Supplementary Materials for:

Distinct tumor-infiltrating lymphocyte landscapes are associated with clinical outcomes in localized non-small cell lung cancer

Lorenzo Federico[†1], Daniel J. McGrail[†2], Salah-Eddine Bentebibel[†3], Cara Haymaker[4], Andrea Ravelli[3], Marie-Andrée Forget[3], Tatiana Karpinets[5], Peixin Jiang[6], Alexandre Reuben[6], Marcelo V. Negrao[6], Jun Li[5], Roohussaba Khairullah[6], Jianhua Zhang[5], Annikka Weissferdt[7], Ara A. Vaporciyan[8], Mara B. Antonoff[8], Garrett Walsh[8], Shiaw-Yih Lin[2], Andrew Futreal[5], Ignacio Wistuba[4], Jack Roth[8], Lauren A. Byers[6], Pierre-Olivier Gaudreau[9], Naohiro Uraoka[4], Alejandro Francisco Cruz[4], Hitoshi Dejima[4], Rossana N. Lazcano[4], Luisa M. Solis[4], Edwin R. Parra[4], J. Jack Lee[10], Stephen Swisher[8], Tina Cascone[6], John V. Heymach[6], Jianjun Zhang[5,6]*, Boris Sepesi[8]*, Don L. Gibbons[6,11]*, and Chantale Bernatchez[3]*

[1]Therapeutics Discovery Division, [2]Department of Systems Biology, [3]Department of Melanoma Medical Oncology, [4]Department of Translational Molecular Pathology, [5]Department of Genomic Medicine, [6]Department of Thoracic/Head and Neck Medical Oncology, [7]Department of Pathology, [8]Department of Thoracic and Cardiovascular Surgery, The University of Texas MD Anderson Cancer Center; Houston, Texas, 77030, USA.
[9]Department of Oncology, Queens' University and the Canadian Cancer Trials Group; Kingston, Ontario, Canada
[10]Department of Biostatistics, [11]Department of Molecular and Cellular Oncology, The University of Texas MD Anderson Cancer Center; Houston, Texas, 77030, USA.

[†]These authors equally contributed to this work.

*Corresponding Authors: jzhang20@mdanderson.org (J.Z.), bsepesi@mdanderson.org (B.S.), dlgibbon@mdanderson.org (D.L.G.), cbernatchez@mdanderson.org (C.B.)

Supplementary Materials and Methods

Supplementary References

Fig. S1. Flow cytometry gating strategy for the lymphocyte characterization in uninvolved lung tissue and tumors.

Fig. S2. Correlation between assays.

Fig. S3. Expression of co-stimulatory/checkpoint proteins by CD103+ status.

Fig. S4. Comparison of tumor and adjacent uninvolved tissue stratified by clinicopathological features.

Fig. S5. Metrics for determining optimal number of clusters for immunotypes.

Fig. S6. tSNE analysis of uninvolved lung tissue and prognostic relevance.

Fig. S7. Exemplary multispectral immunofluorescence images.

Fig. S8. Tumor purity does not influence TMB.

Fig. S9. Relationship between indels and dMMR mutational signatures.

Fig. S10. Overview of IRAPS and immunotype signature validation.

Table S1. Assays performed on samples in this study.

Table S2: Antibodies for flow cytometry.

Table S3. Differential populations between immunotypes

Table S4. Immunotype gene expression signature.

Flow cytometry. Fresh tumor tissue was disaggregated using the BD[TM] Medimachine System (BD biosciences) to make a single cell suspension for subsequent flow cytometry staining. Surface staining was performed in 1X DPBS with 1% bovine serum albumin for 30 min on ice using fluorochrome-conjugated monoclonal antibodies against CD45 (BUV395, Clone HI30), CD3 (PerCP-Cy5.5, Clone SK7), CD3 (PE-Cy[TM]7, Clone UCHT1), CD8 (AF 700, Clone RPA-T8) , CD8 (APC-Cy[TM]7, Clone RPA-T8), CD4 (BUV496, Clone SK3), PD-1 (PerCP-Cy[TM]5.5, Clone EH12.2H7), PD-1 (BV650, Clone EH12) and TIM3 (APC, Clone F38-2E2), TIM3 (BV605, Clone F38-2E2), CD103 (BV711, Clone Ber-Act8), CTLA4 (BV786, Clone BNI3), GITR (AF 488, Clone eBioAITR), LAG3 (PE, Clone 3DS223H), CD56 (PE-Cy7, Clone B159), ICOS (BV421, Clone C398.A4), and CD25 (APCFire/750, Clone BC96). To discriminate between conventional CD4$^+$ T cells and CD4$^+$ Treg cells, intracellular Foxp3 staining with FOXP3 antibody (PE-eFluor610, PCH101) was performed using Foxp3/Transcription Factor kit (eBioscience). To assess cell proliferation, Ki-67 (Ki67, APC, Clone 20Raj1), a cell-cycle marker expressed by cycling or recently divided cells was used. To assess cytotoxic and effector profiles cells were fixed after surface staining, and permeabilized using the BD Cytofix/Cytoperm[TM] solution (BD Biosciences) according to manufacturer's instructions. Intracellular staining was performed in BD Perm/Wash[TM] buffer (554723, BD Bioscences) using Perforin (FITC, Clone dG9), Granzyme B (V450, Clone GB11), and Interferon-γ (PE, Clone B27,) anti-human antibodies. Samples were acquired using the BD FACSCanto II and BD Fortessa X20 and analyzed using FlowJo Software v 10.5.3 (Tree Star, Inc.). Dead cells were stained using LIVE/DEAD[TM] Fixable Yellow Dead Cell Stain dye (Life Technologies) and excluded from the analysis. Flow cytometry gating strategy and antibodies used are shown in Fig. S1 and Table S2, respectively.

