Title

Conserved principles of spatial biology define tumor heterogeneity and response to immunotherapy

Authors

Vivek Behera[1,2], Hannah Giba[2,3], Ue-Yu Pen[2,3], Anna Di Lello[1], Benjamin A. Doran[2,4], Alessandra Esposito[1], Apameh Pezeshk[1], Christine M. Bestvina[1], Justin Kline[1], Marina C. Garassino[1], Arjun S. Raman[2,3,5]

Affiliations

[1]Department of Medicine, Section of Hematology/Oncology, University of Chicago, Chicago, IL, 60637
[2]Duchossois Family Institute, University of Chicago, Chicago, IL, 60637
[3]Department of Pathology, University of Chicago, Chicago, IL, 60637
[4]Pritzker School of Molecular Engineering, University of Chicago, Chicago, IL, 60637
[5]Center for the Physics of Evolving Systems, University of Chicago, Chicago, IL, 60637

Abstract

The tumor microenvironment (TME) is an immensely complex ecosystem[1,2]. This complexity underlies difficulties in elucidating principles of spatial organization and using molecular profiling of the TME for clinical use[3]. Through statistical analysis of 96 spatial transcriptomic (ST-seq) datasets spanning twelve diverse tumor types, we found a conserved distribution of multicellular, transcriptionally covarying units termed 'Spatial Groups' (SGs). SGs were either dependent on a hierarchical local spatial context – enriched for cell-extrinsic processes such as immune regulation and signal transduction – or independent from local spatial context – enriched for cell-intrinsic processes such as protein and RNA metabolism, DNA repair, and cell cycle regulation. We used SGs to define a measure of gene spatial heterogeneity – 'spatial lability' – and categorized all 96 tumors by their TME spatial lability profiles. The resulting classification captured spatial variation in cell-extrinsic versus cell-intrinsic biology and motivated class-specific strategies for therapeutic intervention. Using this classification to characterize pre-treatment biopsy samples of 16 non-small cell lung cancer (NSCLC) patients outside our database distinguished responders and non-responders to immune checkpoint blockade while programmed death-ligand 1 (PD-L1) status and spatially unaware bulk transcriptional markers did not. Our findings show conserved principles of TME spatial biology that are both biologically and clinically significant.

Main

The tumor microenvironment (TME) is a complex milieu of interacting cells, proteins, and other biological components that influences critical properties of tumor biology such as growth, metastasis, and response to therapy[1,2]. Biological variation within the TME reflects clinically relevant differences across genetic, pathway, cellular, and tissue-level scales[4,5]. For instance, recent studies have shown the prognostic and predictive power of TME-specific biomarkers such as tumor infiltrating lymphocyte (TIL) score in melanoma and 'Immunoscore' – the spatial balance of CD3+ and CD8+ T cell density – in colorectal cancer[6–10]. These and other similar findings have motivated significant investment in studying the TME as an ecosystem of cells interacting within the spatial constraints of a tumor, most notably with technologies that couple cellular information about RNA or protein levels with cellular spatial locations[11]. Such spatial molecular profiling studies conducted in a variety of tumor types have revealed a common theme: the substantial heterogeneity within tumors (intratumoral) and across tumors (intertumoral) makes elucidating organizing principles of the TME very challenging[3]. By extension, the clinical utility of TME spatial profiling has been limited in scope.

Recent efforts have begun to outline a strategy to learn conserved aspects of TME spatial biology with the idea that these aspects reflect organizing principles of biological interest. These studies have collectively demonstrated the existence of recurrent multicellular spatial structures associated with tumor biology – somatic mutations, cell cycle synchrony, invasive fronts – and with cancer prognosis[12–17]. Obtaining these insights relied on imaging-based technologies that query tens of proteins to identify phenotypes such as cell type, cell cycle state, and a limited set of cell functional states. While these studies have been invaluable in demonstrating the relevance of spatial organization for TME biology, it has remained unclear whether a broader and more unbiased assessment of cellular phenotypes might demonstrate general principles of TME spatial organization. Spatial transcriptomics ('ST-seq') and related technologies, which provide genome-wide transcriptional information coupled to nearly single-

cell-resolution of spatial coordinates, enable broad and unbiased assessment of TME spatial biology. However, the complexity of such data has precluded moving beyond mere description into an elucidation of spatial biology principles[18].

Advances in statistical inference developed in other fields of biology – protein science, genomics, and microbiome science – provide useful frameworks for addressing this challenge. For instance, at the scale of proteins, analysis of conserved amino acid covariation within ensembles of related proteins has yielded protein 'sectors' – groups of amino acids that are critical for engineering synthetically folded and functional proteins[19–22]. At the scale of genomes, covariation analysis of gene content across extant diversity within kingdoms of life has revealed units of collective protein-protein interactions that are critical for behavior and organismal fitness[23–26]. At the scale of microbiomes, covariation between bacterial taxa across individuals has yielded 'ecogroups' – groups of taxa that are of functional and clinical significance amongst humans[27–29]. Thus, these studies have established a general strategy for parsing organization amongst complex biological systems: first identify an ensemble of systems, then statistically deduce features that are conserved across the ensemble.

Using such studies as inspiration, we hypothesized that statistical analysis of ST-seq data across a diverse ensemble of solid tumors – a 'pan-tumor' database – would reveal conserved patterns of TME spatial biology in an unbiased manner. Our results showed that all TMEs shared the presence of multicellular groups of transcriptionally covarying spots, 'Spatial Groups' (SGs), with expression profiles that are either dependent (defined as 'nested Spatial Groups', NSGs) or independent (defined as 'non-nested Spatial Groups', non-NSGs) on their local spatial environments. We found that NSG biology obeys a characteristic pattern: variation in local-scale biological processes, such as cell adhesion, are nested within the spatial context of larger-scale processes, such as T cell infiltration. We compressed SGs into a tumor-wide measure of spatial heterogeneity in gene expression that we termed 'spatial lability'. This enabled the comparison of spatial biology across our ensemble of tumors. The resulting

classification distinguished biologically and clinically relevant elements of immune regulation, cell signaling, DNA repair, protein and RNA metabolism, and cell cycle regulation. To interrogate the clinical applicability of our findings, we performed ST-seq on 16 'out-of-sample' pre-treatment biopsy samples of patients with metastatic non-small cell lung cancer (NSCLC) who received immune checkpoint blockade (ICB) therapy and were not within our pan-tumor database. Using the pan-tumor spatial lability classification to describe these samples, we found that immune spatial lability distinguished patient response to ICB therapy while standard and previously described spatially unaware markers – PD-L1 status, bulk transcriptional differences, and existing gene sets – did not.

Overall, our findings revealed conserved principles of TME spatial biology that are biologically and clinically meaningful. Our results motivate further interrogation into the nature of collective spatial organization within the TME and open the possibility for interpretable statistical models of clinical endpoints using spatial biology.

Spatial Groups (SGs) define a conserved architecture of TME spatial biology

As our goal was to discover organizing principles of TME spatial biology, we sought to construct a mapping that could infer TME spatial organization from ST-seq transcriptional data. Each dataset we studied was created using 10X Visium technology, which generates transcriptome-wide measurements for up to 14,000 spatial locations (called spots, each of which contains multiple cells) in up to an 11 mm x 11 mm region of biopsy tissue[30]. Previous literature has demonstrated the presence and importance of spatially nested and non-nested biological processes in the TME[17,31]. As such, we wanted our mapping to simultaneously capture and distinguish nested and non-nested biological processes – a quality that currently developed frameworks for ST-seq data do not contain (**Fig. 1A**)[32–37]. We therefore developed a new framework called 'TumorSPACE' (Tumor Spatial Architectures from the Complete Eigenspectrum). While this framework is described in detail in Methods, TumorSPACE first uses

patterns of transcriptional covariation to define hierarchical relationships between ST-seq spots[26,38]. This yields a tree-like relationship between all spots in the TME where each leaf of the tree defines an individual spot and branchpoints in the tree group spots together that are transcriptionally similar. TumorSPACE then removes branches of that tree that do not relate to spatial organization (**Extended Data Fig. 1A**). This resulting tree is a dataset-specific 'TumorSPACE map' between transcriptional information and spatial organization.

We applied TumorSPACE to a diverse database of 96 tumors profiled by ST-seq and used the resulting maps to infer the spatial locations of spots (Methods). Each dataset in our database represented a unique patient sample; the database spanned 12 distinct tumor types, multiple disease stages (localized versus metastatic), and multiple tumor body locations (primary, metastatic lymph node, metastatic organ) (**Extended Data Fig. 1B**, **Supplementary Table 1**). We found that for all datasets, the TumorSPACE maps significantly inferred spot spatial locations ($q < 0.01$) (**Fig. 1B**, Methods). Thus, TumorSPACE maps accurately related transcriptional and spatial information within TMEs.

We next interrogated whether the TumorSPACE maps revealed any underlying conserved principles of TME spatial organization. We first focused on the best-performing TumorSPACE map, a small-cell ovarian cancer dataset 'SCOC-P2'. Branchpoints in this map defined groups of spots that were anisotropically distributed in the biopsy sample and comprised spots that were either (i) physically separated from each other or (ii) were spatially nested within other groups of spots defined by the TumorSPACE map. We therefore termed the branchpoints of TumorSPACE maps 'Spatial Groups' (SGs). We defined any SG that was spatially nested within its parent SG – the SG one layer closer to the root of the map – as a nested Spatial Group (NSG). Any SG that was not spatially nested within its parent was a non-nested Spatial Group (non-NSG) (**Fig. 1C**). We found that in the SCOC-P2 dataset, NSGs could be spatially nested to varying degrees. We therefore defined 'NSG depth' for any NSG as the following: as one moves from an NSG towards the root of the TumorSPACE map, 'NSG depth' is the number

of NSGs that are encountered inclusive of the original NSG prior to arriving at a non-NSG (**Extended Data Fig. 1C**) (Methods). A systematic analysis of all tumors in our database revealed a spatial architecture of the TME that is broadly conserved: SGs are comprised of a consistent distribution of non-NSGs and NSGs that can be nested up to several degrees (**Fig. 1D**).

We next sought to characterize the biology reflected by NSGs and non-NSGs. We described TME biology using cell type distribution and cellular gene pathway usage since these qualities have been implicated in TME spatial biology across many cancer types. At each SG, we detected differential abundance of genes, pathways, and cell types (**Extended Data Fig. 2A**) (Methods). Since each spot consists of multiple cells, we used SpaCET for deconvoluting cell types (Methods)[39]. We found a relationship between the spatial scale of SGs and biological processes: SGs that were larger in spatial distribution displayed changes in cell type abundance (particularly in CD4+ and CD8+ T cells) while SGs that were smaller in spatial distribution displayed changes in pathway usage (particularly in pathways for cell adhesion, cell cycle, and adaptive cytotoxicity) (**Extended Data Fig. 2B**).

As NSGs are nested within the local spatial context of their parent SGs, we asked how much a differential process (pathway or cell type) within an NSG was dependent on biological processes defined by the spatial context of its parent. For this, we quantified contextual dependence as the odds ratio of detecting a change in a biological process within an NSG ('Process B') given a particular change in a biological process ('Process A') at its parent SG. We then computed whether any odds ratio was significantly different than 1. This measured whether the associated set of parent-child processes was contextually dependent or independent (**Fig. 1E**, left) (Methods). We found that 74% of differential processes within an NSG were dependent on the local spatial context defined by the parent SG, illustrating extensive biological nesting within NSGs. Moreover, we found that certain biological processes were associated with stronger local spatial contexts: NSG processes were nearly universally dependent on oncogenic

pathways in parent SGs yet were less frequently dependent on parent SG pathways that largely involved direct cell-cell contacts, such as cell adhesion and immune cytotoxicity (**Fig. 1E**, right; **Extended Data Fig. 2C**). Additionally, the strength of local spatial context associated with a biological process within a parent SG – an averaged odds ratio across all processes – was linearly related to the spatial scale of the parent NSG (**Fig. 1F**, top). Thus, as parent SGs became larger, their influence on biological processes encoded within NSGs became greater. In contrast, no such relationship was present when considering the influence of biological processes in non-NSGs on their local spatial environment (**Fig. 1F**, bottom). These results demonstrated that NSGs reflect nested biological properties.

Overall, our findings revealed a general spatial architecture of TMEs. TMEs are hierarchically organized into multicellular units of transcriptional covariation, 'Spatial Groups', that can be either spatially nested (NSGs) or non-nested (non-NSGs). The spatial organization of NSGs reflects the contextual dependence of smaller-scale biological processes involving cell-cell interactions, amongst larger-scale biological processes such as cell type abundance (**Fig. 1G**). These findings motivated using SGs as a common unit of spatial organization for investigating heterogeneity in TME spatial biology.

Using SGs to define spatial lability in TMEs

To capture variation in gene expression patterns amongst SGs in a holistic manner, we defined gene 'spatial lability' – the extent of change of gene expression when comparing across partitions of the TME. We first identified all SGs for a given TME. Then, for a given gene, we isolated the SGs and associated ST-seq spots where the gene was differentially expressed (**Fig. 2A**, top). Finally, we computed the fraction of the tumor dataset represented by those ST-seq spots and termed this fraction the 'spatial lability' (SLAB) score for the gene of interest (**Fig. 2A**, bottom) (Methods). A comparison of SLAB scores with gene expression for all genes across all tumors in our database illustrated that SLAB scores were positively correlated with average

gene expression but also captured other modes of gene-level spatial variation. For example, genes with low average expression exhibited variation in SLAB score (**Fig. 2B**). Furthermore, 77% of genes had no correlation between bulk gene expression and SLAB score when comparing across tumors (**Extended Data Fig. 3**). As a specific example, the calreticulin gene (CALR) had similar bulk expression levels in 3 selected tumors, yet its SLAB score varied from high to low across these tumors (**Fig. 2C**). Visualization of CALR expression across the TME clearly illustrated that the degree of spatial heterogeneity in gene expression followed the degree of spatial lability across tumors (**Fig. 2D**). Examination of the SGs associated with differential expression of CALR showed that the high spatial lability tumor contained changes in CALR expression at both large- and medium-sized SGs, while the tumors with lower SLAB scores had SG changes restricted to medium and small SGs (**Fig. 2E**). Together, these data demonstrate that the SLAB score captures information about the spatial heterogeneity of gene expression across the TME that is distinct from bulk gene expression.

To validate that SLAB scores captured spatial heterogeneity in the TME, we compared SLAB scores with an orthogonal measure of spatial biology – multiplexed immunofluorescence (mIF) across 51 marker genes – for two diffuse large B cell lymphoma (DLBCL) samples within our pan-tumor database (Methods). We used a grid approach to define spatial domains of varying sizes. Variation in cell type abundance was computed using the coefficient of variation across these grids (**Extended Data Fig. 4A**) (Methods). This approach demonstrated cell types to be more spatially labile in the DLBCL-P2 tumor (Patient 2) than in DLBCL-P1 (Patient 1) (**Extended Data Fig. 4B**). SLAB scores, computed across SpaCET-deconvoluted cell type proportions, recapitulated this finding (**Extended Data Fig. 4C**). Examination of IF intensity distributions for CD3 (a pan T cell marker) and CD21 (an abundant B cell marker) demonstrated that germinal center (GC) effacement might explain the inter-tumoral differences in cell type spatial lability between the two DLBCL samples (**Extended Data Fig. 4D**). Moreover, we found

that SGs with simultaneous enrichment of T cells and depletion of B cells included many H&E-identified GCs in DLBCL-P2 (**Extended Data Fig. 4E**).

Together, these results demonstrate that the SG-based metric we created – SLAB scores – accurately captured information about the spatial heterogeneity of gene expression across TMEs, thereby enabling spatially based comparisons between tumors.

Classification of TMEs by spatial lability

To compare TMEs by their profiles of spatial lability, we aligned the genome-wide SLAB scores across for all tumors in our database and computed the Euclidean distance between each pair of tumors (**Extended Data Fig. 5**). Hierarchical clustering of pairwise distance between tumors defined a tree representing a pan-tumor classification where tumors were grouped by similarity of their spatial lability profiles and branchpoints reflected signatures of differential spatial lability between groups (**Fig. 3A,B**). Interrogation of the tree illustrated two results. First, tumors were approximately ordered by their average spatial lability. In the representation depicted in **Fig. 3B**, tumors ordered from left to right – labeled as groups A through M – reflected a continuum of average SLAB scores from high to low respectively (**Extended Data Fig. 6A**). As an example, while the average gene expression across all genes for the group of tumors on the far right (group M) was higher than the other groups, the spatial lability of genes in group M tumors was significantly lower than the other groups (**Extended Data Fig. 6B**). Second, the resulting clusters illustrated that tumor groups were either uniform (e.g. groups C and F) or varied (e.g. groups E, L, M) in their tissue of origin. For example, we found that tumors originating in breast tissue (91% triple-negative, rest unknown) classified into groups B, D, E, H, I, L, and M. Furthermore, we saw that group E were composed of tumors originating in breast, skin, ovarian, and central nervous system (CNS) tissues. These results suggested that patterns of spatial lability across our pan-tumor database described both tumor-type-specific and tumor-type-agnostic differences in the TME.

We wanted to test whether the pan-tumor classification based on spatial lability captured similarity in spatial organization at the level of individual spots. For this, we tested whether the TumorSPACE map of one tumor (tumor A) could be used to predict the pairwise distances between spots of another tumor (tumor B). We then compared whether the accuracy of such predictions related to tumors being within the same or different spatial lability class. We computed predictions by projecting the transcriptional data from tumor B into the TumorSPACE map previously built for tumor A, which had not incorporated any information about tumor B. As a result, pairwise spot-spot distances were predicted between all spots in tumor B. These predictions were then compared to the actual pairwise distances between spots in tumor B (**Extended Data Fig. 7A**) (Methods). We excluded group M tumors from this analysis since our results showed that these tumors lacked spatial lability altogether. Overall, we found that 52% of such tumor pairs predicted spot distance information better than null models. Additionally, models were more likely to be predictive when selecting two tumors from the same spatial lability class than from different classes (67% versus 49%), similar to the prediction increase when comparing tumors of the same type versus different type (65% versus 47%) (**Extended Data Fig. 7B**). In accord with this finding, spatial lability was an independent contributor to cross-tumor spatial prediction from tumor type, suggesting that classes of tumors defined by profiles of spatial lability reflect shared spatial organization even when composed of diverse tumors (**Extended Data Fig. 7C**). Thus, while our pan-tumor database lacked the IHC information required for comparison to clinical classification schemes such as TIL score and Immunoscore, our analysis of the tree in **Fig. 3B** illustrated a classification of tumors by their spatial biology that was not oriented merely by tumor type. This motivated further investigation into the spatial lability changes that separated tumor groups.

We interrogated differences at branchpoints of the spatial lability classification (e.g. group A versus group nA) using multiple complementary approaches – gene-level SLAB differences, pathway-level SLAB differences using over-representation analysis (ORA), and pathway-level

SLAB differences using gene-set enrichment analysis (GSEA). The results we report for each of these analyses were robust to gene co-linearity and multiple hypothesis testing (Methods). Four branchpoints contained statistically significant differences in spatial lability amongst genes (**Fig. 3C**, left panel). We therefore performed an in-depth analysis at these four branchpoints of the genes and pathways underlying differences in spatial lability (**Supplementary Table 2**).

Two of these branchpoints defined groups of tumors – C and E – that exhibited significant changes in spatial lability associated with TME immune biology. Group C was comprised of a set of exclusively primary CNS tumors and exhibited increased spatial lability of neurotransmitter activity genes (GRIK1, KCNN2) and of complement activation pathways (**Fig. 3C**, top row). Notably, complement activation has been implicated in promoting glioma cell proliferation and neovascularization in the hypoxic TME characteristically found in such tumors as well as in mediating the suppression of anti-tumor immunity in both CNS tumors and non-CNS tumor types[40]. Group E tumors, comprised of a diverse mixture of tumor types, showed increased spatial lability of genes associated with immune exhaustion through diverse mechanisms such as myeloid cell activation (P2RY11), TGF-β signaling (SMAD5), antigen presentation (DPP9), innate immune cell activation (TRIM11, TRIM44), T cell migration (DPP9, ELMO2), and T cell activation (STAT5, STAT5A, NFATC2IP, PLCG1, ORAI1) (**Fig. 3C**, second row)[41–50]. Analysis of pathways demonstrated increased spatial lability in well-studied immune signaling pathways (vesicle transport, solute carrier (SLC) transporters) as well as in pathways linked to antigen generation (RNA metabolism, post-translational protein modifications), suggesting that complementary biological processes collectively reflect immune spatial lability[51–54]. Together, these data illustrated that group C and group E tumors have TMEs with increased immune spatial lability via distinct components of TME immune biology.

The other two branchpoints defined groups of tumors – F and L – with spatial lability in non-immune areas of TME biology as well as group M, the group notable for spatially invariant biology across all studied genes and pathways. Group F was comprised of exclusively ovarian

tumors and had increased spatial lability for pathways related to olfactory receptors – a class of cancer testis antigens that are abundantly expressed in ovarian tumors and are under development as a CAR-T therapeutic target (**Fig. 3C**, third row)[55]. Group L, composed of diverse tumors, demonstrated increased spatial lability for genes involved in mitochondrial biology (MRPL40, LYRM1, MRPS14, NDUFS4, NFU1, MRPL21) as well as in RNA and protein processing (COPS8, TCEAL8, RPAIN, ZCCHC17, SNW1, TTC1, KIAA1191, PEX19, GPN1, PPIE) (**Fig. 3C**, fourth row panel). Accordingly, this group of tumors was enriched in spatial lability for pathways related to cell-intrinsic processes – metabolism, transcriptional regulation, and DNA repair.

We previously observed that cell-extrinsic versus cell-intrinsic processes were enriched in NSGs and non-NSGs respectively (**Fig. 1F, G**). Having now observed that the spatial lability classification varied across groups in enrichment for cell-extrinsic versus cell-intrinsic processes, we hypothesized that the classification in **Fig. 3B** was reliant on information with NSGs and non-NSGs to different degrees depending on tumor group. To test this idea, we performed GSEA pathway enrichment at branchpoints E/nE and L/M using spots found within only NSGs or within only non-NSGs. We examined these branchpoints because they demonstrated enrichment for cell-extrinsic and cell-intrinsic biology respectively. We found that NSGs alone identified 69% of pathways enriched by GSEA for spatial lability in Group E, while non-NSGs alone did not identify any of these pathways. Conversely, non-NSGs alone identified 42% of pathways enriched for spatial lability in Group L, while NSGs alone did not identify any of these pathways (**Fig. 3D**). Notably, 31% and 58% of pathways with altered spatial lability in groups E and L, respectively, required transcriptional information contained within both NSGs and non-NSGs. Furthermore, across all studied branchpoints we found that the likelihood of a pathway exhibiting detectable changes in spatial lability within NSGs versus non-NSGs depended on whether the pathway described cell-extrinsic or cell-intrinsic processes (**Fig. 3E**).

Together, these results illustrated a pan-tumor classification defined by spatial lability. The biological variation associated with this classification distinguished cell-extrinsic processes – i.e. immune signaling – that are found mostly within NSGs versus cell-intrinsic processes like DNA repair found in both NSGs and non-NSGs. Interrogation of genes and pathways distinguishing groups also showed spatial lability in targets with proven therapeutic significance. Together, these findings motivated using our pan-tumor classification schema to predict the clinical outcome of patients whose tumors were not contained within our tumor database.

