 50th (median) and 75th percentiles, with vertical whiskers extending to a maximum distance of 1.5× the interquartile range from the hinge. Data beyond the whisker ends are plotted individually.

density (cells per mm^2) of PDL1$^+$cytokeratin$^+$ cells, as well as PD1$^+$CD8$^+$ cells in *TP53*^{mut} versus *TP53*^{WT} tumor samples (Fig. 7b) and, consistent with the scRNA-seq cell–cell interaction inference (Fig. 6d), an increase in colocalization between PDL1$^+$cytokeratin$^+$ cells and PD1$^+$CD8$^+$ cells in *TP53*^{mut} versus *TP53*^{WT} tumor samples, across multiple regions of interest (ROIs) measuring direct contact, adjacency (six nearest cell neighbors) and proximity (<50 µm) (Fig. 7c). To investigate these observations across a larger independent cohort[73], we reanalyzed mIF data from 139 LUAD tissue ROIs and 28 participants with known *TP53* mutation status. We found increased colocalization of PDL1$^+$cytokeratin$^+$ cells and PD1$^+$CD8$^+$ cells in *TP53*^{mut} compared to *TP53*^{WT} tumor samples, quantified by adjacency and proximity (Fig. 7d), corroborating the results from our cohort. Lastly, using data from a recent study looking at molecular features underlying response to ICIs in advanced NSCLC[74], we found that *TP53* mutations were associated with longer progression-free survival (PFS) in participants treated with anti-PD1 immunotherapy (Fig. 7e) or any form of ICI single-agent or combination treatment (Extended Data Fig. 8f). The survival benefit was most significant for participants with impactful II, DNE and LOF classes of *TP53* mutations[74], while there was no significant association between *TP53* mutations and PFS in lung squamous cell carcinoma (LUSC) (Fig. 7e, Extended Data Fig. 8f).

Spatially defined fibroblast–macrophage niche enriched in hypoxia and EMT programs in *TP53*^{mut} LUAD

To comprehensively map intercompartment cellular colocalization using the spatial data, we deconvolved the proportion of cell subsets across all ST spots and calculated the average correlation of cell type proportions across all LUAD sections and their difference between *TP53*^{mut} and *TP53*^{WT} LUAD tumors (Fig. 8a and Extended Data Fig. 9a). Interestingly, there was a significantly lower colocalization (that is, more negative correlation) of endothelial cells with mesenchymal cells and with myeloid cells in *TP53*^{mut} versus *TP53*^{WT} LUAD (Fig. 8b), consistent with the increase in hypoxia related gene expression in myeloid subsets in *TP53*^{mut} tumors (Fig. 5e). In addition, spatial colocalization of plasma and follicular B cells with malignant cells was increased in *TP53*^{mut} LUAD (Extended Data Fig. 9b, bottom). B cell infiltration has been associated with increased PD1 and PDL1 expression, tumor mutational burden (TMB) and prolonged survival in NSCLC[75] and likely contributes to the heightened immunogenicity of *TP53*^{mut} LUAD tumors.

We further performed NMF of the spatial data to identify 15 multicellular niches (or factors) with varying participation of different cell subsets (Fig. 8c). Factor 7 (NMF7) contained multiple myeloid and mesenchymal subsets with known significance in tumor progression, including TAM.SPP1, CAF.COLs and myofibroblasts, which were highly colocalized with each other, as well as with hallmark hypoxia and EMT program expression (Fig. 8c,d). Spots high in NMF7 were low in malignant cell proportions and high in myeloid and mesenchymal cell proportions (Extended Data Fig. 9c). NMF7 was also more positively correlated with hallmark hypoxia and EMT program expression in the tumor periphery (spots with ≤50% malignant cells) and more negatively correlated with endothelial cells across all spots in *TP53*^{mut} versus *TP53*^{WT} LUAD (Fig. 8e). We confirmed the presence of SPP1^{+}CD68^{+} macrophage and α-smooth muscle actin (α-SMA)$^{+}$ fibroblast multicellular niche by mIF and found fewer CD31^{+} endothelial cells in proximity to SPP1^{+}CD68^{+} macrophages in *TP53*^{mut} versus *TP53*^{WT} samples in our cohort (Fig. 8f,g and Extended Data Fig. 10a,b), supporting our spatial data analysis. In sum, NMF7 represents a fibroblast–macrophage tissue niche that resides on the periphery of the tumor core and likely contributes to increased hypoxia and an early EMT phenotype in *TP53*^{mut} LUAD tumors, consistent with the known roles of CAFs and macrophages in promoting EMT in tumors[76,77].

Discussion

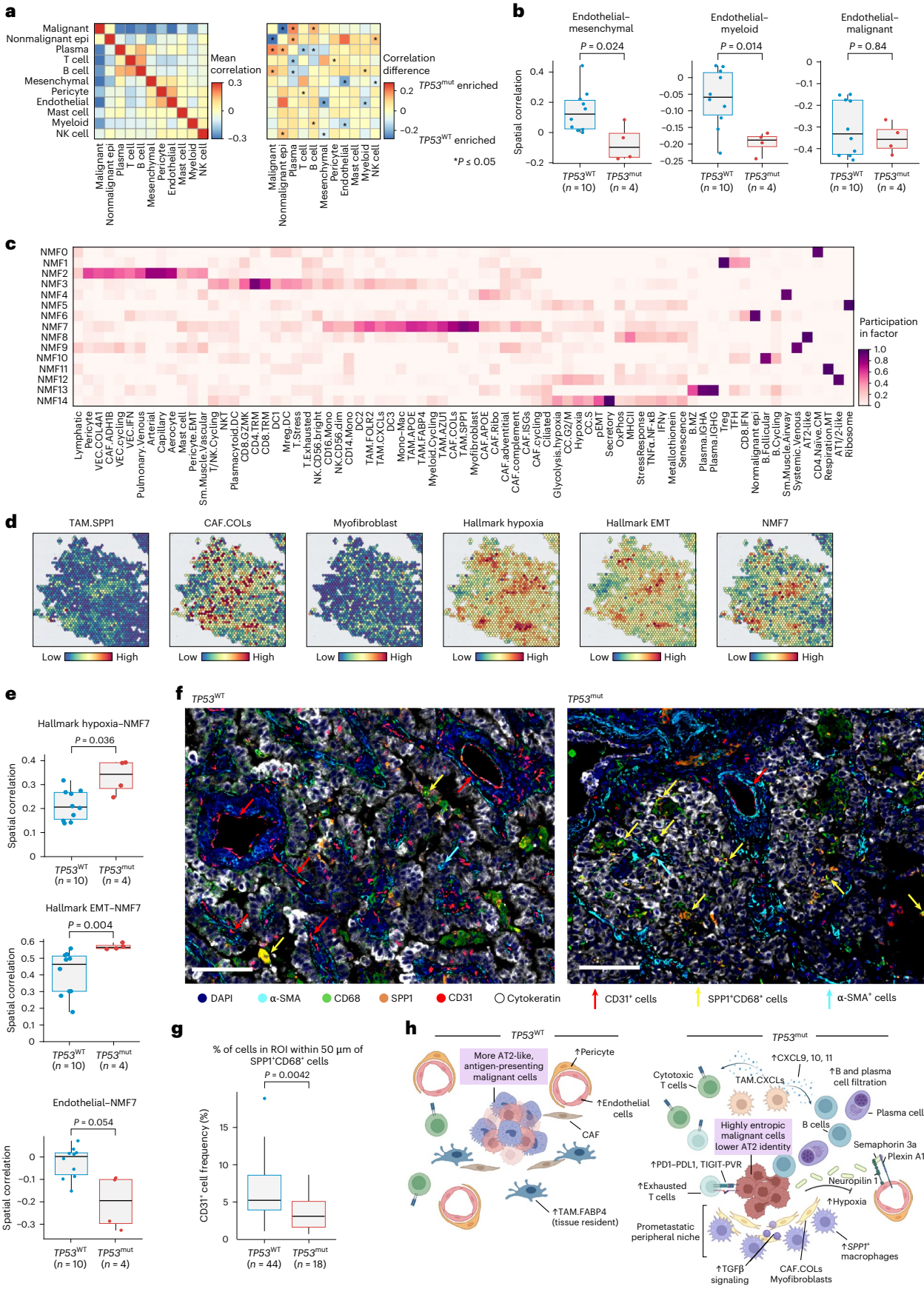
Targeted therapies for NSCLC are currently directed at oncogenic driver mutations and chromosomal rearrangements, yet there are currently no effective therapies targeting *TP53* mutations, despite being common in tumors and extensively studied. A recent landmark study showed that p53 suppresses tumor formation in *KRAS*-transformed AT2 cells by inducing AT2-to-AT1 cell differentiation in mouse models and describe a highly plastic transitional population that expands after *Trp53* loss[78]. Complementing this work, we hypothesized that cellular and spatial profiling of the TME in *TP53*^{mut} human LUAD tumors could further our understanding of how *TP53* mutations contribute to shorter survival and reveal new therapeutic opportunities. To this end, we systematically characterized the LUAD TME of *TP53*^{mut} versus *TP53*^{WT} tumors by combining scRNA-seq of 166,821 cells across 23 participants with NSCLC, 23 matched tumor and normal WES samples, 42 mIF samples and 20 ST tissue sections. Together, this multiomic dataset from matched NSCLC samples provides a valuable resource of TME cell annotations, markers, ligand–receptor interactions and spatial organization that will enable further investigation by the scientific community.

Our integrative analysis revealed a substantial remodeling of the TME associated with *TP53* mutations (Fig. 8h), including decreased presence of pericytes and endothelial aerocytes. The mechanisms of TME remodeling may include ligand–receptor interactions between malignant and endothelial cells (*SEMA3A–NRP1* and *EPHB2–EFNB1*) that were found to be enriched in *TP53*^{mut} tumors and have known roles in inhibiting vascularization, which were significantly more colocalized in the *TP53*^{mut} spatial data. *TP53*^{mut}-specific depletion of endothelial cells was accompanied by increased hypoxia and pEMT program expression in malignant cells, which was altered across multiple functional categories of *TP53* mutations in bulk transcriptomic and proteomic data, most prominently in DNE, LOF and impactful II *TP53* mutations. This effect was independent of presence or absence of *KRAS* mutations but was partially reduced in tumors with *TP53–EGFR* comutations. Hypoxia has been shown to induce EMT through *HIF1A* and other mediators[79,80] and to positively select for p53-deficient cells that have lost apoptotic potential in mice[81]. Thus, a hypoxic TME may provide a survival advantage for highly plastic *TP53*^{mut} LUAD malignant cells, which have lost alveolar identity and experience increased proliferative capacity and metastatic potential. This model is supported by our analysis of data from genetically engineered K and KP mouse models[51] and from the A549 LUAD cell line (*TP53*^{WT}) overexpressing different *TP53* mutational variants[49].

Furthermore, de novo discovery of spatial multicellular communities revealed a highly hypoxic, EMT-promoting spatial niche enriched for *SPP1*$^{+}$ TAMs, myofibroblasts and collagen-producing CAFs. *SPP1* expression was highly upregulated in *TP53*^{mut} monocytes and TAMs and tightly linked with regulation of hypoxia and EMT. Increased transforming growth factor-β signaling likely also helps shape an EMT-supporting niche in *TP53*^{mut} tumor samples, where *TGFB2–TGFBR2* interactions involving malignant cells were significantly enriched in spatial data. Previous studies linking mutant p53 to metastatic potential focused on malignant cell-intrinsic processes[12,13,82], as the effects of mutant p53 on tumor–stromal and tumor–immune crosstalk have been described only recently in model systems[83]. Our holistic study investigates the

Fig. 8 | Spatial multicellular CAF–TAM niche enriched in hypoxia and EMT activity in *TP53*^{mut} LUAD. a, Overall spatial correlation of broad cell classes across ST spots (left) and differential spatial correlation between *TP53*^{mut} and *TP53*^{WT} tumor samples (right). Red indicates either high colocalization (left) or colocalization enrichment (right) in *TP53*^{mut} while blue represents either low colocalization (left) or colocalization enrichment (right) in *TP53*^{WT} tumor samples between pairs of cell subsets. **b**, Box plots comparing spatial correlation of endothelial cells with mesenchymal (left), myeloid (middle) and malignant (right) cells across spots in *TP53*^{WT} ($n = 10$) and *TP53*^{mut} ($n = 4$) tumor sections, where n refers to the number of ST samples. P values (from left to right: 0.024, 0.014 and 0.84) were calculated using a two-sided Mann–Whitney–Wilcoxon test. **c**, Heatmap of the 15 factors (y axis) depicting multicellular programs, as calculated using NMF across different cell subsets (x axis). **d**, Spatial enrichment of myeloid (TAM.SPP1) and mesenchymal (CAF.COLs and myofibroblasts) cell subsets comprising NMF7 (left) and expression programs spatially correlated with NMF7 (right) in a representative *TP53*^{mut} tumor ST sample (participant BWH14). **e**, Box plots comparing spatial correlation of NMF7 with hallmark hypoxia score (top), hallmark EMT score (middle) and endothelial cell proportion (bottom) across tumor periphery spots (spots with ≤50% malignant cells) in *TP53*^{WT} ($n = 10$) and *TP53*^{mut} ($n = 4$) tumor sections, where n refers to the number of ST samples. P values (from top to bottom: 0.036, 0.004 and 0.054) were calculated using a two-sided Mann–Whitney–Wilcoxon test. **f**, Representative mIF images of *TP53*^{WT} (left; BWH04, predominantly papillary histology with 40% lepidic and 20% acinar patterns) and *TP53*^{mut} (right; BWH14, predominantly acinar histology with 15% lepidic and 5% micropapillary pattern) tumors. Red arrows correspond to CD31^{+} cells, yellow arrows correspond to SPP1^{+}CD68^{+} cells and light-blue arrows correspond to α-SMA^{+} cells. Scale bar, 100 µm. **g**, Box plots comparing the proportion of CD31^{+} cells among all cells within 50 µm of SPP1^{+}CD68^{+} cells in 62 ROIs from *TP53*^{WT} ($n = 44$) versus *TP53*^{mut} ($n = 18$) LUAD tumor sections from ten cases in our cohort, where n refers to the number of ROIs. The P value (0.0042) was calculated using a two-sided Mann–Whitney–Wilcoxon test. **h**, Graphical summary of the differences between the *TP53*^{WT} (left) and *TP53*^{mut} (right) LUAD TME presented in this study. Throughout this figure, box plot horizontal lines show the 25th, 50th (median) and 75th percentiles, with vertical whiskers extending to a maximum distance of 1.5× the interquartile range from the hinge. Data beyond the whisker ends are plotted individually.

https://doi.org/10.1038/s43018-025-01053-7



cellular and spatial context of *TP53* mutated LUAD tumors to provide insight into how *TP53* mutations could remodel the TME to promote tumor survival and metastasis, leading to poorer outcomes.

CAFs have a critical role in TME remodeling, influencing tumor progression, immune responses and ECM structure through their diverse subtypes. Interestingly, the CAF.COLs subset that we found to be colocalized with *SPP1*$^{+}$ TAMs most closely matched the CAF_{infla} subtype described in a prior study[59], noted for cytokine secretion and immune modulation, and shared traits with CAFs positive for immunomodulatory fibroblast activation protein in NSCLC[84]. This underscores the potential of specific CAF populations to alter the immune landscape and affect treatment response. Myofibroblasts from our study exhibited features of CAF_{myo}, associated with ECM deposition and tumor stiffness[59], highlighting the functional diversity of CAFs and their complex contributions to TME dynamics.

Interestingly, one aspect of TME remodeling in *TP53*^{mut} LUAD is a higher tumor infiltration by B cells and *CD8*$^{+}$ T cells. Increased lymphocyte recruitment may be caused in part by enriched *CXCL11*–*CXCR3* ligand–receptor interactions between myeloid and B or T cells in *TP53*^{mut} tumors, as we recently reported in more mesenchymal-like pleural mesothelioma tumors[85]. Our findings of a potentially more immunogenic TME in *TP53*^{mut} LUAD are consistent with previous studies reporting longer survival in persons with advanced *TP53*^{mut} versus *TP53*^{WT} NSCLC receiving immune checkpoint therapy[86]. By reanalyzing recently published data[74], our study also shows improved PFS in persons with *TP53*^{mut} versus *TP53*^{WT} LUAD receiving anti-PD1 therapy, especially for DNE, LOF and impactful II classes of *TP53* mutations. In agreement, our computational screen for immune checkpoint receptor–ligand pairs showed enrichment in *PDCD1*–*CD274*, *CTLA4*–*CD86* and *TIGIT*–*PVR* interactions in *TP53*^{mut} versus *TP53*^{WT} LUAD tumors; *CTLA4*–*CD86* and *TIGIT*–*PVR* interactions have not yet been studied in the context of *TP53* mutations to the best of our knowledge. Increased expression of *TIGIT* in T cells and *PVR* in malignant cell both contribute to the spatially resolved enrichment of *TIGIT*–*PVR* coexpression, which we observed in all eight *TP53*^{mut} LUAD tumors in our cohort, as well as in bulk RNA-seq data from TCGA and bulk proteomic data from CPTAC. Recent trials targeting TIGIT alone or in combination with PD1 and PDL1 in lung cancer have yet to demonstrate substantial survival benefit, highlighting the importance of identifying subsets of tumors that could benefit from therapeutic intervention (for example, *TP53*^{mut} PVR-expressing tumors)[87]. Overall, our work provides a rationale for personalizing *TP53*^{mut} LUAD treatment by targeting the TME using interactions and pathways uncovered in this study.

Methods

Participants

Fresh solid primary tumor tissue was collected using an institutional review board protocol (98-063) approved by the Mass General Brigham Human Research Protection Program ethics committee at Harvard Medical School. Participants were all confirmed to be treatment naive and provided written consent to the study for sharing deidentified clinical information before collection. Immediately after resection, the tissue was reviewed by the clinical pathology team and high-quality portions (determined on the basis of tumor content, necrosis, calcification, fat and hemorrhage) were allocated for WES, scRNA-seq, ST and mIF. The protocol was designed to reduce the time between surgical resection, anatomic pathology review, placement in media and processing at the Broad Institute. Blood was drawn from the same participants and cryopreserved at −80 °C for subsequent processing.