Multiplex immunofluorescence. Multiplex immunofluorescence analysis was performed as previously described[1]. Briefly, 4-µm-thick formalin-fixed, paraffin- embedded (FFPE) tumor sections were stained using an automated staining system (BOND-MAX; Leica Biosystems, Wetzlar, Germany). Antibodies against cytokeratins (clone AE1/AE3, dilution 1:300, DAKO, Santa Clara, CA), PD-L1 (clone E1L3N, dilution 1:3000, Cell Signaling Technology, Danvers, MA), CD3 (catalog number A045201-2, dilution 1:100, DAKO), CD8 (clone C8/144B, dilution 1:300, Thermo Fisher Scientific, Bellefonte, PA), and FOXP3 (clone D2W8E, dilution 1:100, Cell Signaling Technology), PD-1 (clone [EPR4877(2)], dilution 1:250, ABCAM, Cambridge, MA), CD68 (clone PG-M1, dilution 1:450, DAKO) and CD20 (clone L26, dilution 1:250, DAKO). All antibodies were linked with one of the fluorophores from the Opal 7 IHC kit, (catalog no. NEL797001KT; Akoya Biosciences, Waltham, MA, USA). Five fields (each $0.3345mm^2$) were selected within the tumor using the phenochart 1.0.4 viewer software (Akoya Biosciences) and imaged at 20x magnification with the Vectra 3.0 scanner (Akoya Biosciences). Imaging analysis was performed with a quantitative image analysis software (InForm, Akoya Biosciences) by using spectral libraries defined with single-marker immunofluorescence detection. The following phenotypes were identified for the T cell panel: total T cells ($CD3^+$), cytotoxic T cells ($CD3^+CD8^+Foxp3^-$), helper T cells ($CD3^+CD8^-Foxp3^-$), Tregs ($CD3^+CD8^-Foxp3^+$), PD-1 expressing by T cells ($CD3^+PD-1^+$).

TLS quantification. TLSs were quantified by a pathologist (RL) using Haematoxylin and Eosin -stained slides obtained from FFPE tumor sections**.** TLS were recognized as a group of cohesive lymphocytes that morphologically resembled follicles of secondary lymphoid organs, excluding the bronchial associated lymphoid tissue (BALT). We evaluated total number of intratumoral TLS

in the entire tumor section from a representative tumor block and then normalized to tumor area to yield TLSs density (TLS per mm^2), as previously described[2].

TCR sequencing. Analysis of the proportion of T cells in tumors and uninvolved lung tissues was performed using immunoSEQ (Adaptive Biotechnologies). Briefly, genomic DNA was extracted from bulk tumor samples and the CDR3 variable region of the T cell receptor involved in antigen binding was PCR amplified using the hsTCRβ immunoSEQ kit (Adaptive Biotechnologies). Following PCR amplification, PCR products were sequenced using a MiSeq 150x kit (Illumina), as recommended. Sequencing data were then uploaded to the immunoSEQ Analyzer using the Analyzer Data Assistant for data deconvolution. T cell fraction, a measure of the proportion of T cells in a sample was calculated based on the assumption that each cell contains 6.6pg of DNA according to the following formula: T cell fraction (% T cells) = T cell DNA / Total DNA where T cell DNA = # productive TCR templates X 6.6pg of DNA per cell and Total DNA = DNA input for assay.

RNA sequencing. RNA sequencing analysis was performed as previously described[3]. Purified double-stranded cDNA (dscDNA) was generated from 150ng RNA from fresh frozen tumor specimens and was amplified using both 3′ poly(A) selection and random priming. Samples were quantified using the NanoDrop ND-3300 spectrophotometer and Invitrogen Quant-iT picogreen DNA quantitation assay (both ThermoFisher Scientific). Then 150 ng of each sample was sheared using the Covaris E220 focused ultrasonicator following the Covaris DNA shearing protocol to obtain a final library insert size of 150–200bp (both Covaris, Woburn, MA, USA). A double-stranded DNA library was created using 100ng of sheared, dscDNA, preparing the fragments for

hybridization onto a flowcell. The ligated products were amplified using 8 cycles of PCR and the resulting libraries were assessed using the Agilent Bioanalyzer High Sensitivity DNA chip to determine successful library construction. A qPCR quantitation was performed on each library to determine the concentration of adapter ligated fragments using an Eppendorf eMotion Real-Time PCR Detection System (Eppendorf, Hauppauge, NY, USA) and a KAPA Library Quant Kit (Kapa Biosystems, Wilmington, MA, USA). Libraries were pooled in equimolar amounts based on the qPCR quantification; a qPCR was run on the pooled libraries to determine the concentration of the pool for bridge amplification. Using the concentration from the Eppendorf qPCR instrument, 15pM of the library pool was loaded onto a flowcell and amplified by bridge amplification using the Illumina cBot instrument (Illumina, San Diego, CA, USA). A paired-end 76-cycle run was used to sequence the flowcell on a HiSeq 2000 or 2500 Sequencing System (Illumina). Transcriptome reads were mapped to the reference human genome hg19, and then normalized and quantified as counts using HTSeq-count (HTSeq). The counts were used to prepare inputs for further analysis by other computational tools.

Whole-exome sequencing. Whole-exome sequencing analysis was performed as previously described[3]. Briefly, exome capture was performed on 200ng of genomic DNA per sample based on KAPA library prep (Kapa Biosystems) using the Agilent SureSelect Human All Exon V4 kit according to the manufacturer's instructions and paired-end multiplex sequencing of samples was performed on the HiSeq 2500 sequencing platform (Illumina). The Genome Analysis Toolkit (modules "IndelRealigner" and "BaseRecalibrator") was applied to perform insertion/ deletion (indel) realignment and base quality recalibration. MuTect and Pindel were applied to each tumor and its matching normal tissue sample to detect somatic single nucleotide variants (SNVs) and

small indels. Tumor mutational burden (TMB) was quantified as the total number of non-synonymous mutations per megabase (mut/Mb). Sequenza was used to quantify tumor purity[4].

ICON cohort immunotype analysis. Patients were separated into immunotypes based on immune populations determined by flow cytometry by first reducing the data dimensionality using t-distributed stochastic neighbor embedding (tSNE), and then performing k-means clustering on the resulting two-dimensional data using built-in functions in Matlab 2019a. Flow parameters not present in over 25% patients and patients lacking over 50% of flow markers were excluded from analysis, yielding a data set of 47 patients. Univariate and multivariate survival were assessed as described above, considering clinical co-variates found to be significant from initial analysis (tumor stage and subtype). To test for significantly different immune cell populations and mutational signatures between immunotypes, a Wilcoxon rank-sum test was used. Differences between immunotypes based on smoking status, sex, and neoadjuvant treatment were determined by Fisher's exact test. Differences between immunotypes based on tumor stage and tumor size were assessed with a Wilcoxon rank-sum test. All comparisons were corrected for multiple comparisons using the Benjamini–Hochberg procedure. To generate a gene expression signature predictive of patient immunotype, we first subdivided the ICON cohort into training (N = 19) and testing (N = 9) groups. The training group was used to generate the signature based on an adaption of our published IRAPS algorithm[5], using immunotype as the binary variable instead of drug sensitivity. Gene set enrichment analysis (GSEA)[6] was performed with GSEA software v4.0.2 using the pre-ranked function based on log2 (gene fold change), canonical pathways (c2.cp.v7.0), 1000 iterations, and meandiv normalization. All other analysis was performed in Matlab R2019a. Heatmap visualization was generated using ComplexHeatmap v2.1.0 in R v3.6.1.