Pan-tumor TME classification by spatial lability distinguishes response to immunotherapy in metastatic NSCLC

As two branchpoints in our pan-tumor classification demonstrated variation in immune spatial lability, we hypothesized that classification of a separate cohort of tumors by immune spatial lability could be used to predict patient response to anti-PD1/anti-PD-L1 immune checkpoint blockade (ICB) – a widely approved therapeutic modality across diverse solid tumors. To this end, we focused our efforts on patients diagnosed with metastatic NSCLC. Despite substantial improvements in overall survival with the use of ICB therapies in the metastatic NSCLC frontline setting, 5-year overall survival remains quite poor at 19%[56]. Moreover, the only clinically approved biomarker of response to ICB therapy, PD-L1 immunohistochemistry (IHC), is weakly predictive of outcomes in the frontline metastatic setting for NSCLC, prompting ongoing studies on whether gene expression or cell type abundance biomarkers might be more predictive of such outcomes[5,56–58].

To address whether spatial lability informs ICB response, we conducted a retrospective pilot study of 16 patients with metastatic NSCLC without targetable mutations who received frontline ICB with or without chemotherapy (Methods). For each patient, we conducted ST-seq on pre-treatment biopsy samples followed by (i) computing genome-wide SLAB profiles for all samples and (ii) contextualizing the resulting data using the pan-tumor classification of tumor

spatial lability defined by the discovery cohort in **Fig. 3B**. We also performed immunohistochemistry (IHC) to determine PD-L1 tumor proportion score on the same pre-treatment diagnostic biopsy sample. We then evaluated whether classification by spatial lability or PD-L1 status could predict Progression-Free Survival (PFS) after ICB treatment (**Fig. 4A**). Two possible variables were identified that could confound an association with ICB response: ICB regimen choice and presence of the somatic mutation KRAS G12C, which is targetable in the second line (**Extended Data Fig. 8A,B**). Univariate analysis found that neither variable was associated with PFS, excluding the possibility that these factors influenced our study (**Extended Data Fig. 8C**).

To evaluate this out-of-sample validation cohort of NSCLC tumors within the context of our pan-tumor classification from **Fig. 3B**, we first used TumorSPACE to identify SGs for each NSCLC tumor. We found a similar distribution of nested and non-nested SGs as within our pan-tumor database, illustrating the generalizability of the distribution of SGs in TMEs (**Extended Data Fig. 9**). Comparison of spatial lability profiles between the NSCLC tumors and the pan-tumor database defined two groups. One group, comprised of twelve NSCLC tumors, exhibited a spatial lability profile similar to group C and group E tumors in our pan-tumor classification – high spatial lability amongst immune-related components ('immune spatially labile', 'ISL'). The other group, comprised of four NSCLC tumors, exhibited a spatial lability profile similar to group L and group M tumors – low spatial lability in immune biology ('immune spatially invariant', 'ISI') (**Fig. 4B**) (Methods). Classification by immune spatial lability (ISL versus ISI) was highly predictive of PFS after ICB treatment (hazard ratio = 0.09, p = 0.00095), unlike classification by PD-L1 using either classical NSCLC groupings – < 1%, 1-49%, ≥ 50% (p = 0.55) – or binary cutoffs of either 1% (p = 0.27) or 50% (p = 0.77) (**Fig. 4C, Extended Data Fig. 10A**). Moreover, classification by bulk expression using either all genes or 8 previously published gene sets for NSCLC IO response was not predictive of PFS. However, notably a DNA damage response gene set was predictive (p = 0.003) only when using SLAB scores instead of gene expression (p

= 0.33) (**Extended Data Fig. 10B,C**) (**Supplementary Table 3**). Moreover, eight out of the twelve patients with measurable disease at treatment onset demonstrated shrinkage in tumor volumes shortly after treatment began, suggesting that classification by PFS was detecting differences in treatment response and durability rather than in treatment-agnostic factors such as disease prognosis (**Extended Data Fig. 10D**) (Methods). Together, these results showed that our pan-tumor classification schema, defined by variation in spatial lability, was sufficient to distinguish variation in response to ICB therapy in our patient cohort in contrast to PD-L1 IHC and previously published bulk expression gene sets.

Since our pan-tumor classification distinguished the NSCLC tumors as immune spatially labile versus immune spatially invariant, we sought to determine which biological processes were relevant to NSCLC ICB response. We tested whether genes with differential SLAB scores at the pan-tumor group E branchpoint also exhibited differential SLAB scores between the ISL and ISI NSCLC datasets. Of the 537 genes distinguishing the group E branchpoint, 398 were also statistically enriched for spatial lability in NSCLC ISL tumors compared to NSCLC ISI tumors, while the other 139 were not (**Fig. 4D, left**). Pathway analysis of these two gene groups demonstrated that both groups related to immune activation and signal transduction (**Fig. 4D, right**). However, the 139 non-differential SLAB genes were enriched for signaling via the VEGF receptor, estrogen receptor, and NTRK receptors – signaling pathways implicated in immune activation in cancers other than lung cancer. On the other hand, the 398 differential genes were enriched for immune signaling pathways (e.g. vesicle transport) and specifically for Notch signaling, a pathway that has been implicated to mediate immune checkpoint exhaustion in lung cancer through a variety of mechanisms[59]. Closer investigation of the Notch pathway demonstrated a set of 11 genes with coordinated SLAB changes between ISL and ISI tumors (**Extended Data Fig. 11**). As an example, an examination of three genes (HDAC6, NOTCH2, and PSEN1) that each promote Notch pathway activation via distinct mechanisms demonstrated spatially coordinated expression changes at large, medium, and small SG scales (**Fig. 4E**).

Discussion

Through statistical analysis of a broad diversity of solid tumors, we have shown that there is a conserved, hierarchical spatial architecture that organizes the apparent biological complexity of the TME. Individual spots group together into either non-nested or nested SGs which hierarchically integrate into the whole biopsy sample, thereby providing a holistic picture of emergent TME organization. The results in **Figs. 3** and **4** suggest a cohesive model that directly links this spatial architecture with clinical response to ICB therapy in patients. SGs are information-dense units of spatial organization encoding complex molecular interactions between cells and variation amongst SG-based TME profiles distinguishes ICB therapy response (**Fig. 4F**, left). Our findings have implications for both tumor biology and for translation towards clinical oncology.

With respect to tumor biology, our findings demonstrate that Spatial Groups can be conceptualized as statistical 'units' of the hierarchical organization in TMEs. A natural next step is to deeply interrogate the biology underlying this statistical structure to elucidate drivers of variation in SG distribution and TME organization (**Fig. 4F**, top right). In general, existing biological knowledge of tumors (e.g. databases reflecting experimental results from cell lines and *in vivo* models) has viewed individual cells as the components of interest with respect to understanding properties of whole tumors. Our results suggest an alternative foundation for biological interrogation: the collective spatial interactions amongst SGs are key to understanding emergent biological qualities of tumors. Elucidating the biology underlying these interactions will likely require interrogating SGs without perturbing their native context rather than isolating and removing them from a tumor. As such, approaches studying SG variation under observable metabolic gradients and pairing SG identification with spatial metabolomics and proteomics may be useful for discovering biological mechanisms influencing TME organization.

Efforts to bring spatial molecular profiling into clinical settings are limited by not having a consensus description of tumor spatial biology. Terms such as “immune inflamed”, “immune excluded”, and “immune desert” have served as a useful paradigm, yet, as recent studies have illustrated, are too broad for describing TME heterogeneity[60,61]. Our results demonstrate how treating the TME as an emergent cellular ecosystem and identifying conserved statistical features of spatial organization results in a holistic, unbiased, and quantitative approach for classifying tumors. The resulting SG-based classification was built on a discovery cohort of 96 tumors spanning twelve tumor types gathered from multiple institutions and countries and tested in a validation cohort on a tumor type (NSCLC) with markedly low representation in the discovery cohort. Importantly, our discovery cohort was not pre-selected to represent variation in ICB response but was assembled in an unbiased manner and studied to characterize the biology reflecting heterogeneity in TME spatial organization. Thus, the success of this classification in delineating responders and non-responders to ICB therapy in the setting of metastatic NSCLC underscores the shared qualities of SGs across tumor types and suggests that variation amongst TME SG profiles may be useful for developing a framework for therapeutic ‘logic’ (**Fig. 4F**, middle right). NSGs may have increased relevance for understanding and targeting key aspects of cell-cell signaling while non-NSGs might reflect elucidating molecular determinants of tumor fitness that are independent of the local environment. The incorporation of more cohort studies into our classification where pre-treatment biopsy samples are coupled with outcomes following therapeutic intervention will address this concept. It is possible that future studies of patient cohorts in both ICB-naïve and ICB-refractory settings could leverage SG-based descriptions for the discovery of therapeutics that augment ICB (**Fig. 4F**, bottom right). We anticipate that describing TMEs using SGs will open the possibility of creating interpretable statistical models of the TME that enable spatially informed precision oncology.

Behera et al., Figure 1

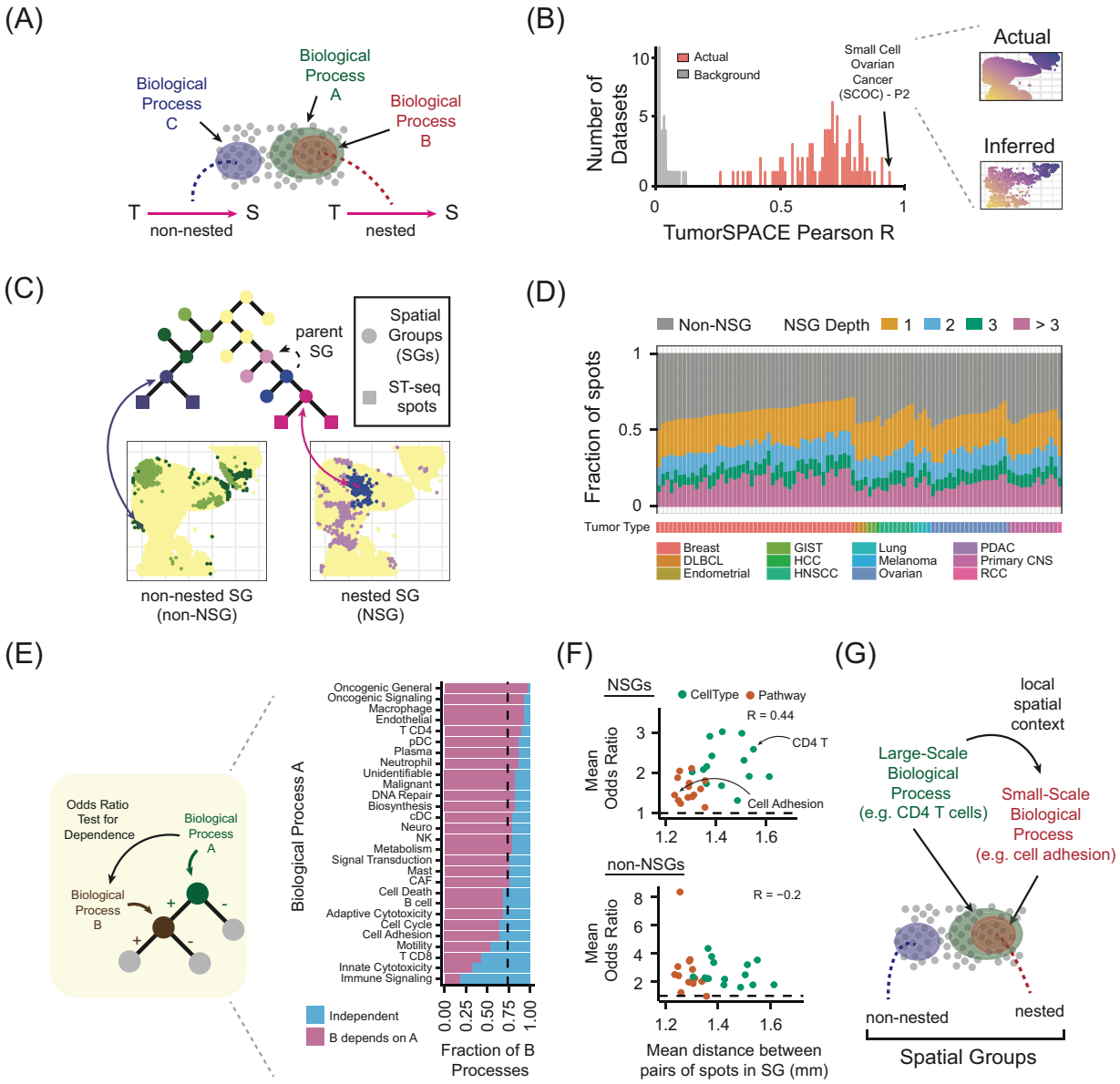


Figure 1. A conserved architecture of TME spatial biology. (A) A map (pink arrows) relating transcriptional ('T') and spatial ('S') information that captures non-nested and nested spatial contexts of biological processes. **(B)** Histogram of correlation values (Pearson R) between actual spatial distances for all pairs of spots within a given ST-seq dataset and pairwise distances inferred by TumorSPACE. Gray distribution reflects a background expectation of correlation values (Methods). (Inset) Actual spot locations and inferred spot locations for the small-cell ovarian cancer patient 2 ST-seq dataset (SCOC-P2). **(C)** (Top) Section of TumorSPACE map for sample SCOC-P2; squares at the bottom of the tree are individual spots, each circle is a Spatial Group (SG). 'Parent SG' is delineated to define the relationship between SGs. (Bottom) Picture of actual SCOC-P2 spot locations with spots colored by SG designation. The left and right panels illustrate examples of non-nested spatial groups (non-NSGs) and nested spatial groups (NSGs) respectively. **(D)** Fraction of spots in an ST-seq dataset (y-axis) belonging to non-NSGs (gray bars) or NSGs of varying depth (colored bars) for all tumors in our pan-tumor database ('Tumor

Type' on x-axis, see color key). **(E)** (Left) Workflow for evaluating if a differentially abundant biological process within a parent SG ('Biological Process A') influences a biological process within an NSG ('Biological Process B'). (Right) The fraction of processes in an NSG (x-axis) that are dependent (purple bar) or independent (blue bar) on processes in a parent SG (y-axis). **(F)** Mean odds ratio (y-axis) of processes (colored dots) versus size of SG (x-axis). **(G)** A model of TME spatial biology: TMEs are comprised of non-nested and nested Spatial Groups. Nested spatial groups encode large-scale processes that influence small-scale processes.

Behera et al., Figure 2

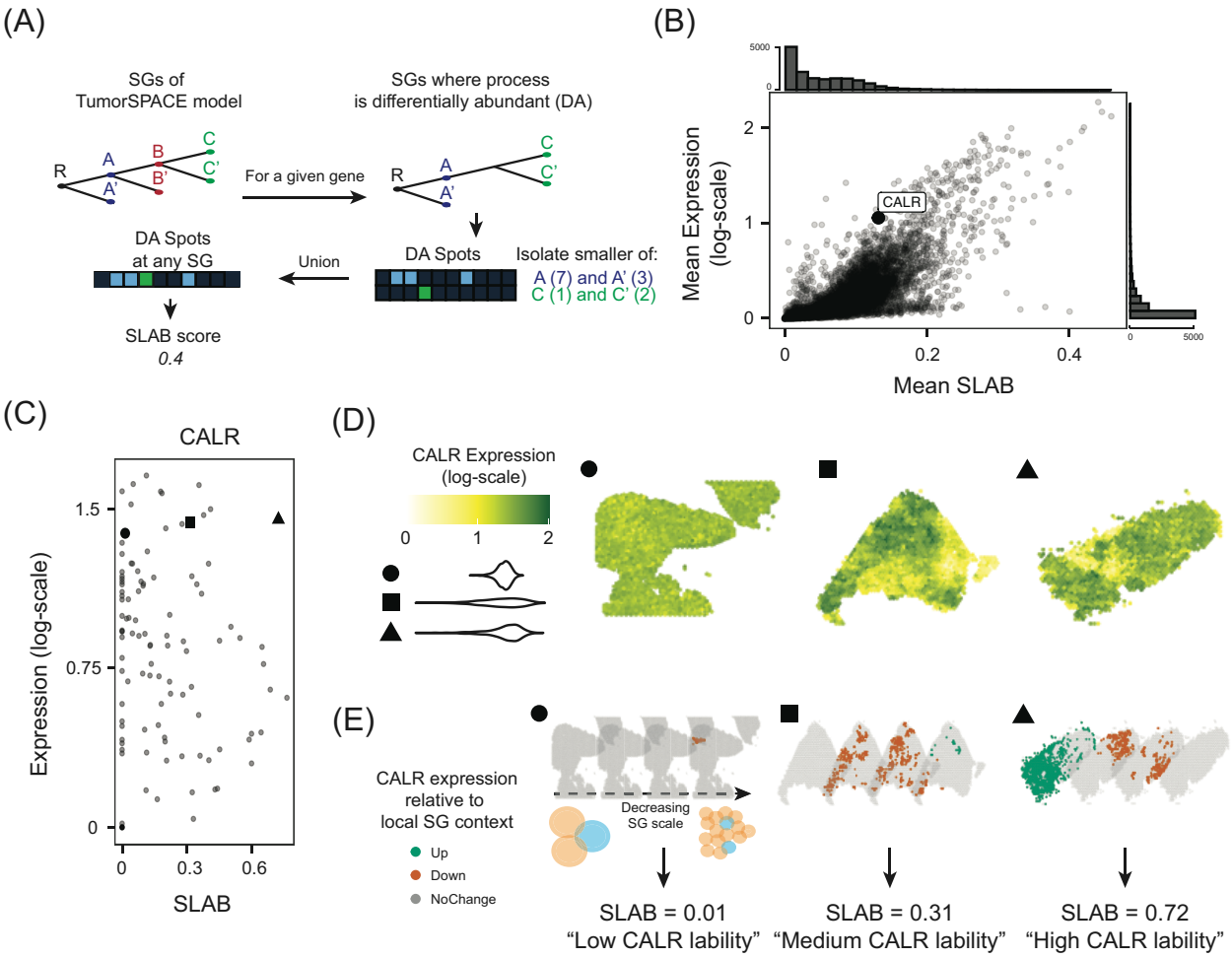


Figure 2. Spatial lability: a measure of spatial heterogeneity based on Spatial Groups. (A) Workflow for computing spatial lability (SLAB) score. R is the root SG in this example dataset consisting of 10 spots and 6 descendant SGs: A, A', B, B', C, C'. Highlighted spots reflect spots belonging to the smaller of the descendant SGs from R (Methods). Numbers in parentheses indicate number of spots within an SG. **(B)** Mean gene expression for all genes across all tumors in our database (y-axis) versus mean SLAB score across all tumors in our database (x-axis). Each point is a single gene; density of points enumerated by histograms on x and y axes. Dot in the center is the gene calreticulin (CALR). **(C)** Expression of CALR averaged across all spots in each tumor (y-axis) versus CALR SLAB score (x-axis). Each dot is a tumor in our database. Three tumors (black circle, square, and triangle) are highlighted that harbor the same mean CALR expression but varying SLAB scores. **(D)** Spatial distribution of CALR expression across tumors highlighted in panel C. CALR expression is represented in log-scale (see colorbar); below colorbar is distribution of CALR expression across all spots in the labeled tumor. Spots in each triangle, square, and circle tumors are colored by CALR expression. **(E)** CALR expression within SGs illustrated as SGs decrease in spatial scale for circle, square, and triangle tumors with corresponding spatial lability scores. Green spots reflect increased CALR expression within SG; brown spots reflect decreased CALR expression within SG; gray spots reflect no difference in expression within SG. SGs are included in plots from left to right if they impact (i) 20 – 50% of

biopsy spots, (ii) 10 – 20% of biopsy spots, (iii) 5 – 10% of biopsy spots, or (iv) less than 5% of biopsy spots. SLAB scores are computed from the union of colored spots and displayed below.

Behera et al., Figure 3

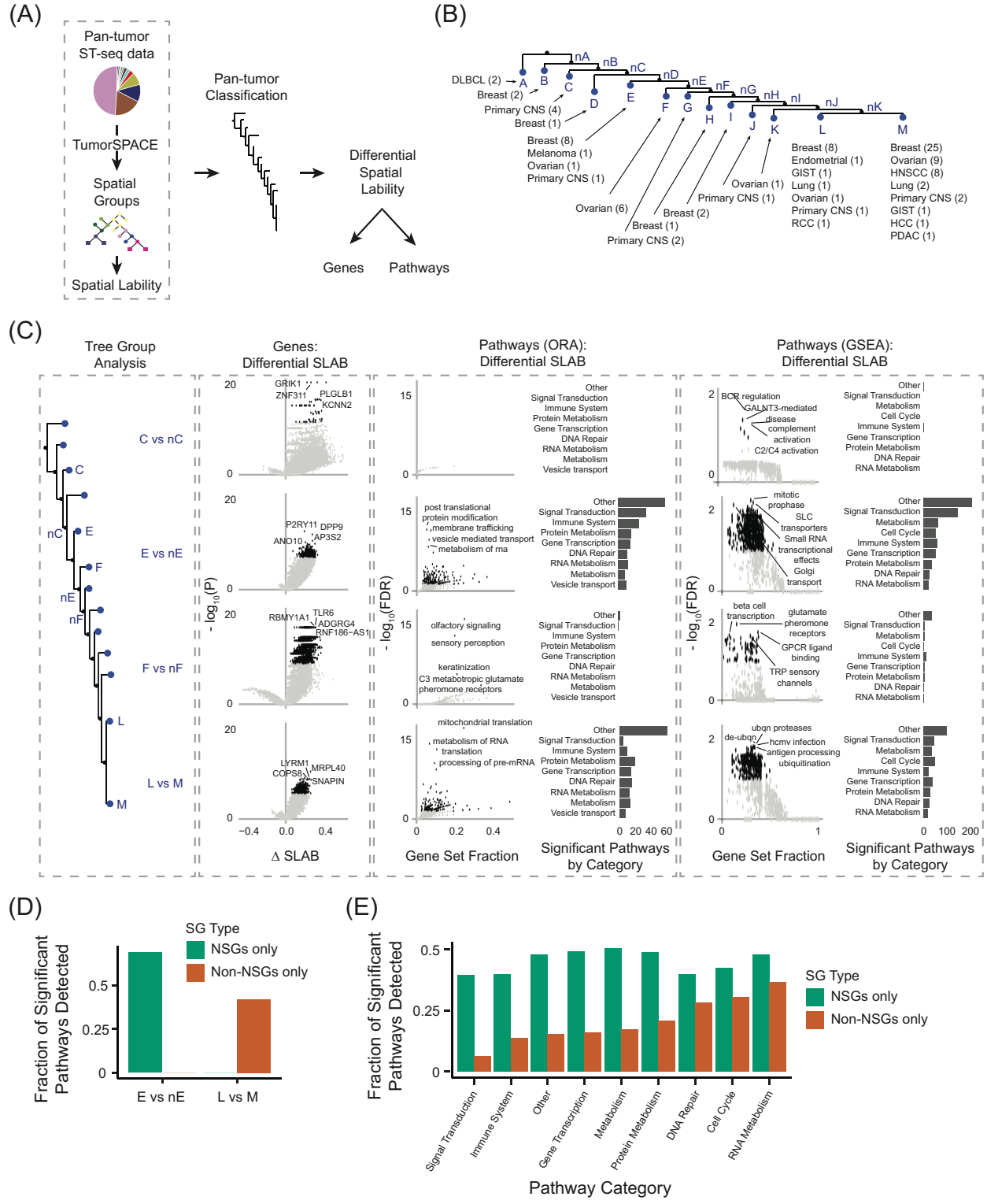


Figure 3. Pan-tumor spatial lability classification distinguishes tumors by spatial heterogeneity amongst cell-intrinsic and cell-extrinsic processes. (A) Workflow for defining and interrogating a classification of tumors based on spatial lability. **(B)** The pan-tumor classification tree. Each leaf is labeled alphabetically and comprised of specific tumors with the remaining tumors labeled to indicate not being a part of the group of tumors in the leaf. For instance, two Diffuse Large B Cell Lymphoma (DLBCL) tumors comprise group A; all other tumors comprise the 'nA' category. Parentheses indicate number of tumors of a specific tumor type in the group. **(C)** (Left) Branchpoints in the spatial lability classification where any statistically significant differences in gene spatial lability were detected. (Middle) Volcano plots describing significant differences in gene spatial lability for each group. (Right) Over-representation analysis (ORA) and gene-set enrichment analysis (GSEA) of pathway-based spatial lability. Within each sub-panel (ORA, GSEA), these results are shown as Volcano plots and histograms grouped by pathway category. **(D)** Fraction of significant pathways detected by GSEA (see **Fig. 3C**, right) that were enriched (y-axis) at branchpoints E vs nE or L vs M (x-axis) when considering spots within only NSGs (green) or only non-NSGs (orange). **(E)** Fraction of significant pathways per pathway category detected by GSEA (see **Fig. 3C**, right; x-axis) that were enriched at any branchpoint (y-axis) when considering spots within only NSGs (green) or only non-NSGs (orange).