WES sample processing

DNA was extracted from fresh-frozen lung cancer tissue embedded into OCT (TissueTek, Sakura) and from PBMCs isolated from preserved participant-matched blood (AllPrep DNA/RNA extraction kit, Qiagen). Library construction was performed as previously described[88]. In brief, DNA input for shearing was diluted to a final concentration of 20–250 ng in 50 µl of solution. Adaptor ligation was performed using palindromic forked adaptors (Integrated DNA Technologies), containing unique dual-indexed molecular barcode sequences to improve downstream pooling. End repair, poly(A) tailing, adaptor ligation and library enrichment PCR were carried out using a 96-reaction kit format (Kapa HyperPrep). During solid-phase reversible immobilization, a final elution volume of 30 µl was produced to maximize library concentration and vortexing was performed to maximize the effluent. Constructed libraries were first pooled and hybridization and capture were performed (Illumina's Nextera Exome Kit) using the recommended protocol from the manufacturer, using a skirted PCR plate to facilitate automation (Agilent Bravo liquid handling system). Library pools then underwent qPCR quantification and libraries were adjusted to a concentration of 2 nM. DNA libraries were cluster-amplified using exclusion amplification chemistry in patterned flow cells (Illumina) according to the manufacturer's recommended protocol. Sequencing of flow cells was performed using sequencing-by-synthesis chemistry and analyzed using RTA (version 2.7.3 or later). Library pools were sequenced on 76 cycle runs using two eight-cycle index reads across the appropriate number of lanes to attain coverage for all libraries.

Tissue processing, CD45 sorting and scRNA-seq

Tumor tissue resections were transferred in RPMI on ice from the operating room and washed in cold PBS in the laboratory and transferred to 5-ml Eppendorf tubes containing 3 ml of dissociation mixture (NSCLC patient-derived explant culture protocol[89]). Samples were minced in the Eppendorf tube using scissors into pieces smaller than ~0.4 mm and incubated at 37 °C, rotating at ~14 rpm. for 10 min. Each sample was pipetted 20 times with a 1-ml pipette tip at room temperature and placed back into incubation for 10 min. The sample was again pipetted 20 times using a 1-ml pipette tip, transferred to a 1.7-ml Eppendorf tube and centrifuged at 300–580*g* for 4–7 min at 4 °C. The pellet was resuspended in 200–500 µl of ammonium–chloride–potassium red blood cell (RBC) lysis buffer (Thermo Fisher Scientific, A1049201) and incubated for 1 min on ice. Twice the volume of cold PBS was added to stop the reaction and cells were pelleted by a short centrifugation for 8 s at 4 °C, using the short spin setting and ramping up to 11,000*g*. If RBCs remained, the RBC lysis step was performed up to two additional times. To assess cell count and viability, 5 µl of Trypan blue (Thermo Fisher Scientific, T10282) was mixed with 5 µl of the sample and loaded onto an INCYTO C-Chip disposable hemocytometer, Neubauer improved (VWR, 82030-468). Depletion of CD45^{+} cells in NSCLC samples was performed using CD45 MicroBeads (Miltenyi Biotec, 130-045-801) according to the manufacturer's protocol. Following CD45^{+} cell depletion, cells were recounted and adjusted to a final range of 200–2,000 cells per µl.

NSCLC tissue dissociation protocol

The digestion mix contained 2,692 µl of HBSS (Thermo Fisher Scientific, 14170112), 187.5 µl of 20 mg ml^{-1} pronase (Sigma Aldrich, 10165921001) diluted to a final concentration of 1,250 µg ml^{-1}, 27.6 µl of 1 mg ml^{-1} elastase (Thermo Fisher Scientific, NC9301601) diluted to a final concentration of 9.2 µg ml^{-1}, 30 µl of 10 mg ml^{-1} DNase I (Sigma Aldrich, 11284932001) diluted to a final concentration of 100 µg ml^{-1}, 30 µl of 10 mg ml^{-1} dispase (Sigma Aldrich, 4942078001) diluted to a final concentration of 100 µg ml^{-1}, 30 µl of 150 mg ml^{-1} collagenase A (Sigma Aldrich, 10103578001) diluted to a final concentration of 1,500 µg ml^{-1} and 3 µl of 100 mg ml^{-1} collagenase IV (Thermo Fisher Scientific, NC9836075) diluted to a final concentration of 100 µg ml^{-1}.

scRNA-seq library preparation and sequencing

A total of 8,000 cells were loaded onto each channel of the Chromium Controller (10X Genomics) to generate single-cell gel beads in emulsion. Libraries were constructed using the Chromium single-cell 3′ library and gel bead kit (version 2 or 3; PN-120237, 10X Genomics);

barcoded reverse transcription of RNA, complementary DNA (cDNA) amplification, fragmentation and adaptor and sample index attachment were all performed according to the manufacturer's recommended protocol. Barcoded libraries from four 10C channels were pooled together and sequenced on either one lane of an Illumina HiSeq X or one flow cell of a NextSeq, with paired-end reads accordingly: read 1, 26 nt; read 2, 55 nt; index 1, 8 nt; index 2, 0 nt.

Tissue processing for ST

Fresh-frozen lung cancer tissue samples were embedded into OCT (TissueTek, Sakura) and shipped to the Broad Institute for cryosectioning and generation of H&E tissue sections. H&E-stained tissues were subject to pathology review to assess tissue quality, structural preservation, cellular viability, tumor content, inflammation, fibrosis and necrosis. Samples were excluded on the basis of small tissue size, low cellularity or extensive fibrosis. ROIs were then selected on the basis of the criteria mentioned and marked on H&E images for subsequent tissue sectioning and mounting on Visium slides. Visium sectioning and processing were performed on the eight samples that passed quality control. Tissues were cryosectioned at 10-µm thickness at −22 °C and placed in the capture areas of Visium tissue optimization slides (3000394, 10X Genomics) and Visium spatial gene expression slides (2000233, 10X Genomics). The tissue sections were adhered by warming the backside of the slides and placed at −80 °C for 1–3 days.

Visium spatial gene expression library generation

The tissue optimization sample slide and spatial gene expression slide were processed according to the manufacturer's protocols (CG000238_VisiumSpatialTissueOptimizationUserGuide_Rev_A.pdf and CG000239_VisiumSpatialGeneExpression_UserGuide_Rev_C.pdf). In short, following tissue methanol fixation and H&E staining, brightfield morphology images were obtained with a Zeiss Axio microscope using the Metafer slide-scanning platform (Metasystems) at ×10 resolution. The images were joined together with the VSlide software (Metasystems) and exported as .tiff files. Optimal tissue permeabilization times were tested with eight different time points (0, 3, 6, 12, 18, 24, 30 and 36 min) on the tissue optimization sample slide for one of the samples. Then, 12 min of permeabilization was set as the optimal time point and used for the spatial gene expression slide for all samples. The released RNAs from the tissue were reverse-transcribed into cDNA by priming to the spatial barcoded primers on the glass in the presence of template switching oligo. After second strand synthesis, a denaturation step released the cDNAs, which were then amplified with 14–15 cycles of PCR amplification. Finally, indexed sequencing-ready, spatial gene expression libraries were constructed according to the manufacturer's protocol. The libraries were pooled together to generate >50,000 reads per spatial spot and sequenced on Illumina NovaSeq sequencers with S1 and SP kits and the following settings: read 1, 28 cycles; read 2, 90 cycles; index 1, 10 cycles; index 2, 10 cycles.

WES data processing

The Picard pipeline (http://picard.sourceforge.net/) was used to align the tumor and normal whole-exome sequences to the hg19 reference human genome build and generate BAM files. Somatic short variant discovery was performed using the Mutect 2 algorithm (https://software.broadinstitute.org/gatk/documentation/tooldocs/current/org_broadinstitute_gatk_tools_walkers_cancer_m2_MuTect2) from the GAT4K pipeline run on the Terra cloud-based platform. Somatic mutations were annotated using Funcotator (https://gatk.broadinstitute.org/hc/en-us/articles/360035889931-Funcotator-Information-and-Tutorial). Copy-number ratios for each exon were calculated by comparing mean exon coverage with expected coverage on the basis of a panel of normal samples and were then segmented for downstream analysis. Then, .maf and .vcf files were generated as outputs, which were subsequently processed with maftools[90] and cnvkit[91] packages.

Large-scale CNA inference and correlation

The inferCNV package (https://github.com/broadinstitute/inferCNV) was applied to infer large-scale CNAs from scRNA-seq data. Normal lung epithelial cells[45] were used as a reference. The function infercnv::run() was performed with a cutoff of 0.1, cluster_by_groups = T and HMM = T. CNA values were predicted using a six-state hidden Markov model and were normalized to the third quantile of all values per sample. Cells with >0.7 normalized CNA values were assigned as malignant cells. Spearman correlation was run between inferred CNAs from scRNA and from WES data across genes and plotted as a heatmap.

scRNA-seq preprocessing

CellRanger (version 3.1) was used to align 10X Chromium reads to the GRCh38 human genome reference and generate barcode, gene and count matrices. Preprocessed matrices were then loaded into Seurat (version 4)[92] in R (version 4.1.0) using the Read10X() function. Cells with fewer than 1,000 unique molecular identifiers (UMIs), fewer than 400 detected genes (1,000 genes for malignant cells) or greater than 25% mitochondrial genes were excluded from further analyses.

scRNA-seq integration, clustering, annotation and doublet removal

The UMI count matrices were batch-corrected by sample, using the Seurat version 4 SCTransform() (using 2,000 variable features and regressing out percentage of mitochondrial genes), FindIntegrationAnchors() and IntegrateData(), using the largest sample as the reference dataset. Principal component analysis was then performed in the variable gene space and 15 principal components were used for Louvain clustering and uniform manifold approximation and projection (UMAP) dimensionality reduction. Cell type markers for broad cell classes were identified using the FindAllMarkers() function, using a maximum of 1,000 cells per cluster. Cell clusters were annotated on the basis of the expression of established markers. Individual cell classes were then subset and subjected to another series of integration, clustering and annotation on the basis of the expression of subset-specific markers. Doublet cell clusters were removed on the basis of the expression of markers specific to more than one broad cell classes.

scRNA-seq signature scoring

We identified malignant program signatures by performing anchor-based integration across samples of malignant cells followed by Louvain clustering at resolution 1.2 and using the FindAllMarkers() function to derive the top ten most differentially expressed genes from each cluster on the basis of the $\log_2$ fold change and Bonferroni-corrected P values. Signatures for normal lung epithelial cells were derived from healthy lung scRNA-seq data[45] in a similar manner (Supplementary Table 4). To define hallmark EMT and hypoxia scores, we used gene sets termed 'Hallmark_epithelial_mesenchymal_transition' and 'Hallmark_hypoxia' from MSigDB R package (version 7.5.1)[93]. For each cell, signature scores were computed using the Seurat AddModuleScore() function.

Correspondence between clustering-based malignant programs and cNMF-derived programs

To evaluate the robustness of our malignant program signature discovery, we performed cNMF[36] on malignant cells using the Python package cNMF (version 1.3.4). Each k was run 100 times, and $k = 30$ was selected for downstream analysis. For each cNMF module, the top 20 unique genes were determined on the basis of the ranked spectra values and used to compute module scores as explained above. Pearson correlation across all malignant cells was used between our clustering-based malignant program scores and cNMF module scores to assess their correspondence.

Correspondence between CAF subset signatures

To assess the correspondence between CAF subsets defined in our study and pan-cancer CAF subsets from a previous study[59], we used the top 20 markers defining each CAF subset to compute module scores for all CAFs. Pearson correlation of CAF subset module scores was used to quantify the correspondence between CAF populations defined in our study and the previous one[59].

scRNA-seq differential expression analysis

The data were first split into different cell subsets and pseudobulk profiles were generated per sample using the Seurat AverageExpression() function. A Mann–Whitney–Wilcoxon test was used to test for significance in gene differential expression between groups. The average $\log_2$ fold change was plotted using the ComplexHeatmap (version 2.10.0) package in R and asterisks were added for genes that were significantly differentially expressed.

Gene set enrichment analysis

Using the top 100 differentially expressed genes per cell subset, we performed gene set enrichment analysis using the R package clusterProfiler (version 4.0)[94], focusing on Gene Ontology terms, Kyoto Encyclopedia of Genes and Genomes pathways, Reactome and Hallmark database gene sets. Gene ratios and adjusted *P* values were visualized as dot plots, after applying Benjamini–Hochberg correction and setting *P*-value and *q*-value cutoffs of 0.05.

Similarity to normal lung and K or KP mouse cells

Each NSCLC cell was assigned a normal lung subset it was most transcriptionally similar to using the Seurat TransferData() function with default parameters, where the cell compartment object in the NSCLC scRNA-seq dataset represented the query and the corresponding compartment in normal lung scRNA-seq datasets[44,45] represented the reference. A confusion matrix was created, displaying the number of cells that overlapped between predicted identities from the normal lung and NSCLC annotations. A similar approach was used to classify K and KP mouse HPCS cells to their closest NSCLC malignant cell annotations.

Signaling entropy calculation

Entropy was inferred for each epithelial cell using the R package SCENT (version 1.0.3)[52], using the default protein–protein interaction network ('net17Jan2016.m'). Raw UMI counts were used as the scRNA-seq input. Differential potency estimation was performed using the DoIntegPPI() function, which generated entropy values per cell. The same method was used to calculate entropy values for bulk RNA-seq data from TCGA and bulk proteomic data from CPTAC. Entropy values were compared using the ggviolin() function using the ggpubr package in R.

Cell–cell interaction analysis

CellPhoneDB (version 2.1.7)[95] was used to predict significant ligand–receptor interactions using normalized raw counts and annotation of broad cell classes as input. The method was run with default parameters separately for each sample ID to account for batch effects between samples[96]. Interactions with a *P* value below 0.05 were classified as significant. For a given ligand–receptor pair, the normalized enrichment difference (ED) was calculated as follows:

$$\mathrm{ED} = n_A / N_A - n_B / N_B,$$

where n_A and n_B represent the number of tumor samples with significant interactions in groups A and B, respectively, while N_A and N_B represent the total number of tumor samples in these groups. Significant enrichment for an interaction in a group was calculated using two-sided two-proportion *Z*-test or Fisher's exact test. The $-\log_{10} P$ values and normalized EDs for each ligand–receptor interaction and cell pair combination were plotted using the ggplot2 package in R.

scRNA-seq regulon analysis

An scRNA-seq transcription factor gene regulatory network was constructed using pySCENIC (version 0.11.2)[57]. Raw UMI counts were used as input and regulons were predicted using gene module coexpression using the GRNBoost2 package. The number of regulons was pruned using the feather ranking databases (hg19-500bp-upstream-7species.mc9nr, hg19-tss-centered-10kb-10species.mc9nr and hg19-tss-centered-5kb-10species.mc9nr). Regulon activity, driving transcription factors and weights for individual target genes were predicted with the cisTarget() function in pySCENIC. Regulon activity enrichment scores were predicted for each cellular subset using AUCell and regulon specificity scores were used to identify regulatory networks specific to each cell subset. The *z*-score-transformed regulon *x* subset matrices were visualized using the ComplexHeatmap package in R.

Additional scRNA-seq datasets

Processed scRNA-seq data from an additional five studies with available *TP53* mutational status were assembled into a validation compendium, consisting of 15 $TP53^{WT}$ and 7 $TP53^{mut}$ tumor samples (ArrayExpress accessions E-MTAB-6149 and E-MTAB-6653 (ref. 39), HTAN-MSK cohort from https://humantumoratlas.org/ (ref. 43), Gene Expression Omnibus (GEO) accession GSE123904 (ref. 40), Genome Sequence Archive accession HRA000154 (ref. 41) and https://doi.org/10.24433/CO.0121060.v1 (ref. 42)). In vivo scRNA-seq data for $KRAS^{mut}TP53^{WT}$ and $KRAS^{mut}TP53^{mut}$ mice (Smart-Seq2) were obtained from GEO accession GSE152607 (ref. 51). In vitro scRNA-seq data from lung cancer A549 cell lines with different induced *TP53* variants was obtained from GEO accession GSE161824 (ref. 49).

Signature scoring of bulk expression profiles

Bulk RNA-seq and somatic mutation data from TCGA LUAD and somatic mutations were downloaded from cBioPortal[97] using the cgdsr (version 1.3.0) package in R. Curated gene sets with highly specific markers for individual cell types, subsets or malignant programs were generated from the scRNA-seq data as described above. For a given gene set, the signature score was defined as the mean $\log_{10}$ expression of all genes in bulk RNA-seq data from TCGA. The same approach was used for calculating expression signatures in bulk proteomic data from CPTAC[98], which were downloaded from the CPTAC data portal (https://cptac-data-portal.georgetown.edu/cptacPublic/). We confirmed the reproducibility of all our bulk deconvolution results using BayesPrism (version 1.4)[99].

Effects of comutations and *TP53* variant classes

TCGA tumor samples were separated into six mutually exclusive categories with sufficiently large sample sizes for comparative analysis on the basis of combinations of mutations in *KRAS*, *EGFR* and *TP53* ($KRAS^{mut}$, $EGFR^{mut}$, $TP53^{mut}$, $KRAS^{mut}TP53^{mut}$, $EGFR^{mut}TP53^{mut}$ and $KRAS^{WT}EGFR^{WT}TP53^{WT}$). TCGA and CPTAC tumor samples were furthermore grouped into different mutually exclusive *TP53* mutational variant classes on the basis of either DNE and LOF status or impact categories as described by prior studies[48,49]. Mann–Whitney–Wilcoxon test *P* values were computed for each group relative to the relevant WT or control group.

