**Resolving the spatial and cellular architecture of lung adenocarcinoma by multi-region single-cell sequencing**

**SUPPLEMENTARY DATA FILE 1**

**Supplementary Methods**

**Single cell isolation from LUADs and normal spatial samples**

Fresh tissues were collected in RPMI medium supplemented with 2% fetal bovine serum (FBS) and maintained on ice for immediate processing. Tissues were placed in a cell culture dish containing Hank’s balanced salt solution (HBSS) on ice and extra-pulmonary airways and connective tissue were removed with scissors. Samples were transferred to a new dish on ice and minced into tiny pieces (approximately 1mm3) followed by enzymatic digestion using a cocktail containing Collagenase A (10103578001, Sigma Aldrich), Collagenase IV (NC9836075, Thermo Fisher Scientific), DNase I, (11284932001, Sigma Aldrich), Dispase II (4942078001, Sigma Aldrich), Elastase 1 (NC9301601, Thermo Fisher Scientific) and Pronase (10165921001, Sigma Aldrich), as previously described (1). Samples were incubated in a 37 °C oven for 10 minutes with gentle rotation, pipet mixed at room temperature for further disaggregation, and then incubated at 37 °C for an additional 10 minutes. Samples were then filtered through 70 μm strainers (Miltenyi biotech, 130-098-462) and washed with ice-cold HBSS. Filtrates were then centrifuged and re-suspended in ice-cold ACK lysis buffer (A1049201, Thermo Fisher Scientific) for red blood cell (RBC) lysis. Following RBC lysis, samples were centrifuged and re-suspended in ice-cold FBS, filtered (using 40 μm FlowMi tip filters; H13680-0040, Millipore), and an aliquot was taken to enumerate cells and check for viability by Trypan blue exclusion analysis.

**Sorting and enrichment of viable lung EPCAM+ epithelial and EPCAM- non-epithelial singlets**

Single cells from patient 1 (P1) were stained with SYTOX Blue viability dye (S34857, Life Technologies), and processed on a fluorescence-activated cell sorting (FACS) Aria I instrument. Cells from patients 2 through 5 (P2-P5) were stained with anti-EPCAM-PE (347198, BD Biosciences;1:50 dilution in ice-cold phosphate-buffered saline, PBS, containing 2% FBS) for 30 minutes with gentle rotation at 4°C. EPCAM-stained cells were then washed, filtered using 40 μm tip filters, stained with SYTOX Blue and processed on a FACS Aria II instrument. Doublets and dead cells were eliminated, and two populations of viable (SYTOX-negative) singlets were separately collected in PBS containing 2% FBS: EPCAM-positive (EPCAM+, epithelial) and EPCAM-negative (EPCAM-, non-epithelial). Cells were washed again to eliminate ambient RNA, and a sample was taken for counting by Trypan Blue exclusion (T8154, Sigma Aldrich) before loading on 10X Chromium microfluidic chips.

**Preparation of single-cell 5’ gene expression libraries**

Up to 10,000 cells per sample were partitioned into nanoliter-scale Gel beads-in-emulsion (GEMs) using Chromium Next GEM Single Cell 5' Gel Bead Kit v1.1 (1000169, 10X Genomics) and by loading onto Chromium Next GEM Chips G (1000127, 10X Genomics). GEMs were then recovered to construct single-cell 5’ gene expression libraries using the Chromium Next GEM Single Cell 5’ Library kit (1000166, 10X Genomics) according to the manufacturer’s protocol. Briefly, recovered barcoded GEMs were broken and pooled, followed by magnetic bead clean-up (Dynabeads MyOne Silane, 37002D, Thermo Fisher Scientific). 10X-barcoded full-length cDNA was then amplified by PCR and analyzed using Bioanalyzer High Sensitivity DNA kit (5067-4626, Agilent). Up to 50 ng of cDNA was carried over to construct gene expression libraries and was enzymatically fragmented and size-selected to optimize the cDNA amplicon size prior to 5’ gene expression library construction. Further, samples were subject to end-repair, A-tailing, adaptor ligation, and sample index PCR using Single Index Kit T Set A (2000240, 10X Genomics) to generate Illumina-ready barcoded gene expression libraries. Library quality and yield was measured using Bioanalyzer High Sensitivity DNA (5067-4626, Agilent) and Qubit dsDNA High Sensitivity Assay (Q32854, Thermo Fisher Scientific) kits. Indexed libraries were normalized by adjusting for the ratio of the targeted cells per library as well as individual library concentration and then pooled to a final concentration of 10 nM. Library pools were then denatured and diluted as recommended for sequencing on the Illumina NovaSeq 6000 platform.