TCGA Data Analysis. TCGA data were downloaded using the TCGA data portal (https://portal.gdc.cancer.gov/) from the Pan-Cancer Atlas release (April 2018). TCGA mutational signatures were acquired from Knijnenburg et al.[7]. Immunotype gene expression scores were calculated from log2-transformed RNAseq data, and patients were divided at the median value. GSEA and survival analysis were performed as described for the ICON cohort. Validation of changes in mutational signatures 1 and 6 between immunotypes was assessed using a Wilcoxon rank-sum test.

Data availability. Raw data for flow cytometry and multispectral immunofluorescence, along with corresponding clinical annotations for the flow cytometry cohort, are deposited at Mendeley Data (DOI: 10.17632/2j3dpx83w6.1). Raw TCR sequencing data is available through the immunoSeq website (DOI: 10.21417/LF2021AO). Data are also available upon request from the authors.

Statistical analysis. Specific details of statistical analyses are provided in corresponding figure legends. In general, comparisons matched variables, such as tumor and normal adjacent tissue, were made using a paired t-test (normally distributed data) or sign-rank test (non-normally distributed data). Comparisons of two groups were made using t-test (normally distributed data) or rank-sum test (non-normally distributed data). Comparison of two continuous variables was made using a Pearson correlation coefficient (normally distributed data) or Spearman rank correlation (non-normally distributed data). Multiple comparisons were accounted for by Benjamini-Hochberg procedure. Survival was assessed with log-rank test (two groups) or Cox proportional hazards model (multiple variables).

Supplementary References

1. Parra ER, Uraoka N, Jiang M, et al. Validation of multiplex immunofluorescence panels using multispectral microscopy for immune profiling of formalin-fixed and paraffin-embedded human tumor tissues. *Sci Rep.* 2017;7:13380.

2. Posch F, Silina K, Leibl S, et al. Maturation of tertiary lymphoid structures and recurrence of stage II and III colorectal cancer. *Oncoimmunology.* 2018;7:e1378844.

3. Mitchell KG, Diao L, Karpinets T, et al. Neutrophil expansion defines an immunoinhibitory peripheral and intratumoral inflammatory milieu in resected non-small cell lung cancer: a descriptive analysis of a prospectively immunoprofiled cohort. *J Immunother Cancer.* 2020;8. https://doi.org/10.1136/jitc-2019-000405.

4. Favero F, Joshi T, Marquard AM, et al. Sequenza: allele-specific copy number and mutation profiles from tumor sequencing data. *Ann Oncol.* 2015;26:64-70.

5. McGrail DJ, Lin CC-J, Garnett J, et al. Improved prediction of PARP inhibitor response and identification of synergizing agents through use of a novel gene expression signature generation algorithm. *NPJ Syst Biol Appl.* 2017;3:8.

6. Subramanian A, Tamayo P, Mootha VK, et al. Gene set enrichment analysis: a knowledge-based approach for interpreting genomewide expression profiles. *Proc Natl Acad Sci USA.* 2005;102:15545-15550.

7. Knijnenburg TA, Wang L, Zimmermann MT, et al. Genomic and molecular landscape of DNA damage repair deficiency across the cancer genome atlas. *Cell Rep.* 2018;23:239-254.e236.
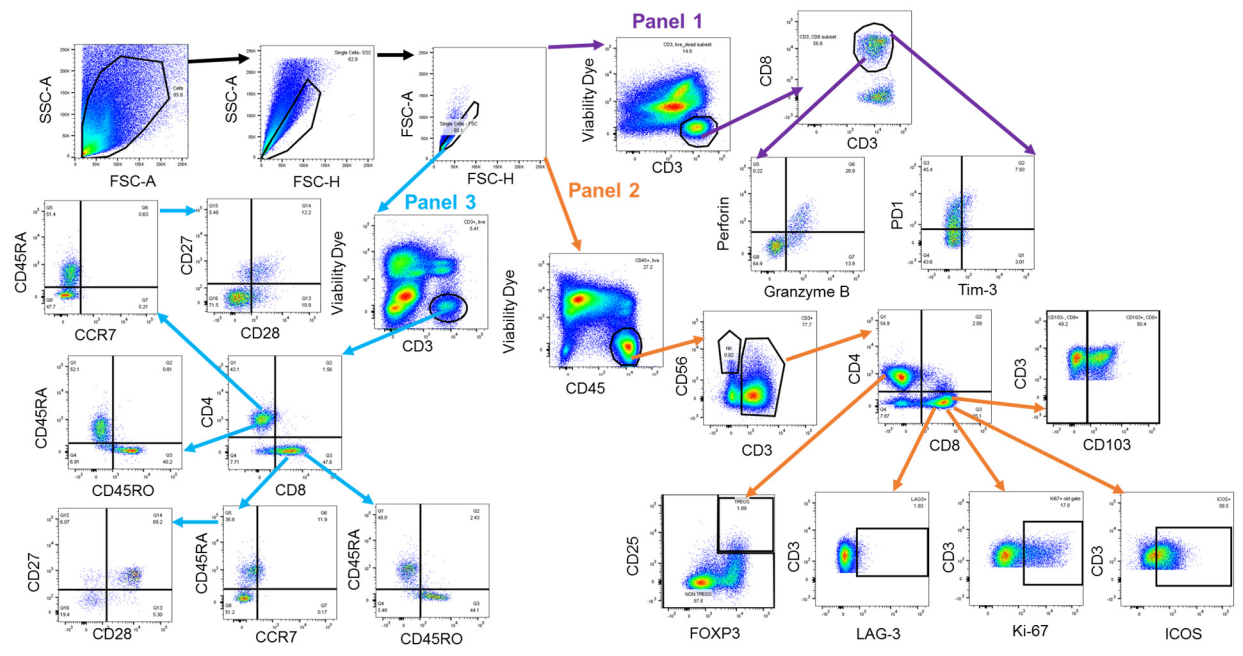
Fig. S1. Flow cytometry gating strategy for the lymphocyte characterization in uninvolved lung tissue and tumors.
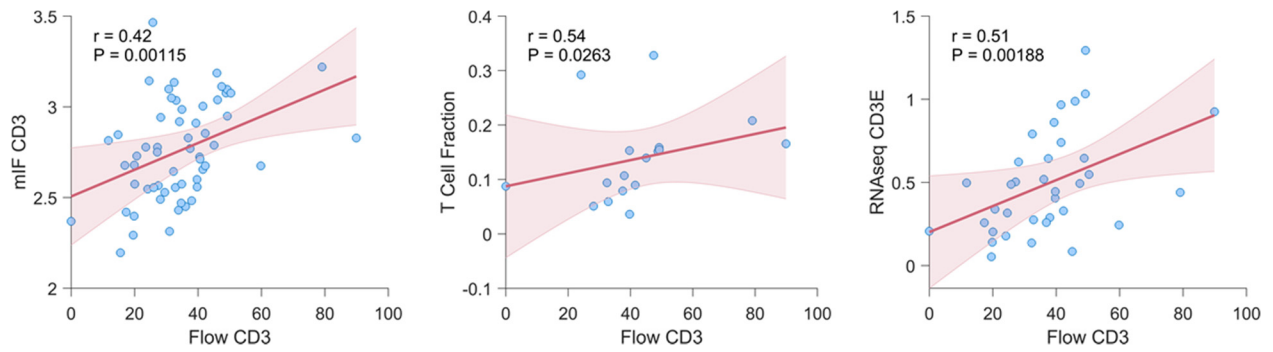
Fig. S2. Correlation between assays.