Survival analysis

To link the expression of individual genes or gene sets with survival, TCGA LUAD tumor samples were classified into groups with high (top 25th percentile), medium (25th–75th percentile) and low (bottom 25th percentile) expression or high (top 50th percentile) and low (bottom 50th percentile) expression. Kaplan–Meier curves were generated using the survival package in R and *P* values were calculated using a log-rank test or bivariate Cox proportional-hazards regression analysis to evaluate the significance of different model parameters (for example,

expression level of genes or malignant signatures and *TP53* mutation status) in relation to overall survival (defined as time to death from any cause). The reported *P* values from the Cox proportional-hazards model reflect the significance of one variable while adjusting for different covariates (for example, histology and stage).

The SU2C-MARK cohort[74] was used to assess the impact of different genetic alterations on ICI treatment response. Clinical and WES data for participants were downloaded from the publication's supplementary materials. The mutation status of *TP53*, *EGFR*, *KRAS*, *KEAP1* and *STK11* was converted to WT or mutant and the *TP53* variant and impact subcategories were determined. Participants were separated on the basis of their subtype (LUAD or LUSC) and treatment (anti-PD1 or all immunotherapy single-agent and combination treatments). Cox proportional-hazards regression analysis was performed to determine the association between PFS and mutation status.

Multiple linear regression of clinical covariates

Expression level of individual genes or gene signatures was predicted on the basis of the mutation status of nine different gene variants (Fig. 3b) and TMB using the following model:

$$Y = \beta_0 + \sum_{i=1}^{10} \beta_i X_i + \epsilon,$$

where Y represents the expression level, β_0 is the y intercept, β_i ($i \in \{1, 2, \ldots, 10\}$) is the regression coefficient associated with each variable, X_i ($i \in \{1, 2, \ldots, 9\}$) is the mutation status of nine different gene variants, X_{10} is the TMB and ϵ is the error term. For each mutation status variable X_i, a P value was computed to test the null hypothesis that the regression coefficient β_i is equal to zero (indicating no effect). A significant P value (≤0.05) suggests that the mutation status of the respective gene is a significant predictor of Y (that is, there exists a significant relationship between the mutation status and the expression level). We further conducted the same analysis using separate linear models, each correcting for different combinations of clinical covariates—including histology, stage, size and their combination—to assess the robustness of associations after covariate adjustment (Fig. 3c).

Pathologist annotation of H&E stains from ST samples

A board-certified pathologist, blinded to all sample identities and ST measurements, performed manual annotation of 13 H&E stains using Loupe browser. Annotations on the H&E were classified into multiple categories (for example, cancer, lymphocytes, myeloid, fibroblast, lepidic adenocarcinoma, tertiary lymphoid structure and vascular endothelium) that were most representative of cells within barcoded spots. Spots in which there were cell types whose identities could not be readily visually resolved were not annotated. These categories were then mapped to three general cell classes (malignant, stromal and immune) to minimize variability in pathology annotations across samples and allow for comparison with the computationally inferred cellular compositions.

ST data processing and analysis

Tangram (version 0.4.0), scanpy (version 1.8.1) and anndata (version 0.7.6) were used to map scRNA-seq data to spatial locations on 10X Visium ST samples as previously described[34]. UMI counts for scRNA-seq and ST data were converted into anndata objects and preprocessed using standard scanpy functions[100]. The top 100 genes that best characterized each broad cell class were selected by sc.tl.rank_genes_groups(). This subset of genes was used afterward as training genes for Tangram alignment. To map single cells more accurately onto ST spots, an estimate of cell density per spot was attained through watershed-based nuclear segmentation of the H&E-stained serial tissue section. Images were first imported by squidpy (version 1.1.0)[101], smoothened using squidpy.im.process with $\sigma = 4$ and then segmented by squidpy.im.segment with method = 'watershed', using parameters channel = 0 (to select the red color channel) and thres = 120. A threshold of 120 was selected to best separate nuclei from background and sq.im.calculate_features was used to count the number of segmented nuclei per spot after segmentation. Finally, mapping was performed by tg.map_cells_to_space using mode = 'constrained' and density_prior = no. of nuclei per spot/total no. of nuclei to constrain alignment based on segmented nuclei density, returning probabilities of spatial location on a per-cell basis, which were used in later analysis. To project individual gene expression onto ST spots, tg.project_genes was used. Cell type and subset proportion were predicted per spot by summing up cell x spot probabilities and normalizing the total probability scores per spot to 1. Spatial colocalization was inferred using Pearson correlation between predicted cell type and subset compositions, as well as additional continuous attributes from the metadata across spots.

NMF analysis was adapted from cell2location[102] using cell subset proportions per ST spot as inputs, enabling the analysis of spatially colocalized cellular programs. NMF was trained five times for a range of $k = 3$–30 factors and $k = 14$ was selected on the basis of stability of training across the five restarts and elucidating discrete spatially delimited compartments across ST samples. Using scanpy, NMF weights were plotted onto ST data to investigate spatial-dependent patterning of NMF factors. NMF weights per subset were visualized using the heatmap function in cell2location.

mIF and analysis

mIF experiments were carried out as previously reported[73] on 5-µm FFPE whole-tissue sections, by staining for nuclear counterstain DAPI and the following antibodies: anti-CD8 (clone 41BB; Leica, CD8-4B11-L-CE; 1:200 dilution), anti-PDL1 (clone E1L3N; Cell Signaling, 13684; 1:300 dilution), anti-PD1 (clone EPR4877(2); Abcam, Ab137132; 1:300 dilution), anti-cytokeratin (clone AE1/AE3; Dako, GA053; 1:100 dilution), anti-α-SMA (clone D4K9N; Cell Signaling, 34105; 1:200 dilution), anti-CD68 (clone D4B9C; Cell Signaling, 76437; 1:3,500 dilution), anti-OPN/SPP1 (clone EPR21139-316; Abcam, ab214050; 1:250 dilution) and anti-CD31 (polyclonal; Abcam, ab28364; 1:250 dilution). Non-overlapping representative ROIs for each tissue section were selected by a pathologist (S.J.R.) blinded to the sample identity and subsequently images were spectrally unmixed and analyzed using Inform 2.6 (Akoya Biociences). Each analyzable ROI was segmented and quantified for expression of each marker using the Inform analysis tools. Inform-analyzed ROIs were then processed using the Pythologist software[103]. Cell colocalization involved three spatial categories: direct contact, adjacency within six nearest neighbors and proximity within 50 µm of PDL1$^+$cytokeratin$^+$ cells. The proportion of PD1$^+$CD8$^+$ cells among all PDL1$^+$cytokeratin$^+$ neighboring cells was computed across these spatially defined categories when at least 20 neighboring cells were present.

Statistics and reproducibility

No statistical methods were used to perform power analysis but our sample sizes are similar to those reported in previous NSCLC publications[39–43]. Given the frequency of *TP53* mutations in LUAD (50% of cases), we expected an even split of samples into *TP53*^{mut} and *TP53*^{WT} tumors. As *TP53* mutations have been found to have histology-dependent effects on outcome[104], we excluded five tumor samples with either squamous or atypical (mucinous, neuroendocrine and colloid) adenocarcinoma histology from all downstream comparisons between *TP53*^{mut} and *TP53*^{WT} tumors. We furthermore excluded Visium data from one participant with *TP53*^{mut} (BWH16) because of an insufficient number of malignant cells in the relevant slides. Experiments were not randomized. scRNA-seq data collection (discovery cohort) was performed blinded to *TP53* status and downstream computational analysis. Moreover, board-certified pathologists were blinded to our data and results for histological annotations and ROI selection.

Statistical tests were performed as described in the respective figure legends. We used nonparametric tests (for example, Mann–Whitney–Wilcoxon) that do not assume normal distribution, except when analyzing TCGA RNA-seq data (for example, *t*-test), as log expression data have been shown to approximate a normal distribution, although this was not formally tested. Our findings were highly reproducible across large independent cohorts, including TCGA ($n = 510$), CPTAC ($n = 110$) and other scRNA-seq studies ($n = 22$) with available *TP53* mutational status, increasing the confidence and statistical power of our study. Moreover, scRNA-seq data from LUAD model systems (mouse models[51] and cell lines[49]) and spatial and mIF data provided orthogonal validation of our main findings.

Reporting summary

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

Data availability

In collaboration with the NIH-funded HTAN Data Coordinating Center (U24), WES, scRNA-seq and ST data are available from an interactive, online platform for download, independent visualization and analysis (https://data.humantumoratlas.org/explore/; select atlas 'HTAN HTAPP' and organ 'lung' from the scroll down menus). Processed scRNA-seq data, spatial transcriptomic data and mutation information can be downloaded from Zenodo (https://doi.org/10.5281/zenodo.16546233)[105]. Source data are provided with this paper.

Code availability

Code for analysis presented in this paper is available from GitHub (https://github.com/TsankovLab/HTAN_Lung) and Zenodo (https://doi.org/10.5281/zenodo.15866295)[106].

References

1. Travis, W. D. et al. The 2015 World Health Organization classification of lung tumors: impact of genetic, clinical and radiologic advances since the 2004 classification. *J. Thorac. Oncol.* **10**, 1243–1260 (2015).
2. The Cancer Genome Atlas Research. Network Comprehensive genomic characterization of squamous cell lung cancers. *Nature* **489**, 519–525 (2012).
3. The Cancer Genome Atlas Research Network. Comprehensive molecular profiling of lung adenocarcinoma. *Nature* **511**, 543–550 (2014).
4. Hanna, N. H. et al. Therapy for stage IV non-small-cell lung cancer with driver alterations: ASCO and OH (CCO) joint guideline update. *J. Clin. Oncol.* **39**, 1040–1091 (2021).
5. Howlader, N. et al. The effect of advances in lung-cancer treatment on population mortality. *N. Engl. J. Med.* **383**, 640–649 (2020).
6. Herbst, R. S., Morgensztern, D. & Boshoff, C. The biology and management of non-small cell lung cancer. *Nature* **553**, 446–454 (2018).
7. Forde, P. M. et al. Neoadjuvant nivolumab plus chemotherapy in resectable lung cancer. *N. Engl. J. Med.* **386**, 1973–1985 (2022).
8. Non-small cell lung cancer treatment. *National Cancer Institute* https://www.cancer.gov/types/lung/patient/non-small-cell-lung-treatment-pdq (2021).
9. Campbell, J. D. et al. Distinct patterns of somatic genome alterations in lung adenocarcinomas and squamous cell carcinomas. *Nat. Genet.* **48**, 607–616 (2016).
10. Halvorsen, A. R. et al. TP53 mutation spectrum in smokers and never smoking lung cancer patients. *Front. Genet.* **7**, 85 (2016).
11. Kaubryte, J. & Lai, A. G. Pan-cancer prognostic genetic mutations and clinicopathological factors associated with survival outcomes: a systematic review. *npj Precis. Oncol.* **6**, 27 (2022).
12. Al Bakir, M. et al. The evolution of non-small cell lung cancer metastases in TRACERx. *Nature* **616**, 534–542 (2023).
13. Van Egeren, D. et al. Genomic analysis of early-stage lung cancer reveals a role for *TP53* mutations in distant metastasis. *Sci. Rep.* **12**, 19055 (2022).
14. Duffy, M. J., Synnott, N. C., O'Grady, S. & Crown, J. Targeting p53 for the treatment of cancer. *Semin. Cancer Biol.* **79**, 58–67 (2022).
15. Bykov, V. J. N., Eriksson, S. E., Bianchi, J. & Wiman, K. G. Targeting mutant p53 for efficient cancer therapy. *Nat. Rev. Cancer* **18**, 89–102 (2018).
16. Zhu, G. et al. Mutant p53 in cancer progression and targeted therapies. *Front. Oncol.* **10**, 595187 (2020).
17. Uehara, I. & Tanaka, N. Role of p53 in the regulation of the inflammatory tumor microenvironment and tumor suppression. *Cancers* **10**, 219 (2018).
18. Shi, D. & Jiang, P. A different facet of p53 function: regulation of immunity and inflammation during tumor development. *Front. Cell Dev. Biol.* **9**, 762651 (2021).
19. Hodis, E. et al. Stepwise-edited, human melanoma models reveal mutations' effect on tumor and microenvironment. *Science* **376**, eabi8175 (2022).
20. Efe, G., Rustgi, A. K. & Prives, C. p53 at the crossroads of tumor immunity. *Nat. Cancer* **5**, 983–995 (2024).
21. Skoulidis, F. et al. *STK11/LKB1* mutations and PD-1 inhibitor resistance in *KRAS*-mutant lung adenocarcinoma. *Cancer Discov.* **8**, 822–835 (2018).
22. Biton, J. et al. *TP53*, *STK11*, and *EGFR* mutations predict tumor immune profile and the response to anti-PD-1 in lung adenocarcinoma. *Clin. Cancer Res.* **24**, 5710–5723 (2018).
23. Dong, Z. Y. et al. Potential predictive value of *TP53* and *KRAS* mutation status for response to PD-1 blockade immunotherapy in lung adenocarcinoma. *Clin. Cancer Res.* **23**, 3012–3024 (2017).
24. Leader, A. M. et al. Single-cell analysis of human non-small cell lung cancer lesions refines tumor classification and patient stratification. *Cancer Cell* **39**, 1594–1609 (2021).
25. Davis, A. P. et al. Efficacy of immunotherapy in *KRAS*-mutant non-small-cell lung cancer with comutations. *Immunotherapy* **13**, 941–952 (2021).
26. Scalera, S. et al. *KEAP1* and *TP53* frame genomic, evolutionary, and immunologic subtypes of lung adenocarcinoma with different sensitivity to immunotherapy. *J. Thorac. Oncol.* **16**, 2065–2077 (2021).
27. Stuart, T. et al. Comprehensive integration of single-cell data. *Cell* **177**, 1888–1902 (2019).
28. Dudley, A. C. Tumor endothelial cells. *Cold Spring Harb. Perspect. Med.* **2**, a006536 (2012).
29. Cooke, V. G. et al. Pericyte depletion results in hypoxia-associated epithelial-to-mesenchymal transition and metastasis mediated by Met signaling pathway. *Cancer Cell* **21**, 66–81 (2012).
30. De Zuani, M. et al. Single-cell and spatial transcriptomics analysis of non-small cell lung cancer. *Nat. Commun.* **15**, 4388 (2024).
31. Takano, Y. et al. Spatially resolved gene expression profiling of tumor microenvironment reveals key steps of lung adenocarcinoma development. *Nat. Commun.* **15**, 10637 (2024).
32. Greenwald, A. C. et al. Integrative spatial analysis reveals a multi-layered organization of glioblastoma. *Cell* **187**, 2485–2501 (2024).
33. Li, R. et al. Mapping single-cell transcriptomes in the intra-tumoral and associated territories of kidney cancer. *Cancer Cell* **40**, 1583–1599 (2022).
34. Biancalani, T. et al. Deep learning and alignment of spatially resolved single-cell transcriptomes with Tangram. *Nat. Methods* **18**, 1352–1362 (2021).