**Raw single-cell sequencing data processing and quality control**

Detailed quality control metrics were generated and carefully evaluated. Cells with low complexity libraries or likely cellular debris (in which detected transcripts are aligned to less than 200 genes) were filtered out. Low-quality cells where >15% of transcripts were derived from the mitochondrial genome were also excluded (**Supplementary Fig. S2A** middle panel). Doublets were identified for exclusion using a multi-step approach: 1) library complexity: cells with highly complex libraries (in which detected transcripts are aligned to more than 6500 genes); 2) cluster distribution: doublets or multiplets likely forming distinct clusters with hybrid expression features and exhibiting an aberrantly high gene count; 3) cluster marker gene expression: cells of a cluster expressing markers from distinct lineages (e.g., cells in the T-cell cluster showing expression of epithelial cell markers). Step 3 was performed after batch effect assessment and correction (see below). The above steps were repeated multiple times to ensure elimination of most barcodes associated with cell doublets.

We then used k-BET (2) for batch effect assessment and applied Harmony (3) for actual batch effect correction. As outlined below, these tools were applied to clustering of major lineages and specific subsets before any clustering analysis or cell type identification/annotation was performed. Statistical assessment of possible batch effects was performed using the R package k-BET (a robust and sensitive k-nearest neighbor batch-effect test) (2). k-BET was run on cells from all patient samples including P1, and on major lymphoid cell types including B cells, CD4 T cells, CD8 T cells, myeloid cells, and epithelial cells separately with default parameters. Each cell type was down sampled to 500 cells. We chose the k input value from 1% to 100% of the sample size. In each run, the number of tested neighborhoods was 10% of the sample size. The mean and maximal rejection rates were then calculated based on a total of 100 repeated k-BET runs. A low rejection rate indicates homogeneous mixing of samples from different batches. Following estimation of sample processing- or sequencing-related batch effects using k-BET, and to remove batch effects, we next employed Harmony (3) before clustering analysis (**Supplementary Fig. S2A, Fig. S5A, Fig. S11B, Fig. S16B, Fig. S20A**).

**Determination of major cell types and states as well as inference of cell cycle stage**

Cell type identification was performed on Harmony-defined clusters following batch effect correction. To determine the major cell type of each cell, differentially expressed genes (DEGs) were identified for each cell cluster using the *FindAllMarkers* function in Seurat and the top 50 most significant DEGs were reviewed. In parallel, feature plots were generated for top 50 DEGs and for a suggested set of canonical immune and stromal cell markers, as described in previous reports (4-7), followed by a manual review process. Enrichment of these markers (e.g., *EPCAM* for epithelial cells; *PTPRC* for immune cells; *CD3D/E* for T cells; *CD8A/B* for CD8 T cells, *CD4/CD40LG* for CD4 T cells; *CD19/MS4A1/CD79A* for B cells, *COL1A1/COL1A2* for fibroblasts) was considered a strong indication of the clusters representing the corresponding cell types. Top-ranked DEGs and the enrichment of canonical marker genes were integrated with global cluster distribution to infer major cell types and cellular states, as previously described (8,9). Cell cycle stage was computationally assigned for each individual cell by the function *CellCycleScoring* that is implemented in Seurat. Cell cycle stage was inferred based on expression profiles of the cell cycle related signature genes, as previously described (10). For transcriptional signature analysis, single-sample GSVA (ssGSVA) was applied to the scRNA-seq data and pathway scores were calculated for each cell using *gsva* function in GSVA software package (11).