Behera et al., Figure 4

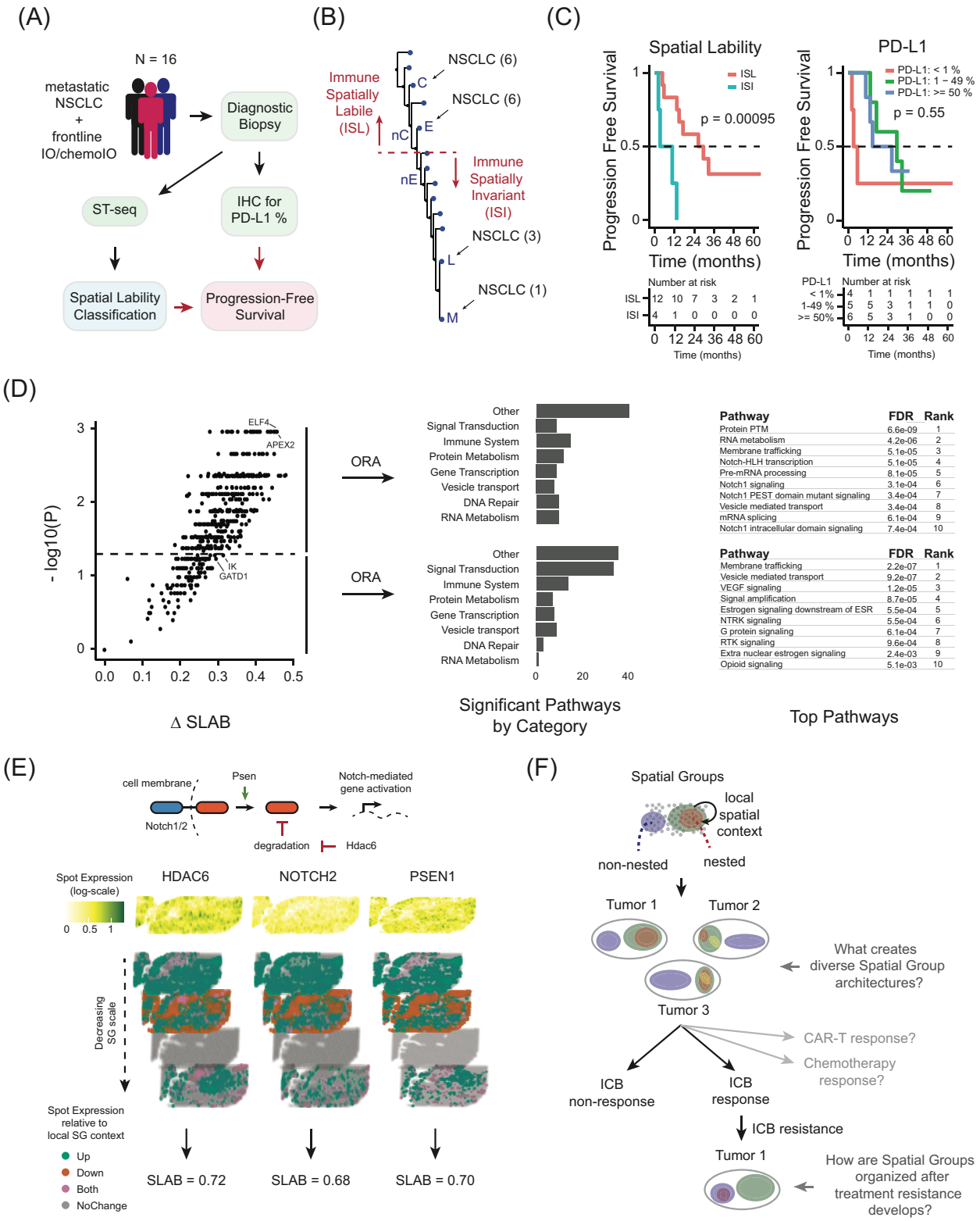
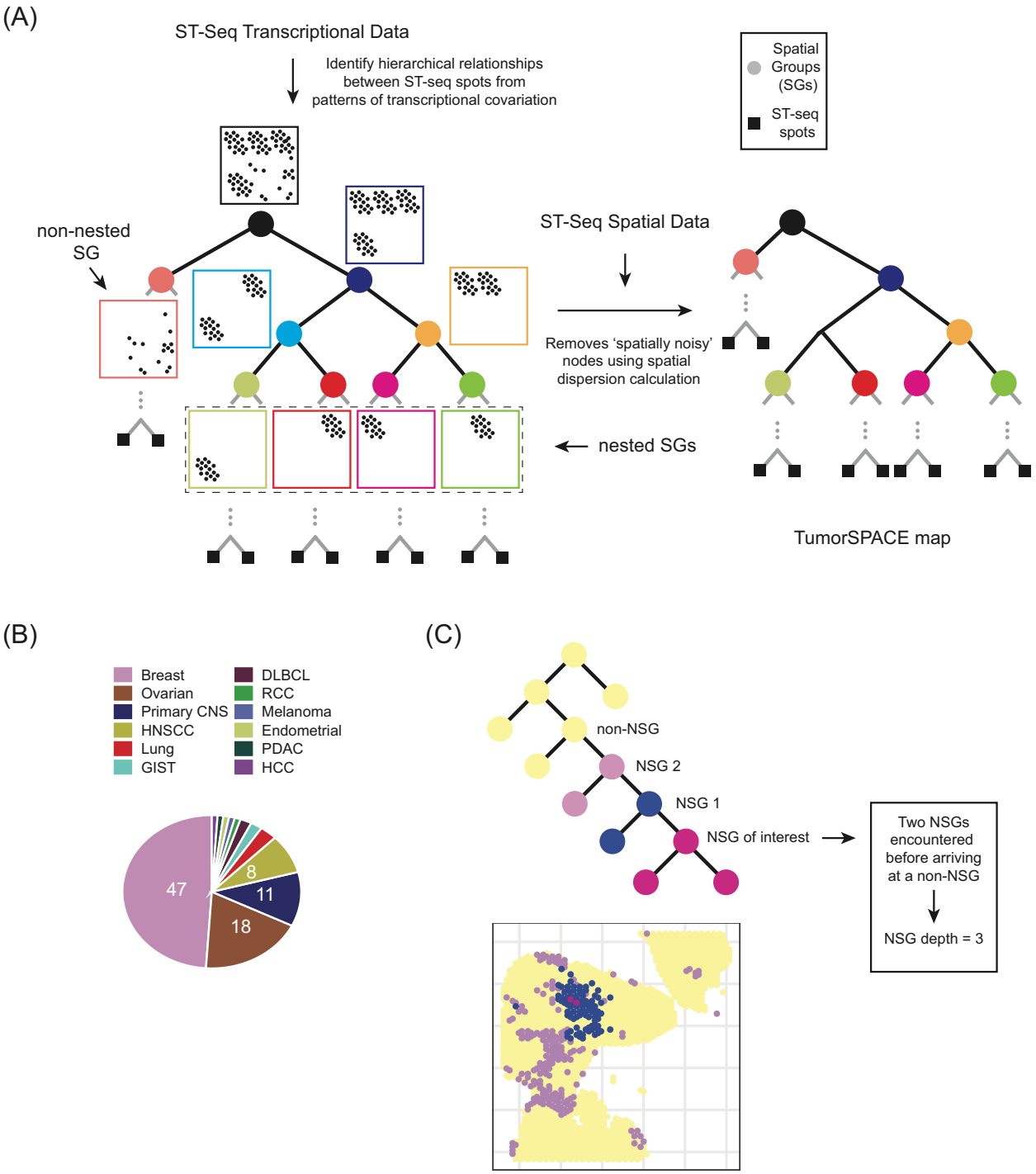


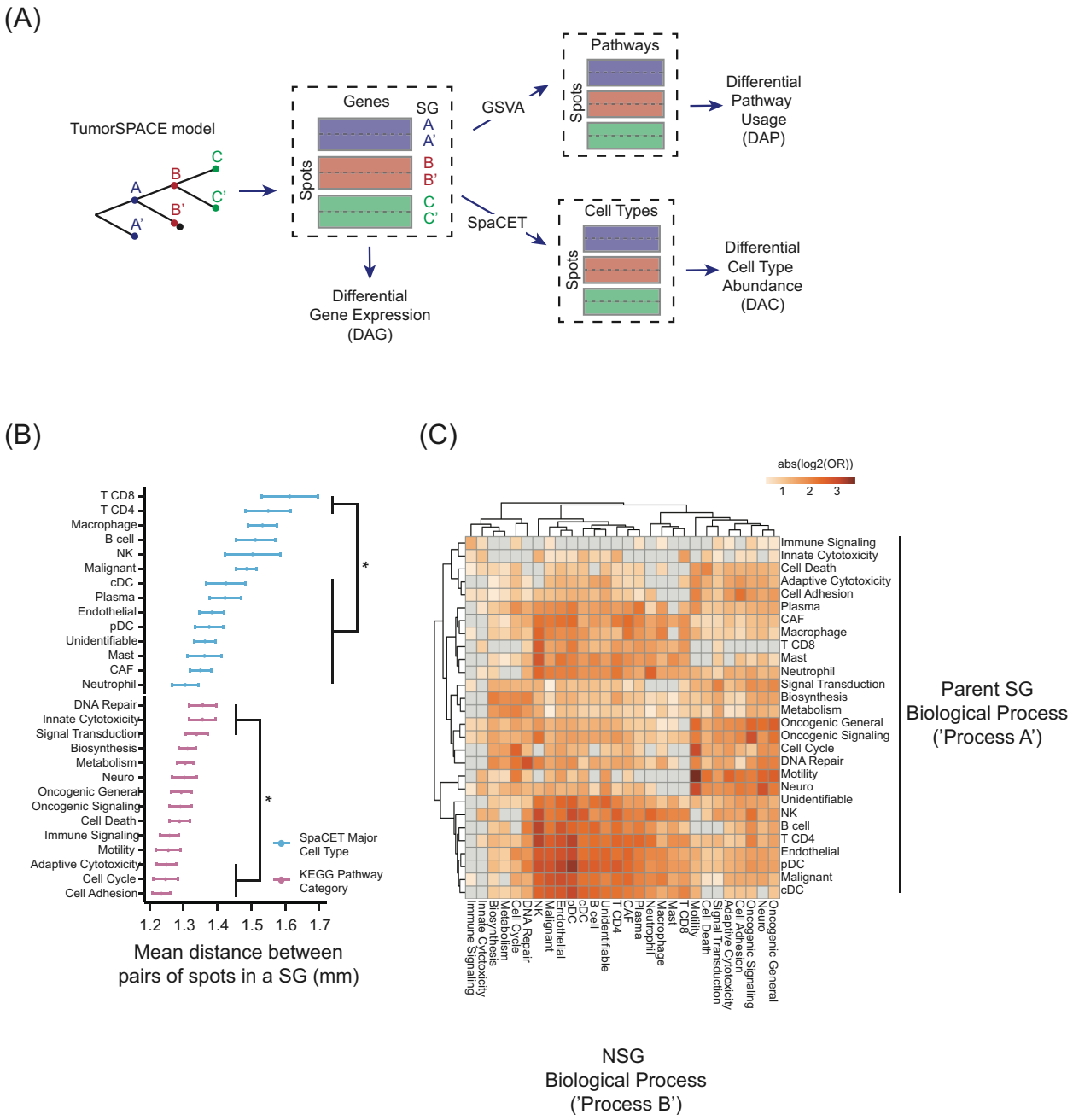
Figure 4. Pan-tumor classification distinguishes responders to immune checkpoint blockade in metastatic non-small cell lung cancer (NSCLC). (A) Sixteen patients with metastatic NSCLC underwent a diagnostic biopsy and were given immunotherapy (IO) or a combination of IO and chemotherapy. The diagnostic biopsy was subjected to ST-seq and immunohistochemistry (IHC) for PD-L1 status. From the ST-seq data, spatial lability classification was performed, and progression-free survival (PFS) was compared between groups defined by (i) spatial lability and (ii) PD-L1 status. **(B)** Comparison of NSCLC samples from panel A with the pan-tumor classification from **Fig. 3B**. Twelve samples had similar spatial lability profiles to tumors in groups C and E ('Immune Spatially Labile', 'ISL'). Four samples had similar spatial lability profiles to tumors in groups L and M ('Immune Spatially Invariant', 'ISI'). **(C)** Kaplan-Meier curves for PFS (y-axis) in months (x-axis) stratified by ISL/ISI (left) or by PD-L1 status (right). Number at risk tables show the number of patients remaining uncensored at each time point. **(D)** SLAB scores amongst the 537 genes defining the branchpoint of group E versus nE in **Fig. 3B** were computed for all NSCLC tumors. (Left) Volcano plot depicts difference in spatial lability (x-axis) and Wilcoxon p-value (y-axis, log-transformed) for NSCLC tumors grouped by ISL versus ISI. Dashed line indicates p = 0.05. Over-representation analysis (ORA) of the 537 genes stratified by p <= 0.05 (upper) or p > 0.05 (lower) is represented as number of significant pathways grouped by category (Middle) and top pathways (Right). **(E)** Part of the NOTCH signaling cascade (top panel) highlighting three proteins: NOTCH2, PSEN1, and HDAC6. Gene expression of these three proteins across an NSCLC sample (colorbar in white to green); gene expression changes of these three proteins in the depicted NSCLC sample across SGs (bottom panel with associated color key). **(F)** A depiction of our model that relates tumor SG profiles to NSCLC ICB response (left) and future directions motivated by these results (right).

Behera et al., Extended Data Figure 1



Extended Data Figure 1. (A) Workflow for generating a TumorSPACE map involves first identifying hierarchical relationships between ST-seq spots using transcriptional data alone (left) and then performing 'spatial de-noising' by removing tree nodes with high spatial dispersion values (right) (Methods). These maps can capture both spatially nested and spatially non-nested spot relationships. Grey lines at the bottom of each branchpoint indicate that trees continue to branch until terminating at the individual ST-seq spots (black squares). **(B)** Description of our pan-tumor ST-seq database. Number of datasets for each tumor type (color key) is delineated in the pie graph. **(C)** Description of how NSG depth is calculated for an example set of SGs.

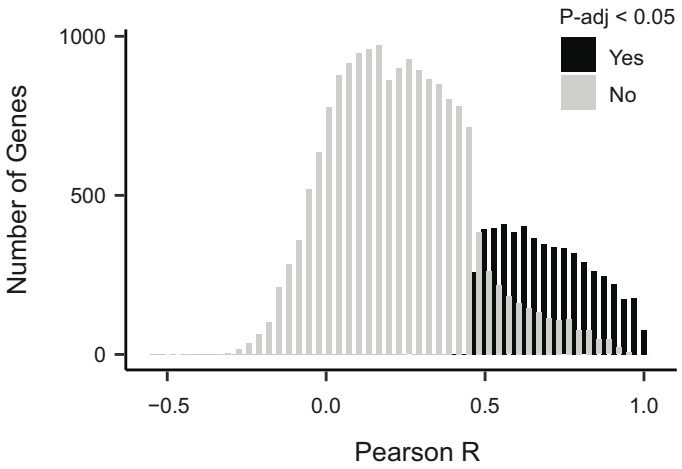
Behera et al., Extended Data Figure 2



Extended Data Figure 2. (A) Using TumorSPACE models to conduct differential analysis of gene expression, pathway usage with gene set variation analysis (GSVA), and cell type differences using SpaCET for spot deconvolution into cells[39]. SGs are labeled as A, A', B, B', C, and C'. **(B)** Mean distance between pairs of spots within a SG across SGs for all ST-seq datasets in our database (x-axis) versus all KEGG pathway categories (purple) and all major cell types as defined by SpaCET (blue) (y-axis). Error bars reflect 95% confidence intervals. *Wilcoxon p-value < 1e-9. **(C)** Odds ratio (absolute value, log-scaled) that a parent SG biological process ('A', rows) is associated with a coordinated direction of change in a second biological process ('B', columns)

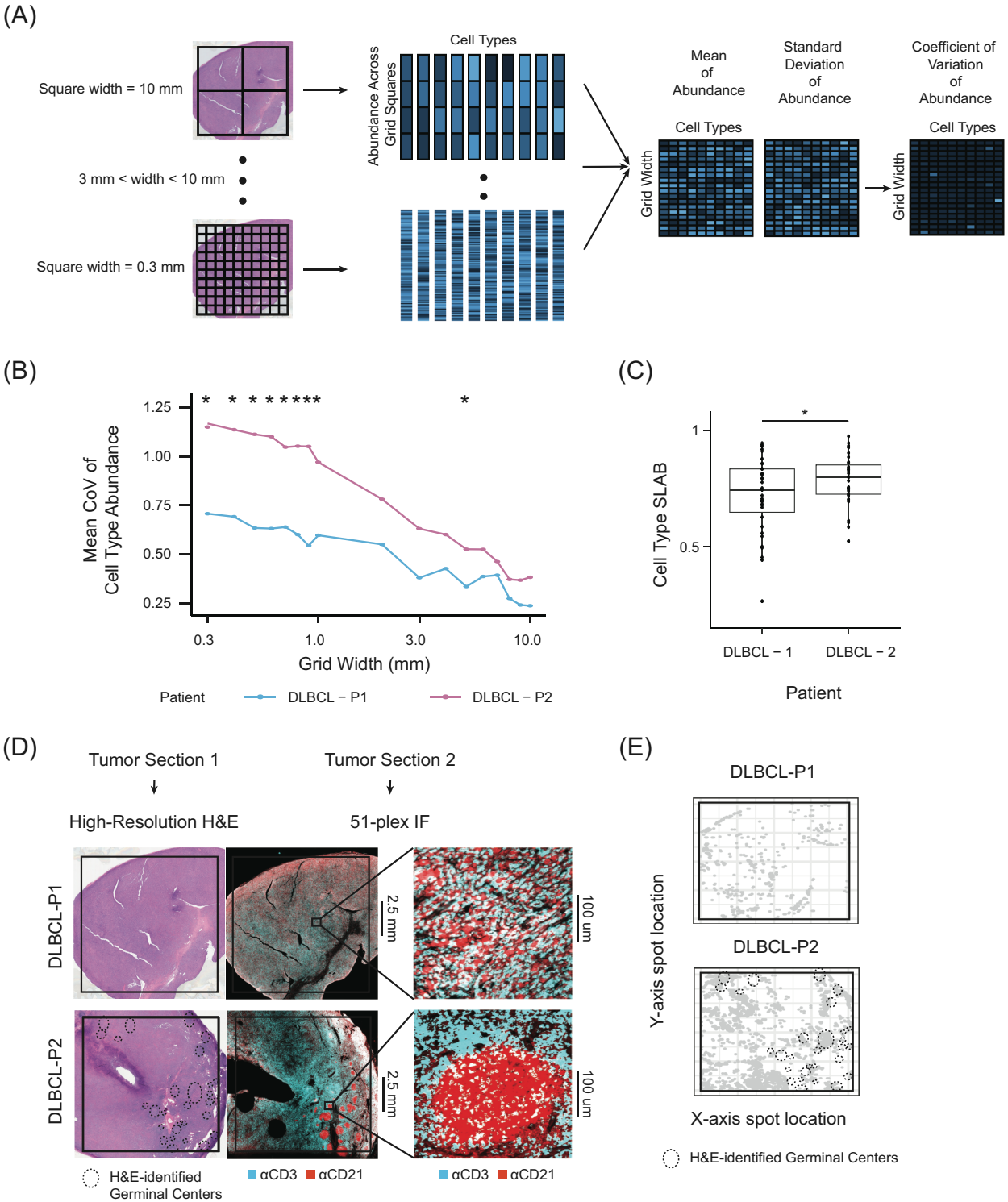
565 reflected within a daughter NSG. Color key indicates magnitude of effect where 1 indicates no
566 effect. Gray cells indicate biological process pairs that were not observed. Rows and columns are
567 hierarchically clustered.
568

Behera et al., Extended Data Figure 3



Extended Data Figure 3. Histogram of correlations (Pearson R) between bulk gene expression and SLAB scores across all tumors in our pan-tumor database for all genes. Genes are stratified by whether correlation was statistically significant (black) or not (grey) compared to an empirical null distribution.