35. Butler, A., Hoffman, P., Smibert, P., Papalexi, E. & Satija, R. Integrating single-cell transcriptomic data across different conditions, technologies, and species. *Nat. Biotechnol.* **36**, 411–420 (2018).
36. Kotliar, D. Identifying gene expression programs of cell-type identity and cellular activity with single-cell RNA-seq. *eLife* **8**, e43803 (2019).
37. Tang, H. et al. Comprehensive evaluation of published gene expression prognostic signatures for biomarker-based lung cancer clinical studies. *Ann. Oncol.* **28**, 733–740 (2017).
38. Nagy, A., Munkacsy, G. & Gyorffy, B. Pancancer survival analysis of cancer hallmark genes. *Sci. Rep.* **11**, 6047 (2021).
39. Lambrechts, D. et al. Phenotype molding of stromal cells in the lung tumor microenvironment. *Nat. Med.* **24**, 1277–1289 (2018).
40. Laughney, A. M. et al. Regenerative lineages and immune-mediated pruning in lung cancer metastasis. *Nat. Med.* **26**, 259–269 (2020).
41. Xing, X. et al. Decoding the multicellular ecosystem of lung adenocarcinoma manifested as pulmonary subsolid nodules by single-cell RNA sequencing. *Sci. Adv.* **7**, eabd9738 (2021).
42. Bischoff, P. et al. Single-cell RNA sequencing reveals distinct tumor microenvironmental patterns in lung adenocarcinoma. *Oncogene* **40**, 6748–6758 (2021).
43. Chan, J. M. et al. Signatures of plasticity, metastasis, and immunosuppression in an atlas of human small cell lung cancer. *Cancer Cell* **39**, 1479–1496.e18 (2021).
44. Travaglini, K. J. et al. A molecular cell atlas of the human lung from single-cell RNA sequencing. *Nature* **587**, 619–625 (2020).
45. Shah, V. S. et al. Single cell profiling of human airway identifies tuft-ionocyte progenitor cells displaying cytokine-dependent differentiation bias in vitro. *Nat. Commun.* **16**, 5180 (2025).
46. LaFave, L. M. et al. Epigenomic state transitions characterize tumor progression in mouse lung adenocarcinoma. *Cancer Cell* **38**, 212–228 (2020).
47. Yang, L. et al. Single-cell transcriptome analysis revealed a suppressive tumor immune microenvironment in EGFR mutant lung adenocarcinoma. *J. Immunother. Cancer* **10**, e003534 (2022).
48. Giacomelli, A. O. et al. Mutational processes shape the landscape of *TP53* mutations in human cancer. *Nat. Genet.* **50**, 1381–1387 (2018).
49. Ursu, O. et al. Massively parallel phenotyping of coding variants in cancer with Perturb-seq. *Nat. Biotechnol.* **40**, 896–905 (2022).
50. Baslan, T. et al. Ordered and deterministic cancer genome evolution after p53 loss. *Nature* **608**, 795–802 (2022).
51. Marjanovic, N. D. et al. Emergence of a high-plasticity cell state during lung cancer evolution. *Cancer Cell* **38**, 229–246.e13 (2020).
52. Teschendorff, A. E. & Enver, T. Single-cell entropy for accurate estimation of differentiation potency from a cell's transcriptome. *Nat. Commun.* **8**, 15599 (2017).
53. Acevedo, L. M., Barillas, S., Weis, S. M., Gothert, J. R. & Cheresh, D. A. Semaphorin 3A suppresses VEGF-mediated angiogenesis yet acts as a vascular permeability factor. *Blood* **111**, 2674–2680 (2008).
54. Guttmann-Raviv, N. et al. Semaphorin-3A and semaphorin-3F work together to repel endothelial cells and to inhibit their survival by induction of apoptosis. *J. Biol. Chem.* **282**, 26294–26305 (2007).
55. Pasquale, E. B. Eph receptors and ephrins in cancer: bidirectional signalling and beyond. *Nat. Rev. Cancer* **10**, 165–180 (2010).
56. Buechler, M. B. et al. Cross-tissue organization of the fibroblast lineage. *Nature* **593**, 575–579 (2021).
57. Van de Sande, B. et al. A scalable SCENIC workflow for single-cell gene regulatory network analysis. *Nat. Protoc.* **15**, 2247–2276 (2020).
58. Lee, K. W., Yeo, S. Y., Sung, C. O. & Kim, S. H. Twist1 is a key regulator of cancer-associated fibroblasts. *Cancer Res.* **75**, 73–85 (2015).
59. Luo, H. et al. Pan-cancer single-cell analysis reveals the heterogeneity and plasticity of cancer-associated fibroblasts in the tumor microenvironment. *Nat. Commun.* **13**, 6619 (2022).
60. Xu, J., Lamouille, S. & Derynck, R. TGF-β-induced epithelial to mesenchymal transition. *Cell Res.* **19**, 156–172 (2009).
61. Szabo, P. M. et al. Cancer-associated fibroblasts are the main contributors to epithelial-to-mesenchymal signatures in the tumor microenvironment. *Sci. Rep.* **13**, 3051 (2023).
62. De Wever, O., Demetter, P., Mareel, M. & Bracke, M. Stromal myofibroblasts are drivers of invasive cancer growth. *Int. J. Cancer* **123**, 2229–2238 (2008).
63. Dong, B., Wu, C., Huang, L. & Qi, Y. Macrophage-related SPP1 as a potential biomarker for early lymph node metastasis in lung adenocarcinoma. *Front. Cell Dev. Biol.* **9**, 739358 (2021).
64. Casazza, A. et al. Impeding macrophage entry into hypoxic tumor areas by Sema3A/Nrp1 signaling blockade inhibits angiogenesis and restores antitumor immunity. *Cancer Cell* **24**, 695–709 (2013).
65. Zhang, F. et al. TGF-β induces M2-like macrophage polarization via SNAIL-mediated suppression of a pro-inflammatory phenotype. *Oncotarget* **7**, 52294–52306 (2016).
66. van der Leun, A. M., Thommen, D. S. & Schumacher, T. N. $CD8^+$ T cell states in human cancer: insights from single-cell analysis. *Nat. Rev. Cancer* **20**, 218–232 (2020).
67. Litchfield, K. et al. Meta-analysis of tumor- and T cell-intrinsic mechanisms of sensitization to checkpoint inhibition. *Cell* **184**, 596–614 (2021).
68. Zhou, X. et al. A pan-cancer analysis of CD161, a potential new immune checkpoint. *Front. Immunol.* **12**, 688215 (2021).
69. Hollern, D. P. et al. B cells and T follicular helper cells mediate response to checkpoint inhibitors in high mutation burden mouse models of breast cancer. *Cell* **179**, 1191–1206 (2019).
70. Gide, T. N. et al. Distinct immune cell populations define response to anti-PD-1 monotherapy and anti-PD-1/anti-CTLA-4 combined therapy. *Cancer Cell* **35**, 238–255 (2019).
71. Thommen, D. S. & Schumacher, T. N. T cell dysfunction in cancer. *Cancer Cell* **33**, 547–562 (2018).
72. Upadhaya, S., Hubbard-Lucey, V. M. & Yu, J. X. Immuno-oncology drug development forges on despite COVID-19. *Nat. Rev. Drug Discov.* **19**, 751–752 (2020).
73. Ricciuti, B. et al. Genomic and immunophenotypic landscape of acquired resistance to PD-(L)1 blockade in non-small-cell lung cancer. *J. Clin. Oncol.* **42**, 1311–1321 (2024).
74. Ravi, A. et al. Genomic and transcriptomic analysis of checkpoint blockade response in advanced non-small cell lung cancer. *Nat. Genet.* **55**, 807–819 (2023).
75. Chen, J. et al. Single-cell transcriptome and antigen-immunoglobin analysis reveals the diversity of B cells in non-small cell lung cancer. *Genome Biol.* **21**, 152 (2020).
76. Yu, Y. et al. Cancer-associated fibroblasts induce epithelial-mesenchymal transition of breast cancer cells through paracrine TGF-β signalling. *Br. J. Cancer* **110**, 724–732 (2014).
77. Bonde, A. K., Tischler, V., Kumar, S., Soltermann, A. & Schwendener, R. A. Intratumoral macrophages contribute to epithelial–mesenchymal transition in solid tumors. *BMC Cancer* **12**, 35 (2012).
78. Kaiser, A. M. et al. p53 governs an AT1 differentiation programme in lung cancer suppression. *Nature* **619**, 851–859 (2023).
79. Tam, S. Y., Wu, V. W. C. & Law, H. K. W. Hypoxia-induced epithelial–mesenchymal transition in cancers: HIF-1α and beyond. *Front. Oncol.* **10**, 486 (2020).
80. Joseph, J. P., Harishankar, M. K., Pillai, A. A. & Devi, A. Hypoxia induced EMT: a review on the mechanism of tumor progression and metastasis in OSCC. *Oral Oncol.* **80**, 23–32 (2018).

81. Graeber, T. G. et al. Hypoxia-mediated selection of cells with diminished apoptotic potential in solid tumours. *Nature* **379**, 88–91 (1996).
82. Tang, Q., Su, Z., Gu, W. & Rustgi, A. K. Mutant p53 on the path to metastasis. *Trends Cancer* **6**, 62–73 (2020).
83. Capaci, V., Mantovani, F. & Del Sal, G. Amplifying tumor–stroma communication: an emerging oncogenic function of mutant p53. *Front. Oncol.* **10**, 614230 (2020).
84. Grout, J. A. et al. Spatial positioning and matrix programs of cancer-associated fibroblasts promote T-cell exclusion in human lung tumors. *Cancer Discov.* **12**, 2606–2625 (2022).
85. Giotti, B. et al. Single cell view of tumor microenvironment gradients in pleural mesothelioma. *Cancer Discov.* **14**, 2262–2278 (2024).
86. Assoun, S. et al. Association of *TP53* mutations with response and longer survival under immune checkpoint inhibitors in advanced non-small-cell lung cancer. *Lung Cancer* **132**, 65–71 (2019).
87. Rousseau, A., Parisi, C. & Barlesi, F. Anti-TIGIT therapies for solid tumors: a systematic review. *ESMO Open* **8**, 101184 (2023).
88. Jerby-Arnon, L. et al. Opposing immune and genetic mechanisms shape oncogenic programs in synovial sarcoma. *Nat. Med.* **27**, 289–300 (2021).
89. Slyper, M. et al. A single-cell and single-nucleus RNA-seq toolbox for fresh and frozen human tumors. *Nat. Med.* **26**, 792–802 (2020).
90. Mayakonda, A., Lin, D. C., Assenov, Y., Plass, C. & Koeffler, H. P. Maftools: efficient and comprehensive analysis of somatic variants in cancer. *Genome Res.* **28**, 1747–1756 (2018).
91. Talevich, E., Shain, A. H., Botton, T. & Bastian, B. C. CNVkit: genome-wide copy number detection and visualization from targeted DNA sequencing. *PLoS Comput. Biol.* **12**, e1004873 (2016).
92. Hao, Y. et al. Integrated analysis of multimodal single-cell data. *Cell* **184**, 3573–3587 (2021).
93. Liberzon, A. et al. The Molecular Signatures Database (MSigDB) hallmark gene set collection. *Cell Syst.* **1**, 417–425 (2015).
94. Wu, T. et al. clusterProfiler 4.0: a universal enrichment tool for interpreting omics data. *Innovation (Camb.)* **2**, 100141 (2021).
95. Efremova, M., Vento-Tormo, M., Teichmann, S. A. & Vento-Tormo, R. CellPhoneDB: inferring cell–cell communication from combined expression of multi-subunit ligand–receptor complexes. *Nat. Protoc.* **15**, 1484–1506 (2020).
96. Soni, N. et al. Single-cell dissection of the genotype–immuno-phenotype relationship in glioblastoma. *Brain* **148**, 3153–3169 (2025).
97. Gao, J. et al. Integrative analysis of complex cancer genomics and clinical profiles using the cBioPortal. *Sci. Signal.* **6**, pl1 (2013).
98. Gillette, M. A. et al. Proteogenomic characterization reveals therapeutic vulnerabilities in lung adenocarcinoma. *Cell* **182**, 200–225 (2020).
99. Chu, T., Wang, Z., Pe'er, D. & Danko, C. G. Cell type and gene expression deconvolution with BayesPrism enables Bayesian integrative analysis across bulk and single-cell RNA sequencing in oncology. *Nat. Cancer* **3**, 505–517 (2022).
100. Wolf, F. A., Angerer, P. & Theis, F. J. SCANPY: large-scale single-cell gene expression data analysis. *Genome Biol.* **19**, 15 (2018).
101. Palla, G. et al. Squidpy: a scalable framework for spatial omics analysis. *Nat. Methods* **19**, 171–178 (2022).
102. Kleshchevnikov, V. et al. Cell2location maps fine-grained cell types in spatial transcriptomics. *Nat. Biotechnol.* **40**, 661–671 (2022).
103. Griffin, G. K. et al. Spatial signatures identify immune escape via PD-1 as a defining feature of T-cell/histiocyte-rich large B-cell lymphoma. *Blood* **137**, 1353–1364 (2021).
104. Wang, P. et al. *TP53* and *CDKN2A* mutations in patients with early-stage lung squamous cell carcinoma: an analysis of the correlations and prognostic outcomes. *Ann. Transl. Med.* **9**, 1330 (2021).
105. Zhao, W. et al. A cellular and spatial atlas of TP53-associated tissue remodeling defines a multicellular tumor ecosystem in lung adenocarcinoma. *Zenodo* https://doi.org/10.5281/zenodo.16546233 (2025).
106. Zhao, W. et al. TsankovLab/HTAN_Lung: v1.0.1 (v1.0.1). *Zenodo* https://doi.org/10.5281/zenodo.15866295 (2025).

Acknowledgements

We thank L. Gaffney, affiliated with the Broad Institute of Massachusetts Institute of Technology and Harvard, for extensive help in editing and graphical design of the figures in the paper, the Broad Genomics Platform, Broad Flow Cytometry Facility, Pathology and Surgery Departments at Massachusetts General Hospital and Brigham and Women's Hospital and members of the R.B., K.P., S.R., A.N.H., A. Regev, B.E.J. and A.M.T. labs. The project is part of the Human Tumor Atlas Pilot Project (HTAPP) consortium and National Institutes of Health (NIH) HTAN (Human Tumor Atlas Network) consortium paper package. Data collection was supported by the Klarman Cell Observatory and also funded in part by federal funds from the National Cancer Institute (NCI), NIH task order HHSN261100039 under contract HHSN261201500003, and U2CCA233195 (to O.R.-R. and A. Regev). Computational analysis and validations were supported by Department of the Army Lung Cancer Research Program Career Development Award W81XWH2210079, American Cancer Society Research Scholar Grant RSG-23-1039063-01-MM, and Icahn School of Medicine at Mount Sinai (ISMMS) seed funding to A.M.T. NIH grant 2T32GM007280-41 partially supported W.Z. and B.Y.S. as part of the ISMMS Medical Scientist Training Program. The graphical summary was created with BioRender.com.

Author contributions

W.Z., A. Rotem, R.B., O.R.-R., K.P., S.R., A.N.H., A. Regev, B.E.J. and A.M.T. designed the experiments. N.R.M., A.S., N.R.B., S.M.J., C.L., S.V., M.S., M-J.S., I.W. and A.M.T. performed the experiments. W.Z., T.T.N., A.B., A.K., B.G., B.K., J.L.W., B.Y.S. and A.M.T. performed the computational analyses. K.H., A.F., J.J.-V., K.D., J.A-G., O.A., and M.A. provided resources. W.Z. and A.M.T. wrote the paper with input and approval from all authors.

Competing interests

A. Rotem is an equity holder in Celsius Therapeutics and Nucleai. A.N.H. has received research support from Amgen, Blueprint Medicines, BridgeBio, Bristol Myers Squibb, C4 Therapeutics, Pfizer, Eli Lilly, Novartis, Nuvalent, Roche/Genentech and Scorpion Therapeutics and served as a paid consultant for Engine Biosciences, Nuvalent, Oncovalent, Tolremo Therapeutics and TigaTx. R.B. has received research support from the NCI, National Institute of Biomedical Imaging and Bioengineering, National Heart, Lung and Blood Institute and US Department of Defense and industry grants from Genentech, Roche, Merck, Siemens, NorthPond and Bicycles Therapeutics. R.B. holds equity and patents licensed to Navigation Sciences. O.R.-R. has given numerous lectures on the subject of single-cell genomics to a wide variety of audiences and, in some cases, has received remuneration to cover time and costs. A. Regev is a cofounder and equity holder of Celsius Therapeutics and an equity holder in Immunitas. A. Regev was also a scientific advisory board member of Thermo Fisher Scientific, Syros Pharmaceuticals, Neogene Therapeutics and Asimov until July 31, 2020. A. Regev, O.R.-R., J.J.-V. and M.S. have been employees of Genentech since 2020 and have equity in Roche. B.E.J serves as a paid consultant to Novartis, Checkpoint Therapeutics, Astra Zeneca, Daichi Sankyo, GSK, Hummingbird Diagnostics, Genentech, Bluedot Bio, G1 Therapeutics, Jazz Pharmaceuticals, Merus, Abdera and Simcere Pharmaceutical and is a paid member of a data safety monitoring committee for Merck and Revolution Medicine. The other authors declare no competing interests.

Additional information

Extended data is available for this paper at https://doi.org/10.1038/s43018-025-01053-7.

Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s43018-025-01053-7.

Correspondence and requests for materials should be addressed to Alexander M. Tsankov.

Peer review information *Nature Cancer* thanks Laurent Le Cam, Christian Schürch and the other, anonymous, reviewer(s) for their contribution to the peer review of this work.

Reprints and permissions information is available at www.nature.com/reprints.

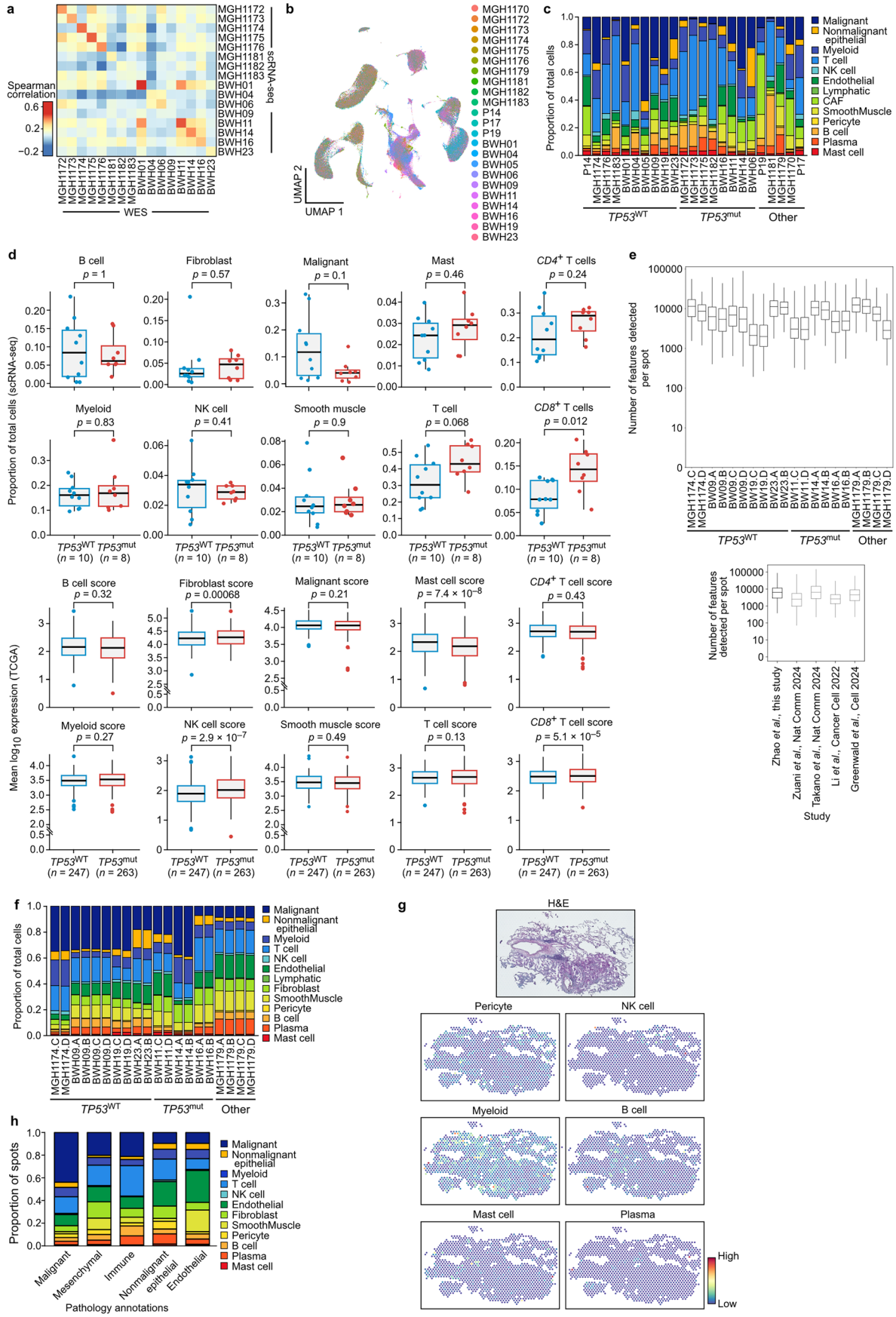
Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Springer Nature or its licensor (e.g. a society or other partner) holds exclusive rights to this article under a publishing agreement with the author(s) or other rightsholder(s); author self-archiving of the accepted manuscript version of this article is solely governed by the terms of such publishing agreement and applicable law.