**Analysis of clustering robustness**

Multiple methods were applied to assess the robustness of the clustering results in this study. In addition to Harmony, reciprocal PCA (rPCA) (12) was used as an independent integration method to cluster and subcluster multiple cellular subsets. Clustering consistency was quantified by calculating the Jaccard index as a measure of the overlap between the results obtained with Harmony and those obtained with rPCA. Due to the large number of cells in the analysis of the five major lineages (n = 186,916 cells in total), we randomly sampled 75% of all cells in rPCA analysis for computational efficiency (**Supplementary Fig. S3A**). To evaluate the extent with which the clustering results relied on individual cells being included, we analyzed randomly down-sampled cells: 25%, 50% and 75% from the original dataset for analysis of major lineages, EPCAM+ epithelial cells, CD8+ T cell subsets, and monocytes/macrophages. Results from clustering of down-sampled cells were compared with those derived from all cells using Jaccard index (**Supplementary Fig. S3B, Fig. S5C, Fig. S12, Fig. S18**).

To improve the computation efficiency, 10% of cells were subsampled for k-means clustering. K-means clustering was performed using *kmeans* R function with the top 10 most significant principal components (PCs). The number of PCs used was determined by visually checking the elbow plot showing the accumulated variations by PCs. We used a range of k = 2 to k = 20. The optimal k was determined by considering two points whilst keeping k as small as possible as well as quantifying cluster purity using the silhouette score and cluster heterogeneity by the within-cluster sum of square (wss) score. The final optimal k was set to 7. Results from k-means clustering were compared to those derived from Harmony (all cells), using Jaccard index (**Supplementary Fig. S3C**).

**Analysis of large-scale copy number variations and *KRAS* mutation status**

To quantify the level of aneuploidy, profiles of copy number variation (CNV) generated by inferCNV (<https://github.com/broadinstitute/inferCNV> and as described in Methods section) were aggregated using a strategy similar to that described in a previous study (13). We first computed arm-level CNV scores as the mean of the squares of CNV values across each chromosomal arm. The arm-level CNV scores were further aggregated across all chromosomal arms by taking the arithmetic mean value of the arm-level scores using R function *mean*. To associate CNV clusters with tumor clonal architecture, we constructed phylogenetic neighbor-joining trees using the R package phangorn for cells from P3 LUAD (14). The consistency between the phylogenetic neighbor-joining tree and the original hierarchical clustering tree was quantified by calculating the Pearson’s correlation coefficient between the orders of cells (clades and clusters) in both trees. *P* – values were calculated using *cor.test* function.

For single-cell somatic mutation analysis, we were able to specifically capture and analyze *KRAS* codon 12 mutation status in our cohort since *KRAS* codon 12 falls within the genomic bandwidth of our 5’ single-cell gene expression assay, i.e., its locus is present within the first 100 bp from the 5’ end of transcripts sequenced. Reads from P2 samples were extracted from the original BAM files using cell-specific barcodes and genomic coordinates of *KRAS* codon 12 mutations and then subjected to quality filtering and duplicates removal, mutation identification and annotation. Extracted alignments were manually evaluated using IGV (15).

**Analysis of LUAD expression datasets**

Gene signatures (B cell C0 signature, non-inflammatory versus inflammatory DC signature from Villani et al (8)) were validated in a previously published dataset by our group comprised of normal lung tissues, precursor atypical adenomatous hyperplasias (AAHs) and matched LUADs (n = 15, each) (16). Gene expression matrix with log transformed TPM values was downloaded from GSE102511. Pearson’s correlation coefficients between different genes were calculated using log2 transformed TPM values.