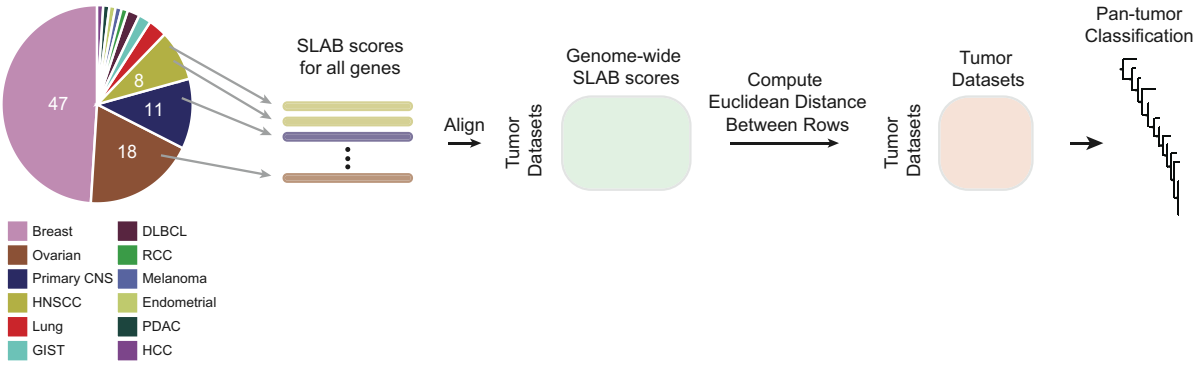
Behera et al., Extended Data Figure 4



Extended Data Figure 4. (A) Schematic for computing the coefficient of variation (CoV) for cell type abundance across a slide from immunofluorescence (IF) data. **(B)** Mean CoV across all cell types (y-axis) versus grid width of biopsy region (x-axis) for DLBCL-P1 (blue) and DLBCL-

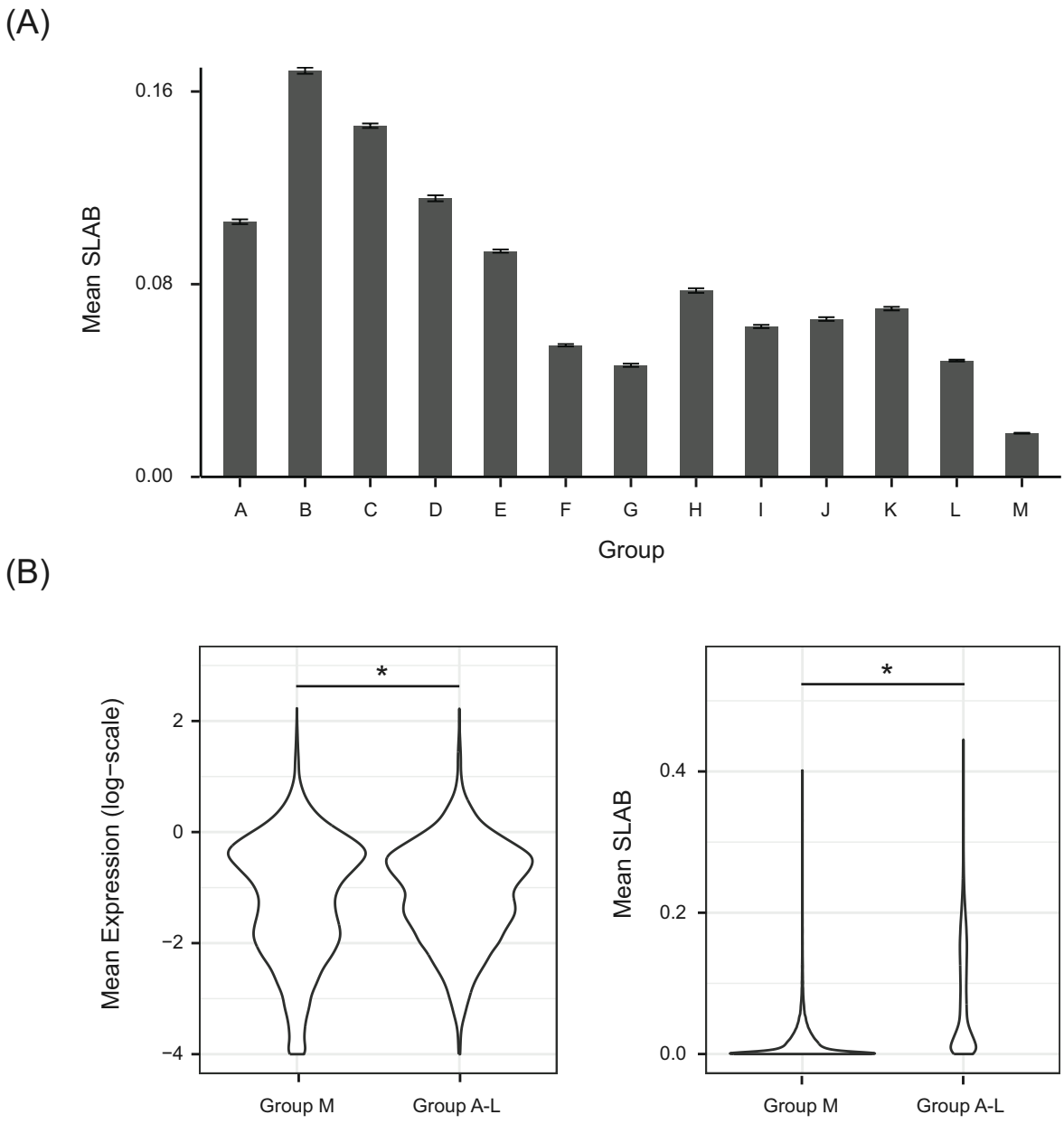
P2 (purple). (* Wilcoxon $p < 0.05$) **(C)** Distribution of SLAB scores (y-axis) for cell types (y-axis) in DLBCL-P1 and DLBCL-P2 (x-axis). **(D)** High-resolution H&E (left) and 51-plex IF (right) images for DLBCL Patients 1 (top) and 2 (bottom). Colors in IF images represent staining for T cells (anti-CD3, blue) and B cells (anti-CD21, red). **(E)** Spatial distribution of spots (grey dots) in DLBCL-P1 (upper) and DLBCL-P2 (lower) within SGs with simultaneous B cell enrichment and T cell depletion. **(D, E)** Dashed circles indicate germinal centers identified from the corresponding H&E images.

Behera et al., Extended Data Figure 5



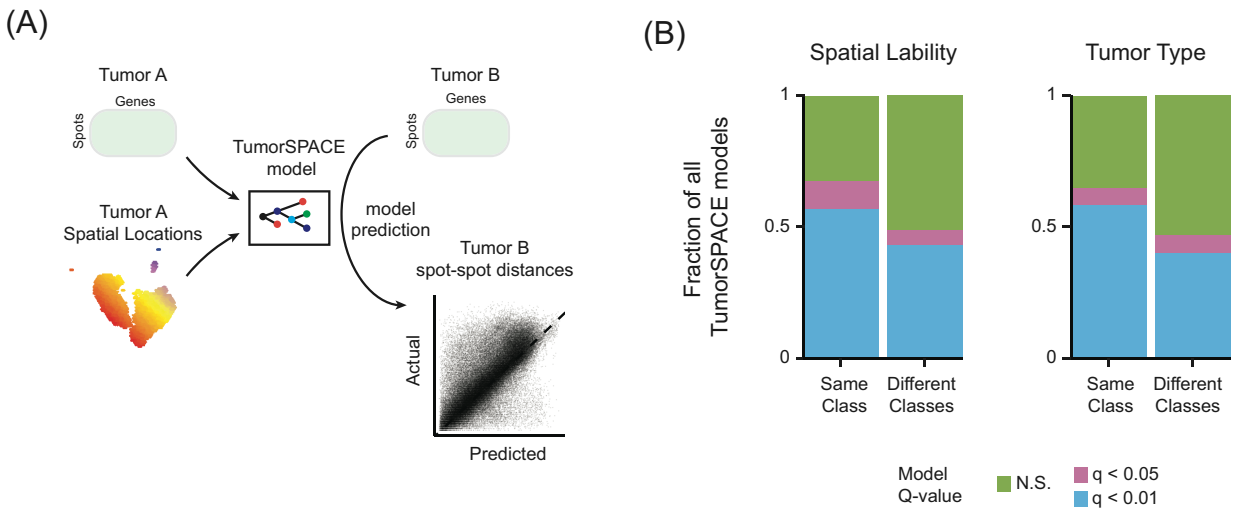
Extended Data Figure 5. Workflow for pan-tumor classification by SLAB scores. First, SLAB scores are computed for each gene for all tumors. This creates a genome-wide profile of SLAB scores for each ST-seq dataset. The datasets are aligned by their genome-wide SLAB profiles creating a matrix where rows are ST-seq datasets, columns are genes, and each entry is the SLAB score for a gene in an ST-seq dataset. Euclidean distance based on genome-wide SLAB scores is computed for all pairs of ST-seq datasets. Hierarchical clustering of pairwise SLAB-based distance results in a pan-tumor classification where tumors that are close together share a similar genome-wide SLAB profile.

Behera et al., Extended Data Figure 6



Extended Data Figure 6. (A) First, for a given gene, the mean SLAB score is computed grouped by spatial lability as in **Fig. 3B** (x-axis). Second, the mean of these group-wise SLAB scores was computed across all genes (y-axis). Error bars depict standard error of the mean. **(B)** Violin plots depicting mean gene expression (log-scale, left) and mean SLAB (right) for all genes where tumors in the pan-tumor database were grouped by whether they belonged to Group M or Groups A-L. * Paired Wilcoxon p < 1e-100.

Behera et al., Extended Data Figure 7



(C)

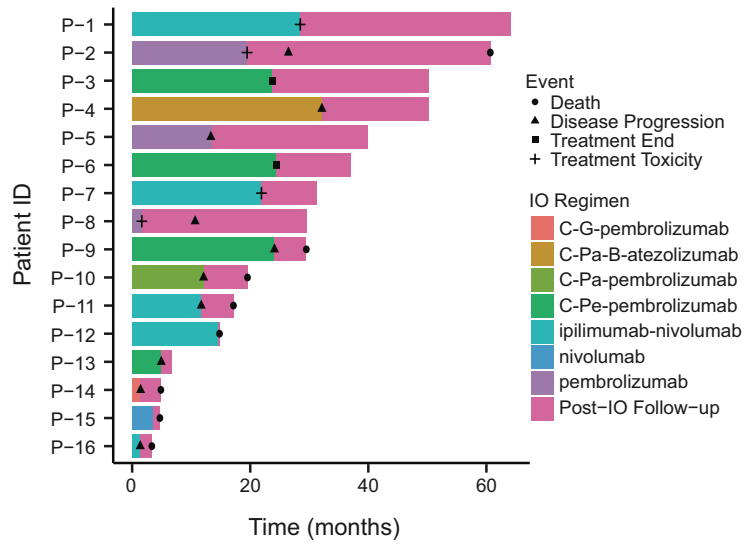
*[A:B Prediction] ~ N(Intercept + ß$_1$*TumorType + ß$_2$*SpatialLabilityType)*

Coefficient	Estimate	Std. Error	Pr(>| *t* |)
Intercept	0.072325	0.002326	< 2e-16
Tumor Type = Same	0.034787	0.004307	1.09e-15
Spatial Lability Class = Same	0.034831	0.005313	6.92e-11

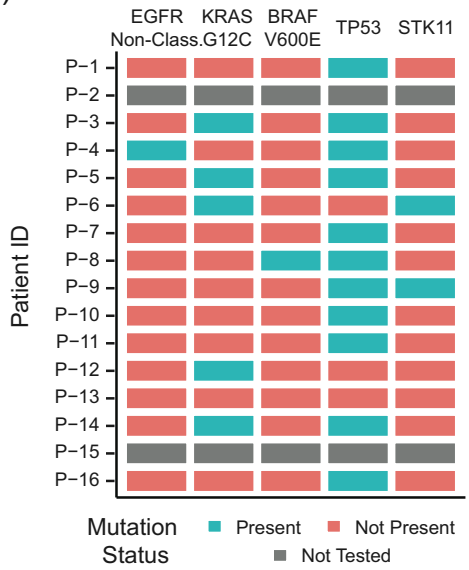
Extended Data Figure 7. (A) Workflow for testing if a TumorSPACE model built for tumor A could predict the spatial organization of tumor B. **(B)** Proportion of all pairs of non-Group M TumorSPACE models that are predictive for spot-spot distances (y-axis) when pairs are stratified as being within the same class or different classes (x-axis). Classes were defined by either spatial lability (left) or by tumor type (right). (**C**) Linear modeling of cross-tumor spatial prediction using either tumor type or spatial lability class as independent variables.

Behera et al., Extended Data Figure 8

(A)



(B)



(C)

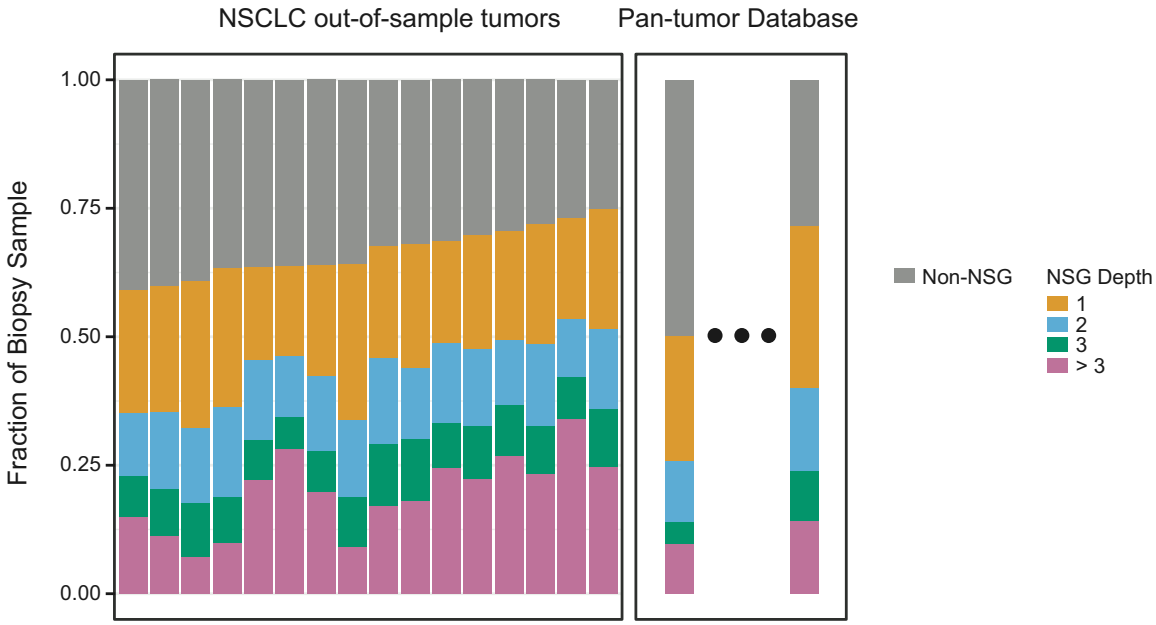
Characteristic	HR[1]	95% CI[1]	p-value
RegimenType			
ChemoICB	—	—	
MonoICB	2.78	0.67, 11.5	0.2
DualICB	0.92	0.22, 3.91	>0.9
KRAS_G12C			
N	—	—	
Y	0.70	0.19, 2.60	0.6

[1] HR = Hazard Ratio, CI = Confidence Interval

Extended Data Figure 8. (A) Swimmer plot illustrating patient treatment courses starting when patients began frontline immunotherapy treatment in the metastatic NSCLC setting. Colors indicate immunotherapy (IO)/chemo-IO regimen, shapes indicate significant events. Chemotherapies are abbreviated as follows: C = carboplatin, G = gemcitabine, Pa = paclitaxel, B = bevacizumab, Pe = pemetrexed. IO therapies include anti-PD1 (pembrolizumab, nivolumab), anti-PD-L1 (atezolizumab), and anti-CTLA-4 (ipilimumab) therapies. **(B)** Mutation status for clinically relevant mutations amongst the 16-patient cohort at the time of pre-treatment diagnostic

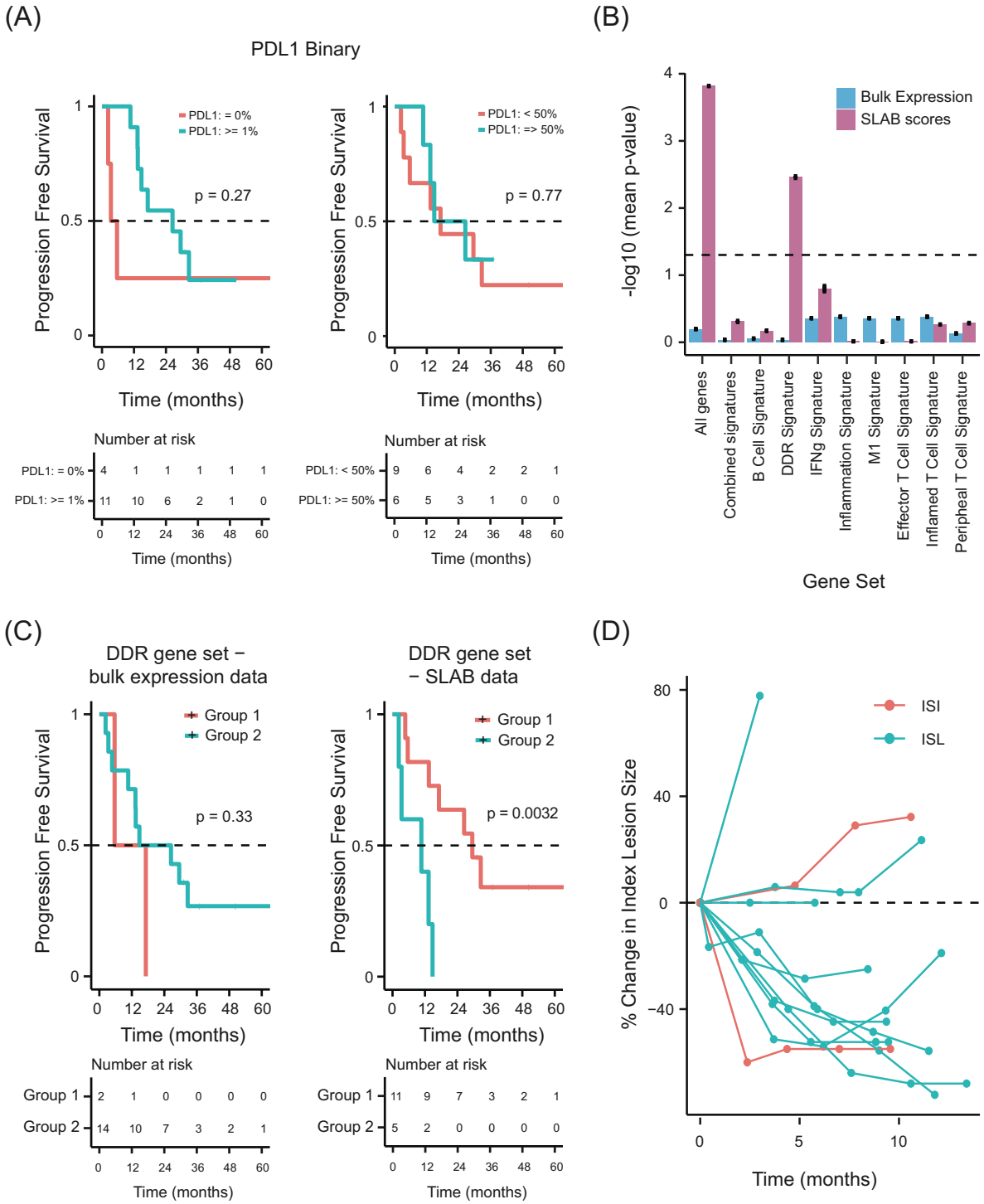
627 biopsy. **(C)** Univariate analysis between (i) ICB regimen type, and (ii) KRAS G12C status and
628 progression-free survival (PFS).
629

Behera et al., Extended Data Figure 9



Extended Data Figure 9. Fraction of spots in an ST-seq dataset (y-axis) belonging to non-NSGs (gray bars) or NSGs of varying depth (colored bars) for NSCLC out-of-sample tumors (left) and for the two tumors in our pan-tumor database representing the highest and lowest non-NSG fraction (right).

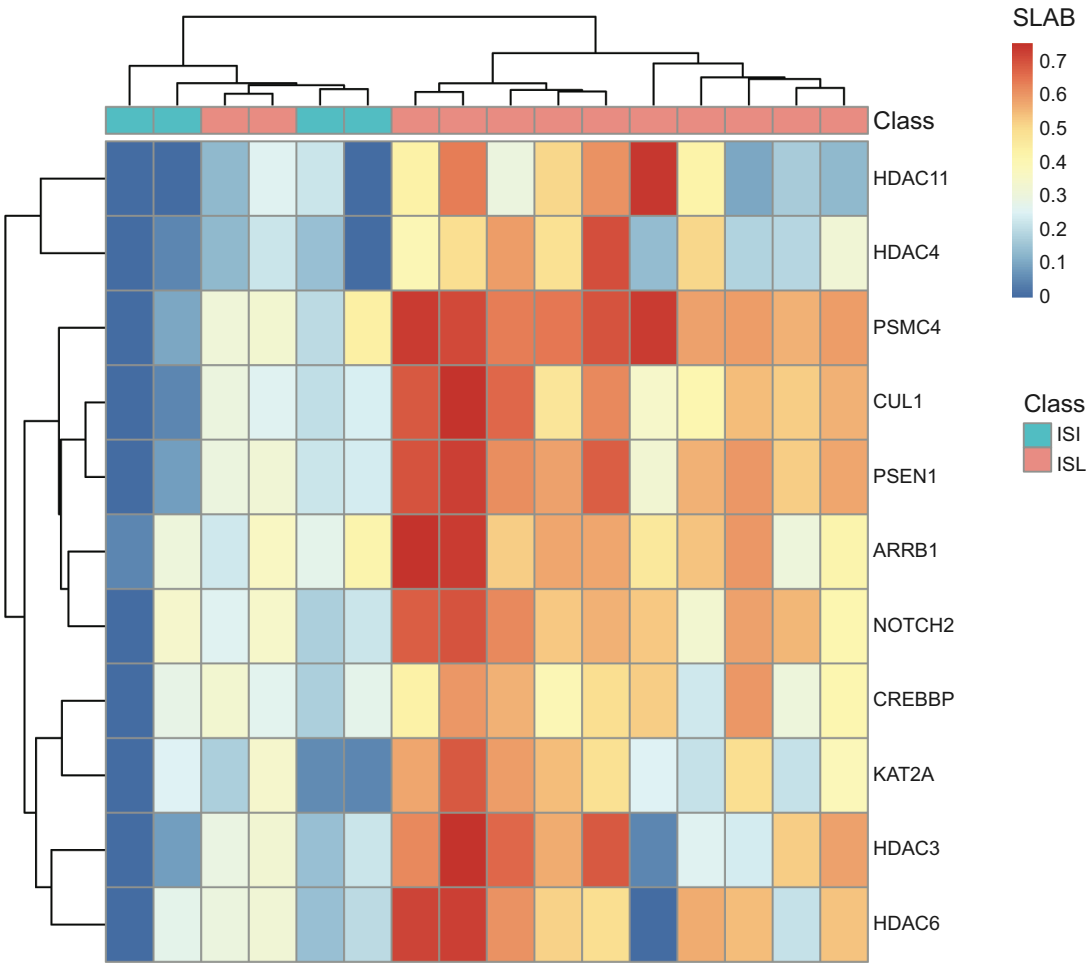
Behera et al., Extended Data Figure 10



Extended Data Figure 10. (A) Kaplan-Meier curve of progression-free survival comparing NSCLC patients of PD-L1 status with binary cutoffs of either 1% (left) or 50% (right) by IHC. (**B**) Mean p-values (log10-transformed, y-axis) of log rank statistical tests using several gene sets (x-axis) to predict progression-free survival amongst NSCLC patients using either bulk gene