© The Author(s), under exclusive licence to Springer Nature America, Inc. 2025

[1]Department of Genetics and Genomic Sciences, Icahn School of Medicine at Mount Sinai, New York, NY, USA. [2]Lipschultz Precision Immunology Institute, Icahn School of Medicine at Mount Sinai, New York, NY, USA. [3]Department of Data Science, Dana-Farber Cancer Institute, Boston, MA, USA. [4]Department of Medical Oncology, Dana-Farber Cancer Institute, Boston, MA, USA. [5]Department of Pathology, Brigham and Women's Hospital, Boston, MA, USA. [6]Klarman Cell Observatory, Broad Institute of Massachusetts Institute of Technology and Harvard, Cambridge, MA, USA. [7]Department of Pathology, Molecular and Cell-Based Medicine, Icahn School of Medicine at Mount Sinai, New York, NY, USA. [8]Center for Immuno-Oncology, Dana-Farber Cancer Institute, Boston, MA, USA. [9]Massachusetts General Hospital Cancer Center, Charlestown, MA, USA. [10]Department of Medicine, Harvard Medical School, Boston, MA, USA. [11]Center for Cancer Genomics, Department of Medical Oncology, Dana-Farber Cancer Institute, Boston, MA, USA. [12]Lowe Center for Thoracic Oncology, Dana-Farber Cancer Institute, Boston, MA, USA. [13]Department of Thoracic Surgery, Brigham and Women's Hospital, Boston, MA, USA. [14]Department of Biology, Massachusetts Institute of Technology, Cambridge, MA, USA. [15]Tisch Cancer Institute, Icahn School of Medicine at Mount Sinai, New York, NY, USA. [16]Present address: Ginkgo Bioworks, Boston, MA, USA. [17]Present address: Genentech, San Francisco, CA, USA. [18]Present address: Louis V. Gerstner Jr. Graduate School of Biomedical Sciences at Memorial Sloan Kettering Cancer Center, New York, NY, USA. [19]Present address: Astrazeneca, Waltham, MA, USA. [20]These authors contributed equally: William Zhao, Thinh T. Nguyen, Atharva Bhagwat, Akhil Kumar, Bruno Giotti. ✉e-mail: alexander.tsankov@mssm.edu

https://doi.org/10.1038/s43018-025-01053-7

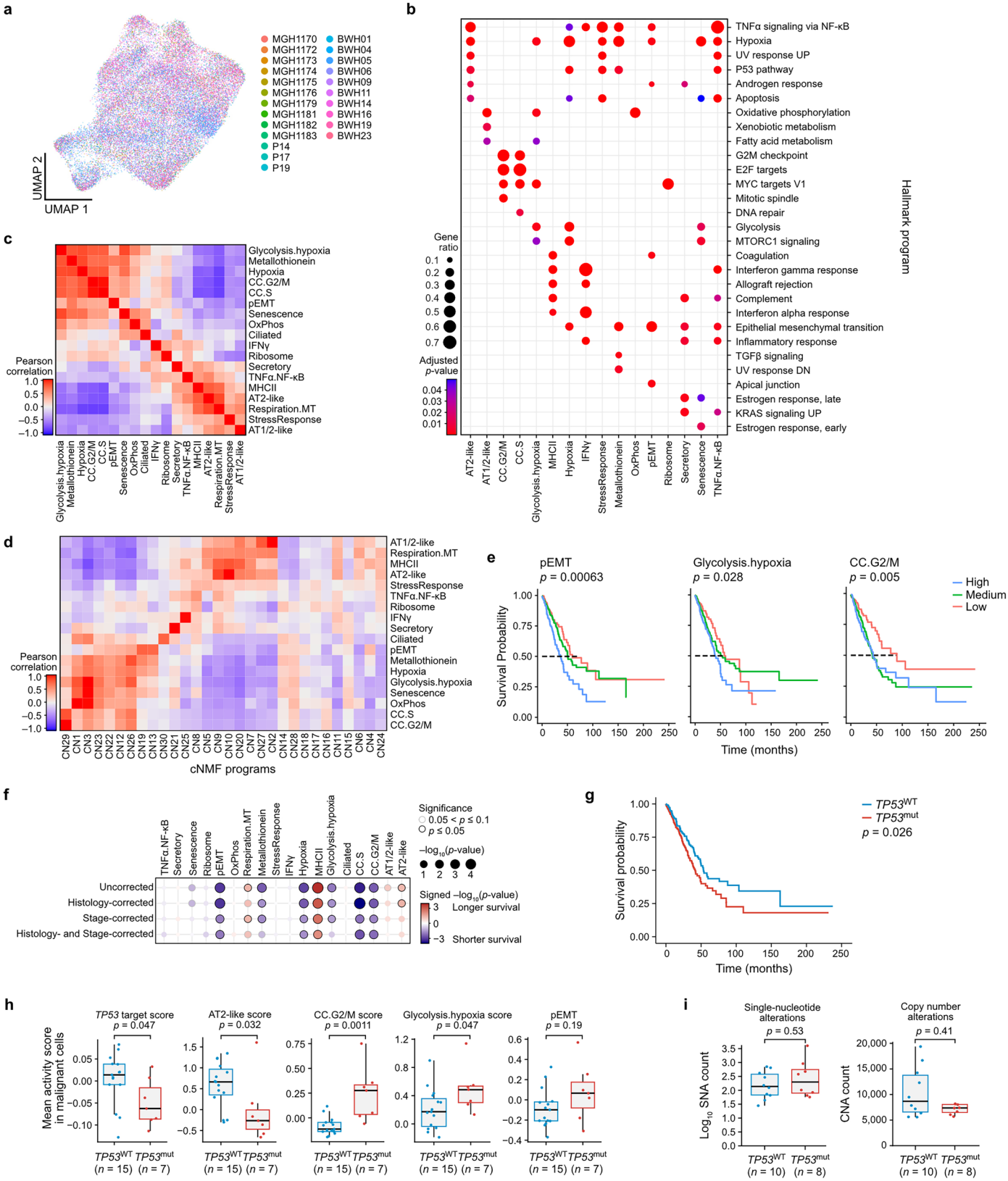


Extended Data Fig. 1 | See next page for caption.

https://doi.org/10.1038/s43018-025-01053-7

Extended Data Fig. 1 | Integration and analysis of WES, scRNA-seq, and spatial transcriptomics data across NSCLC tumors. a, Spearman correlation showing correspondence between the $\log_2$ CNA ratio predicted by scRNA-seq and WES data from the same patients. Correlation was computed across genes for each patient. Low correspondence in two tumor samples (MGH1183 and BWH09) likely resulted from small number of malignant cells (MGH1183) and intratumoral heterogeneity (BWH09). **b**, UMAP of 166,821 cells, colored by patient identity shows integration between samples. **c**, Stacked bar plot displaying the composition of broad cell classes for each scRNA-seq patient samples. **d**, Top: Proportion of broad cell classes (not shown in Fig. 1f) out of all cells found in the scRNA-seq data from *TP53*^{WT} ($n = 10$, blue) and *TP53*^{mut} ($n = 8$, red) LUAD tumors, where *n* refers to number of patients. *P*-values for each plot from left to right, top to bottom: 1, 0.57, 0.1, 0.46, 0.24, 0.83, 0.41, 0.9, 0.068, and 0.012. Bottom: Average $\log_{10}$-transformed expression of cell type specific marker genes (derived from the scRNA-seq data) in TCGA LUAD bulk RNA-seq data separated into *TP53*^{WT} ($n = 247$, blue) and *TP53*^{mut} ($n = 263$, red) tumor samples, where *n* refers to number of patients. *P*-values for each plot from left to right, top to bottom: 0.32, 0.00068, 0.21, 7.4×10^{-8}, 0.43, 0.27, 2.9×10^{-7}, 0.49, 0.13, and 5.1×10^{-5}. *P*-values were calculated using a two-sided Mann-Whitney-Wilcoxon test (scRNA-seq, top) or a two-tailed multiple regression t-test (TCGA, bottom). **e**, Distribution of detected gene features per spot across Visium spatial transcriptomics slides. Top: Box plots showing the distribution of features detected per spot for each Visium spatial transcriptomics (ST) slide (x-axis). Each tumor was sectioned and placed on 2-4 slides with different slides denoted by suffixes A-D. Number of spots for each slide from left to right: 940; 1,076; 4,299; 2,931; 4,359; 4,520; 2,104; 2,576; 960; 1,012; 1,313; 1,304; 3,687; 2,983; 3,335; 2,840; 1,715; 1,740; 1,802; 1,913. Bottom: Number of unique gene features detected per spatial spot in cancer tissues analyzed by Visium ST technology across multiple independent studies. Number of spots from each study from left to right: 47,409; 88,520; 100,082; 38,836; 39,566. Visium ST data in this manuscript is of comparable or superior quality as Visium ST data from several other recently published studies. **f**, Stacked bar plot depicting mean cell composition for each Visium ST slide (x-axis), predicted using Tangram cell deconvolution. **g**, Spatial feature plots of broad cell classes not shown in Fig. 1g, predicted by Tangram mapping. **h**, Cell type composition across pathologist-annotated tissue regions. Stacked bar plots show the mean proportions of Tangram-predicted cell types (colored bars) within spots grouped by pathological annotations. Annotations of H&E stains overlapping ST spots were performed by a board-certified pathologist (N.R.M.) without knowledge of sample and spot identities and classified into five distinct categories: malignant, mesenchymal, immune, nonmalignant epithelial, and endothelial spots. Throughout this figure, box plot horizontal lines show 25th, 50th (median) and 75th percentiles, with vertical whiskers extending to a maximum distance of 1.5 · interquartile range from the hinge. Data beyond the whisker ends are plotted individually.

https://doi.org/10.1038/s43018-025-01053-7

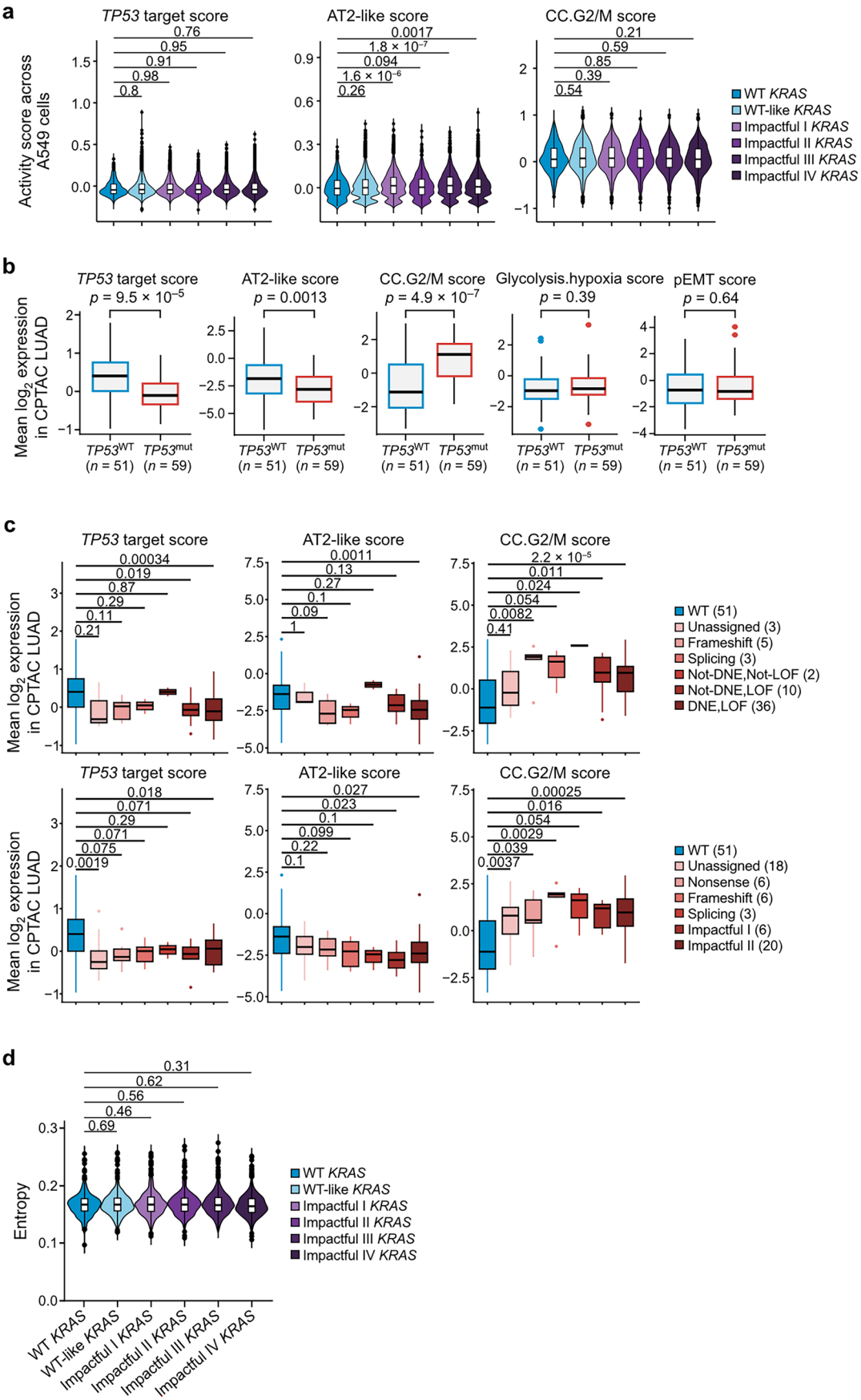


Extended Data Fig. 2 | See next page for caption.

https://doi.org/10.1038/s43018-025-01053-7

Extended Data Fig. 2 | Characterization of the malignant compartment in NSCLC. a, UMAP of 33,377 malignant cells, integrated across tumors and colored by patient. **b**, Dot plot of gene set enrichment outputs (Hallmark) of 100 most differentially expressed markers for each malignant subcluster. **c**, Pearson correlation of average malignant program scores across tumors, analogous to Fig. 2c but performed across patients instead of across cells. **d**, Pearson correlation of malignant program (identified through *de novo* clustering) scores and cNMF program scores across malignant cells. **e**, Kaplan Meier survival curves showing association of malignant program expression with overall survival in bulk TCGA LUAD. Tumors ($n = 510$) were stratified by program expression into high, medium, and low expression categories. *P*-values from left to right: 0.00063, 0.028, and 0.005. *P*-values were calculated using a log-rank test. **f**, Association of malignant program expression with overall survival in TCGA, without correction (top row) as well as after correcting for histology, stage, and histology plus stage (bottom row). Dot color represents the direction and strength of association, computed as the signed significance score ($-\log_{10} p$-value × sign of the Cox *z*-statistic). Red color indicates association with longer overall survival while blue color indicates association with shorter overall survival. Dot size reflects the $-\log_{10} p$-value from the Cox model. *P*-values were calculated using a univariable (uncorrected model) or multivariate Cox regression models. **g**, Kaplan Meier survival curves showing association of *TP53* mutation status with overall survival in bulk TCGA LUAD. $n = 247$ and 263 for $TP53^{WT}$ and $TP53^{mut}$ LUAD patients, respectively. *P*-value (0.026) was calculated using a log-rank test. **h**, Comparison of malignant program mean expression scores between $TP53^{mut}$ ($n = 7$) and $TP53^{WT}$ ($n = 15$) tumors in the scRNA-seq validation cohort, where *n* refers to number of patients. *P*-values for each plot from left to right: 0.047, 0.032, 0.0011, 0.047, and 0.19. *P*-values were calculated using a two-sided Mann-Whitney-Wilcoxon test. **i**, Comparison of $\log_{10}$ SNA and CNA counts between $TP53^{mut}$ ($n = 8$) *vs.* $TP53^{WT}$ ($n = 10$) tumors shows no significant difference in our cohort, where *n* refers to number of patients. *P*-values for each plot from left to right: 0.53 and 0.41. *P*-values were calculated using a two-sided Mann-Whitney-Wilcoxon test. Throughout this figure, box plot horizontal lines show 25th, 50th (median) and 75th percentiles, with vertical whiskers extending to a maximum distance of 1.5 · interquartile range from the hinge. Data beyond the whisker ends are plotted individually.