Correlation between expression of *CD24* and that of other immune-related genes or *EPCAM* was validated in three independent datasets. Firstly, we analyzed select genes of interest in a cohort of premalignant and invasive lung lesions (n = 9 AAHs, n = 6 LUADs) and normal lung tissues (n = 38) that were profiled using the nCounter PanCancer Immune Profiling Panel (NanoString Technologies) (17). Secondly, we interrogated bulk RNA-seq expression data of 51 treatment-naïve early-stage LUADs (disease stages: I, II and III) LUADs with matched adjacent normal lung tissues from The Cancer Genome Atlas (TCGA) cohort (18) and downloaded from the NCI Cancer Genomic Data Commons (NCI-GDC: <https://gdc.cancer.gov)>. For these two datasets, Pearson’s correlation coefficients were computed using normalized expression data (log2 transformed FPKM values). Thirdly, expression of select genes was validated in an in-house LUAD cohort (MDACC; n = 109) analyzed by targeted immune profiling of 1,392 genes using the HTG EdgeSeq Precision Immuno-Oncology Panel (HTG Molecular Diagnostics, Inc.). The R package MCP-counter (19) was applied to the normalized data of primary LUADs in both TCGA and MDACC cohorts. For MDACC cohort, we also performed correlation analysis of *CD24* with a computed immune cytotoxicity score (20), and Pearson’s correlation coefficient was calculated using median normalized expression.

**Immunohistochemical analysis of CD24 in LUAD**

Archived and surgically resected tumor specimens were collected from treatment-naïve LUAD patients (n = 240) with early-stage (stages I-III) disease and incorporated into a tissue microarray (TMA). Patients in this cohort were evaluated at The University of Texas MD Anderson Cancer Center and consents were obtained from all patients and were approved by the institution’s review board. Protein expression of CD24 was evaluated by immunohistochemistry (IHC) using a Leica Bond Max automated stainer (Leica Biosystems Nussloch GmbH) and using the same conditions as simultaneously stained positive and negative controls. Briefly, 4 μm-thick sections from this TMA were deparaffinized and rehydrated with standard Leica Bond protocol followed by antigen retrieval at 100°C for 20 minutes with Bond ER Solution #1, (Leica Biosystems, equivalent Citrate Buffer, pH6). Anti-CD24 primary antibody (Clone SN3, dilution 1:200, Invitrogen Cat# MA5-11828) was incubated for 15 minutes at room temperature and detected using the Bond Polymer Refine Detection kit (DS9800, Leica Biosystems) with DAB as chromogen. Slides were then counterstained with hematoxylin, dehydrated, and mounted. The stained slides were digitally scanned using the Aperio ScanScope Turbo slide scanner (Leica Microsystems Inc.) and images were captured at ×200 magnification. The images were visualized using ImageScope software (Leica Microsystems, Inc.), and digital image analysis was performed using the Aperio Image Toolbox (Leica Microsystems Inc.) by an experienced pathologist (JF). Cytoplasmic expression of CD24 was quantified using a 4-value intensity score (0, none; 1, weak; 2, moderate; and 3, strong) and the percentage (0%–100%) of the extent of reactivity. A final expression score (H-score) was obtained by multiplying the intensity and reactivity extension values (range, 0–300) as described previously (21,22). TMA specimens were simultaneously stained with positive and negative controls using the same IHC conditions with a Bond Max automated staining system (Leica Microsystems, Inc.).

**Survival analysis**

Survival analysis was performed in treatment-naïve early-stage LUAD patients from TCGA (n = 468) as well as in 56 patients (MDACC cohort) who had not received adjuvant therapy following surgery. Patients were dichotomized into upper (Q1) and lower (Q4) quartiles expressing immune cell signatures analyzed in our single-cell cohort, namely B cell C0 signature score and cDC2 signature score of the expression of a non-inflammatory/inflammatory ratio. For survival analysis based on CD24 expression, patients were dichotomized based on median *CD24* mRNA or protein expression. Statistical differences in survival between different groups were evaluated in R using the log-rank test and the Kaplan–Meier method for estimation of survival probability.