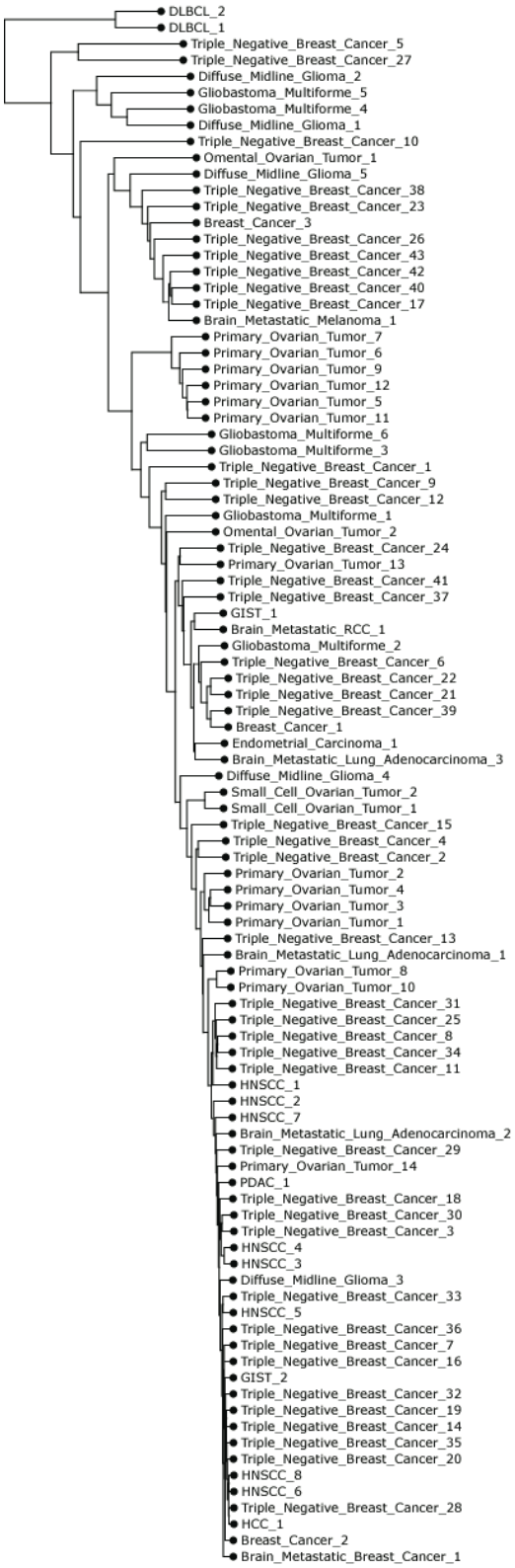
643 expression data (blue) or SLAB score data (purple) (see **Supplementary Table 3** for gene sets).
644 Error bars represent standard error of the mean when performing classification 100 times
645 (Methods). Dashed line indicates p = 0.05. (**C**) Kaplan-Meier curve of progression-free survival
646 comparing NSCLC patients stratified by the DNA Damage Response gene set using either bulk
647 expression data (left) or SLAB score data (right). (**D**) Spider plot depicting percent change in
648 volume of index tumor lesion using serial computed tomography (CT) scans (y-axis) in the months
649 following treatment start (x-axis). Each line describes a single patient classified as either ISL
650 (blue) or ISI (red), and each point on a line indicates a CT scan measurement at that time.
651

Behera et al., Extended Data Figure 11



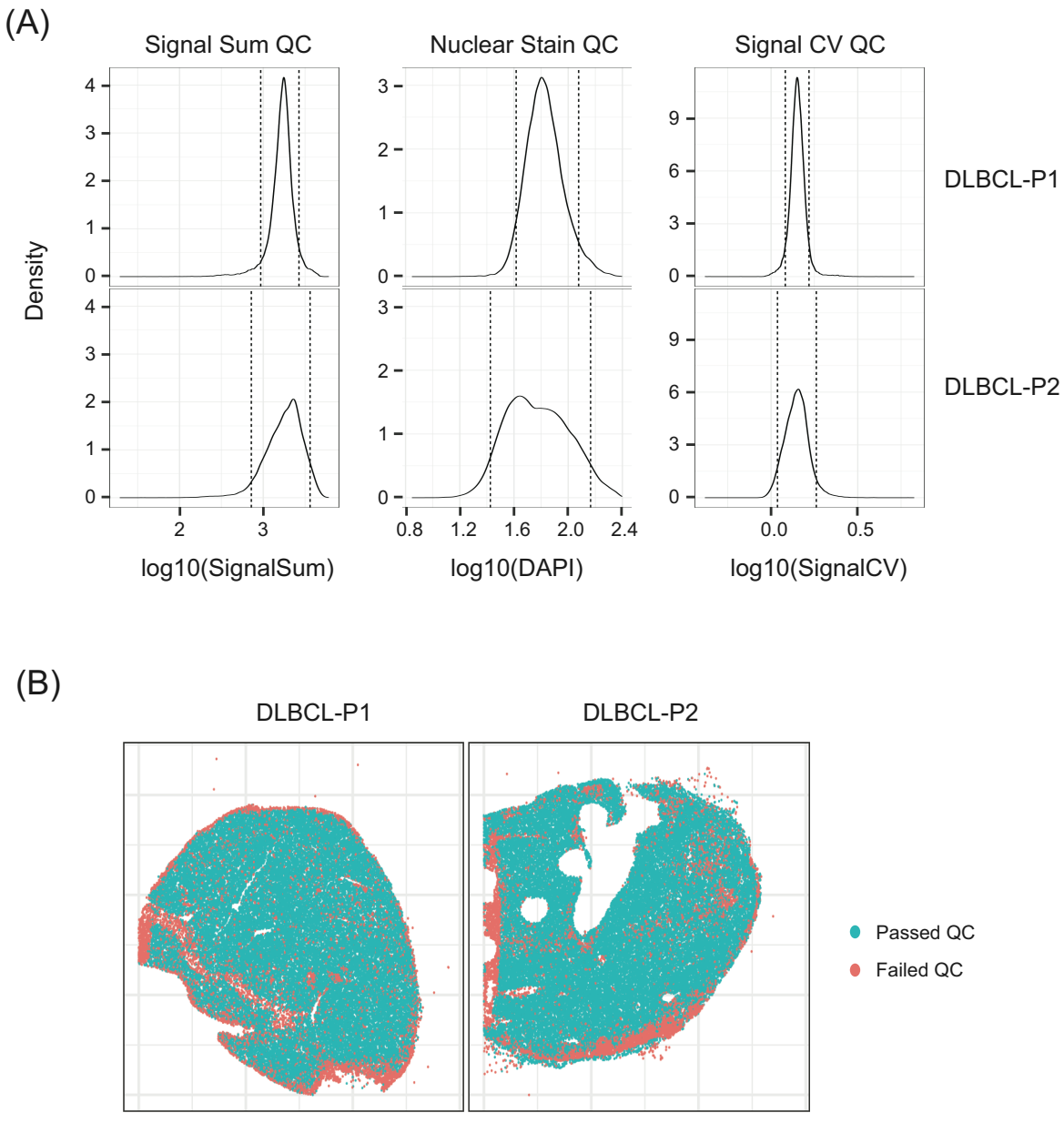
Extended Data Figure 11. Heatmap of SLAB scores (cells, see color key) for 11 NOTCH-pathway genes (rows) in the 16 NSCLC datasets (columns). Both rows and columns are hierarchically clustered by Euclidean Distance. Patients are labeled as either immune spatially labile (ISL, red) or immune spatially invariant (ISI, blue).

Behera et al., Extended Data Figure 12



660 **Extended Data Figure 12.** Pan-tumor dataset classification by SLAB scores. Each leaf in the tree
661 is a distinct patient dataset.
662

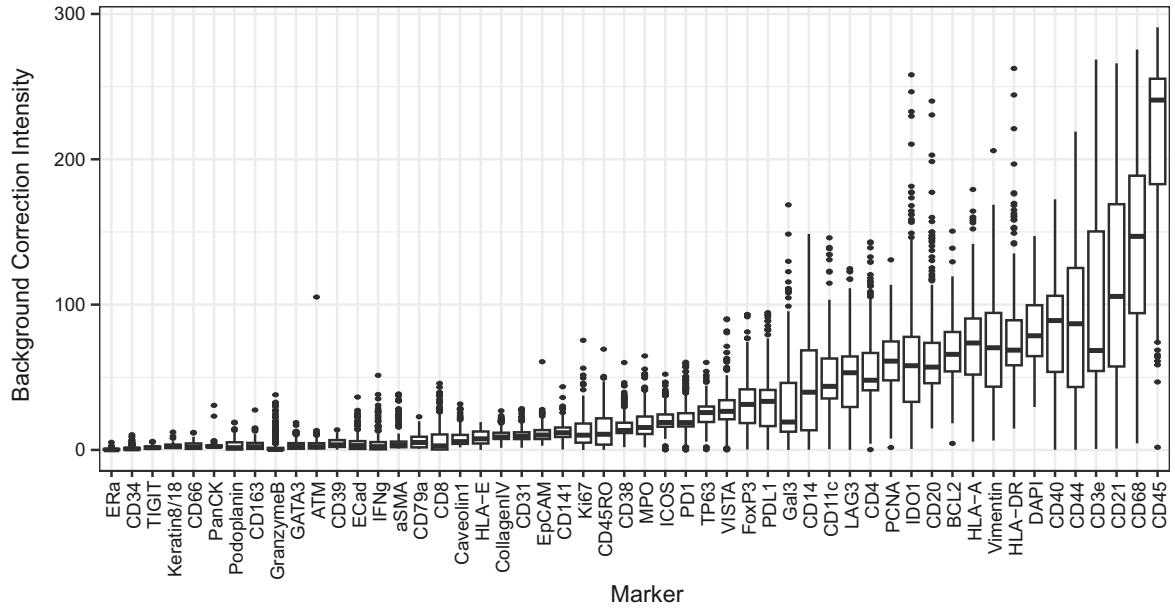
Behera et al., Extended Data Figure 13



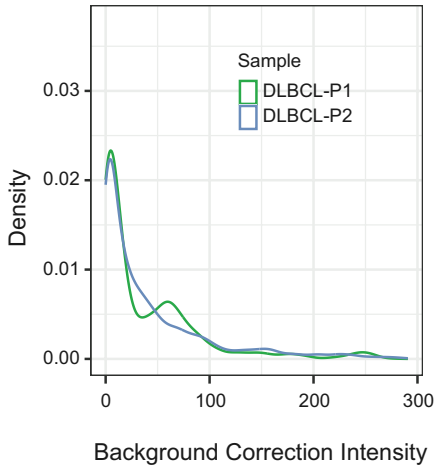
Extended Data Figure 13. (A) Density plots of QC metrics – signal sum across IF markers, signal coefficient of variation (CV) across IF markers, and DAPI intensity – for DLBCL Patients 1 (top) and 2 (bottom). Dotted lines represent 95% (right) and 5% (left) quantile boundaries used to remove outlier cells. **(B)** Spatial locations of spots that passed (green) versus failed (red) the QC thresholds in **(A)**.

Behera et al., Extended Data Figure 14

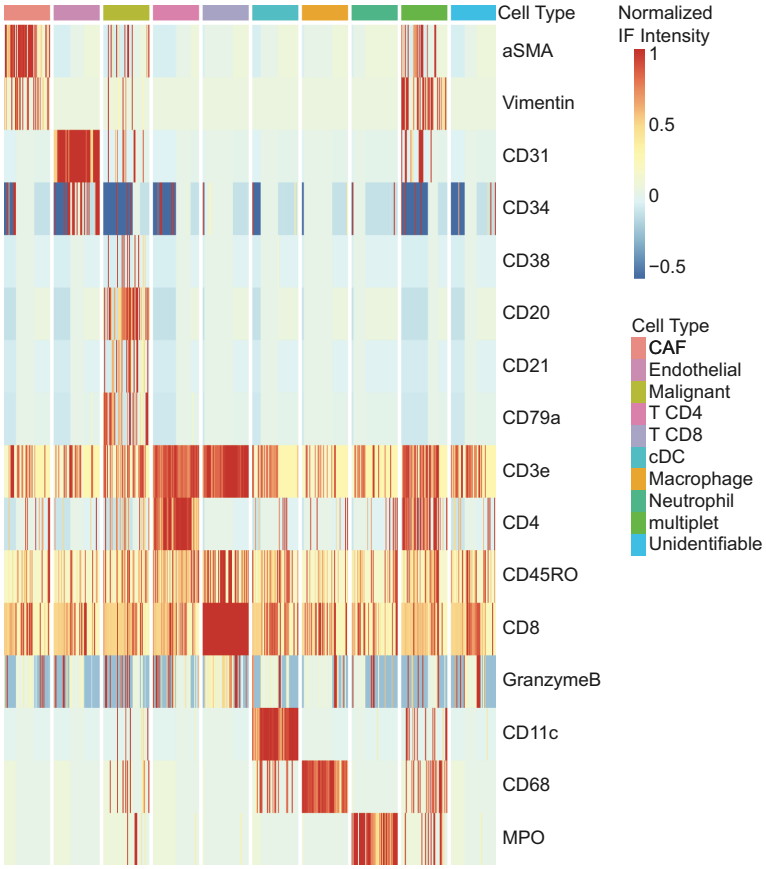
(A)



(B)

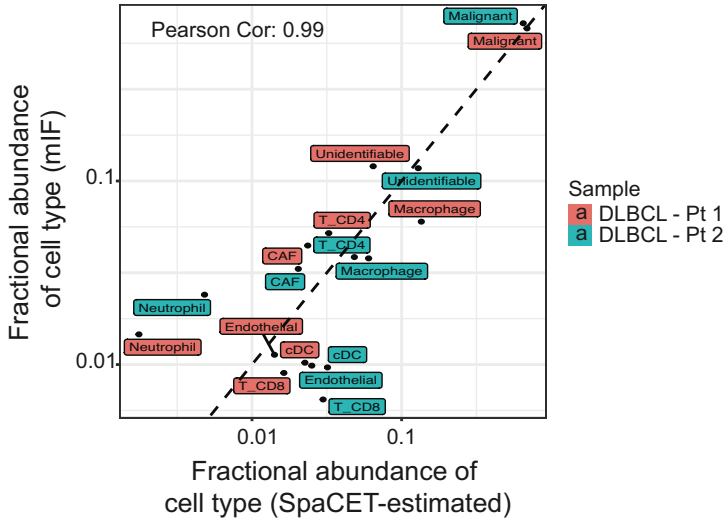


(C)



672 **Extended Data Figure 14. (A)** Background correction intensities (y-axis) for specific IF markers
673 (x-axis). Each boxplot comprises the set of region-specific intensities where each point is the
674 background correction intensity for a given tumor region. **(B)** Density plot of background correction
675 values from (A) for DLBCL Patient 1 (green) and 2 (blue). **(C)** Heatmap depicting normalized
676 intensities for representative markers (rows) from 50 cells (columns) in each cell type
677 classification group (see color key).
678

Behera et al., Extended Data Figure 15



Extended Data Figure 15. Fraction abundance of cell type as determined by multiplexed immunofluorescence (mIF) (y-axis) versus fraction abundance of cell type as determined by SpaCET-estimated deconvolution from ST-seq transcriptional data (x-axis) for two DLBCL patients shown in **Extended Data Fig. 4**. Dashed line indicates linear with associated Pearson correlation.

Methods

Computational method details

ST-seq dataset download and alignment

Previously deposited ST-seq datasets (**Supplementary Table 1**) were downloaded for integration from GEO (https://www.ncbi.nlm.nih.gov/geo/) into the pan-tumor ST-seq database as long as they had the following SpaceRanger outputs available: 1) a spot-by-UMI gene count matrix, 2) a spot-by-pixel location matrix, and 3) a scalefactors_json.json file containing 'spot_diameter_fullres'. For analyses including physical distance rather than pixel distance, pixel distance was converted to physical distance by computing a $\frac{pixel}{\mu m} = \frac{\text{'spot\_diameter\_fullres'}}{55}$ scaling factor that compares spot diameter in pixels to the known spot diameter of 55 μm.

SpaceRanger

For internally generated ST-seq datasets, reads were aligned and mapped to the hg38 (GRCh38) human genome reference using the SpaceRanger v2.0.0 count pipeline (**Supplementary Table 4**). This pipeline generates a raw unique molecular identifier (UMI) gene count matrix in which each row consists of a spot that has X/Y coordinates in pixels that correspond to the aligned H&E image. The SpaceRanger algorithm also identifies spots within or outside of detectable tissue, and for all subsequent analyses only spots within tissue were used.

TumorSPACE: models and associated analysis

The sub-sections within this section will introduce a number of variables. As such, below is a table of variable definitions.

Variable	Definition	Section Where First Referenced
M	SpaceRanger UMI gene count matrix	*Creating a latent space*
m	Number of ST-seq spots in M	*Creating a latent space*
n	Number of genes in M	*Creating a latent space*
U	SVD left singular matrix	*Creating a latent space*
Σ	SVD singular value matrix	*Creating a latent space*
V^T	Transpose of SVD right singular matrix	*Creating a latent space*
D	Spectral Distance matrix	*Creating a latent space*
p	PC depth hyperparameter	*Creating a latent space*
T	TumorSPACE tree model	*Creating a latent space*
G	The set of tree internal nodes	*Creating a latent space*
g	A single tree internal node	*Creating a latent space*
M_B	Bootstrapped gene count matrix	*Bootstrapping the latent space…*
X	Statistical random variable	*Bootstrapping the latent space…*
N	Normal Distribution	*Bootstrapping the latent space…*
μ	Mean	*Bootstrapping the latent space…*
σ	Standard deviation	*Bootstrapping the latent space…*
T_B	Bootstrapped TumorSPACE tree	*Bootstrapping the latent space…*
b	Node TBE support	*Bootstrapping the latent space…*
k	Node spatial dispersion	*Calculating physical spatial dispersion…*
K	Ripley's reduced second moment function with border correction	*Calculating physical spatial dispersion…*

g_s	The set of spots in node *g*	*Calculating physical spatial dispersion…*
len_x	The x-axis range of values for a set of spots	*Calculating physical spatial dispersion…*
len_y	The y-axis range of values for a set of spots	*Calculating physical spatial dispersion…*
λ	Ripley intensity normalization factor	*Calculating physical spatial dispersion…*
r_{max}	Ripley maximum spot distance	*Calculating physical spatial dispersion…*
R	The set of spot distances used to compute spatial dispersion	*Calculating physical spatial dispersion…*
r	A particular spot distance for computing spatial dispersion	*Calculating physical spatial dispersion…*
δ	Spot pairwise physical distance	*Calculating physical spatial dispersion…*
Δ	Matrix of pairwise spot-spot physical distances	*Calculating physical spatial dispersion…*
X_{spot}	X axis physical location of a spot	*Calculating physical spatial dispersion…*
Y_{spot}	Y axis physical location of a spot	*Calculating physical spatial dispersion…*
κ	The number of spot KNN matches	*Calculating physical spatial dispersion…*
P	The set of *p* being tested	*Calculating physical spatial dispersion…*
n_p	The length of set *P*	*Calculating physical spatial dispersion…*
Uniform([a,b])	The Uniform Distribution between a and b	*Calculating physical spatial dispersion…*
H	The set of hyperparameters for *b, k, and* κ	*Calculating physical spatial dispersion…*
n_H	The length of set H	*Calculating physical spatial dispersion…*
h_i	A given choice of hyperparameter values for *b, k, and* κ	*Calculating physical spatial dispersion…*
G_{filt}	The set of internal nodes in *G* that meet a given set of hyperparameter bounds on *b and k*	*Calculating physical spatial dispersion…*
G_{filt}'	The set of G_{filt} plus the parent nodes of G_{filt} within *T*	*Calculating physical spatial dispersion…*
T_{filt}	Tree *T* filtered for internal nodes within G_{filt}'	*Calculating physical spatial dispersion…*
d	Spot pairwise spectral distance	*Prediction accuracy calculation*
I	The set of all spots in a biopsy	*Prediction accuracy calculation*
$NN_{i,\kappa}$	The set of κ nearest neighbors to spot *i* in latent space *T*	*Prediction accuracy calculation*
$spot_{i,z}$	A given nearest neighbor spot within $NN_{i,\kappa}$	*Prediction accuracy calculation*
ρ	Pearson Correlation Coefficient	*Prediction accuracy calculation*

corr	Pearson Correlation function	*Prediction accuracy calculation*
vec	Matrix vectorization function	*Prediction accuracy calculation*
S_{spot}	The tuple of spot locations (X_{spot}, Y_{spot})	*Prediction accuracy calculation*
T_{filt}^{opt}	The optimized TumorSPACE model for a given tumor biopsy	*Prediction accuracy calculation*
$A_{g_i}^{path}$	The full ancestral node path for an internal node g_i in any tree *T*	*Spatial Group (SG) depth*
$a_{g_i}^{k}$	The k[th] ancestor of an internal node g_i in any tree *T*	*Spatial Group (SG) depth*
σ_{node_g}	The spatial domain size of an internal node g_i in any tree *T*	*Spatial Group (SG) depth*
λ_{node_g}	The SG depth of an internal node g_i in any tree *T*	*Spatial Group (SG) depth*
G_{DA}	The subset of internal nodes within tree *T* that will be used for differential abundance computation	*SG-based differential abundance*
W(a,b)	Wilcoxon rank-sum test between *a* and *b*	*SG-based differential abundance*
$p_{g_{DA},f}$	Differential Abundance Probability at node g_{DA} for process *f*	*SG-based differential abundance*
$q_{g_{DA},f}$	Differential Abundance Probability at node g_{DA} for process *f*, empirically bootstrapped and multiple hypothesis adjusted	*SG-based differential abundance*
DA	The set of differentially abundant spots for process *i* at node *j* in a given tumor biopsy	*SG-based differential abundance*
$OR_{i,j}$	The odds ratio of independence between process *i* in an NSG and process *j* in a parent SG	*Contextual dependence of processes…*
$SLAB$	The SLAB Score for process *i* in a given tumor biopsy	*SG-based spatial lability (SLAB) score*
L^k	The matrix of SLAB scores composed of samples (rows) in groups K and nK	*Differential SLAB score analysis*
MW(a,b)	Mann-Whitney U Test between *a* and *b*	*Differential SLAB score analysis*
$p_{K,f}$	p-value of the MW test between samples in groups K vs nK for process *f*	*Differential SLAB score analysis*
$p_{K,f}^{shuffled}$	p-value of the MW test between samples shuffled between groups K vs nK for process *f*	*Differential SLAB score analysis*

Overview

Building a TumorSPACE model requires spatial transcriptomic data and two inputs from SpaceRanger: (1) the raw gene UMI count matrix and (2) the spot spatial coordinate matrix. Model building subsequently operates on the gene count matrix to build many models that vary in hyperparameter choice. The spot spatial coordinates are then used for selecting the optimal hyperparameter set that maximizes accurate recovery of spatial spot organization.

Four hyperparameters are tuned during this process: (1) the number of principal components (PCs) of data-variance used for creating a latent space of the transcriptional data, (2) the limit of statistical robustness for spot-spot relatedness in the latent space, (3) the spatial dispersion of the nodes in the latent space hierarchical tree model, and (4) the number of KNN matches used for spot spatial prediction. The following sections will first establish the model latent space and compute statistical robustness and spatial dispersion properties of that latent space. Subsequently, all hyperparameters will be tuned to define the optimal model for mapping transcriptional content from TME spots to TME spatial organization.

Creating a latent space

The first step in building a TumorSPACE model is to create a latent space representation of the gene count data that incorporates statistical bootstrapping. TumorSPACE first embeds ST-seq spots into a latent space by applying singular value decomposition (SVD) to the gene count matrix[62]:

$$M = U\Sigma V^T \quad (1)$$

M is the SpaceRanger gene count matrix (*m* spots as rows, *n* genes as columns), *U* is the left singular matrix, Σ is the singular value matrix, and *V* is the right singular matrix. *U* is defined by cell spots (rows) and left singular vectors (columns), where each entry is the projection of a cell spot onto a left singular vector. Σ is a diagonal matrix where entries are singular values. V^T is defined by genes (rows) and right singular vectors (columns) where each entry is the projection of a gene onto a right singular vector.

First, from (1), a metric termed 'spectral distance' (*D*) between all spots is calculated. This metric was previously developed by our laboratory in the context of analyzing phylogenetic bacterial proteome content[9]. As implemented for spatial data in this manuscript, performing SVD on the gene count matrix determines the extent to which each cell spot projects onto each left singular vector. Therefore, a distance considering the transcriptomes of two spots can be computed by measuring the difference in the projections of two spots onto a left singular vector. Note, this definition of distance does not consider any information about spatial spot distribution.