https://doi.org/10.1038/s43018-025-01053-7

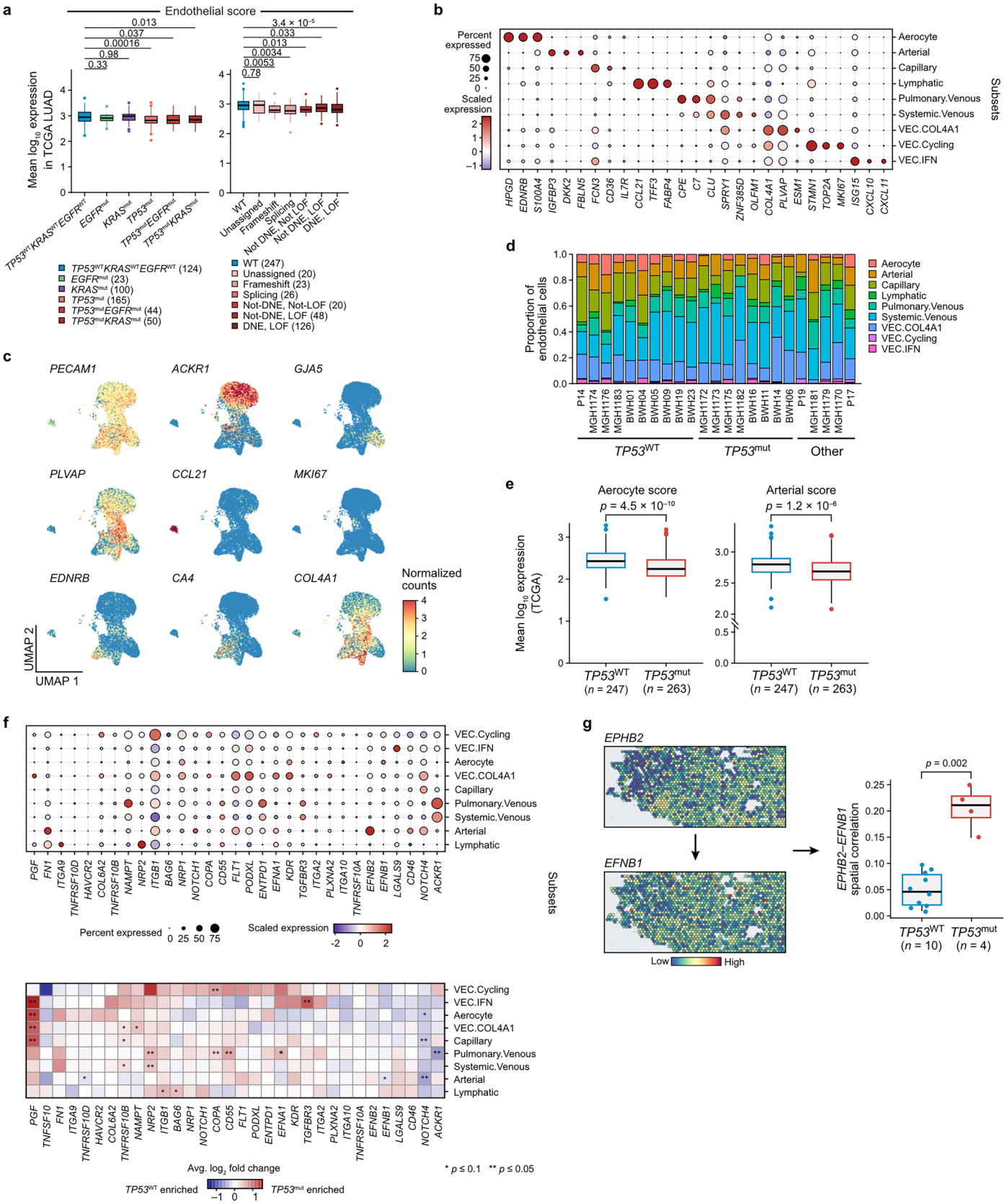


Extended Data Fig. 3 | See next page for caption.

https://doi.org/10.1038/s43018-025-01053-7

Extended Data Fig. 3 | Consistent changes in malignant program expression in *TP53*^{mut} LUAD across variant classes in bulk proteomic data. a, Comparison of malignant program mean expression scores between *KRAS*^{WT} and different *KRAS*^{mut} Impactful categories in A549 cells (from left to right: $n = 644$; 40,104; 11,093; 2,214; 8,740; 19,711, where n refers to number of cells). *P*-values were calculated using a two-sided Mann-Whitney-Wilcoxon test. *P*-values are indicated within the plot. **b**, Comparison of malignant program mean expression between *TP53*^{mut} ($n = 59$) and *TP53*^{WT} ($n = 51$) tumors using bulk proteomic data from CPTAC, where n refers to number of patients. *P*-values for each plot from left to right: 9.5×10^{-5}, 0.0013, 4.9×10^{-7}, 0.39, and 0.64. *P*-values were calculated using a two-sided Mann-Whitney-Wilcoxon test. **c**, Comparison of malignant program mean $\log_2$ expression between *TP53*^{WT} and different dominant-negative/loss of function (top, from left to right: $n = 51, 3, 5, 3, 2, 10$, and 36) and Impactful categories (bottom, from left to right: $n = 51, 18, 6, 6, 3, 6$, and 20) of *TP53*^{mut} tumor samples in bulk proteomic data from CPTAC, where n refers to number of patients. *P*-values were calculated using a two-sided Mann-Whitney-Wilcoxon test. *P*-values are indicated within the plot. **d**, Comparison of entropy scores between *KRAS*^{WT} ($n = 356$ cells subsampled) and different *KRAS*^{mut} Impactful categories ($n = 517$ cells subsampled within each variant category) in A549 cell lines. *P*-values were calculated using a two-sided Mann-Whitney-Wilcoxon test. *P*-values are indicated within the plot. Throughout this figure, box plot horizontal lines show 25th, 50th (median) and 75th percentiles, with vertical whiskers extending to a maximum distance of 1.5 · interquartile range from the hinge. Data beyond the whisker ends are plotted individually.

https://doi.org/10.1038/s43018-025-01053-7

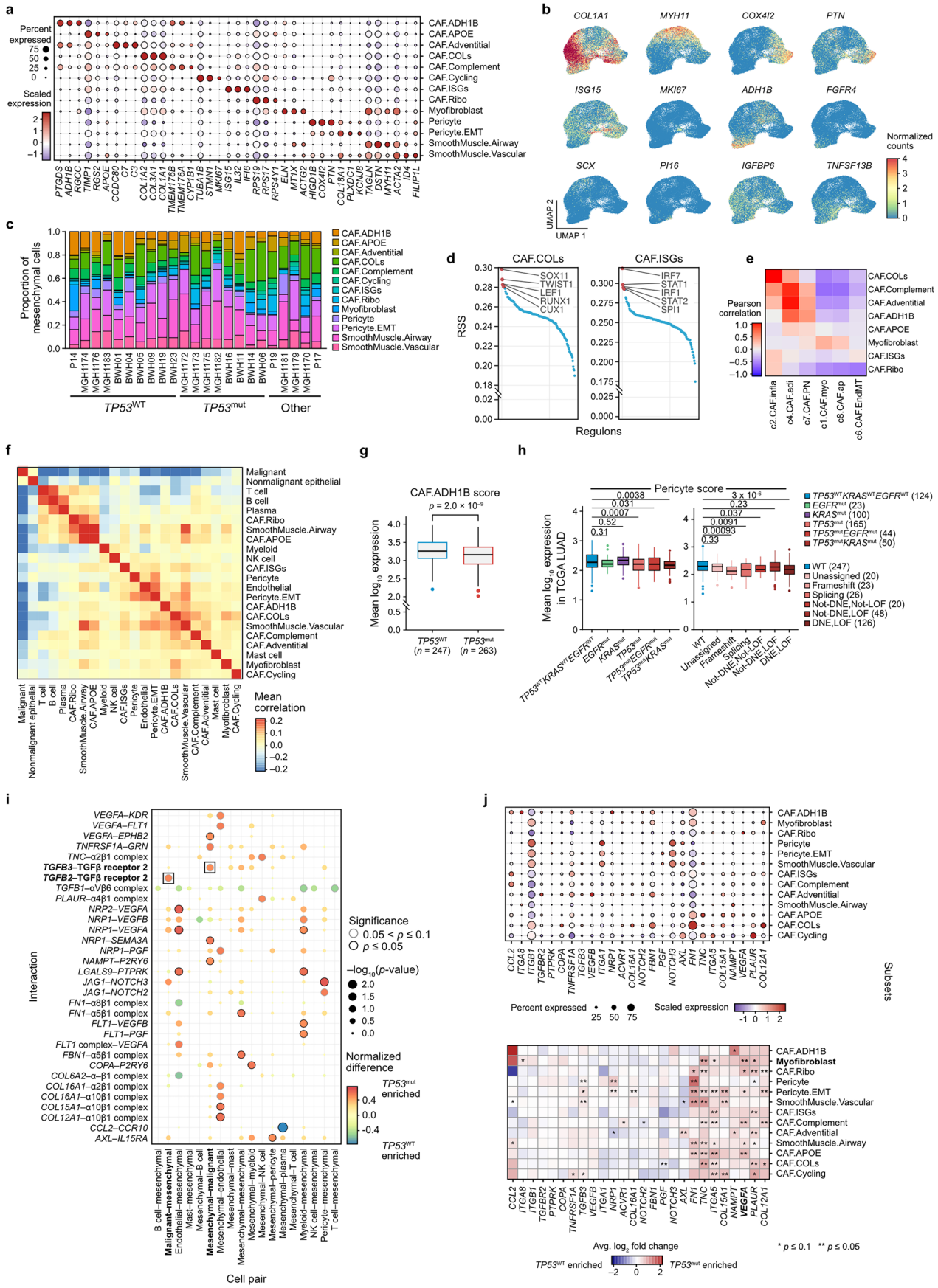


Extended Data Fig. 4 | See next page for caption.

https://doi.org/10.1038/s43018-025-01053-7

Extended Data Fig. 4 | Characterizing endothelial subsets and their differential roles in *TP53*^{WT} and *TP53*^{mut} LUAD. a, Left: Comparison of mean log_{10} endothelial marker expression between *TP53*^{WT} and different single- or co-mutations of *EGFR*, *KRAS*, and *TP53* in bulk RNA-seq data from TCGA tumor samples (from left to right: $n = 124, 23, 100, 165, 44$, and 50, where n refers to number of patients). *P*-values from left to right: 0.33, 0.98, 0.00016, 0.037, and 0.013. Right: Comparison of mean log_{10} endothelial marker expression between *TP53*^{WT} and different dominant-negative/loss of function *TP53*^{mut} variant categories in bulk RNA-seq LUAD tumor samples from TCGA (from left to right: $n = 247, 20, 23, 26, 20, 48$, and 126, where n refers to number of patients). *P*-values from left to right: 0.78, 0.0053, 0.0034, 0.013, 0.033, and 3.4×10^{-5}. *P*-values were calculated using a two-sided Mann-Whitney-Wilcoxon test. **b**, Dot plot showing the expression of three marker genes for each annotated endothelial subset. **c**, Feature plots of the expression of nine selected endothelial subset markers. **d**, Stacked bar plot displaying the cell composition of endothelial subsets for each scRNA-seq patient sample. From left to right: $n = 1{,}892; 377; 241; 209; 284; 206; 628; 1{,}246; 1{,}320; 1{,}841; 114; 116; 358; 12; 1{,}304; 2{,}276; 165; 145; 29; 71; 2{,}022; 322; 142$, where n represents the number of endothelial cells for each patient. **e**, Average log_{10} expression of aerocytes (left) and arterial (right) markers derived from the scRNA-seq data for *TP53*^{WT} ($n = 247$, blue) and *TP53*^{mut} ($n = 263$, red) LUAD bulk RNA-seq tumor samples from TCGA, where n refers to number of patients. *P*-values from left to right: 4.5×10^{-10} and 1.2×10^{-6}. *P*-values were calculated using a two-tailed multiple regression t-test. **f**, Top: Dot plot showing the expression level (color) and percent cells expressing (dot size) genes involved in ligand-receptor interactions displayed in Fig. 4c. Bottom: Pseudobulk differential expression analysis performed on the same genes comparing *TP53*^{mut} *vs.* *TP53*^{WT} tumor samples in our scRNA-seq data. *P*-values were calculated using a two-sided Mann-Whitney-Wilcoxon test. Red indicates enrichment in *TP53*^{mut} and blue enrichment in *TP53*^{WT} tumor samples. **g**, Left: Spatial expression of *EFHB2* and *EFNB1* in a representative *TP53*^{mut} ST sample. Right: Box plot comparing spatial correlation of *EFHB2* and *EFNB1* between *TP53*^{mut} ($n = 4$) and *TP53*^{WT} ($n = 10$) tumor sections, where n refers to number of ST samples. *P*-value (0.002) was calculated using a two-sided Mann-Whitney-Wilcoxon test. Throughout this figure, box plot horizontal lines show 25th, 50th (median) and 75th percentiles, with vertical whiskers extending to a maximum distance of 1.5 · interquartile range from the hinge. Data beyond the whisker ends are plotted individually.

https://doi.org/10.1038/s43018-025-01053-7

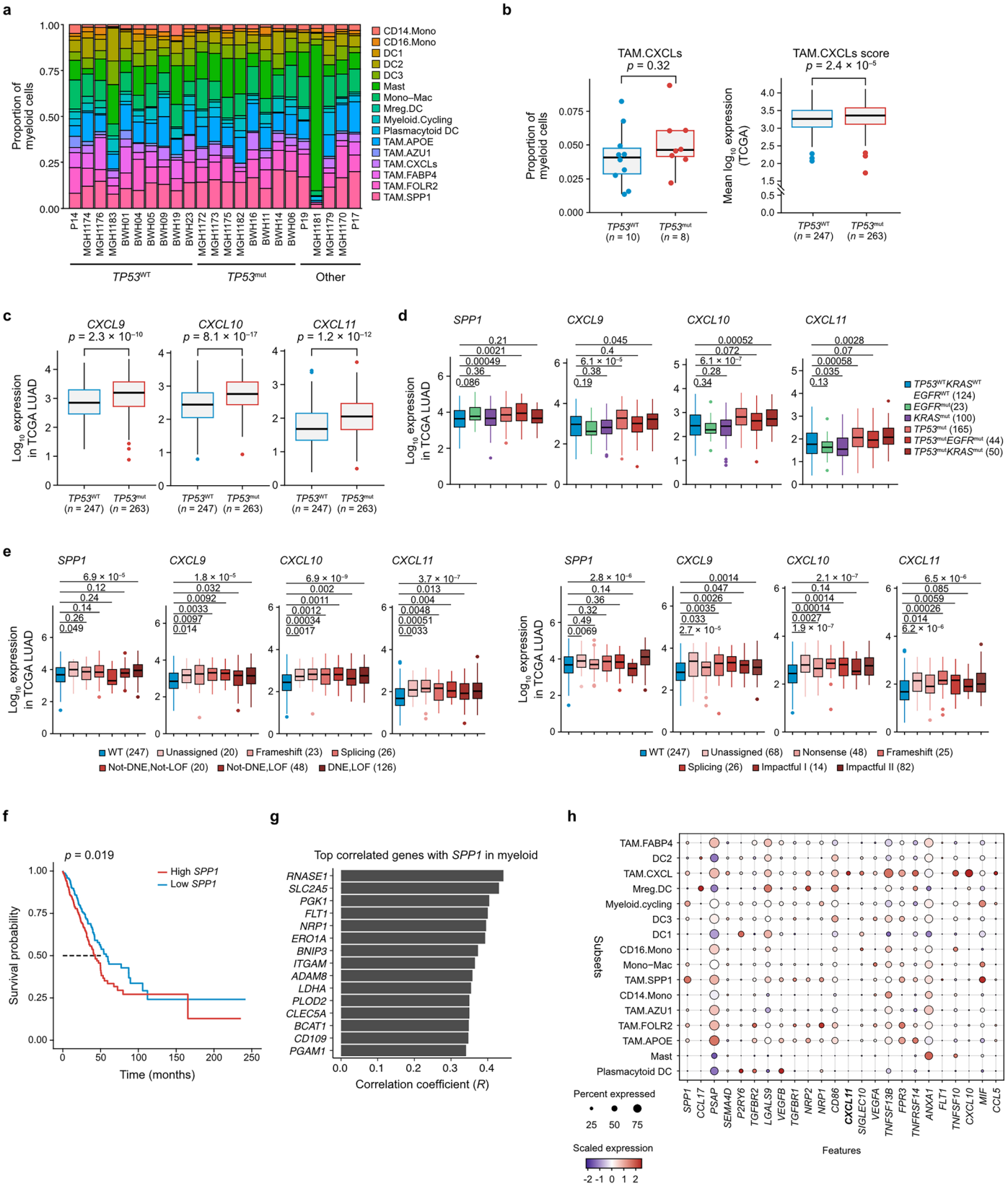


Extended Data Fig. 5 | See next page for caption.

https://doi.org/10.1038/s43018-025-01053-7

Extended Data Fig. 5 | Mesenchymal cell compartment differences between *TP53*^{mut} and *TP53*^{WT} LUAD. a, Dot plot showing the expression of three marker genes for each annotated mesenchymal subset. **b**, Feature plot showing the expression of 12 selected mesenchymal subset markers. **c**, Stacked bar plot displaying the cell composition of mesenchymal subsets for each patient sample. From left to right: n = 2,821; 377; 167; 176; 360; 149; 1,519; 2,226; 1,358; 1,047; 182; 558; 875; 37; 2,502; 2,428; 1,780; 1,244; 8,009; 80; 2,432; 2,015; 375, where n represents the number of mesenchymal cells for each patient. **d**, Regulon specificity scores (RSS) plotted for two CAF subsets of interest. **e**, Pearson correlation of our CAF subset scores (y-axis) and published pan-cancer CAF subset scores (x-axis) across mesenchymal cells. **f**, Spatial correlation (colocalization) of mesenchymal cell subsets with other broad cell classes across all ST tumor samples. Red indicates high colocalization and blue low colocalization between pairs of cell subsets. **g**, Average $\log_{10}$ expression of CAF. ADH1B markers (derived from the scRNA-seq data) for *TP53*^{WT} (n = 247, blue) and *TP53*^{mut} (n = 263, red) LUAD bulk RNA-seq tumor samples from TCGA, where n refers to number of patients. P-value (2.0×10^{-9}) was calculated using a two-tailed multiple regression t-test. **h**, Left: Comparison of mean $\log_{10}$ pericyte marker expression between *TP53*^{WT} and different single- or co-mutations of *EGFR*, *KRAS*, and *TP53* in bulk RNA-seq data from TCGA tumor samples (from left to right: n = 124, 23, 100, 165, 44, and 50, where n refers to number of patients). P-values from left to right: 0.31, 0.52, 0.0007, 0.031, and 0.0038. Right: Comparison of mean $\log_{10}$ pericyte marker expression between *TP53*^{WT} and different dominant-negative/loss of function *TP53*^{mut} variant categories in bulk RNA-seq LUAD tumor samples from TCGA (from left to right: n = 247, 20, 23, 26, 20, 48, and 126, where n refers to number of patients). P-values from left to right: 0.33, 0.00093, 0.0091, 0.037, 0.23, and 3×10^{-6}. P-values were calculated using a two-sided Mann-Whitney-Wilcoxon test. **i**, Dot plot of differential ligand-receptor (rows) interactions between mesenchymal cells and other cell classes (columns) in scRNA-seq data. Red indicates enrichment in *TP53*^{mut} and blue enrichment in *TP53*^{WT} tumor samples. Black outline indicates p-value ≤ 0.05, as assessed by two-sided two-proportion Z-test. **j**, Top: Dot plot showing the expression level (color) and percent cells expressing (dot size) genes involved in ligand-receptor interactions displayed in Extended Data Fig. 5i. Bottom: Pseudobulk differential expression analysis performed on the same genes comparing *TP53*^{mut} *vs*. *TP53*^{WT} tumor samples in scRNA-seq data. P-values were calculated using a two-sided Mann-Whitney-Wilcoxon test. Red indicates enrichment in *TP53*^{mut} and blue enrichment in *TP53*^{WT} tumor samples. Throughout this figure, box plot horizontal lines show 25th, 50th (median) and 75th percentiles, with vertical whiskers extending to a maximum distance of 1.5 · interquartile range from the hinge. Data beyond the whisker ends are plotted individually.