Correlation of T cell percentage within CD45[+] cells determined by flow cytometry (Flow CD3) and multispectral immunofluorescence (mIF CD3), TCRseq (T Cell Fraction), or RNAseq (RNAseq CD3E). r = Pearson correlation coefficient.
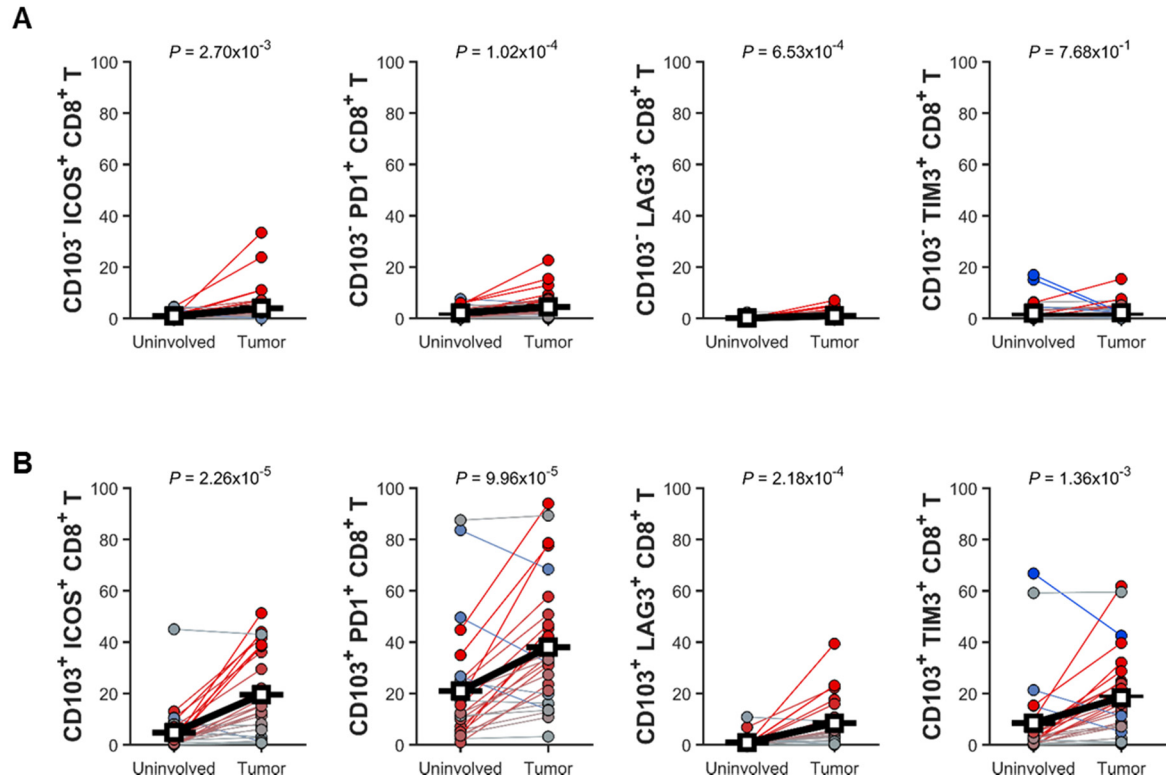
Fig. S3. Expression of co-stimulatory/checkpoint proteins by CD103+ status.

(A) Percentage of CD103-ICOS+, CD103-PD1+, CD103-TIM3+, and CD103-LAG3+CD8+ T cells within CD3+ T cells in tumor and uninvolved matched tissues. **(B)** Percentage of CD103+ICOS+, CD103+PD1+, CD103+TIM3+, and CD103+LAG3+CD8+ T cells within CD3+ T cells in tumor and uninvolved matched tissues.

Fig. S4. Comparison of tumor and adjacent uninvolved tissue stratified by clinicopathological features.