***In vitro* knockdown of *Cd24a* expression in mouse LUAD cells**

Early passage and mycoplasma-free MDA-F471 cells previously derived from LUADs of *Gprc5a-/-* mice (23) were maintained in a humidified incubator in DMEM-F12 medium supplemented with 10% fetal bovine serum (FBS) and antibiotics. Knockout of *Cd24a* expression in MDA-F471 cells was established using the CRISPR-Cas9 system, using lentiviral supernatant from predesigned Edit-R Lentiviral small guide RNAs (sgRNAs) for guiding Cas9 nuclease to create double-strand breaks in the target DNA. Cells were transduced with either control non-targeting sgRNA (VSGC11954; sgCt) or a combination of two different target sequences of mouse *Cd24a* (VSGM11945-15EG12484; sg*Cd24a*; both Horizon Discovery, Boyertown, PA) according to manufacturer’s protocol. At 48 hours post-transduction, cells were treated with puromycin (5 µg/ml) for selection of cells with integrated sgRNA, and selection medium was replenished every three days. At day 10 post-transduction, cells with high transduction efficiency were sorted by FACS.

**Fluorescence-activated cell sorting of transduced cells based on CD24 surface protein expression**

sgCt- and sg*Cd24a-*transduced MDA-F471 cells were trypsinized and washed with PBS containing 2% FBS. Cells were stained with mouse-specific Pe-Cy7-conjugated anti-CD24 antibody (1:200 dilution, clone M1/69, eBioscience) for 30 mins in the dark and with gentle rotation at 4°C. Cells were then washed twice, filtered, stained with SYTOX Blue viability dye, and processed on a FACS Aria I instrument. Under sterile conditions, CD24high-expressing MDA-F471 were sorted from cells transduced with sgCt, and CD24low cells were sorted from cells transduced with sg*Cd24* targeting lentiviruses. A fraction of the sorted cells was used for quantitative real time PCR analysis, and the rest of the cells were cultured in a humidified incubator and maintained in puromycin selection media.

**Quantitative real time-PCR** (**qRT-PCR**) **analysis of *Cd24a* expression in sorted mouse LUAD cells**

RNA was extracted from sorted MDA-F471 cells using the RNeasy extraction kit (Qiagen) following the manufacturer’s instructions, and the purified RNA was assessed on Nanodrop (RNA quality) and quantified using Qubit RNA Broad Range Assay (Q10211). cDNA was reverse transcribed from total RNA using the High-Capacity cDNA reverse transcription kit (Applied Biosystems, Foster City, CA). *Cd24* was amplified using Taqman probes specific for mouse *Cd24a* (Mm00782538\_sH FAM-MGB) and TaqMan Fast Advance Master Mix (4444963, Applied Biosystems) on QuantStudio 7 Flex Real-Time PCR System (Applied Biosystems), and according to the following cycling conditions: 1s denaturation at 95°C, 20s annealing and extension at 60°C for 40 cycles. Gene expression was normalized to both *Gapdh* (Mm99999915\_g1 FAM-MGB) and *Actin* (#Mm00607939\_s1 FAM-MGB)using the 2-ΔΔCt method.

**Analysis of LUAD transplant growth in syngeneic mice following genetic deletion of *Cd24* or inhibition with neutralizing antibodies**

*Gprc5a−/−* mice with a mixed C57BL/6 x 129/Sv genetic background were generated as previously described (24,25). For all experiments, 8-week old age-matched *Gprc5a−/−* female mice were divided into starting groups of 7 - 10 mice for each experimental arm. Before injection, parental MDA-F471, MDA-F471-CD24hi or MDA-F471-CD24neg cells were trypsinized, washed in PBS, and re-suspended at an average density of 1 x 106 cells in 50 µl PBS (calcium- and magnesium-free) supplemented with 50% phenol red-free Matrigel Matrix with high protein concentration (Corning, NY). Female mice received analgesic (Buprenorphine SR, 5 µg/ml, intraperitoneally) and were anesthetized with isoflurane. Under aseptic conditions and using a 3/10-cc insulin syringe equipped with a 30-G hypodermic needle, cells in Matrigel suspension were engrafted into the right flank of each mouse. Subcutaneous tumors were measured twice a week starting at day 7 post-engraftment using digital calipers, and tumor volume was calculated using the formula: (length x width2)/2. Mice engrafted with parental MDA-F471 cells were randomized into treatment groups (7-10 mice per group), receiving either anti-CD24 monoclonal antibody or rat IgG1 isotype control (200 µg), starting at day 7 post-engraftment, twice a week for 3 weeks (6 doses in total). Mice were sacrificed at day 28 post-engraftment.