Next, groups of left singular vectors are combined to create 'spectral groups'. These groups are defined based on the eigenvalues associated with each left singular vector: left singular vectors with similar eigenvalues are grouped together:

$$SG = \{sg_1, sg_2, \ldots\} \quad (2)$$

where *SG* is the total set of spectral groups, $sg_1$is first set of columns extracted from *U*, sg_2 is the second set of columns extracted from *U*, and so on. The concept of spectral groups was also previously developed by our laboratory[38]. Defining sg_i and sg_{i+1} is done by identifying larger than expected decreases in singular values between consecutive left singular vectors. To compute spectral groups, a vector of differences between consecutive singular values is computed for all left singular vectors. We use the upper and lower quartiles of this distribution in combination with a scaling parameter alpha to define the 'expected difference' bounds between singular values. Any difference in singular values outside of these bounds deviates from expectation and therefore defines a spectral group (see associated GitHub code for specification of parameters). The spectral distance for a pair of spots within a spectral group is then computed as the Euclidean distance between spot projections onto left singular vectors comprising the spectral group weighted by the eigenvalue associated with each left singular vector. The summation of these distances across all spectral groups is the spectral distance, $d_{i,j}$, between spots i and j.

After computing $d_{i,j}$ for all spots, the resulting construct is a spectral distance matrix *D* comprised of *m* rows and *m* columns where *m* is the number of spots in the original gene count matrix and each entry in *D* is the spectral distance between two spots. *D* is then used as input for hierarchical clustering with complete linkage to result in a tree *T* that relates all spots in a tumor sample to each other. *T* has *m* leaves and (*m-1*) internal vertices (nodes). The leaves are the ST-seq spots and the nodes $g \in G$ represent *G* hierarchically ordered groupings of these spots. The resulting network is the TumorSPACE latent space of the original gene count matrix.

The number of spectral groups is dependent on how many of the total left singular vectors are considered. An increasing number of left singular vectors being included corresponds directly to the inclusion of deeper principal components when computing the latent space. For TumorSPACE models, the depth of principal components, '*p*', is a hyperparameter that is tuned for embedding the gene count matrix into a latent space.

Bootstrapping the latent space to evaluate statistical robustness

TumorSPACE does not assume that each node *g* arises from biological signal. Instead, TumorSPACE bootstraps *T* using the Booster package's implementation of transfer bootstrap expectation (TBE), the probability that node *g* appears in an empirically bootstrapped tree (default settings used for Booster)[63]. For generating empirically bootstrapped trees, we applied Gaussian multiplicative noise injection to the initial gene count matrix *M* to create a "bootstrapped" gene count matrix M_B.

$$M_B = M \odot X \tag{3}$$

such that $\odot$ indicates element-wise multiplication by a normally distributed random variable $X \sim N(\mu, \sigma^2)$ with $\mu = 1$ and $\sigma = 0.2$. This matrix was then used as an input to (1) and a tree was created following the steps outlined in '*Creating a latent space*' to generate a bootstrapped tree T_B. Bootstrapping was done 10 times for a given dataset, followed by input of the original tree *T* and the bootstrapped trees T_B into Booster for TBE computation. This results in a labeling of the original tree *T*'s set of nodes *G* with TBE support values b_G such that $b_G \in [0,1]$.

Calculating physical spatial dispersion in latent space

The final property of *T* that is computed is the spatial dispersion *k* for each node comprising *T*. Spatial dispersion is estimated for each node using Ripley's reduced second moment function *K(r)* with border correction[64,65]. Let g_s be the set of ST-seq spots within node *g* in *T*. The window of physical tumor space is defined by the spot spatial coordinate matrix such that len_x indicates the x-axis window length and len_y indicates the y-axis window length. We then compute λ, a normalization factor for spot intensity within a spatial region, and r_{max}, a factor that incorporates lambda to determine the maximum spatial distance being assessed.

$$\lambda = \frac{|g_s|}{len_x * len_y} \tag{4}$$

$$r_{max} = \min\left\{\min\left\{len_x, len_y\right\}, \sqrt{\frac{1000}{\pi * \lambda}}\right\} \tag{5}$$

where *min* denotes the minimum between a set of values. Let *R* be the set of spot spatial distances that will be assessed, such that

$$r \in R \mid R = \left\{0, \frac{r_{max}}{512}, \frac{2 * r_{max}}{512}, \ldots, r_{max}\right\} \tag{6}$$

We define the physical distance δ between any two spots as

$$\delta(spot_i, spot_j) = \Delta_{i,j} = \Delta_{j,i} = \sqrt{(X_{spot_i} - X_{spot_j})^2 + (Y_{spot_i} - Y_{spot_j})^2} \quad (7)$$

where (X_{spot},Y_{spot}) denote the physical space coordinates for a given spot. Spatial dispersion *K(r)* with border correction is then computed for all spots $g_{s,i}, \in g_s$ as

$$t^r_{g_{s,.i}} = f(g_{s,i}, r) = \sum_{j=1}^{m} \mathbf{1}\{0 < \delta(g_{s,i}, spot_j) < r\} \quad (8)$$

$$K(r) = \frac{\sum_{i=1}^{card(G)} (\mathbf{1}\{b_i \geq r\} * t^r_{g_{s,.i}}}{\lambda * \sum_{i=1}^{card(G)} \mathbf{1}\{b_i \geq r\}} - \pi r^2 \quad (9)$$

where $t(g_{s,i}, r)$ is the number of spots within distance *r* of a given $g_{s,i}$ and b_i is the distance from spot $g_{s,i}$ to the window boundary. The general notation *card(S)* indicates the number of elements in a set S, and the general notation ***1****{f(x)}* signifies a value of 1 when f(x) is true and a value of 0 when *f(x)* is false. Finally, spatial dispersion *k* is computed by summing the absolute value of *K(r)* over $r \in R$ as follows.

$$k = \sum_{r \in R} abs(K(r)) \quad (10)$$

This calculation labels all nodes *G* in tree *T* with spatial dispersion values k_G such that $k_G \in_{\mathbb{R}} [0, \infty]$.

Hyperparameter optimization to create a TumorSPACE map

TumorSPACE model optimization involves selecting the values of four hyperparameters that maximize model prediction accuracy (described in '*Prediction Accuracy Calculation*') for a given dataset. These hyperparameters tune three properties of tree *T* – principal component depth *p* (from '*Creating a latent space*'), node TBE support *b* (from '*Bootstrapping the latent space to evaluate statistical robustness*'), node spatial dispersion *k* (from '*Calculating physical spatial dispersion in latent space*') – as well as one property of accuracy computation, the number of spot KNN matches κ. We perform hyperparameter tuning as a nested grid search by tuning *p* as an outer layer and then optimizing [*b, k,* κ] for a given value of *p*.

First, a set of PC depth values (n_p where default is set to 10) is randomly selected to create a set $P = \{p_1, p_2, \ldots, p_{n_p}\}$. The PCs termed p_i are chosen on a logarithmic interval between a minimum and maximum PC depth, which is the rank of the gene count matrix *M*. Next, a matrix of three hyperparameter values, $H = \{h^{(1)}_{1\ldots n_H}, h^{(2)}_{1\ldots n_H}, h^{(3)}_{1\ldots n_H}\}$, are created where the vectors $h^{(1)}$, $h^{(2)}$, and $h^{(3)}$ are independently sampled from distributions as follows.

$$h^{(1)} \in_{\mathbb{R}} 10^{X * log10\left(\frac{b_{max}+1}{b_{min}+1}\right) + log10(b_{min}+1)} - 1 \quad (11)$$

$$h^{(2)} \in_{\mathbb{R}} 10^{X * log10\left(\frac{k_{max}+1}{k_{min}+1}\right) + log10(k_{min}+1)} \quad (12)$$

$$h^{(3)} \in_{\mathbb{Z}} round(10^{X * log10\left(\frac{\kappa_{max}+1}{\kappa_{min}+1}\right) + log10(\kappa_{min}+1)}) \quad (13)$$

In (*11-13*), X is a random variable drawn from $Uniform([0,1])$. Default values for hyperparameter bounds are $b_{min} = 0$, $b_{max} = 0.5$, $k_{min} = 0$, $k_{max} = 1$ $\kappa_{min} = 5$, $\kappa_{max} = 300$. A minimum of $n_H = 100$ sets of $\{h^{(1)}, h^{(2)}, h^{(3)}\}$ are initially sampled, after which additional sets are sampled until prediction accuracy optimization has converged. Prediction accuracy convergence

is reached when the difference in prediction accuracy (defined below in '*Prediction Accuracy Calculation'*) for the top 2 scoring hyperparameter sets is less than 0.05. For a given hyperparameter set h_i, the TumorSPACE tree T is filtered for the set of nodes G_{filt} such that each node in G_{filt} satisfies

$$b \geq h_i^{(1)} \qquad \text{AND} \qquad k \geq h_i^{(2)} \tag{14}$$

The final filtered tree, T_{filt}, comprises the set of nodes G_{filt}', which consists of G_{filt} as well as the complete set of parent nodes from which G_{filt} descend even if those parent nodes do not meet the criteria in *(14)*, along with all ST-seq spots.

Prediction accuracy calculation

To identify the TumorSPACE model properties that were optimized for predicting spot spatial locations from transcriptomic data, we masked the physical location of each ST-seq spot and identified its *k* nearest neighbors in the TumorSPACE latent space by minimizing spectral distance.

For any masked spot *i* amongst all spots *I*, we can define its κ nearest neighbors $NN_{i,\kappa}$ as

$$NN_{i,\kappa} = argmin_{j \in J}^{\kappa} \left(d\left(spot_i, spot_j\right) \right) \tag{15}$$

where $i \in I$, $\kappa \in h^{(3)}$ as defined in (*13*), *J* is the set of all spots other than spot *i,* and *argmin*$^{\kappa}$ selects the set of κ spots with the smallest spectral distance relative to spot *i*. To prevent overfitting, we identified for each $spot_{i,z} \in NN_{i,\kappa}$ a randomly chosen $spot'_{i,z}$ that belongs to the internal node $g^{i,z}$ within T_{filt} immediately ancestral to $spot_{i,z}$.

We then estimated the location of masked *spot i* based on the x and y locations of the corresponding $spot'_{i,z}$ spots.

$$\hat{X}_{spot_i} = \frac{1}{\kappa}\sum_{z=1}^{\kappa} X_{spot'_{i,z}} \text{ and } \hat{Y}_{spot_i} = \frac{1}{\kappa}\sum_{z=1}^{\kappa} Y_{spot'_{i,z}} \tag{16}$$

Finally, we computed the Pearson Correlation ρ between the vectorized matrix Δ_{actual} of pairwise actual spot-spot physical distances and the vectorized matrix $\Delta_{predicted}$ of pairwise predicted spot-spot physical distances.

$$\Delta_{actual}[i,j] = \sqrt{(X_{spot_i} - X_{spot_j})^2 + (Y_{spot_i} - Y_{spot_j})^2} \tag{17}$$

$$\Delta_{predicted}[i,j] = \sqrt{(\hat{X}_{spot_i} - \hat{X}_{spot_j})^2 + (\hat{Y}_{spot_i} - \hat{Y}_{spot_j})^2} \tag{18}$$

$$\rho = corr\left(vec(\Delta_{actual}), vec(\Delta_{predicted})\right) \tag{19}$$

where *vec()* indicates matrix vectorization to a single column and *corr()* indicates Pearson Correlation. To compute a null distribution for ρ using empirical bootstrapping of actual versus predicted spot locations in a given dataset, we shuffled the vector $\hat{S}_{spot_i} = (\hat{X}_{spot_i}, \hat{Y}_{spot_i})$ without replacement and then re-computed *(17-18)* using this shuffled vector $\hat{S}_{spot_i}^{shuffled}$ of predicted spot locations.

$$\hat{S}_{spot}^{shuffled} = \{K \subseteq \hat{S}_{spot_i} | \, card(K) = card(I)\} = (\hat{X}_{spot}^{shuffled}, \hat{Y}_{spot}^{shuffled}) \quad (20)$$

$$\Delta_{predicted}^{shuffled}[i,j] = \sqrt{(\hat{X}_{spot_i}^{shuffled} - \hat{X}_{spot_j}^{shuffled})^2 + (\hat{Y}_{spot_i}^{shuffled} - \hat{Y}_{spot_j}^{shuffled})^2} \quad (21)$$

$$\rho^{shuffled} = corr\left(vec(\Delta_{actual}), vec(\Delta_{predicted}^{shuffled})\right) \quad (22)$$

For **Fig. 1B**, $\rho^{shuffled}$ is computed for 100 shuffles and the maximum ρ is taken as the 'null' prediction value. The null distribution is plotted in the grey distribution in **Fig. 1B**.

Finally, the optimal TumorSPACE model T_{filt}^{opt} is found that maximizes ρ across hyperparameter sets *P* and *H*.

$$T_{filt}^{opt} = argmax_{p \in P, h \in H}^{1} (\rho_{p,h}) \quad (23)$$

TumorSPACE model outputs

For a given input tumor ST-seq dataset, the output from TumorSPACE includes: (1) the TumorSPACE model T_{filt}^{opt}, (2) the Pearson Correlation estimate ρ, and (3) the set of predicted spot locations $(\hat{X}_{spot_i}, \hat{Y}_{spot_i})$ for all ST-seq spots. The final set of internal nodes within T_{filt}^{opt} are termed Spatial Groups (SGs).

Spatial Group (SG) depth

We computed SG depth as a measurable quantity that describes how a given SG relates to the other parts within a TumorSPACE model. As such, we first define 'SG depth' as a property of all SGs within a TumorSPACE model, and next define 'spot SG depth' as a property of all spots within the gene count matrix.

To first define SG depth, we compute the complete ancestral node path for any internal node g_i within T_{filt}^{opt} as

$$A_{g_i}^{path} = \{a_{g_i}^0, a_{g_i}^1, a_{g_i}^2, \ldots, a_{g_i}^n\} \quad (24)$$

such that

$$a_{g_i}^0 = g_i \quad (24)$$

$$a_{g_i}^{k+1} = A\left(a_{g_i}^k\right) for \; k \in [1, n-1] \quad (25)$$

$$a_{g_i}^n \in G \quad (26)$$

$a_{g_i}^k$ indicates the k^{th} ancestral node of node g_i, *A(node)* denotes the immediate ancestral node of a given node in T_{filt}^{opt}, and *G* is the set of internal nodes in T_{filt}^{opt}. By definition, $a_{g_i}^n$ will be the root node of T_{filt}^{opt}. We next define the spatial domain size σ_{node_g} for a given node *g* as the mean spot-spot physical distance between all spots within *g*.

$$\sigma_{node_g} = \frac{\sum_{i=1}^{I}\sum_{j=1}^{J}\delta(spot_i, spot_j)}{I*J} \; where \; I = J = card(g_s) \quad (27)$$

Finally, we identify the subset of nodes $A_{g_i}^{nested}$ within $A_{g_i}^{path}$ that satisfy the condition whereby the *(k+1)*th node is equal to or larger in spatial domain size than the *k*th node in that path.

$$A_{g_i}^{nested} \in A_{g_i}^{path} \mid \begin{cases} A_{g_i}^{path} = \{a_{g_i}^0, a_{g_i}^1, a_{g_i}^2, \dots, a_{g_i}^n\} \\ A_{g_i}^{nested} = \{a_{g_i}^0, a_{g_i}^1, a_{g_i}^2, \dots, a_{g_i}^k\} \\ \sigma_{a_{g_i}^{l+1}} \geq \sigma_{a_{g_i}^l} \end{cases} \quad (28)$$

where $k \leq n$ and $0 \leq l \leq k$. The SG depth, λ_{node_g}, for a given internal node g_i is defined to be the number of ancestral generations that satisfy this condition of spatial domain nesting.

$$\lambda_{node_g} = card\left(A_{g_i}^{nested}\right) - 1 \quad (29)$$

SG-based differential abundance

Differential abundance calculation requires two inputs: (1) an optimized TumorSPACE model T_{filt}^{opt} and (2) a spot-by-feature matrix *F*. We computed differential abundance using three types of biological processes: genes, pathways, and deconvoluted cell type proportions. Computation of gene count, pathway usage, and cell type proportion matrices are described in the *'SpaceRanger'*, *'GSVA'*, and *'SpaCET'* Methods sections, respectively. The gene count matrix is normalized by the spot-wise total UMI count.

First, we identified a subset of SGs $G_{DA} \in G$ at which DA will be computed. We set a minimum of 10 spots that must be present in both a given SG $g_{DA} \in G_{DA}$ and in its sibling node g'_{DA} (e.g. *A* and *A'* in **Fig. 2A**) for inclusion within G_{DA}.

$$G_{DA} = \left[g_{DA} \in G'_{filt} \mid card(C(g_{DA})) \geq 10 \;\&\; card(C(g'_{DA})) \geq 10\right] \quad (30)$$

where *C(n)* indicates the row indices within matrix *F* of the spots descending from SG g_{DA}. Subsequently, for each node g_{DA} and process *f*, the spot-wise process values between g_{DA} and g'_{DA} are compared using a two-sided Wilcoxon Rank Sum Test, where the test p-value is given by *W(a,b)*.

$$p_{g_{DA},f} = W(F_{C(g_{DA}),f}, F_{C(g'_{DA}),f}) \quad (31)$$

To facilitate empirical correction for multiple hypothesis testing, we perform 20 shuffles of the process values between g_{DA} and g'_{DA}, followed by computation of the Wilcoxon p-value between these shuffles. Let $C_{n_{total}}$ be the concatenation of spot indices $C(g_{DA})$ and $C(g'_{DA})$.

$$C_{g_{total}} = C(g_{DA}) + C(g'_{DA}) \quad (32)$$

$$C_{g_{total}}^{shuffled} = \{K \subseteq C_{g_{total}} \mid card(K) = card(C_{g_{total}})\} = C_{g_{DA}}^{shuffled} + C_{g'_{DA}}^{shuffled} \quad (33)$$

$$C_{g_{DA}}^{shuffled} = C_{g_{total}}^{shuffled}[1: card(C(g_{DA}))] \quad (34)$$

$$C_{g'_{DA}}^{shuffled} = C_{g_{total}}^{shuffled}[(card\left(C(g_{DA})\right) + 1): (card(C(g_{DA})) + card(C(g'_{DA})))] \quad (35)$$

$$p_{g_{DA},f}^{shuffled} = W(F_{C_{g_{DA}}^{shuffled},f}, F_{C_{g'_{DA}}^{shuffled},f}) \quad (36)$$

Let $P_{g_{DA},j}^{shuffled} = \{p_{g_{DA},j,1}^{shuffled}, p_{g_{DA},j,2}^{shuffled}, \dots, p_{g_{DA},j,n}^{shuffled}\}$ be the set of *n* DA probabilities for node g_{DA} and shuffle *j*, where *n* is the number of processes in *F*. Then, a given process is found to be differentially abundant at a given node if its unadjusted p-value, $p_{g_{DA},f}$, is less than the minimum of all shuffled probabilities for that node. To assign the direction of process abundance change for nodes with significant abundance changes, given that our test examines relative changes in expression between $g_{DA}$and in its sibling node g'_{DA}, we defined the larger of the two nodes as having a "baseline expression profile" for that shared local transcriptional and spatial context. Conversely, the smaller of the two nodes was defined as having either increased or decreased abundance relative to the larger node.

Contextual dependence of processes based on architecture of SGs

To determine whether differentially abundant processes within NSGs were impacted by the differentially abundant processes of their parent SGs, we computed the odds ratio test for independence as follows. Let $f_i \in F$ and $f_j \in F$ denote two biological processes drawn from the set of all pathways and cell types identified (see *'GSVA'* and *'SpaCET'* sections). Across all TumorSPACE models, we identified the set of NSG-parent SG pairs – denoted by (N_i, P_j) – such that $N_i^+$and N_i^- indicate the subset of NSGs where process i was increased or decreased in abundance, respectively, and $P_i^+$and P_i^- indicate the subset of parent SGs where process j was increased or decreased in abundance, respectively. Then, the odds ratio of independence $OR_{i,j}$ was defined as,

$$OR_{i,j} = \frac{\left({card(N_i^+, P_i^+)}/{card(N_i^-, P_i^+)}\right)}{\left({card(N_i^+, P_i^-)}/{card(N_i^-, P_i^-)}\right)} \quad (37)$$

Standard definitions were used for calculation of odds ratio standard error and p-values[66].

SG-based spatial lability (SLAB) score

Given a single TumorSPACE model T_{filt}^{opt} and a process f_i for which differential abundance has been computed in T_{filt}^{opt}, we define the SLAB score as follows. Let G^{f_i} be the set of SGs in T_{filt}^{opt} in which the process f_i is differentially abundant ($q < 0.05$). For each node $g_k^{f_i} \in G^{f_i}$, this means that process f_i is differentially abundant between $g_k^{f_i}$ and its sister node $g_k^{f_i\prime}$ in T_{filt}^{opt}. First, we identify which of the nodes, either $g_k^{f_i}$ or $g_k^{f_i\prime}$, contains the fewer number of spots. This node is defined as the node with either increased or decreased abundance of process f_i, while the node with the greater number of spots is considered to be the 'baseline' abundance state for process f_i in that subset of the tumor biopsy. $DA\left(T_{filt}^{opt}, f_i, g_k^{f_i}\right)$ describes the set of spots with differential abundance in process f_i for TumorSPACE model T_{filt}^{opt} at node $g_k^{f_i}$:

$$DA\left(T_{filt}^{opt}, f_i, g_k^{f_i}\right) = \{\min\left(\left[C\left(g_k^{f_i}\right), C\left(g_k^{f_i\prime}\right)\right]\right) \mid g_k^{f_i} \in G^{f_i}\} \quad (38)$$

Next, we compute the union of those differentially abundant spots and compute the fraction that these spots constitute compared to the total set of spots *I* in the biopsy as a whole.

$$SLAB\left(T_{filt}^{opt}, f_i\right) = \frac{card(U\{DA\left(T_{filt}^{opt}, f_i, g_k^{f_i}\right) \mid g_k^{f_i} \in G^{f_i}\})}{card(I)} \quad (39)$$

For **Fig. 3D,E**, SLAB scores were computed using either only NSGs or only non-NSGs as the input set of SGs used for computing G^{f_i}.

SLAB score correlation with bulk expression

For correlation of genome-wide SLAB scores with bulk gene expression, as in **Fig. 2B** and **Extended Data Fig. 3**, we did the following. First, we identified the set of all dataset-gene pairs for which the gene had greater than 0 UMIs detected per spot and a non-zero SLAB score in that dataset. Next, to enable computing correlation statistics, we identified genes with greater than 10 dataset entries in the filtered dataset-gene pair list. For these genes, we computed the Pearson Correlation estimate and p-value between SLAB score and mean spot UMI count across datasets. Correction for multiple hypothesis testing was done using the Benjamini-Hochberg method with a corrected q-value threshold of 0.05[67].