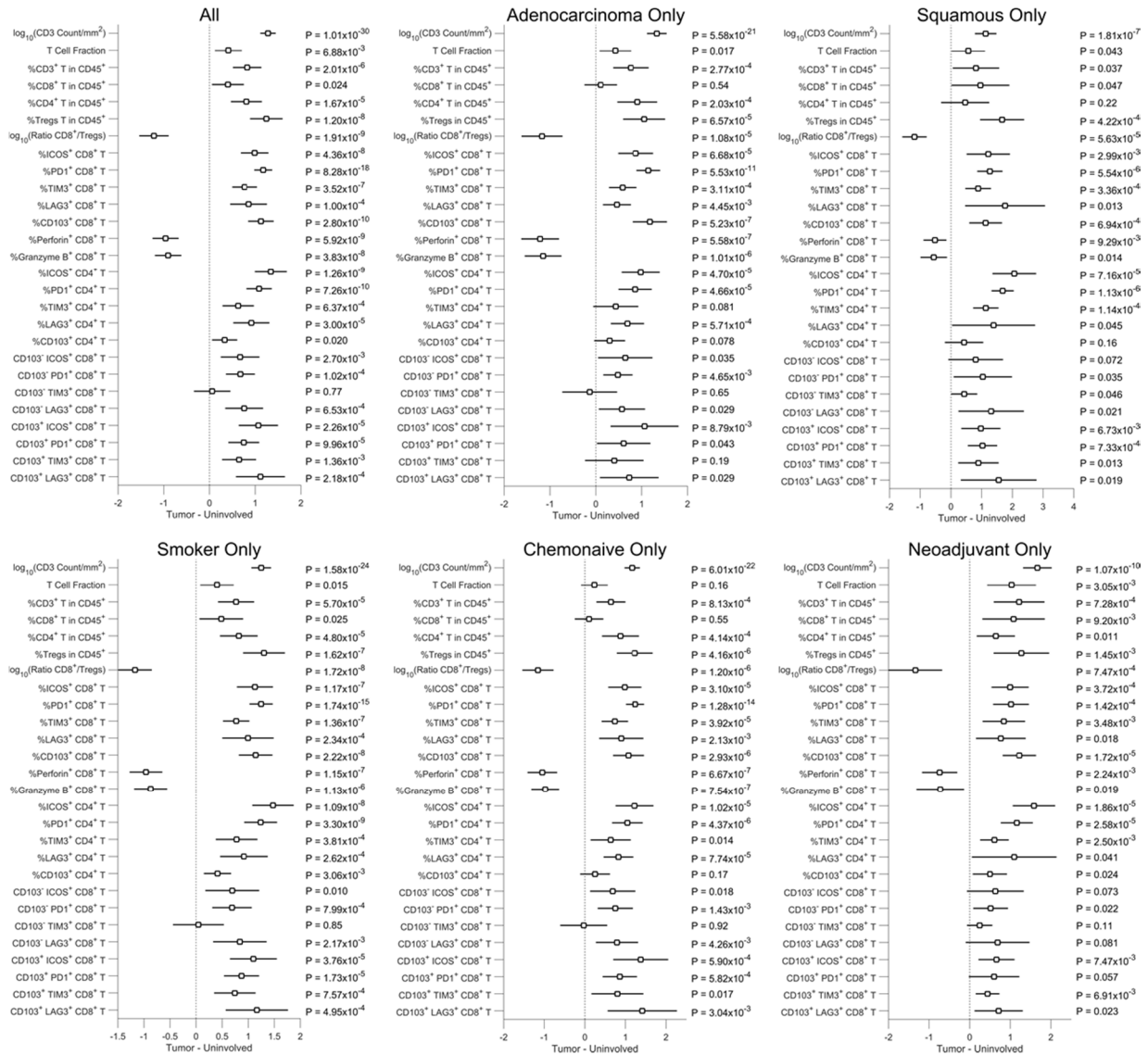| Feature	All	Adenocarcinoma Only	Squamous Only	Smoker Only	Chemonaive Only	Neoadjuvant Only
$\log_{10}(\text{CD3 Count/mm}^2)$	$P = 1.01 \times 10^{-30}$	$P = 5.58 \times 10^{-21}$	$P = 1.81 \times 10^{-7}$	$P = 1.58 \times 10^{-24}$	$P = 6.01 \times 10^{-22}$	$P = 1.07 \times 10^{-10}$
T Cell Fraction	$P = 6.88 \times 10^{-3}$	$P = 0.017$	$P = 0.043$	$P = 0.015$	$P = 0.16$	$P = 3.05 \times 10^{-3}$
$\%\text{CD3}^+$ T in CD45^+	$P = 2.01 \times 10^{-6}$	$P = 2.77 \times 10^{-4}$	$P = 0.037$	$P = 5.70 \times 10^{-5}$	$P = 8.13 \times 10^{-4}$	$P = 7.28 \times 10^{-4}$
$\%\text{CD8}^+$ T in CD45^+	$P = 0.024$	$P = 0.54$	$P = 0.047$	$P = 0.025$	$P = 0.55$	$P = 9.20 \times 10^{-3}$
$\%\text{CD4}^+$ T in CD45^+	$P = 1.67 \times 10^{-5}$	$P = 2.03 \times 10^{-4}$	$P = 0.22$	$P = 4.80 \times 10^{-5}$	$P = 4.14 \times 10^{-4}$	$P = 0.011$
$\%$Tregs in CD45^+	$P = 1.20 \times 10^{-8}$	$P = 6.57 \times 10^{-5}$	$P = 4.22 \times 10^{-4}$	$P = 1.62 \times 10^{-7}$	$P = 4.16 \times 10^{-6}$	$P = 1.45 \times 10^{-3}$
$\log_{10}(\text{Ratio CD8}^+/\text{Tregs})$	$P = 1.91 \times 10^{-9}$	$P = 1.08 \times 10^{-5}$	$P = 5.63 \times 10^{-5}$	$P = 1.72 \times 10^{-9}$	$P = 1.20 \times 10^{-6}$	$P = 7.47 \times 10^{-4}$
$\%\text{ICOS}^+ \text{CD8}^+$ T	$P = 4.36 \times 10^{-8}$	$P = 6.68 \times 10^{-5}$	$P = 2.99 \times 10^{-3}$	$P = 1.17 \times 10^{-7}$	$P = 3.10 \times 10^{-5}$	$P = 3.72 \times 10^{-4}$
$\%\text{PD1}^+ \text{CD8}^+$ T	$P = 8.28 \times 10^{-18}$	$P = 5.53 \times 10^{-11}$	$P = 5.54 \times 10^{-6}$	$P = 1.74 \times 10^{-15}$	$P = 1.28 \times 10^{-14}$	$P = 1.42 \times 10^{-4}$
$\%\text{TIM3}^+ \text{CD8}^+$ T	$P = 3.52 \times 10^{-7}$	$P = 3.11 \times 10^{-4}$	$P = 3.36 \times 10^{-4}$	$P = 1.36 \times 10^{-7}$	$P = 3.92 \times 10^{-5}$	$P = 3.48 \times 10^{-3}$
$\%\text{LAG3}^+ \text{CD8}^+$ T	$P = 1.00 \times 10^{-4}$	$P = 4.45 \times 10^{-3}$	$P = 0.013$	$P = 2.34 \times 10^{-4}$	$P = 2.13 \times 10^{-3}$	$P = 0.018$
$\%\text{CD103}^+ \text{CD8}^+$ T	$P = 2.80 \times 10^{-10}$	$P = 5.23 \times 10^{-7}$	$P = 6.94 \times 10^{-4}$	$P = 2.22 \times 10^{-8}$	$P = 2.93 \times 10^{-6}$	$P = 1.72 \times 10^{-5}$
$\%\text{Perforin}^+ \text{CD8}^+$ T	$P = 5.92 \times 10^{-9}$	$P = 5.58 \times 10^{-7}$	$P = 9.29 \times 10^{-3}$	$P = 1.15 \times 10^{-7}$	$P = 6.67 \times 10^{-7}$	$P = 2.24 \times 10^{-3}$
$\%\text{Granzyme B}^+ \text{CD8}^+$ T	$P = 3.83 \times 10^{-8}$	$P = 1.01 \times 10^{-6}$	$P = 0.014$	$P = 1.13 \times 10^{-6}$	$P = 7.54 \times 10^{-7}$	$P = 0.019$
$\%\text{ICOS}^+ \text{CD4}^+$ T	$P = 1.26 \times 10^{-9}$	$P = 4.70 \times 10^{-4}$	$P = 7.16 \times 10^{-5}$	$P = 1.09 \times 10^{-8}$	$P = 1.02 \times 10^{-5}$	$P = 1.86 \times 10^{-5}$
$\%\text{PD1}^+ \text{CD4}^+$ T	$P = 7.26 \times 10^{-10}$	$P = 4.66 \times 10^{-5}$	$P = 1.13 \times 10^{-6}$	$P = 3.30 \times 10^{-9}$	$P = 4.37 \times 10^{-6}$	$P = 2.58 \times 10^{-5}$
$\%\text{TIM3}^+ \text{CD4}^+$ T	$P = 6.37 \times 10^{-4}$	$P = 0.081$	$P = 1.14 \times 10^{-4}$	$P = 3.81 \times 10^{-4}$	$P = 0.014$	$P = 2.50 \times 10^{-3}$
$\%\text{LAG3}^+ \text{CD4}^+$ T	$P = 3.00 \times 10^{-5}$	$P = 5.71 \times 10^{-4}$	$P = 0.045$	$P = 2.62 \times 10^{-4}$	$P = 7.74 \times 10^{-5}$	$P = 0.041$
$\%\text{CD103}^+ \text{CD4}^+$ T	$P = 0.020$	$P = 0.078$	$P = 0.16$	$P = 3.06 \times 10^{-3}$	$P = 0.17$	$P = 0.024$
$\text{CD103}^- \text{ICOS}^+ \text{CD8}^+$ T	$P = 2.70 \times 10^{-3}$	$P = 0.035$	$P = 0.072$	$P = 0.010$	$P = 0.018$	$P = 0.073$
$\text{CD103}^- \text{PD1}^+ \text{CD8}^+$ T	$P = 1.02 \times 10^{-4}$	$P = 4.65 \times 10^{-3}$	$P = 0.035$	$P = 7.99 \times 10^{-4}$	$P = 1.43 \times 10^{-3}$	$P = 0.022$
$\text{CD103}^- \text{TIM3}^+ \text{CD8}^+$ T	$P = 0.77$	$P = 0.65$	$P = 0.046$	$P = 0.85$	$P = 0.92$	$P = 0.11$
$\text{CD103}^- \text{LAG3}^+ \text{CD8}^+$ T	$P = 6.53 \times 10^{-4}$	$P = 0.029$	$P = 0.021$	$P = 2.17 \times 10^{-3}$	$P = 4.26 \times 10^{-3}$	$P = 0.081$
$\text{CD103}^+ \text{ICOS}^+ \text{CD8}^+$ T	$P = 2.26 \times 10^{-5}$	$P = 8.79 \times 10^{-3}$	$P = 6.73 \times 10^{-3}$	$P = 3.76 \times 10^{-5}$	$P = 5.90 \times 10^{-4}$	$P = 7.47 \times 10^{-3}$
$\text{CD103}^+ \text{PD1}^+ \text{CD8}^+$ T	$P = 9.96 \times 10^{-5}$	$P = 0.043$	$P = 7.33 \times 10^{-4}$	$P = 1.73 \times 10^{-5}$	$P = 5.82 \times 10^{-4}$	$P = 0.057$
$\text{CD103}^+ \text{TIM3}^+ \text{CD8}^+$ T	$P = 1.36 \times 10^{-3}$	$P = 0.19$	$P = 0.013$	$P = 7.57 \times 10^{-4}$	$P = 0.017$	$P = 6.91 \times 10^{-3}$
$\text{CD103}^+ \text{LAG3}^+ \text{CD8}^+$ T	$P = 2.18 \times 10^{-4}$	$P = 0.029$	$P = 0.019$	$P = 4.95 \times 10^{-4}$	$P = 3.04 \times 10^{-3}$	$P = 0.023$