**Supplementary References**

1. Slyper M, Porter CBM, Ashenberg O, Waldman J, Drokhlyansky E, Wakiro I*, et al.* A single-cell and single-nucleus RNA-Seq toolbox for fresh and frozen human tumors. Nature Medicine **2020**;26(5):792-802 doi 10.1038/s41591-020-0844-1.

2. Büttner M, Miao Z, Wolf FA, Teichmann SA, Theis FJ. A test metric for assessing single-cell RNA-seq batch correction. Nat Methods **2019**;16(1):43-9 doi 10.1038/s41592-018-0254-1.

3. Korsunsky I, Millard N, Fan J, Slowikowski K, Zhang F, Wei K*, et al.* Fast, sensitive and accurate integration of single-cell data with Harmony. Nat Methods **2019**;16(12):1289-96 doi 10.1038/s41592-019-0619-0.

4. Björklund Å K, Forkel M, Picelli S, Konya V, Theorell J, Friberg D*, et al.* The heterogeneity of human CD127(+) innate lymphoid cells revealed by single-cell RNA sequencing. Nat Immunol **2016**;17(4):451-60 doi 10.1038/ni.3368.

5. Lambrechts D, Wauters E, Boeckx B, Aibar S, Nittner D, Burton O*, et al.* Phenotype molding of stromal cells in the lung tumor microenvironment. Nature medicine **2018**;24(8):1277-89 doi 10.1038/s41591-018-0096-5.

6. Li S, Morita H, Sokolowska M, Tan G, Boonpiyathad T, Opitz L*, et al.* Gene expression signatures of circulating human type 1, 2, and 3 innate lymphoid cells. J Allergy Clin Immunol **2019**;143(6):2321-5 doi 10.1016/j.jaci.2019.01.047.

7. Hashimoto K, Kouno T, Ikawa T, Hayatsu N, Miyajima Y, Yabukami H*, et al.* Single-cell transcriptomics reveals expansion of cytotoxic CD4 T cells in supercentenarians. Proc Natl Acad Sci U S A **2019**;116(48):24242-51 doi 10.1073/pnas.1907883116.

8. Villani AC, Satija R, Reynolds G, Sarkizova S, Shekhar K, Fletcher J*, et al.* Single-cell RNA-seq reveals new types of human blood dendritic cells, monocytes, and progenitors. Science (New York, NY) **2017**;356(6335) doi 10.1126/science.aah4573.

9. Sade-Feldman M, Yizhak K, Bjorgaard SL, Ray JP, de Boer CG, Jenkins RW*, et al.* Defining T Cell States Associated with Response to Checkpoint Immunotherapy in Melanoma. Cell **2018**;175(4):998-1013.e20 doi 10.1016/j.cell.2018.10.038.

10. Jerby-Arnon L, Shah P, Cuoco MS, Rodman C, Su MJ, Melms JC*, et al.* A Cancer Cell Program Promotes T Cell Exclusion and Resistance to Checkpoint Blockade. Cell **2018**;175(4):984-97.e24 doi 10.1016/j.cell.2018.09.006.

11. Hänzelmann S, Castelo R, Guinney J. GSVA: gene set variation analysis for microarray and RNA-Seq data. BMC Bioinformatics **2013**;14(1):7 doi 10.1186/1471-2105-14-7.

12. Stuart T, Butler A, Hoffman P, Hafemeister C, Papalexi E, Mauck WM, 3rd*, et al.* Comprehensive Integration of Single-Cell Data. Cell **2019**;177(7):1888-902.e21 doi 10.1016/j.cell.2019.05.031.