Spatial lability pan-tumor classification

Given the set of SLAB scores that were computed for all available genes for each of the 96 datasets within the pan-tumor ST-seq database (see 'ST-seq dataset download and alignment'), we aligned these score vectors into a matrix M_{SLAB}^{Pan} such that each dataset was a row and each gene was a column. For any instances where a gene had mapped reads in one dataset but not another – thus resulting in blank cells in this matrix – the score within this matrix was set to zero. Next, Euclidean distance was computed between each pair of rows, resulting in a distance matrix of dimensions 96 x 96 that compared all datasets to each other. Finally, the Unweighted Pair Group Method with Arithmetic mean (UPGMA) algorithm was used for constructing a hierarchical tree relating datasets to each other (**Extended Data Fig. 12**)[68].

For defining tumor groups from this tree, we used the path of tree connections containing the highest number of tumors as our reference point. From this 'main path', we labeled any diverging branchpoints with labels A, B, C, … as shown in **Fig. 3B**. Whenever a group was defined (e.g. group 'A'), the remaining 'main path' tumors were defined as *not* in that group (e.g. group 'nA', where 'n' indicates 'not').

Building across-tumor spatial models

To evaluate the ability of spatial organization in one tumor biopsy (training) to predict spatial organization in another tumor biopsy (testing), we applied the TumorSPACE workflow in the following manner. First, an alignment was performed between the training tumor gene count matrix and the testing tumor gene count matrix since the experiments may have used distinct probe sets and thus mapped reads to non-identical sets of genes. Next, the testing tumor spot transcriptomes were projected into the latent space of the training data, after which spectral distances were computed to determine similarity between training tumor spots and testing tumor spots. Finally, the training tissue TumorSPACE model *T* was optimized on the same properties as before (see '*Hyperparameter optimization to create a TumorSPACE map*') by tuning spatial predictions on the training tissue and then evaluating prediction quality on the testing tissue.

For latent space projection between the aligned gene count matrices, we denote the aligned gene count matrices for training and testing tissues as M_{tr} *(*m_1 *x g)* and M_{te} *(*m_2 *x g)* respectively. As follows from (1), we computed SVD on M_{tr} as

$$M_{tr} = U_{tr} * \Sigma_{tr} * V_{tr}^{T} \quad (40)$$

We then projected M_{te} into the latent space U_{tr} as follows. The value of the hyperparameter *p* is the value of *p* that maximizes spatial prediction in the training tumor dataset.

$$U_{te} = M_{te} * (V_{tr}^{T\,(1:p)})^{-1} * (\Sigma_{tr}^{(1:p)})^{-1} \quad (41)$$

where $(V_{tr}^{T\,(1:p)})^{-1}$ is computed using the pseudo-inverse. Vertical concatenation of U_{tr} and U_{te} yields a joint *U* matrix $U_{tr,te}$. From $U_{tr,te}$ and Σ_{tr}, we compute spectral distance between all spots in M_{tr} and M_{te} as $D_{tr,te}$.

$$D_{tr,te} = spectraldistances(U_{tr,te}^{(1:p)}, \Sigma_{tr}^{(1:p)}, getintervals\left(U_{tr,te}^{(1:p)}\right)) \quad (42)$$

Finally, we filtered matrix $D_{tr,te}$ for the matrix of spectral distances $D'_{tr,te}$ of shape m_2 x m_1 that contains pairwise distances of spots only between M_{tr} and M_{te}. This operation removes intra-group spectral distance comparisons for both M_{tr} and M_{te} and keeps only inter-group spectral distance comparisons between pairs of spots in M_{tr} and M_{te}.

$$D'_{tr,te} = D_{tr,te}^{((m_1+1):(m_1+m_2),1:m_1)} \quad (43)$$

Differential SLAB score analysis

Using the tumor groups as defined in '*Spatial lability pan-tumor classification*', we compared each group *K* to the set *nK* of 'main path' tumors divergent from that group.

To compare gene-level SLAB scores, we first compose the matrix L^k of SLAB scores where L^k has $k + k_n$ rows corresponding to tumors $k \in K$ and $k_n \in nK$ and *F* columns where $f \in F$ constitutes the full set of genes. Next, for each gene *f*, we compare the tumors in *K* and *nK* where *C(X)* indicates the row indices within matrix L^k that correspond to tumors in either group. Comparison is performed using a Mann-Whitney U Test, where the test p-value is given by *MW(a,b)*.

$$p_{K,f} = MW(L^k_{C(K),f}, L^k_{C(nK),f}) \quad (44)$$

To facilitate empirical correction for multiple hypothesis testing, we perform 1000 shuffles of the SLAB counts between *K* and *nK*, followed by computation of the MW p-value between these shuffles. Let $C_{K_{total}}$ be the concatenation of row indices $C(K)$ and $C(nK)$.

$$C_{K_{total}} = C(K) + C(nK) \quad (45)$$

$$C_{K_{total}}^{shuffled} = \{J \subseteq C_{K_{total}} \mid card(J) = card(C_{K_{total}})\} = C_K^{shuffled} + C_{nK}^{shuffled} \quad (46)$$

$$C_K^{shuffled} = C_{K_{total}}^{shuffled}[1: card(C(K))] \quad (47)$$

$$C_{nK}^{shuffled} = C_{K_{total}}^{shuffled}[(card(C(K)) + 1): (card(C(K)) + card(C(nK)))] \quad (48)$$

$$p_{K,f}^{shuffled} = MW(L^k_{C_K^{shuffled},f}, L^k_{C_{nK}^{shuffled},f}) \quad (49)$$

Let $P_{K,j}^{shuffled} = \{p_{K,j,1}^{shuffled}, p_{K,j,2}^{shuffled}, \ldots, p_{K,j,n}^{shuffled}\}$ be the set of *n* probabilities for group *K* and shuffle *j*, where *n* is the number of genes in *F*. Then, a given gene is found to have a differential SLAB score between a given grouping *K* vs *nK* if its unadjusted p-value, $p_{K,f}$, is less than the 5th percentile (*q = 0.05*) of all shuffled probabilities.

To compare pathway-level SLAB scores, we used either over-representation analysis (ORA) or gene-set enrichment analysis (GSEA). For both analyses, we computed enrichment for the set of Reactome pathways within the MSigDB database[69,70] (**Supplementary Table 5**). ORA was performed using Enrichr with default parameters, which uses a Fisher exact test to compute enrichment of a gene list for a given pathway[71]. The background gene set used was the set of all genes with mapped reads in any sample. GSEA was performed using 20,000 permutations with the "signal-to-noise" ratio used for ranking[69]. For both ORA and GSEA, correction for multiple hypothesis testing was implemented by using a false discovery rate threshold of < 0.1.

Classification of NSCLC datasets by pan-tumor immune spatial lability

For comparison of out-of-sample NSCLC tumors to pan-tumor spatial lability groups shown in **Fig. 3B**, we first computed SLAB scores for all genes and aligned the score vectors to match the columns (genes) of the pan-tumor SLAB score matrix M_{SLAB}^{Pan} . Any genes with no detectable reads for a given sample had their SLAB score set to zero. We called this new matrix M_{SLAB}^{NSCLC}. For every pair of rows $(r_{p_i}^{Pan}, r_{n_j}^{NSCLC})$, where $r_{p_i}^{Pan}$ indicates the score vector for sample $p_i \in P$ in the pan-tumor database and $r_{n_j}^{NSCLC}$ indicates the score vector for sample $n_j \in N$ in the NSCLC out-of-sample dataset, we computed the Euclidean distance D_{p_i,n_j}^{SLAB} that describes the similarity between these two samples with respect to their SLAB scores. For a given NSCLC sample n_j, we identified the pan-tumor dataset p_i with the lowest Euclidean distance to n_j and assigned n_j to the same spatial lability class as p_i.

For defining immune spatial lability, since tumor groups 'C' and 'E' both demonstrated enrichment in SLAB score for immune biology components, we defined 'nE' tumors as immune spatially invariant (ISI) and tumors in groups A, B, C, D, or E as immune spatially labile (ISL).

Classification of NSCLC datasets using bulk expression and published gene sets

To determine whether classification of tumor datasets by either (1) bulk expression versus SLAB score or (2) previously published gene sets for NSCLC IO response was predictive of PFS in our NSCLC cohort, we performed the following analysis.

First, we computed aligned matrices as described in '*Spatial lability pan-tumor classification'* for both the pan-tumor datasets and the NSCLC datasets where matrices contained either bulk expression data or SLAB score data. For bulk expression, we computed the mean spot-wise UMI count for any given gene. Second, we filtered the aligned matrices for subset of columns (genes) described by a particular gene set or used all columns for the 'all genes' analysis. Third, we computed a hierarchical tree using the pan-tumor data and identified the best matches to datasets within that tree for all NSCLC datasets as described in '*Spatial lability pan-tumor classification'.* Fourth, K-means clustering with K = 2 was applied to the Euclidean distances between all pairs of 'best match' pan-tumor datasets. K-means clustering was performed 100 times for each condition using different random seeds each time. Finally, the two classes that were defined were used to classify the NSCLC cohort based on the matching performed between the NSCLC tumors and the pan-tumor datasets in the third step described above.

These classes were subsequently applied to survival analysis (described below in '*Survival analysis')* to determine if they were predictive of NSCLC ICB outcomes.

Survival analysis

For survival analysis we used the R 'survival' package to model progression-free survival (PFS) as a function of possible confounder variables (Treatment regimen, KRAS mutation status) or classification variables (PD-L1 multi-class, PD-L1 binary, ISL/ISI, bulk expression- and SLAB score- gene sets). For confounder analysis, outcomes were modeled using Cox's univariate proportional hazards model. For Kaplan-Meier survival curves stratified by classification variables, survival was estimated using the Kaplan-Meier method and reported p-values were calculated using the log rank statistical test. For censored data labeling, 1 indicates that PFS was observed while 0 indicates the patient was censored for PFS.

GSVA

Gene set variation analysis (GSVA) estimates GSVA pathway enrichment scores for a given set of pathways from gene expression data[72]. It requires (1) the spot-by-gene count matrix from SpaceRanger and (2) a list of pathway gene sets. We used a subset of the KEGG pathways from the MSigDB database[69,73] (**Supplementary Table 6**). The 'KCDF' parameter was set to "none", which enforces a direct estimation of cumulative density function without assuming a kernel function. Otherwise, default parameters were used.

SpaCET

SpaCET estimates deconvoluted cell type proportions within spots of an ST-seq experiment[39]. It requires the user to supply (1) the SpaceRanger gene count matrix as input and (2) a value for the 'cancerType' parameter to define the SpaCET library scRNA-seq datasets used for cell type definition. The 'cancerType' values chosen for each ST-seq dataset are listed in **Supplementary Table 7**. Otherwise, default parameters and commands were used as per the repository instructions (https://data2intelligence.github.io/SpaCET/articles/visium_BC.html).

CODEX multiplexed immunofluorescence - analysis

Cell segmentation

Following image acquisition and pre-processing (see '*Experimental method details: CODEX multiplexed immunofluorescence'*), we applied the neural network-based cell segmentation tool, DeepCell, on the DAPI channel for nuclei identification[74]. Next, these nuclei segmentation masks were used to estimate whole cell segmentation boundaries using the 'skimage.morphology.binary_dilation' function in the Python scikit-image package[75]. This function dilates nuclear segmentation boundaries by stochastically flipping pixels into the mask boundary with a probability equal to the fraction of positive neighboring pixels for 9 cycles. We then computed mean expression for each antibody across pixels within each whole cell segmentation boundary, which we define as the signal intensity $Signal_i^t$ for cell *i and* target *t*.

Cell-level quality control

Since there is technical variation in CODEX staining and imaging quality, we applied multiple quality control filters to eliminate cells with atypical quality characteristics. First, we defined for cell *i* the signal sum Σ_i, mean μ_i, standard deviation σ_i, and coefficient of variation CoV_i across the set of targets *T*, composed of DAPI + all antibodies in **Supplementary Table 8**.

$$\Sigma_i = \sum_t Signal_i^t \; for \; t \in T \tag{50}$$

$$\mu_i = \frac{\Sigma_i}{card(T)} \quad (51)$$

$$\sigma_i = \sqrt{\frac{\sum_t (Signal_i^t - \mu_i)^2}{card(T)}} \quad (52)$$

$$\mathrm{CoV}_i = \frac{\sigma_i}{\mu_i} \quad (53)$$

We then filtered cells for analysis only when Σ_i, CoV_i, and $Signal_i^{DAPI}$ fall within the 5 – 95% of values for cells within that particular sample (**Extended Data Fig. 13A**). Let *I* indicate the set of cells within a single sample. Then,

$$cells_{filt} \in cells_I \mid \begin{cases} quantile(\Sigma_I, 0.05) \leq \Sigma_f \leq quantile(\Sigma_I, 0.95) \\ quantile(\mathrm{CoV}_I, 0.05) \leq \mathrm{CoV}_f \leq quantile(\mathrm{CoV}_I, 0.95) \\ quantile(Signal_I^D, 0.05) \leq Signal_f^D \leq quantile(S\square\square gnal_I^D, 0.95) \end{cases}$$

where Σ_f, CoV_f, and $Signal_f^D$ denote the signal sum, signal CoV, and DAPI intensity signal for a given cell *f* in $cells_{filt}$. We found that excluded cells tended to be found along tissue borders (**Extended Data Fig. 13B**).

Signal normalization

We normalized signal intensities for (1) variation in local background and (2) variation in signal distribution between samples.

To correct for variation in local background, we divided each sample into 100 equally sized bins and used multi-Gaussian modeling for each target $t \in T$ to identify the upper limits of that marker's local null distribution. Let *i* and *j* represent the bin numbers in the *x* and *y* directions respectively. Then we denote $cells_{i,j}$ as the set of cells in a given sample bin *(i,j)* and $Signal_{i,j}^t$ as the set of signal intensities for marker *t* for $cells_{i,j}$. We used the 'mclust' R package to fit 2 Gaussians to $Signal_{i,j}^t$ for all values of *i*, *j*, and *t*. Then we defined the upper bound $Background_{i,j}^t$ of the null distribution as the 95% percentile of that distribution for a given bin and marker. We found wide variation in the distributions of $Background_{i,j}^t$ for different targets *t* and for different samples, underscoring the need to use target-specific background correction (**Extended Data Fig. 14A, B**). Finally, we subtracted $Background_{i,j}^t$ from $Signal_{i,j}^t$ as a correction for local background signal variation.

$$i, j \in_{\mathbb{Z}} \{1, 2, 3, \ldots, 10\} \quad (55)$$

$$Signal_{i,j}^t \sim N(\mu_1, \sigma_1^2) + N(\mu_2, \sigma_2^2) \text{ where } \mu_1 < \mu_2 \quad (56)$$

$$Background_{i,j}^t = \mu_1 + 1.645 * \frac{\sigma_1}{\sqrt{card(cells_{i,j})}} \quad (57)$$

$$Norm_{i,j}^t = Signal_{i,j}^t - Background_{i,j}^t \quad (58)$$

To minimize variation in signal quantitation between samples, we then scaled the intensity distribution for each target to match across both DLBCL samples, using 'DLBCL Patient 1' as a reference for scaling. Let $Norm_i^t$ and $Scale_i^t$ indicate the distribution of normalized and scaled

intensities, respectively, from (*54*) for target *t* and DLBCL patient *i*. Let *mean* and *sd* indicate the mean and standard deviations of these intensity distributions.

$$Scale_i^t = \frac{Norm_i^t - mean(Norm_i^t)}{sd(Norm_i^t)} * sd(Norm_1^t) + mean(Norm_1^t) \quad (59)$$

Cell type classification

Cell types were identified from CODEX data using the following thresholds on the scaled intensities $Scale_i^t$. Let *cells*$_{all}$ indicate the set of all cells across both DLBCL samples. Let $cells_c \in cells_{all}$ indicate the subset of these cells classified as class *c* for all cell types $c \in C$. The limits for each cell-type defining marker in the following definitions were identified and tested iteratively to minimize the fraction of 'Unidentifiable' cells while maintaining specific classifications for each cell type (**Extended Data Fig. 14C**).

$$cells_{CAF} \in cells_{all} \mid \begin{cases} Scale_i^{aSMA} > 0.25 \; or \\ Scale_i^{Vimentin} > 10 \end{cases} for\ i \in cells_{CAF} \quad (60)$$

$$cells_{endothelial} \in cells_{all} \mid \begin{cases} Scale_i^{CD31} > 2.5 \; or \\ Scale_i^{CD141} > 2 \end{cases} for\ i \in cells_{endothelial} \quad (61)$$

$$cells_{DLBCL} \in cells_{all} \mid \begin{cases} Scale_i^{CD38} > 4 \; or \\ Scale_i^{CD20} > 10 \; or \\ Scale_i^{CD21} > 20 \; or \\ Scale_i^{CD79a} > 0.25 \end{cases} for\ i \in cells_{DLBCL} \quad (62)$$

$$cells_{CD4\,T} \in cells_{all} \mid \begin{cases} \begin{cases} Scale_i^{CD3e} > 50 \; and \\ Scale_i^{CD4} > 15 \end{cases} for\ i \in cells_{CD4\,T} \end{cases} \quad (63)$$

$$cells_{CD8\,T} \in cells_{all} \mid \begin{cases} \begin{cases} Scale_i^{CD3e} > 50 \; and \\ Scale_i^{CD4} < 15 \; and \; for\ i \in cells_{CD8\,T} \\ Scale_i^{CD8} > 5 \end{cases} \end{cases} \quad (64)$$

$$cells_{cDC} \in cells_{all} \mid \begin{cases} \begin{cases} Scale_i^{CD11c} > 15 \; or \\ Scale_i^{CD141} > 2 \end{cases} for\ i \in cells_{cDC} \end{cases} \quad (65)$$

$$cells_{macrophage} \in cells_{all} \mid \begin{cases} \begin{cases} Scale_i^{CD11c} < 15 \; and \\ Scale_i^{CD68} > 50 \end{cases} for\ i \in cells_{macrophage} \end{cases} \quad (66)$$

$$cells_{neutrophil} \in cells_{all} \mid \begin{cases} \begin{cases} Scale_i^{MPO} > 5 \; or \\ Scale_i^{CD66} > 0.25 \end{cases} for\ i \in cells_{neutrophil} \end{cases} \quad (67)$$

For cells that were classified into multiple classes by these criteria, we labeled cells as "DLBCL" if one of their multiple class labels was "DLBCL" and otherwise labeled them as "Unidentifiable".

As an orthogonal validation of SpaCET-based cell type deconvolution, we observed high concordance (R = 0.99) in cell type classification between transcriptional inference and IF-based classification (**Extended Data Fig. 15, Supplementary Table 9**).

CoV analysis of cell type abundance

For estimating variation in cell type abundance at a variety of distance scales, we computed the coefficient of variation (CoV) in cell type abundance across a grid of regions in each tumor biopsy. First, we divided each tumor region into bins of size ranging from 0.3 x 0.3 mm to 10 x 10 mm. Let $Bins_I^k$ indicate the set of *I* bins for a given width *k* for a tumor sample. We then selected the subset $Bins_I^{k\prime}$ in which every bin $bin_i^{k\prime} \in Bins_I^{k\prime}$ that has at least 100 annotated cells. We computed the fractional abundance of each cell type in $bin_i^{k\prime}$ as $frac_{i,c}^{k\prime}$ where $c \in C$ is the set of all annotated cell types. Finally, we computed CoV_c^k to be coefficient of variation for a given cell type *c* at length scale *k*.

$$\Sigma_c^k = \sum_i frac_{i,c}^{k\prime} \; for \; i \in I \tag{68}$$

$$\mu_c^k = \frac{\Sigma_c^k}{card(I)} \tag{69}$$

$$\sigma_c^k = \sqrt{\frac{\sum_i (frac_{i,c}^{k\prime} - \mu_c^k)^2}{card(I)}} \tag{70}$$

$$CoV_c^k = \frac{\sigma_c^k}{\mu_c^k} \tag{71}$$

Experimental method details

10X Visium CytAssist spatial transcriptomics (ST-seq)

Tissue quality was determined by isolation of RNA from FFPE using the Qiagen RNeasy FFPE kit. Samples were then analyzed for tissue extraction quality using the Agilent 2100 bio-analyzer and Agilent RNA-6000 pico kit. For each sample, a DV200 score – the fraction of RNA fragments > 200 nucleotides in length – was calculated. Tissue quality for all samples was tested on unstained sections adjacent to the section used for ST-seq.

DLBCL samples were previously H&E stained. Imaging and coverslip removal were completed as described by 10X Protocol CG000518-Rev A and decrosslinking was performed according to 10X Protocol CG000520-Rev A[76,77]. NSCLC samples underwent deparaffinization, H&E staining, imaging, and decrosslinking according to CG000520-Rev B[78]. Sample imaging for all samples was performed using the Akoya Biosciences Vectra Polaris at 20X magnification.

We next performed the following steps as per either 10X Protocol CG000495-Rev A for the DLBCL samples or 10X Protocol CG000495-Rev E for the NSCLC samples[79,80]. First, samples underwent probe hybridization with Visium Human Transcriptome Probe Set v2.0, followed by probe ligation, and associated washes (**Supplementary Table 10**). Two native tissue slides and one Visium CytAssist 11 x 11 mm slide were then placed within the Visium CytAssist to enable RNA digestion, tissue removal, and transfer of ligated products onto the two fiducial frames of the Visium Slide. Next, we performed probe extension and elution off the Visium Slide, followed by pre-amplification and SPRIselect cleanup. For SPRIselect cleanup, DLBCL samples placed in only the 'High' position of the 10X magnetic separator, while NSCLC samples were placed in

both 'High' and 'Low' positions according to CG000495-Rev E. To identify the optimal number of cycles for library amplification, we performed qPCR using Applied Biosciences QuantStudio 6 Pro as per CG000495-Rev E (**Supplementary Table 4**). For this step, we included 0.5 μl of carboxy-X-rhodamine (ROX) with the DLBCL samples and not with the NSCLC samples. Sample Index PCR was run using the sample-specific optimal number of cycles, followed by: cleanup, Agilent TapeStation QC, sequencing, and demultiplexing using Bcl2fastq. Sample sequencing was performed on a NovaSeq 6000 for DLBCL samples and a NovaSeqX for NSCLC samples. Sample-specific parameters and QC are listed in **Supplementary Table 4**. For DLBCL experiments, we used the Applied Biosystems Veriti 96 well thermocycler, while for NSCLC samples we used the Eppendorf Mastercycler X50a and X50l.

CODEX multiplexed immunofluorescence

Slide preparation

DLBCL samples were previously obtained as unstained slides mounted with 5 μm thickness formaldehyde-fixed, paraffin-embedded (FFPE) sections from the same patient biopsies as described in *'Patient samples'*. Coverslips were coated with 0.1% poly-l-lysine solution prior to mounting tissue sections to enhance adherence. The prepared coverslips were washed and stored according to guidelines from the CODEX user manual.

Antibody preparation

Custom conjugated antibodies were conjugated using the CODEX conjugation kit as per the CODEX user manual (**Supplementary Table 8**). Briefly, the antibody is (1) partially reduced to expose thiol ends of the antibody heavy chains, (2) conjugated with a CODEX barcode, (3) purified, and (4) added to Antibody Storage Solution for long-term stabilization. Subsequently, antibody conjugation is verified using sodium dodecyl sulfate-polyacrylamide gel electrophoresis and with QC staining.

Staining and data acquisition

Sample slides are stained following protocols in the CODEX User Manual. Briefly, samples are pretreated by heating at 60°C overnight, followed by deparaffinization, rehydration using ethanol washes, and antigen retrieval via immersion in Tris-EDTA pH 9.0 for 20 minutes. Samples are then blocked in staining buffer and incubated with the antibody cocktail for 3 hours at room temperature. After incubation, samples are washed and fixed following the CODEX User Manual. Data acquisition was performed using the PhenoCycler-Fusion 2.0 with a 20X objective, resulting in a resolution of 0.5 μm/pixel.