x-axis label for all panels: Tumor - Uninvolved

Fig. S4. Comparison of tumor and adjacent uninvolved tissue stratified by clinicopathological features.

Difference in indicated immunological features (flow cytometry data) between tumor and adjacent uninvolved tissue either in all patient samples or indicated patient subsets. Values were z-normalized prior to determining difference to facilitate visualization across multiple different immunological features. Bars indicate 95% confidence interval.
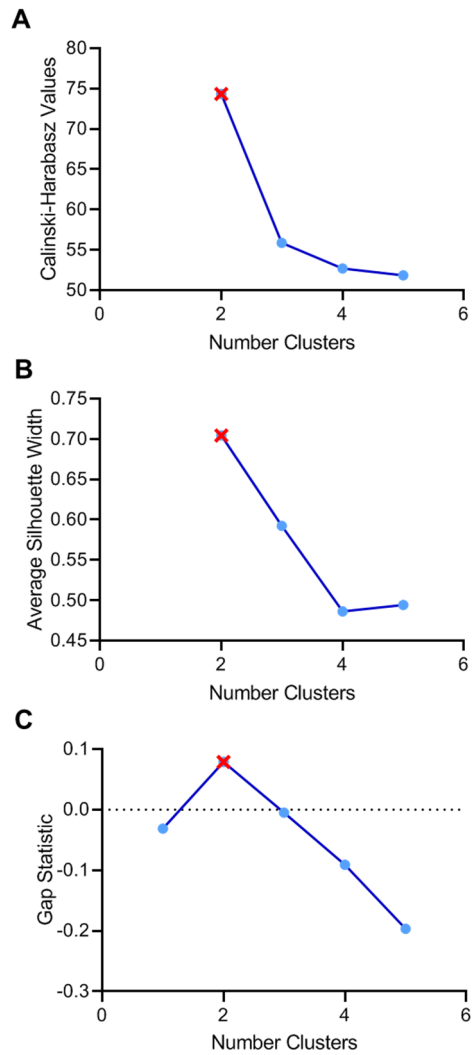
Fig. S5. Metrics for determining optimal number of clusters for immunotypes.