https://doi.org/10.1038/s43018-025-01053-7

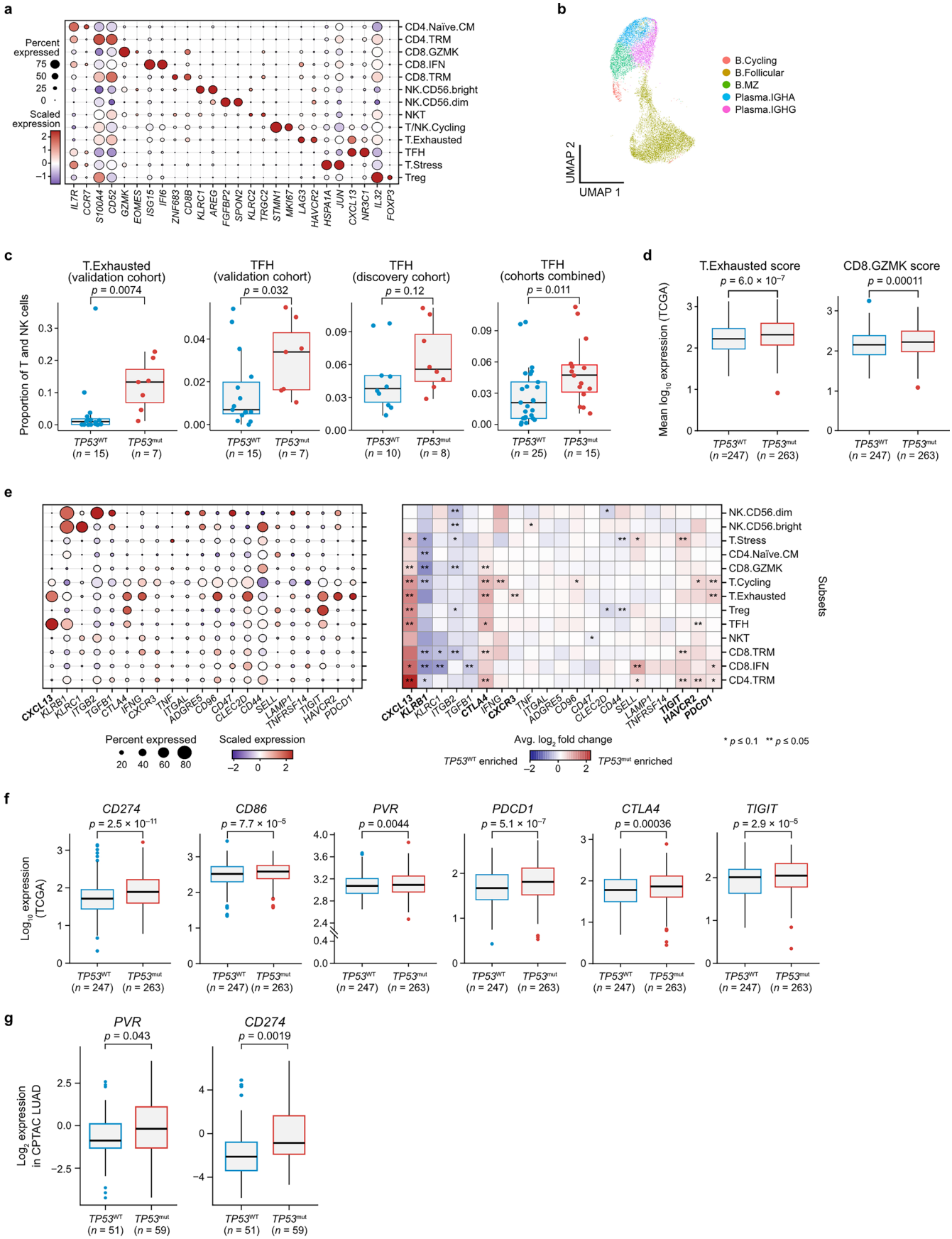


Extended Data Fig. 6 | See next page for caption.

https://doi.org/10.1038/s43018-025-01053-7

Extended Data Fig. 6 | Characterization of myeloid subsets, expression, and ligand-receptor interaction differences between *TP53*^{mut} and *TP53*^{WT} LUAD. **a**, Stacked bar plot displaying the cell composition of myeloid subsets for each patient sample. From left to right: n = 1,828; 1,708; 437; 510; 1,088; 488; 818; 578; 924; 863; 1,307; 1,413; 830; 135; 940; 909; 1,728; 545; 1,466; 189; 680; 1,068; 673, where n represents the number of myeloid cells for each patient. **b**, Left: Proportion of TAM.CXCLs subset relative to all myeloid cells in *TP53*^{WT} (n = 10, blue) *TP53*^{mut} (n = 8, red) LUAD scRNA-seq tumor samples, where n refers to number of patients (P-value = 0.32). Right: Average $\log_{10}$ expression of TAM. CXCLs markers in *TP53*^{WT} (n = 247) and *TP53*^{mut} (n = 263) bulk RNA-seq tumor samples from TCGA, where n refers to number of patients (P-value = 2.4×10^{-5}). P-values were calculated using a two-sided Mann-Whitney-Wilcoxon test (scRNA-seq, left) or a two-tailed multiple regression t-test (TCGA, right). **c**, Comparison of *CXCL9/10/11* expression in *TP53*^{mut} (n = 263) *vs.* *TP53*^{WT} (n = 247) LUAD tumor samples using bulk RNA-seq data from TCGA, where n refers to number of patients, P-values from left to right: 2.3×10^{-10}, 8.1×10^{-17}, and 1.2×10^{-12}. P-values were calculated using a two-tailed multiple regression t-test. **d**, Comparison of *SPP1*, *CXCL9/10/11* gene expression among *TP53*^{WT} and different single- or co-mutations in *EGFR*, *KRAS*, and *TP53* (from left to right: n = 124, 23, 100, 165, 44, and 50, where n refers to number of patients). P-values were calculated using a two-sided Mann-Whitney-Wilcoxon test. P-values are indicated in the plot. **e**, Comparison of gene expression among dominant-negative/loss of function (left, from left to right: n = 247, 20, 23, 26, 20, 48, and 126), and Impactful categories (right, from left to right: n = 247, 68, 48, 25, 26, 14, and 82) of *TP53*^{mut} LUAD tumor samples in TCGA, where n refers to number of patients. P-values were calculated using a two-sided Mann-Whitney-Wilcoxon test. P-values are indicated in the plot. **f**, Kaplan-Meier survival curve showing association between *SPP1* expression and overall survival in bulk RNA-seq LUAD data from TCGA. Tumors (n = 510) were stratified into high or low expression of *SPP*1. P-value (0.019) was calculated using a log-rank test. **g**, Horizontal bar plot showing the most correlated genes with *SPP1* expression in myeloid cells, ranked by Pearson's correlation coefficient. **h**, Dot plot showing the expression level (color) and percent cells expressing (dot size) genes involved in ligand-receptor interactions shown in Fig. 5f. Throughout this figure, box plot horizontal lines show 25th, 50th (median) and 75th percentiles, with vertical whiskers extending to a maximum distance of 1.5 · interquartile range from the hinge. Data beyond the whisker ends are plotted individually.

https://doi.org/10.1038/s43018-025-01053-7

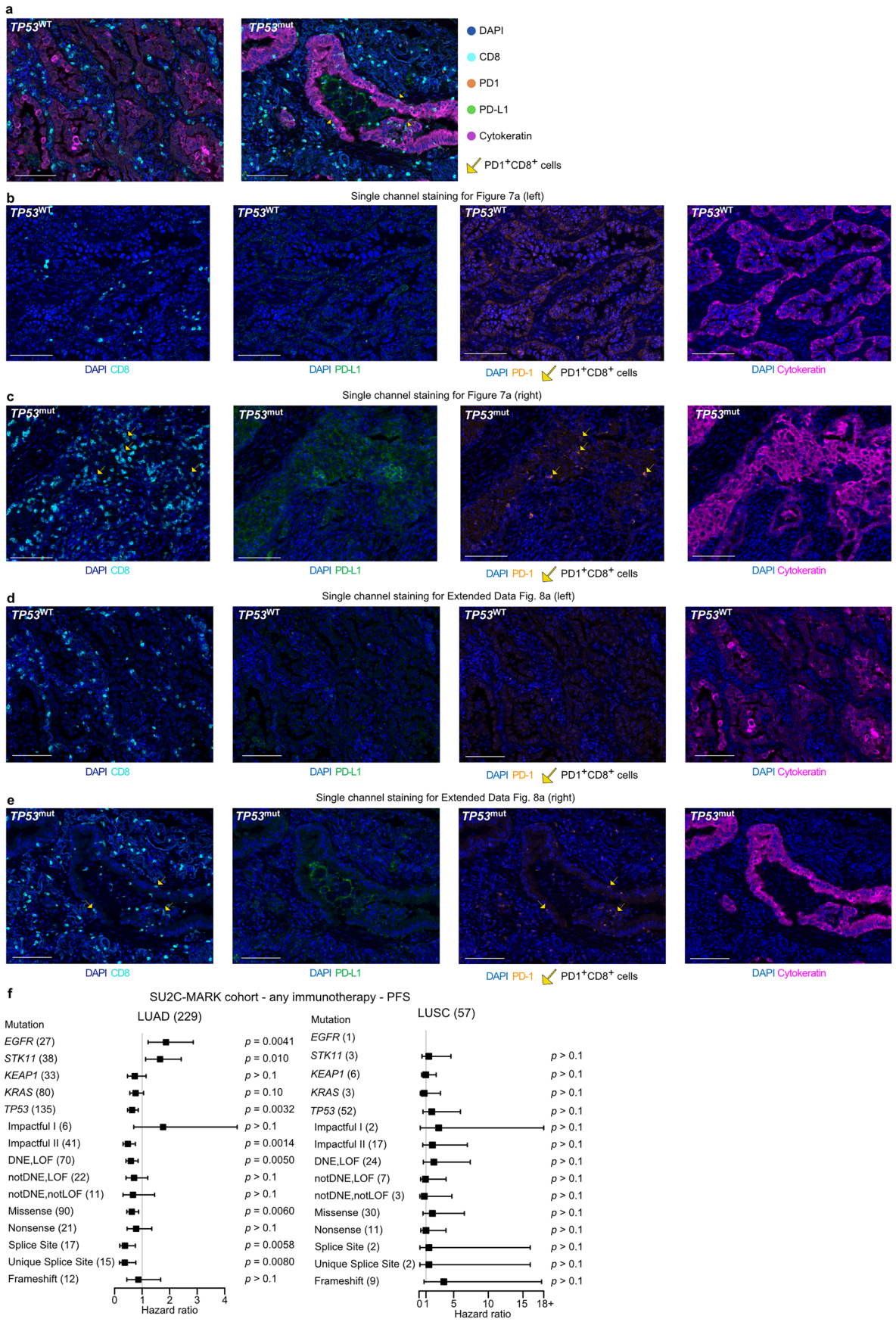


Extended Data Fig. 7 | See next page for caption.

https://doi.org/10.1038/s43018-025-01053-7

Extended Data Fig. 7 | Lymphoid cell diversity, spatial colocalization, and immune checkpoint gene expression differences between *TP53*^{mut} and *TP53*^{WT} LUAD. a, Dot plot showing the expression of two representative marker genes for each annotated T and NK cell subset. **b**, UMAP of all B and plasma cells, integrated across tumors and colored by annotated subset. **c**, Comparison of T.Exhausted (left; validation cohort only) and TFH cell proportions (right; discovery, validation, and combined cohorts) relative to all T and NK cells between *TP53*^{mut} and *TP53*^{WT} tumors profiled by scRNA-seq (from left to right: $n = 15, 15, 10, 25$ for *TP53*^{WT} and $n = 7, 7, 8, 15$ for *TP53*^{mut}, where n is number of patients). *P*-values from left to right: 0.0074, 0.032, 0.12, and 0.011. *P*-values were calculated using a two-sided Mann-Whitney-Wilcoxon test. **d**, Average $\log_{10}$ expression of specific T.Exhausted, and CD8.GZMK markers (derived from the scRNA-seq data) in *TP53*^{WT} ($n = 247$, blue) and *TP53*^{mut} ($n = 263$, red) LUAD bulk RNA-seq tumor samples from TCGA, where n refers to number of patients. *P*-values from left to right: 6.0×10^{-7} and 0.00011. *P*-values were calculated using a two-tailed multiple regression t-test. **e**, Left: Dot plot showing the expression level (color) and percent cells expressing (dot size) genes involved in ligand-receptor analysis displayed in Fig. 6d. Right: Pseudobulk differential expression analysis performed on the same genes comparing *TP53*^{mut} *vs*. *TP53*^{WT} tumor samples in scRNA-seq data. *P*-values were calculated using a two-sided Mann-Whitney-Wilcoxon test. Red indicates enrichment in *TP53*^{mut} and blue enrichment in *TP53*^{WT} tumor samples. **f**, Log_{10} expression of *CD274*, *CD86*, *PVR*, *PDCD1*, *CTLA4*, and *TIGIT* in *TP53*^{WT} ($n = 247$, blue) and *TP53*^{mut} ($n = 263$, red) LUAD bulk RNA-seq tumor samples from TCGA, where n refers to number of patients. *P*-values from left to right: 2.5×10^{-11}, 7.7×10^{-5}, 0.0044, 5.1×10^{-7}, 0.00036, and 2.9×10^{-5}. *P*-values were calculated using a two-tailed multiple regression t-test. **g**, Box plots comparing PVR and CD274 (PD-L1) protein expression between *TP53*^{WT} ($n = 51$) and *TP53*^{mut} ($n = 59$) tumors from CPTAC, where n refers to number of patients. *P*-values from left to right: 0.043 and 0.0019. *P*-values were calculated using a two-sided Mann-Whitney-Wilcoxon test. Throughout this figure, box plot horizontal lines show 25th, 50th (median) and 75th percentiles, with vertical whiskers extending to a maximum distance of 1.5 · interquartile range from the hinge. Data beyond the whisker ends are plotted individually.

https://doi.org/10.1038/s43018-025-01053-7

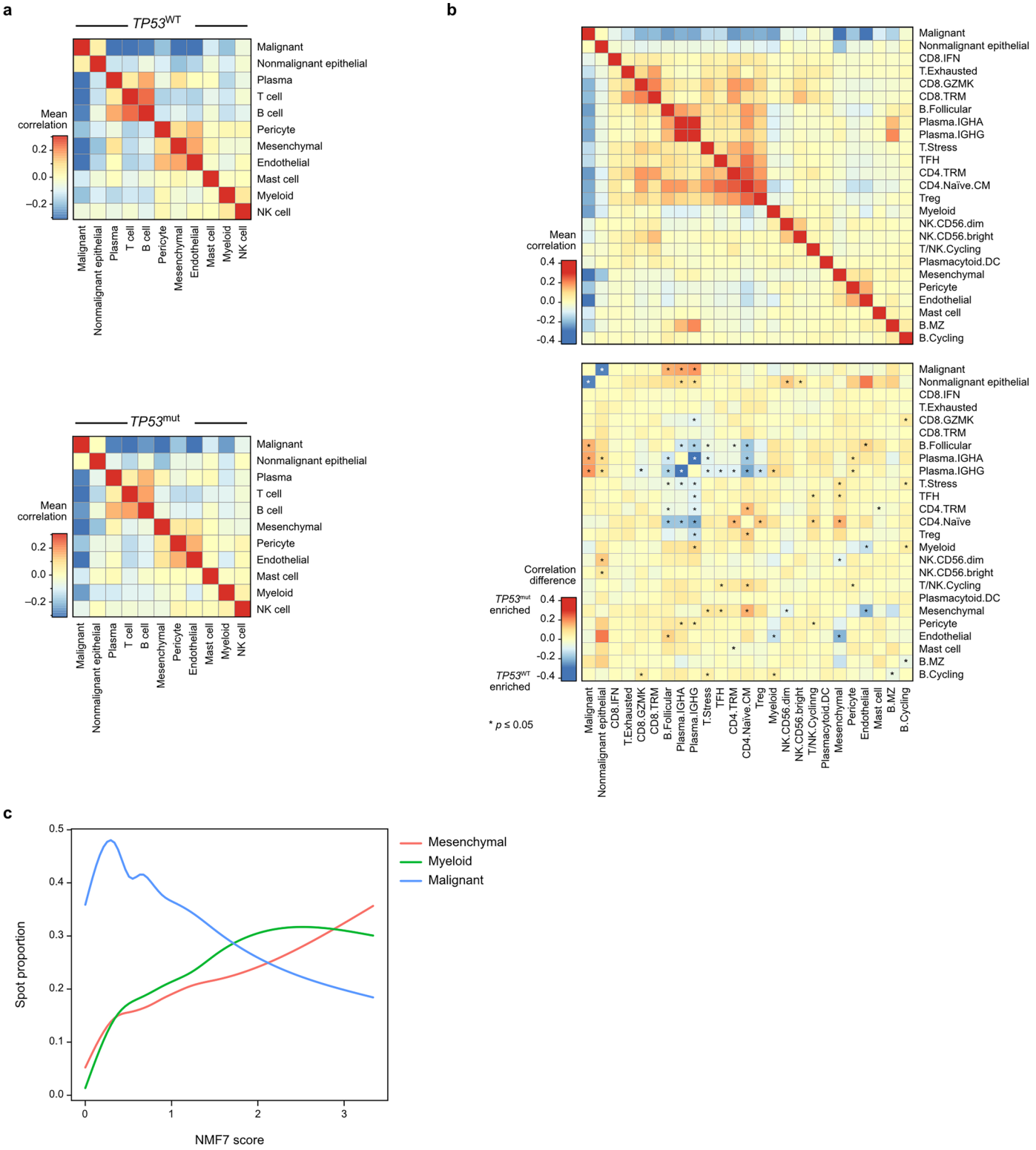


Extended Data Fig. 8 | See next page for caption.