Patient tumor PD-L1 IHC

FFPE biopsy samples were probed for PD-L1 expression using a qualitative immunohistochemical assay with the Dako 22C3 antibody (Pharm Dx kit). PD-L1 expression was classified using the Tumor Proportion Score (TPS), which represents the percentage of viable tumor cells that show partial or complete membrane staining. Normal background histiocytes served as internal controls to ensure quality of the PD-L1 staining. Quantification was performed by a board-certified pathologist as part of routine clinical care.

Patient somatic mutation testing

The molecular profiles of the tumor biopsies were analyzed using Oncoplus or Oncoscreen, two Next Generation Sequencing (NGS) assays[81]. A description of patient mutation status can be found in **Supplementary Table 11**. Since the list of targeted genomic regions varied by the year in which testing was performed, a list of Oncoplus/Oncoscreen versions used for each patient

as well as a list of the targeted genomic regions for each version can be found in **Supplementary Tables 11 and 12**, respectively.

For the Oncoplus analysis, DNA was isolated from the samples using the QIAamp DNA Blood Mini Kit (Qiagen), fragmented, and prepared into a sequencing library with patient-specific indexes (HTP Library Preparation Kit, Kapa Biosystems). Targeted genomic regions were enriched using a panel of biotinylated oligonucleotides (SeqCap EZ, Roche Nimblegen) supplemented with additional oligonucleotides (xGen Lockdown Probes, IDT). The enriched libraries were then sequenced on an Illumina HiSeq 2500 system, and the data was analyzed via bioinformatics pipelines against the hg19 (GRCh37) human genome reference sequence.

For Oncoscreen, DNA was isolated from formalin-fixed paraffin-embedded (FFPE) tumor tissue using the QIAamp DNA FFPE Tissue Kit (Qiagen). DNA was quantified using the Qubit fluorometric assay (Thermo Fisher Scientific) and a quantitative PCR assay (hgDNA Quantitation and QC kit, KAPA Biosystems). Targeted genomic regions were amplified using multiplex PCR (Thermo Fisher Scientific); PCR products were used to prepare NGS libraries with patient-specific adapter index sequences (HTP Library Preparation Kit, KAPA Biosystems). The enriched libraries were then sequenced on an Illumina MiSeq system, and the data was analyzed via bioinformatics pipelines against the hg19 (GRCh37) human genome reference sequence.

Patient tumor volume measurements

For measurement of tumor volume changes over time, computed tomography (CT) imaging reports were obtained for patients in the NSCLC as permitted by the IRBs referenced in 'Patient samples'. For patients with measurable disease at the time of treatment start (denoted as month zero), the largest lesion was identified and labeled the 'index lesion'. Changes in index lesions were collected when described in serial reports by a board-certified radiologist as part of routine clinical care.

Subject details

Patient samples

Non-small cell lung cancer (NSCLC) patients were treated with immune checkpoint blockade therapy +/- chemotherapy at the University of Chicago Medical Center (Chicago, IL). All patients provided written informed consent for the collection and study of pre-treatment diagnostic tumor biopsy samples and for clinical outcomes including treatment regimen, treatment-related toxicities, and disease outcomes, as approved by the University of Chicago Institutional Review Board (IRB 9571 and IRB 24-0063). For the ST-seq analysis, 16 tumor samples were collected prior to therapy initiation, each from a separate patient. Inclusion criteria for these patients included (1) NSCLC stage IV patients either at initial presentation or as progression from previously treated early-stage disease, (2) biopsy of either the primary tumor or a metastatic tumor performed and stored within 6 months prior to treatment in the metastatic setting, (3) subsequent first line treatment with anti-PD1/anti-PD-L1 immune checkpoint blockade (ICB) with or without platinum-based chemotherapy. Exclusion criteria included (1) no prior therapy in the metastatic setting and (2) less than 2 doses of ICB therapy administered. We selected the first 16 patients that met these criteria and that had an available FFPE tumor biopsy block. From the archival block, a fresh 5 μm section was cut and placed on a standard slide for use in ST-seq protocols (see '*10X Visium spatial transcriptomics (ST-seq)*'). Progression was defined as time from the first dose of ICB until either radiographic or symptom-based evidence of disease progression. ICB regimen, ICB treatment duration, reason for ICB discontinuation, time to

progression following ICB start, and time to death following ICB start are listed for all patients (**Supplementary Table 11**).

Diffuse large B-cell lymphoma (DLBCL) patients were treated at the University of Chicago Medical Center (Chicago, IL). All patients provided written informed consent for the collection and study of pre-treatment diagnostic tumor biopsy samples and for clinical outcomes including treatment regimen, treatment-related toxicities, and disease outcomes, as approved by the University of Chicago Institutional Review Board (IRB 13-1297). Each biopsy was reviewed by 2 hematopathologists for diagnostic confirmation. Biopsy slides were previously cut from FFPE sections and H&E stained for prior studies[82].

Supplementary Tables

Supplementary Table 1. Pan-tumor database properties

Supplementary Table 2. Differential gene and pathway spatial lability amongst pan-tumor SLAB classes

Supplementary Table 3. Bulk gene sets for prediction of NSCLC ICB response

Supplementary Table 4. ST-seq QC Metrics

Supplementary Table 5. Reactome pathway list

Supplementary Table 6. KEGG pathway list

Supplementary Table 7. SpaCET input cancer types

Supplementary Table 8. CODEX Antibody Targets

Supplementary Table 9. ST versus IF Cell Type Classification

Supplementary Table 10. Visium Human Transcriptome Probe Set v2.0 - Probe Set Reference CSV file

Supplementary Table 11. NSCLC Clinical Metadata (oncoplus version)

Supplementary Table 12. Oncoplus/Oncoscreen Gene Panels

References

1. Arneth, B. Tumor microenvironment. *Med.* **56**, (2020).
2. Anderson, N. M. & Simon, M. C. The tumor microenvironment. *Curr. Biol.* **30**, R921–R925 (2020).
3. Bressan, D., Battistoni, G. & Hannon, G. J. The dawn of spatial omics. *Science (New York, N.Y.)* vol. 381 eabq4964 (2023).
4. Burgos-Panadero, R. *et al.* The tumour microenvironment as an integrated framework to understand cancer biology. *Cancer Lett.* **461**, 112–122 (2019).
5. Giraldo, N. A. *et al.* The clinical role of the TME in solid cancer. *Br. J. Cancer* **120**, 45–53 (2019).
6. Azimi, F. *et al.* Tumor-infiltrating lymphocyte grade is an independent predictor of sentinel lymph node status and survival in patients with cutaneous melanoma. *J. Clin. Oncol.* **30**, 2678–2683 (2012).
7. Thomas, N. E. *et al.* Tumor-infiltrating lymphocyte grade in primary melanomas is independently associated with melanoma-specific survival in the population-based genes, environment and melanoma study. *J. Clin. Oncol.* **31**, 4252–4259 (2013).
8. Pagès, F. *et al.* International validation of the consensus Immunoscore for the classification of colon cancer: a prognostic and accuracy study. *Lancet* **391**, 2128–2139 (2018).
9. Marliot, F. *et al.* Analytical validation of the Immunoscore and its associated prognostic value in patients with colon cancer. *J. Immunother. Cancer* **8**, 13–15 (2020).
10. Bruni, D., Angell, H. K. & Galon, J. The immune contexture and Immunoscore in cancer prognosis and therapeutic efficacy. *Nat. Rev. Cancer* **20**, 662–680 (2020).
11. Park, Y. M. & Lin, D. C. Moving closer towards a comprehensive view of tumor biology and microarchitecture using spatial transcriptomics. *Nat. Commun.* **14**, 14–16 (2023).
12. Ali, H. R. *et al.* Imaging mass cytometry and multiplatform genomics define the

phenogenomic landscape of breast cancer. *Nat. cancer* **1**, 163–175 (2020).

13. Danenberg, E. *et al.* Breast tumor microenvironment structures are associated with genomic features and clinical outcome. *Nat. Genet.* **54**, 660–669 (2022).

14. Schürch, C. M. *et al.* Coordinated Cellular Neighborhoods Orchestrate Antitumoral Immunity at the Colorectal Cancer Invasive Front. *Cell* **182**, 1341-1359.e19 (2020).

15. Gaglia, G. *et al.* Temporal and spatial topography of cell proliferation in cancer. *Nat. Cell Biol.* **24**, 316–326 (2022).

16. Nirmal, A. J. *et al.* The Spatial Landscape of Progression and Immunoediting in Primary Melanoma at Single-Cell Resolution. *Cancer Discov.* **12**, 1518–1541 (2022).

17. Greenwald, A. C. *et al.* Integrative spatial analysis reveals a multi-layered organization of glioblastoma. *Cell* **187**, 2485-2501.e26 (2024).

18. Du, J. *et al.* Advances in spatial transcriptomics and related data analysis strategies. *J. Transl. Med.* **21**, 1–21 (2023).

19. Halabi, N., Rivoire, O., Leibler, S. & Ranganathan, R. Protein Sectors: Evolutionary Units of Three-Dimensional Structure. *Cell* **138**, 774–786 (2009).

20. Russ, W. P., Lowery, D. M., Mishra, P., Yaffe, M. B. & Ranganathan, R. Natural-like function in artificial WW domains. *Nature* **437**, 579–583 (2005).

21. Russ, W. P. *et al.* An evolution-based model for designing chorismate mutase enzymes. *Science (80-.).* **369**, 440–445 (2020).

22. McLaughlin, R. N., Poelwijk, F. J., Raman, A., Gosal, W. S. & Ranganathan, R. The spatial architecture of protein function and adaptation. *Nature* **491**, 138–142 (2012).

23. Eisen, J. A. Phylogenomics: Improving functional predictions for uncharacterized genes by evolutionary analysis. *Genome Res.* **8**, 163–167 (1998).

24. Pellegrini, M., Marcotte, E. M., Thompson, M. J., Eisenberg, D. & Yeates, T. O. Assigning protein functions by comparative genome analysis: Protein phylogenetic profiles. *Proc. Natl. Acad. Sci. U. S. A.* **96**, 4285–4288 (1999).

25. Cong, Q., Anishchenko, I., Ovchinnikov, S. & Baker, D. Protein interaction networks revealed by proteome coevolution. *Science (80-.).* **365**, 185–189 (2019).

26. Zaydman, M. A. *et al.* Defining hierarchical protein interaction networks from spectral analysis of bacterial proteomes. *Elife* **11**, (2022).

27. Jia, J. *et al.* Conserved Covarying Gut Microbial Network in Preterm Infants and Childhood Growth During the First 5 Years of Life: A Prospective Cohort Study. *Am. J. Clin. Nutr.* **118**, 561–571 (2023).

28. Raman, A. S. *et al.* A sparse covarying unit that describes healthy and impaired human gut microbiota development. *Science (80-.).* **365**, (2019).

29. Geesink, P., Horst, J. ter & Ettema, T. J. G. More than the sum of its parts: uncovering emerging effects of microbial interactions in complex communities. *FEMS Microbiol. Ecol.* **100**, 1–7 (2024).

30. 10x Genomics. Visium Spatial Gene Expression for FFPE Reagent Kits Reagent Kits, Document Number CG000047 Rev E. 1–47 (2024).

31. Walker, B. L. NeST : nested hierarchical structure identification in spatial transcriptomic data. *Nat. Commun.* 1–17 (2023) doi:10.1038/s41467-023-42343-x.

32. Lin, X., Gao, L., Whitener, N., Ahmed, A. & Wei, Z. A model-based constrained deep learning clustering approach for spatially resolved single-cell data. *Genome Res.* **32**, 1906–1917 (2022).

33. Dong, K. & Zhang, S. Deciphering spatial domains from spatially resolved transcriptomics with an adaptive graph attention auto-encoder. *Nat. Commun.* **13**, 1–12 (2022).

34. Long, Y. *et al.* Spatially informed clustering, integration, and deconvolution of spatial transcriptomics with GraphST. *Nat. Commun.* **14**, 1–19 (2023).

35. Hu, J. *et al.* SpaGCN: Integrating gene expression, spatial location and histology to identify spatial domains and spatially variable genes by graph convolutional network. *Nat. Methods* **18**, 1342–1351 (2021).

36. Tian, T., Zhang, J., Lin, X., Wei, Z. & Hakonarson, H. Dependency-aware deep generative models for multitasking analysis of spatial omics data. *Nat. Methods* (2024) doi:10.1038/s41592-024-02257-y.

37. Haviv, D., Gatie, M., Hadjantonakis, A.-K., Nawy, T. & Pe'er, D. The covariance environment defines cellular niches for spatial inference. *Nat. Biotechnol.* (2024) doi:10.1038/s41587-024-02193-4.

38. Doran, B. A. *et al.* An evolution-based framework for describing human gut bacteria. *bioRxiv* 2023.12.04.569969 (2023) doi:10.1101/2023.12.04.569969.

39. Ru, B., Huang, J., Zhang, Y., Aldape, K. & Jiang, P. Estimation of cell lineages in tumors from spatial transcriptomics data. *Nat. Commun.* **14**, (2023).

40. Zhu, H., Yu, X., Zhang, S. & Shu, K. Targeting the Complement Pathway in Malignant Glioma Microenvironments. *Front. Cell Dev. Biol.* **9**, 1–16 (2021).

41. Gruenbacher, G. *et al.* The human G protein-coupled ATP receptor P2Y11 is associated with IL-10 driven macrophage differentiation. *Front. Immunol.* **10**, 1–13 (2019).

42. Batlle, E. & Massagué, J. Transforming Growth Factor-β Signaling in Immunity and Cancer. *Immunity* **50**, 924–940 (2019).

43. Sharif, H. *et al.* Dipeptidyl peptidase 9 sets a threshold for CARD8 inflammasome formation by sequestering its active C-terminal fragment. *Immunity* **54**, 1392-1404.e10 (2021).

44. Geiss-Friedlander, R. *et al.* The cytoplasmic peptidase DPP9 is rate-limiting for degradation of proline-containing peptides. *J. Biol. Chem.* **284**, 27211–27219 (2009).

45. Huang, N. *et al.* TRIM family contribute to tumorigenesis, cancer development, and drug resistance. *Exp. Hematol. Oncol.* **11**, 1–22 (2022).

46. Xu, X. & Jin, T. ELMO proteins transduce G protein-coupled receptor signal to control reorganization of actin cytoskeleton in chemotaxis of eukaryotic cells. *Small GTPases* **10**, 271–279 (2019).

47. Ding, Z. C. *et al.* Persistent STAT5 activation reprograms the epigenetic landscape in CD4+ T cells to drive polyfunctionality and antitumor immunity. *Sci. Immunol.* **5**, 1–18 (2020).

48. Zhu, L. *et al.* Dapl1 controls NFATc2 activation to regulate CD8+ T cell exhaustion and responses in chronic infection and cancer. *Nat. Cell Biol.* **24**, 1165–1176 (2022).

49. Richard, A. C., Frazer, G. L., Ma, C. Y. & Griffiths, G. M. Staggered starts in the race to T cell activation. *Trends Immunol.* **42**, 994–1008 (2021).

50. Voros, O., Panyi, G. & Hajdu, P. Immune synapse residency of orai1 alters ca2+ response of t cells. *Int. J. Mol. Sci.* **22**, (2021).

51. Yim, K. H. W., Hrout, A. Al, Borgoni, S. & Chahwan, R. Extracellular vesicles orchestrate immune and tumor interaction networks. *Cancers (Basel).* **12**, 1–23 (2020).

52. Chen, R. & Chen, L. Solute carrier transporters: emerging central players in tumour immunotherapy. *Trends Cell Biol.* **32**, 186–201 (2022).

53. Pan, Y. *et al.* RNA Dysregulation: An Expanding Source of Cancer Immunotherapy Targets. *Trends Pharmacol. Sci.* **42**, 268–282 (2021).

54. Srivastava, A. K., Guadagnin, G., Cappello, P. & Novelli, F. Post-Translational Modifications in Tumor-Associated Antigens as a Platform for Novel Immuno-Oncology Therapies. *Cancers (Basel).* **15**, (2023).

55. Martin, A. L. *et al.* Olfactory Receptor OR2H1 Is an Effective Target for CAR T Cells in Human Epithelial Tumors. *Mol. Cancer Ther.* **21**, 1184–1194 (2022).

56. Garassino, M. C. *et al.* Pembrolizumab Plus Pemetrexed and Platinum in Nonsquamous Non-Small-Cell Lung Cancer: 5-Year Outcomes from the Phase 3 KEYNOTE-189 Study. *J. Clin. Oncol.* **41**, 1992–1998 (2023).

57. Hallqvist, A., Rohlin, A. & Raghavan, S. Immune checkpoint blockade and biomarkers of clinical response in non–small cell lung cancer. *Scand. J. Immunol.* **92**, 1–10 (2020).

58. Mino-Kenudson, M. *et al.* Predictive Biomarkers for Immunotherapy in Lung Cancer:

Perspective From the International Association for the Study of Lung Cancer Pathology Committee. *J. Thorac. Oncol.* **17**, 1335–1354 (2022).

59. Li, X. *et al.* The Notch signaling pathway: a potential target for cancer immunotherapy. *J. Hematol. Oncol.* **16**, 1–31 (2023).

60. Hegde, P. S. & Chen, D. S. Top 10 Challenges in Cancer Immunotherapy. *Immunity* **52**, 17–35 (2020).

61. Clifton, G. T. *et al.* Developing a definition of immune exclusion in cancer: results of a modified Delphi workshop. *J. Immunother. Cancer* **11**, e006773 (2023).

62. Klema, V. & Laub, A. The singular value decomposition: Its computation and some applications. *IEEE Trans. Automat. Contr.* **25**, 164–176 (1980).

63. Ehman, E. C. *et al.* Renewing Felsenstein's phylogenetic bootstrap in the era of Big Data. *Nature* **556**, 452–456 (2018).

64. Ripley, B. D. The Second-Order Analysis of Stationary Point Processes. *J. Appl. Probab.* **13**, 255–266 (1976).

65. Ripley, B. D. *Statistical Inference for Spatial Processes*. (Cambridge University Press, 1988). doi:DOI: 10.1017/CBO9780511624131.

66. Morris, J. A. & Gardner, M. J. Calculating confidence intervals for relative risks (odds ratios) and standardised ratios and rates. *Br. Med. J. (Clin. Res. Ed).* **296**, 1313–1316 (1988).

67. Benjamini, Y. & Hochberg, Y. Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. *J. R. Stat. Soc. Ser. B* **57**, 289–300 (1995).

68. Sokal, R. R. & C.D.Michener. A Statistical Method for Evaluating Systematic Relationships. *Univ. Kansas Sci. Bull.* **28**, 1409–1438 (1958).

69. Subramanian, A. *et al.* Gene set enrichment analysis: A knowledge-based approach for interpreting genome-wide expression profiles. *Proc. Natl. Acad. Sci. U. S. A.* **102**, 15545–15550 (2005).

70. Fabregat, A. *et al.* The Reactome pathway Knowledgebase. *Nucleic Acids Res.* **44**, D481–D487 (2016).

71. Chen, E. Y. *et al.* Enrichr: interactive and collaborative HTML5 gene list enrichment analysis tool. *BMC Bioinformatics* **14**, 128 (2013).

72. Hänzelmann, S., Castelo, R. & Guinney, J. GSVA: Gene set variation analysis for microarray and RNA-Seq data. *BMC Bioinformatics* **14**, (2013).

73. Kanehisa, M. & Goto, S. KEGG: Kyoto Encyclopedia of Genes and Genomes. *Nucleic Acids Res.* **28**, 27–30 (2000).

74. Greenwald, N. F. *et al.* Whole-cell segmentation of tissue images with human-level performance using large-scale data annotation and deep learning. *Nat. Biotechnol.* **40**, 555–565 (2022).

75. van der Walt, S. *et al.* scikit-image: image processing in Python. *PeerJ* **2**, e453 (2014).

76. 10X Genomics. Visium Spatial Gene Expression for FFPE - Tissue Preparation Guide. *Doc. Number CG000518 Rev A* (2022).

77. 10X Genomics. Visium Spatial Gene Expression for FFPE – Deparaffinization, H&E Staining, Imaging & Decrosslinking. *Doc. Number CG000520 Rev A* (2022).

78. 10X Genomics. Visium Spatial Gene Expression for FFPE – Deparaffinization, H&E Staining, Imaging & Decrosslinking. *Doc. Number CG000520 Rev B* (2022).

79. 10X Genomics. Visium CytAssist Spatial Gene Expression Reagent Kits. *Doc. Number CG000495 Rev A* (2022).

80. 10X Genomics. Visium CytAssist Spatial Gene Expression Reagent Kits. *Doc. Number CG000495 Rev E* (2023).

81. Kadri, S. *et al.* Clinical Validation of a Next-Generation Sequencing Genomic Oncology Panel via Cross-Platform Benchmarking against Established Amplicon Sequencing Assays. *J. Mol. Diagnostics* **19**, 43–56 (2017).

82. Godfrey, J. *et al.* PD-L1 gene alterations identify a subset of diffuse large B-cell

lymphoma harboring a T-cell–inflamed phenotype. *Blood* **133**, 2279–2290 (2019).

Acknowledgements

We thank D. Pincus, M. Mani, M. Lingen, D. Zemmour, I. Moskowitz, and R. Ranganathan for helpful discussions. We thank the genomics core and human immunologic monitoring core facilities at the University of Chicago for their aid in sequencing and imaging our ST-seq samples. This work was supported by the Duchossois Family Institute, the Department of Pathology, and the Center for the Physics of Evolving Systems at the University of Chicago. Funding for V.B. was provided by a T32 NIH training grant within the Department of Medicine, Section of Hematology/Oncology at the University of Chicago.

Author contributions

H.G. performed all ST-seq data collection including preparing samples and running the 10X Visium platform. U.P. provided critical conceptual guidance in writing the manuscript and conducted portions of analysis related to Figs. 3, 4, and Extended Data Fig. 10. A.D.L., A.E., A.P., C.M.B. aided in data collection of the NSCLC samples. A.D.L. aided in the writing of the Methods section. C.M.B. provided critical feedback for our manuscript. B.A.D. provided critical conceptual guidance as well as technical support in writing of the code used for TumorSPACE. J.K. provided diffuse large B cell lymphoma samples for ST-seq and conceptual guidance. M.C.G. leads the lung cancer biobank at the University of Chicago, established the IRB required for collecting ST-seq data on the NSCLC samples, and provided critical feedback for our manuscript. V.B. performed all analysis, wrote all code, and coordinated all data collection efforts. V.B. and A.S.R. conceived of the project and A.S.R. provided supervision for all aspects of analysis and data collection. V.B. and A.S.R. wrote the paper.

Competing interests

All authors declare no competing interests.

Data and materials availability

All data relevant to our manuscript can be found within associated Supplementary Tables. All ST-seq data related to the cohort of DLBCL and NSCLC patients will be available for download in the Gene Expression Omnibus (GEO) database upon assignment of an accession number.

Code availability

All code, along with annotations and step-wise instructions, will be available for download via github repository upon publication of our manuscript:

https://github.com/aramanlab/TumorSPACE.jl

Ethics Declaration

Patents (63/572,XXX) related to this research have been filed by the University of Chicago with V.B. and A.S.R. as inventors.

Materials and Correspondence

Author to whom correspondence and materials requested should be addressed is A.S.R.