Red X indicates optimal number of clusters (k=2 for all metrics). **(A)** Calinski-Harabasz values.

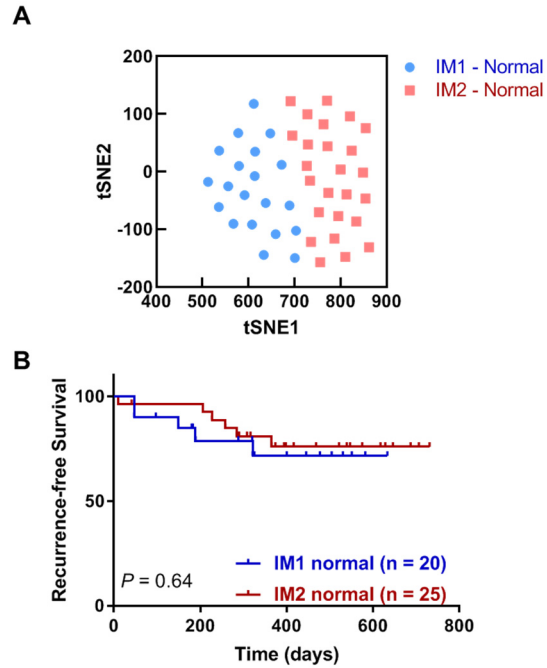**(B)** Average silhouette width. **(C)** Gap statistic.

Fig. S6. tSNE analysis of uninvolved lung tissue and prognostic relevance. (A) tSNE showed uninvolved lung tissue from NSCLC patients (n=45) could be assigned to two major IMs (IM1-normal and IM2-normal). **(B)** Kaplan-Meier curves showing the lack of association between IM1-normal (blue), IM2-normal signature (red) and recurrence-free survival in uninvolved lung tissue.

Panel 1 · Panel 2

DAPI | Cytokeratin | CD3 | CD8 | CD68 | PD-1 | PD-L1

DAPI | Cytokeratin | CD3 | CD8 | CD45RO | FoxP3 | Granzyme B
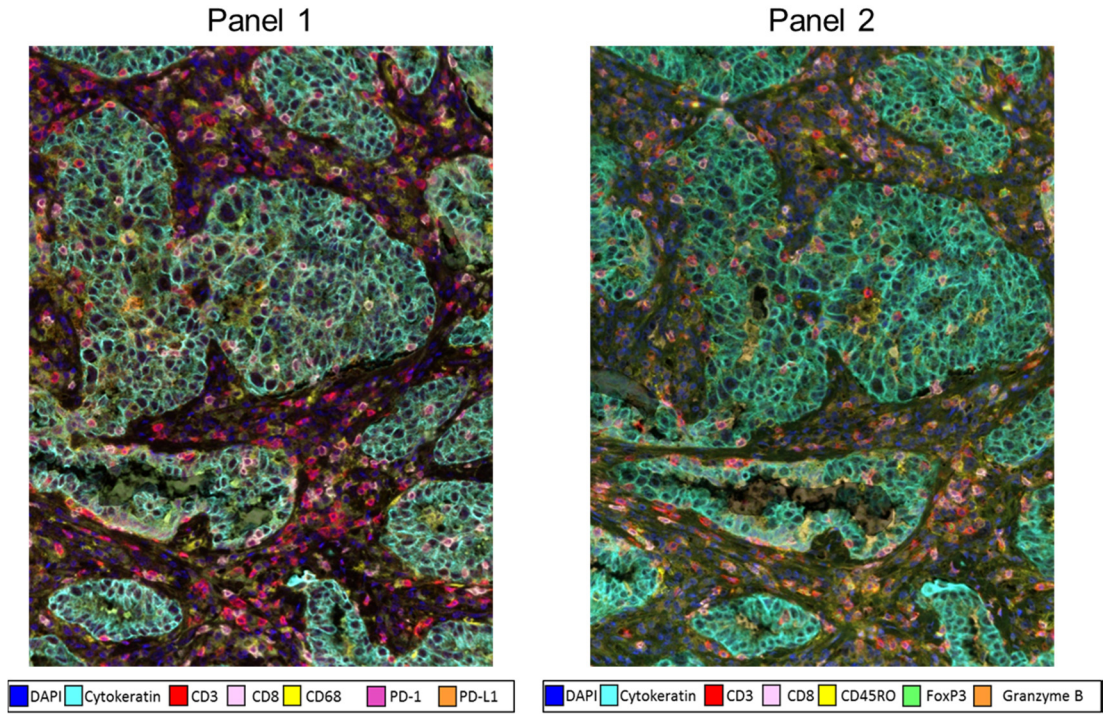
Fig. S7. Exemplary multispectral immunofluorescence images.

Representative mIF images for both panels utilized, with corresponding color annotations shown below.
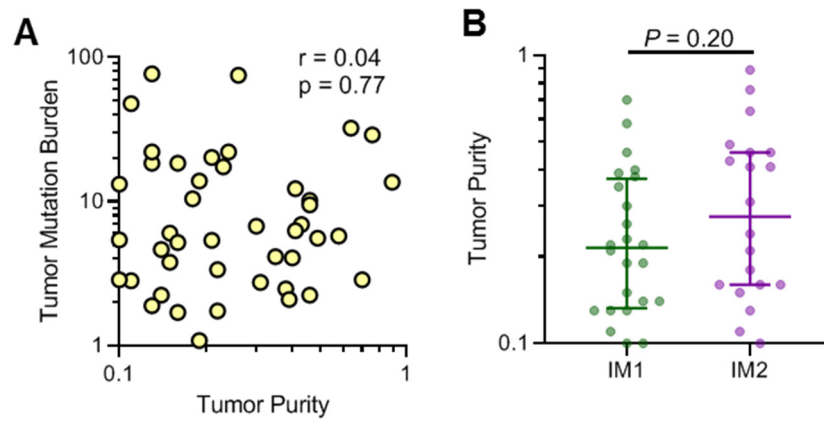
Fig. S8. Tumor purity does not influence TMB.