https://doi.org/10.1038/s43018-025-01053-7

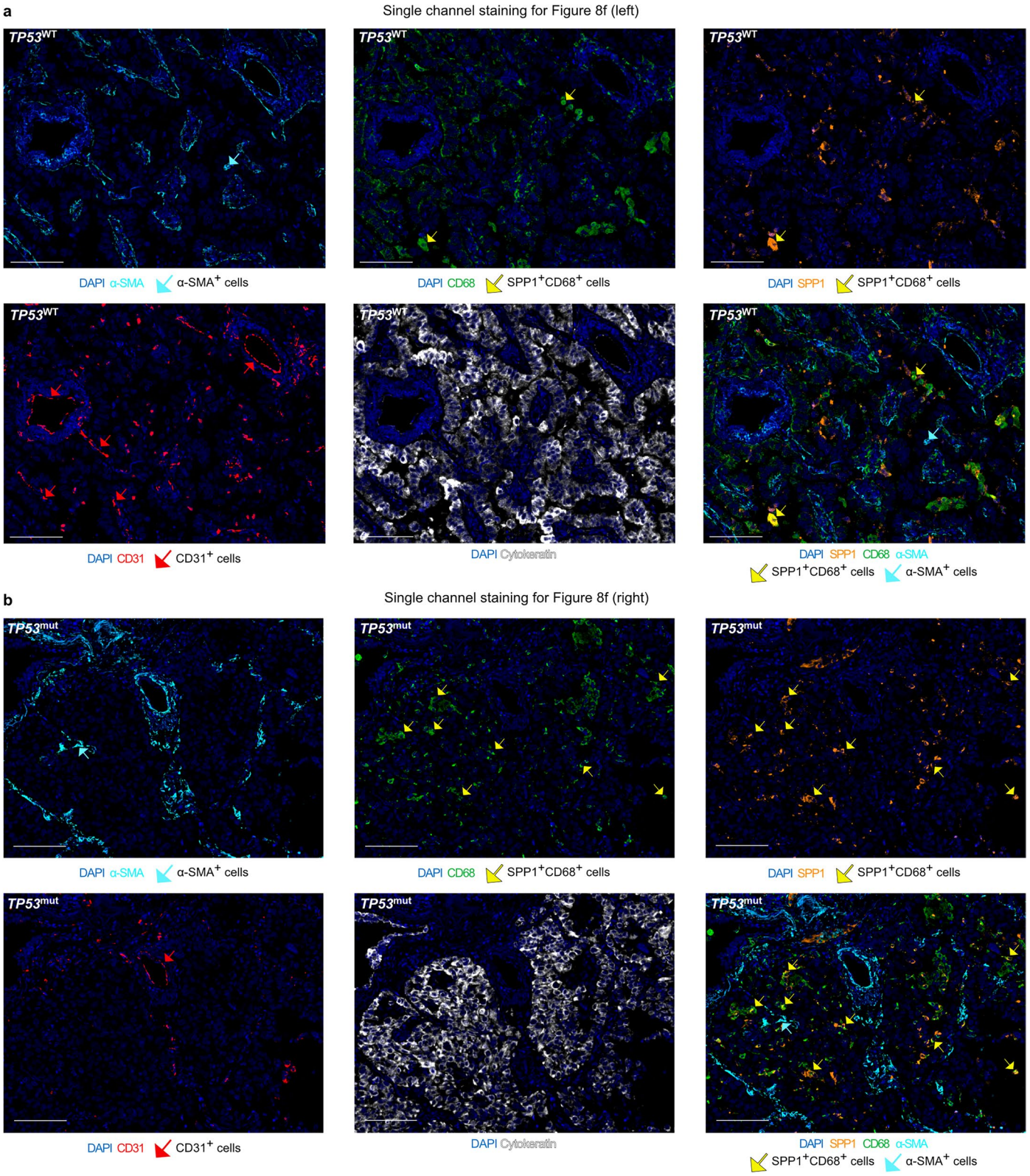
Extended Data Fig. 8 | PD-1+CD8+ cells show increased colocalization with PD-L1+cytokeratin+ cells in *TP53*^{mut} LUAD tumors in additional mIF samples, with SU2C-MARK cohort analysis revealing improved immunotherapy response in *TP53*^{mut} LUAD. a, Representative mIF images of *TP53*^{WT} (left; BWH04) and *TP53*^{mut} (right; BWH11) tumors. Yellow arrows correspond to PD-1^{+}CD8^{+} cells. Scale bar, 100 µm. **b**, Single channel staining for image shown in Fig. 7a, left panel (BWH01). Scale bar, 100 µm. **c**, Single channel staining for image shown in Fig. 7a, right panel (BWH06). Yellow arrows correspond to the same PD1^{+}CD8^{+} cells as in Fig. 7a (right). Scale bar, 100 µm. **d**, Single channel staining for image shown in Extended Data Fig. 8a left panel (BWH04). Scale bar, 100 µm. **e**, Single channel staining for image shown in Extended Data Fig. 8a right panel (BWH11). Yellow arrows correspond to the same PD1^{+}CD8^{+} cells as in Extended Data Fig. 8a (right). Scale bar, 100 µm. **f**, Association with progression free survival (PFS) for patients with different mutations frequently encountered in LUAD (*EGFR*, *STK11*, *KEAP1*, *KRAS, TP53* also split by classes of mutations) following any immune checkpoint blockade therapy in the SU2C-MARK cohort. The analysis was performed separately for LUAD (left) and LUSC (right) patients. Hazard ratios (squares, where lines indicate 95% confidence intervals) and *p*-values were determined using univariate Cox regression models comparing each mutation category and its respective wildtype category. The numbers in brackets indicate the number of patients within a mutation category.

https://doi.org/10.1038/s43018-025-01053-7



Extended Data Fig. 9 | Spatial organization and cellular landmarks in *TP53*^{mut} and *TP53*^{WT} LUAD. a, Mean spatial correlation of broad cell classes across ST spots averaged across *TP53*^{WT} (top) and *TP53*^{mut} (bottom) samples. **b**, Mean (top) and differential (bottom) spatial correlation heatmap of lymphoid (B, plasma, T, NK) subsets with other broad cell classes between *TP53*^{mut} and *TP53*^{WT} in the Visium ST data. Red indicates either high overall colocalization (top) or colocalization enrichment in *TP53*^{mut} (bottom) while blue represents either low overall colocalization (top) or colocalization enrichment in *TP53*^{WT} (bottom) tumor samples between pairs of cell subsets. **c**, Line plot showing proportion of malignant, myeloid, and mesenchymal cells ordered from NMF7 low spots to NMF7 high spots (x-axis) in the Visium ST data.

https://doi.org/10.1038/s43018-025-01053-7



Extended Data Fig. 10 | Representative mIF single channel staining, related to Fig. 8f. a, Single channel staining for image shown in Fig. 8f left panel (BWH04). Red arrows correspond to the same $CD31^+$ cells, yellow arrows correspond to the same $SPP1^+CD68^+$ cells and light blue arrows correspond to the same $\alpha\text{-}SMA^+$ cells as in Fig. 8f left panel (BWH04). Scale bar, 100 µm. **b**, Single channel staining for image shown in Fig. 8f right panel (BWH14). Red arrows correspond to the same $CD31^+$ cells, yellow arrows correspond to the same $SPP1^+CD68^+$ cells and light blue arrows correspond to the same $\alpha\text{-}SMA^+$ cells as in Fig. 8f right panel (BWH14). Scale bar, 100 µm.

nature portfolio

Corresponding author(s): Tsankov, Alexander M.

Last updated by author(s): Jul 31, 2025

Reporting Summary

Nature Portfolio wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Portfolio policies, see our Editorial Policies and the Editorial Policy Checklist.

Statistics

For all statistical analyses, confirm that the following items are present in the figure legend, table legend, main text, or Methods section.

n/a	Confirmed	
☐	☒	The exact sample size (*n*) for each experimental group/condition, given as a discrete number and unit of measurement
☐	☒	A statement on whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
☐	☒	The statistical test(s) used AND whether they are one- or two-sided *Only common tests should be described solely by name; describe more complex techniques in the Methods section.*
☐	☒	A description of all covariates tested
☐	☒	A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons
☐	☒	A full description of the statistical parameters including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals)
☐	☒	For null hypothesis testing, the test statistic (e.g. *F*, *t*, *r*) with confidence intervals, effect sizes, degrees of freedom and *P* value noted *Give P values as exact values whenever suitable.*
☒	☐	For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings
☒	☐	For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes
☐	☒	Estimates of effect sizes (e.g. Cohen's *d*, Pearson's *r*), indicating how they were calculated

Our web collection on statistics for biologists contains articles on many of the points above.

Software and code

Policy information about availability of computer code

Data collection	CellRanger 3.1.0
Data analysis	R (v4.1.0) Python (v3.7) Picard (http://broadinstitute.github.io/picard/) Mutect2 (https://software.broadinstitute.org/gatk/documentation/tooldocs/current/org_broadinstitute_gatk_tools_walkers_cancer_m2_MuTect2) Funcotator (https://gatk.broadinstitute.org/hc/en-us/articles/360035889931-Funcotator-Information-and-Tutorial) maftools (https://bioconductor.org/packages/release/bioc/html/maftools.html) cnvkit (https://cnvkit.readthedocs.io/en/stable/) inferCNV (https://github.com/broadinstitute/inferCNV) Seurat (v4) R package (https://satijalab.org/seurat/) cNMF (v1.3.4) (https://github.com/dylkot/cNMF) MSigDB R package (https://www.gsea-msigdb.org/gsea/msigdb/) ComplexHeatmap (v2.10.0) R package (https://bioconductor.org/packages/release/bioc/html/ComplexHeatmap.html) clusterProfiler (v4.2.2) R package (https://bioconductor.org/packages/release/bioc/html/clusterProfiler.html) SCENT (https://github.com/aet21/SCENT) ggpubr (https://rdrr.io/cran/ggpubr/) ggplot2 (https://ggplot2.tidyverse.org) cgdsr R package (https://rdrr.io/cran/cgdsr) survival R package (https://rdrr.io/cran/survival/)

CellPhoneDB (v2.1.7) (https://github.com/Teichlab/cellphonedb)
pySCENIC (v0.11.2) (https://www.10xgenomics.com/products/loupe-browser)
scanpy (v1.8.1) (https://github.com/theislab/scanpy)
squidpy (v1.1.0) (https://github.com/scverse/squidpy)
cell2location (v0.1) (https://github.com/BayraktarLab/cell2location)
pythologist (https://github.com/jason-weirather/pythologist)
Custom analysis code, which is available on Github (https://github.com/TsankovLab/HTAN_Lung)

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors and reviewers. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Portfolio guidelines for submitting code & software for further information.

Data

Policy information about availability of data

All manuscripts must include a data availability statement. This statement should provide the following information, where applicable:

- Accession codes, unique identifiers, or web links for publicly available datasets
- A description of any restrictions on data availability
- For clinical datasets or third party data, please ensure that the statement adheres to our policy

In collaboration with the NIH-funded HTAN Data Coordinating Center (U24), WES, scRNA-seq, and spatial transcriptomics data is available as an interactive, online platform for download, independent visualization, and analysis at https://data.humantumoratlas.org/explore/, select Atlas "HTAN HTAPP" and Organ "Lung" from the scroll down menus. Processed scRNA-seq data, spatial transcriptomic data, and mutation information can be downloaded at https://doi.org/10.5281/zenodo.16546233.

Research involving human participants, their data, or biological material

Policy information about studies with human participants or human data. See also policy information about sex, gender (identity/presentation), and sexual orientation and race, ethnicity and racism.

Reporting on sex and gender	Sex was determined by self-report, where 52% (12/23) of patients in our cohort and 54% (274/510) of patients in the bulk RNA-seq validation cohort were female. Sex/gender were not considered in the study design.
Reporting on race, ethnicity, or other socially relevant groupings	Information on race, ethnicity, or other socially relevant groupings was not collected in this study.
Population characteristics	Figure 1A and Supplementary Table 1.
Recruitment	Patients undergoing NSCLC surgery have provided written consent for use of disposed treatment naive tissue for this study.
Ethics oversight	Mass General Brigham Human Research Protection Program (HRPP) ethics committee at Harvard Medical school approved this research study. Written informed consent was obtained from all human participants for sharing de-identified clinical information.

Note that full information on the approval of the study protocol must also be provided in the manuscript.

Field-specific reporting

Please select the one below that is the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

☒ Life sciences ☐ Behavioural & social sciences ☐ Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see nature.com/documents/nr-reporting-summary-flat.pdf

Life sciences study design

All studies must disclose on these points even when the disclosure is negative.

Sample size	23 NSCLC tumors, including 8 TP53mut and 10 TP53WT LUAD tumors.
Data exclusions	We excluded scRNA-seq and Visium samples with LUSC or atypical adenocarcinoma histological subtypes (mucinous, neuroendocrine, colloid). We furthermore excluded one TP53mut Visium sample (BWH16) from our analyses due to insufficient predicted proportion of malignant cells.
Replication	We believe in the reproducibility of our findings due to the consistent results in large independent cohorts from TCGA, CPTAC, and other published scRNA-seq studies. Furthermore, our findings in scRNA-seq data are also supported by spatial transcriptomic and mIF data, as well as by data from LUAD cell lines with various TP53 and KRAS mutations, K, and KP LUAD mouse models.
Randomization	We did not use randomization in this study, as control and experimental groups were determined by TP53 mutation status in the human tumors, which was unknown at the time of tumor resection.

Blinding | A board-certified pathologist was blinded to our data/results for their histological annotations and pathology report.

Reporting for specific materials, systems and methods

We require information from authors about some types of materials, experimental systems and methods used in many studies. Here, indicate whether each material, system or method listed is relevant to your study. If you are not sure if a list item applies to your research, read the appropriate section before selecting a response.

Materials & experimental systems

n/a	Involved in the study
☐	☒ Antibodies
☒	☐ Eukaryotic cell lines
☒	☐ Palaeontology and archaeology
☒	☐ Animals and other organisms
☒	☐ Clinical data
☒	☐ Dual use research of concern
☒	☐ Plants

Methods

n/a	Involved in the study
☒	☐ ChIP-seq
☒	☐ Flow cytometry
☒	☐ MRI-based neuroimaging

Antibodies

Antibodies used

We used antibodies for our mIF validation experiments:
CD8 (clone 41BB); Leica; Cat#CD8-4B11-L-CE; 1:200 dilution
PD-L1 (clone E1L3N); Cell Signaling; Cat#13684; 1:300 dilution
PD-1 (clone EPR4877(2)); Abcam; Cat#Ab137132; 1:300 dilution
Cytokeratin (clone AE1/AE3); Dako; Cat#GA053; 1:100 dilution
a-SMA (clone D4K9N) Cell Signaling; Cat#34105; 1:200 dilution
CD68 (clone D4B9C); Cell Signaling; Cat#76437; 1:3500 dilution
OPN/SPP1 (clone EPR21139-316); Abcam; Cat#ab214050; 1:250 dilution
CD31 (polyclonal); Abcam; Cat#ab28364; 1:250 dilution

Validation

Below we include respective information on antibody, cat. manufacturer, species, and validation information:
CD8 (clone 41BB); Leica; Cat#CD8-4B11-L-CE; mouse mAb; tested by Leica; normal tissue (T-cells in thymus, spleen, lymph nodes and tonsil); malignant tissue (numerous solid tumors)
PD-L1 (clone E1L3N); Cell Signaling; Cat#13684; rabbit mAb; tested by us; malignant tissue (LUAD)
PD-1 (clone EPR4877(2)); Abcam; Cat#Ab137132; rabbit mAb; tested by Abcam; normal tissue (tonsil, brain, skeletal muscle); malignant tissue (lymphoma)
Cytokeratin (clone AE1/AE3); Dako; Cat#GA053; mouse mAb; tested by Dako; normal tissue (tonsil)
a-SMA (clone D4K9N) Cell Signaling; Cat#34105; rabbit mAb;tested by Cell Signaling; malignant tissue (NSCLC)
CD68 (clone D4B9C); Cell Signaling; Cat#76437; rabbit mAb; tested by Cell Signaling; malignant tissue (LUAD)
OPN/SPP1 (clone EPR21139-316); Abcam; Cat#ab214050; rabbit mAb; tested by us and Abcam; malignant tissue (LUAD), U-87 MG cell line
CD31 (polyclonal); Abcam; Cat#ab28364; rabbit pAbs; tested by Abcam; normal tissue (tonsil, lung, skin)

Plants

Seed stocks

Report on the source of all seed stocks or other plant material used. If applicable, state the seed stock centre and catalogue number. If plant specimens were collected from the field, describe the collection location, date and sampling procedures.

Novel plant genotypes

Describe the methods by which all novel plant genotypes were produced. This includes those generated by transgenic approaches, gene editing, chemical/radiation-based mutagenesis and hybridization. For transgenic lines, describe the transformation method, the number of independent lines analyzed and the generation upon which experiments were performed. For gene-edited lines, describe the editor used, the endogenous sequence targeted for editing, the targeting guide RNA sequence (if applicable) and how the editor was applied.

Authentication

Describe any authentication procedures for each seed stock used or novel genotype generated. Describe any experiments used to assess the effect of a mutation and, where applicable, how potential secondary effects (e.g. second site T-DNA insertions, mosiacism, off-target gene editing) were examined.