13. Tirosh I, Venteicher AS, Hebert C, Escalante LE, Patel AP, Yizhak K*, et al.* Single-cell RNA-seq supports a developmental hierarchy in human oligodendroglioma. Nature **2016**;539(7628):309-13 doi 10.1038/nature20123.

14. Schliep KP. phangorn: phylogenetic analysis in R. Bioinformatics **2011**;27(4):592-3 doi 10.1093/bioinformatics/btq706.

15. Robinson JT, Thorvaldsdóttir H, Wenger AM, Zehir A, Mesirov JP. Variant Review with the Integrative Genomics Viewer. Cancer Res **2017**;77(21):e31-e4 doi 10.1158/0008-5472.Can-17-0337.

16. Sivakumar S, Lucas FAS, McDowell TL, Lang W, Xu L, Fujimoto J*, et al.* Genomic Landscape of Atypical Adenomatous Hyperplasia Reveals Divergent Modes to Lung Adenocarcinoma. Cancer research **2017**;77(22):6119-30 doi 10.1158/0008-5472.Can-17-1605.

17. Kudo Y, Haymaker C, Zhang J, Reuben A, Duose DY, Fujimoto J*, et al.* Suppressed immune microenvironment and repertoire in brain metastases from patients with resected non-small-cell lung cancer. Annals of oncology : official journal of the European Society for Medical Oncology **2019**;30(9):1521-30 doi 10.1093/annonc/mdz207.

18. Cancer Genome Atlas Research N. Comprehensive molecular profiling of lung adenocarcinoma. Nature **2014**;511(7511):543-50 doi 10.1038/nature13385.

19. Becht E, Giraldo NA, Lacroix L, Buttard B, Elarouci N, Petitprez F*, et al.* Estimating the population abundance of tissue-infiltrating immune and stromal cell populations using gene expression. Genome Biol **2016**;17(1):218 doi 10.1186/s13059-016-1070-5.

20. Rooney MS, Shukla SA, Wu CJ, Getz G, Hacohen N. Molecular and genetic properties of tumors associated with local immune cytolytic activity. Cell **2015**;160(1-2):48-61 doi 10.1016/j.cell.2014.12.033.

21. Parra ER, Behrens C, Rodriguez-Canales J, Lin H, Mino B, Blando J*, et al.* Image Analysis-based Assessment of PD-L1 and Tumor-Associated Immune Cells Density Supports Distinct Intratumoral Microenvironment Groups in Non-small Cell Lung Carcinoma Patients. Clin Cancer Res **2016**;22(24):6278-89 doi 10.1158/1078-0432.CCR-15-2443.

22. Treekitkarnmongkol W, Hassane M, Sinjab A, Chang K, Hara K, Rahal Z*, et al.* Augmented Lipocalin-2 is Associated with COPD and Counteracts Lung Adenocarcinoma Development. American journal of respiratory and critical care medicine **2020** doi 10.1164/rccm.202004-1079OC.

23. Fujimoto J, Kadara H, Men T, van Pelt C, Lotan D, Lotan R. Comparative functional genomics analysis of NNK tobacco-carcinogen induced lung adenocarcinoma development in Gprc5a-knockout mice. PLoS One **2010**;5(7):e11847 doi 10.1371/journal.pone.0011847.

24. Fujimoto J, Nunomura-Nakamura S, Liu Y, Lang W, McDowell T, Jakubek Y*, et al.* Development of Kras mutant lung adenocarcinoma in mice with knockout of the airway lineage-specific gene Gprc5a. Int J Cancer **2017**;141(8):1589-99 doi 10.1002/ijc.30851.

25. Tao Q, Fujimoto J, Men T, Ye X, Deng J, Lacroix L*, et al.* Identification of the retinoic acid-inducible Gprc5a as a new lung tumor suppressor gene. J Natl Cancer Inst **2007**;99(22):1668-82 doi 10.1093/jnci/djm208.