(A) Correlation between tumor purity determined from whole-exome sequencing and tumor mutation burden. Spearman rank correlation coefficient. **(B)** Tumor purity determined from whole-exome sequencing in IM1 and IM2. Rank-sum test.
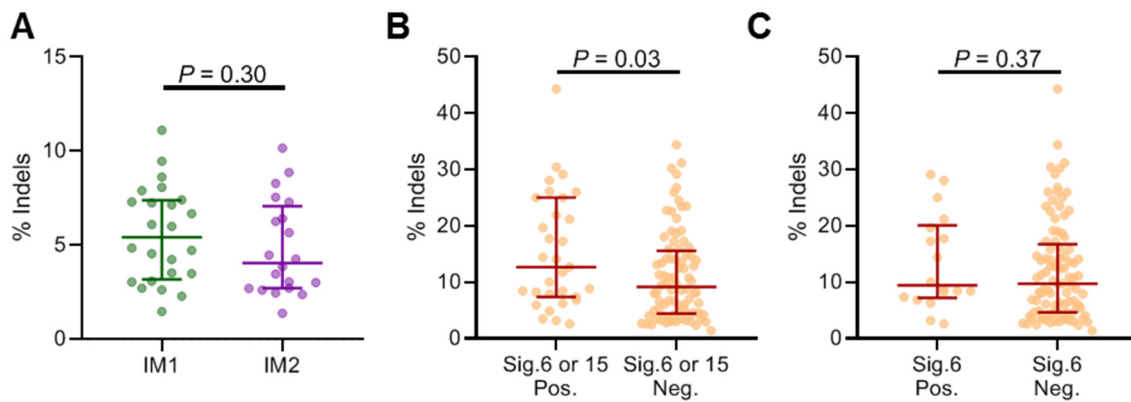
Fig. S9. Relationship between indels and dMMR mutational signatures.

(A) Percentage of mutations that are indels in IM1 and IM2. Rank-sum test. **(B)** Percentage of mutations that are indels in tumors negative or positive for mutational signature 6 or 15, both of which are associated with dMMR. Rank-sum test. **(C)** Percentage of mutations that are indels in tumors negative or positive for mutational signature 6 associated with dMMR. Rank-sum test.
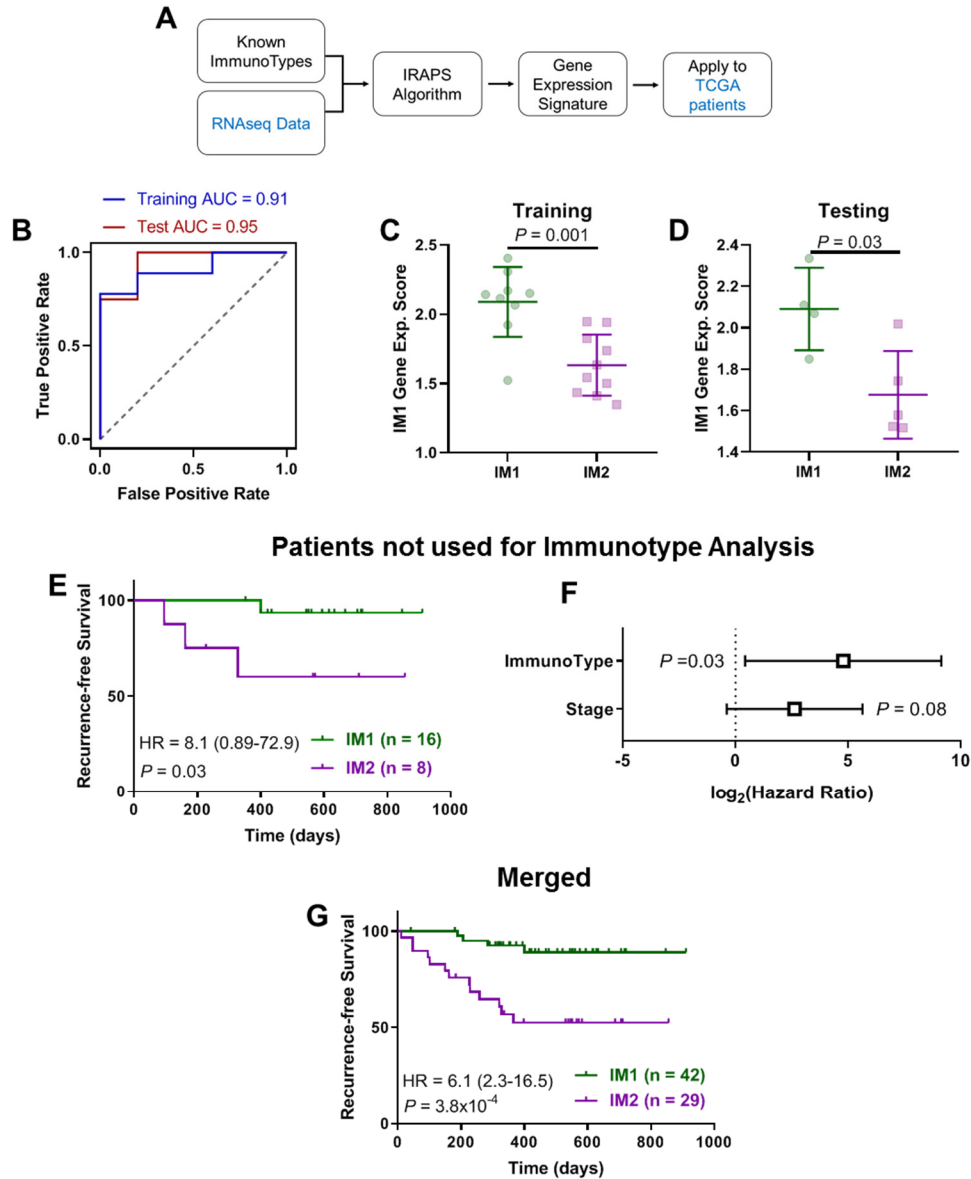
Fig. S10. Overview of IRAPS and immunotype signature validation.

(A) Schematic illustrating the IRAPS algorithm for identification of validation of immunotypes gene expression signature. **(B)** Receiver-operating characteristic (ROC) curves for the prediction of training sets with the flow data with area under the curve (AUC) values indicated. A ROC AUC value of 1 represents perfect prediction and 0.5 represents random chance. **(C-D)** IRAPS scores IMs in both training (n=19) and testing groups (n=9). **(E)** Kaplan-Meier curves showing association between IM1 (green, n=16), IM2 signature (purple, n=8) and RFS, with immunotype

determined by RNAseq in samples not used for flow cytometry. Log-rank test. **(F)** Multivariate survival analysis of samples in (E), controlling for tumor stage using Cox proportional hazards model. **(G)** Kaplan-Meier curve showing RFS in combination of patients in panel E and Fig. 4B. Log-rank test.