**Supplementary Methods**

**Immunohistochemistry staining**

Formalin fixed, paraffin embedded tumor blocks from the PROSPECT dataset were selected for immunohistochemistry (IHC) staining. Thirty-five tumor tissues with “mesenchymal” lung adenocarcinomas and 33 tumor tissues with “epithelial” lung adenocarcinomas, as defined by the mRNA expression, from the PROSPECT tissue bank were found to have sufficient quality material for additional analyses and were blinded to the pathologists who performed the IHC study. Four micron thick tissue sections were cut for IHC. The IHC staining was performed in an automated staining system (Leica Bond Max, Leica microsystems, Vista, CA, USA) using previously optimized IHC parameters. Leica Bond Retrieval Solution #1 was used in most antibody staining studies, except CD4, Granzyme B, and FOXP3, where Bond Retrieval Solution #2 was used. The antibodies used in this study included: PD-L1 (Clone #: E1L3N, Cell Signaling, 1:100 dilution), PD1 (Clone #: EPR4877, Abcam, 1:250 dilution), CD3 (Clone #: F7.2.38, Dako, 1:100 dilution), CD4 (Clone #: 4B12, Novocastra, 1:80 dilution), CD8 (Clone #: C8/144B, Thermo Scientific, 1:25 dilution), CD45RO (Clone #:UCHL1, Leica Biosystems, no dilution required), CD57 (Clone #: HNK-1, BD Biosciences, 1:40 dilution), CD68 (Clone #: PG-M1, Dako, 1:450 dilution), Granzyme B (Clone #: 11F1, Leica Biosystems, no dilution required) and FOXP3 (Clone #: 206D, BioLegend, 1:50 dilution). All antibodies were detected with the Leica Bond Polymer Refine detection kit (Leica Microsystems), including diaminobenzidine reaction to detect the antibody labeling and hematoxylin counterstaining. The stained slides were digitally scanned using the Aperio® ScanScope Turbo slide scanner (Leica Microsystems). The digital images were captured at ×200 magnification. The images were visualized using ImageScope™ software (Leica Microsystems,) and the digital image analysis was performed using Aperio Image Toolbox and Genie™ (Aperio, Leica Microsystems). Each case was analyzed in the whole section of tumor using the cell membrane staining algorithm to obtain the H-score. Non-tumor cells and cytoplasmic signals were excluded from this current study. Positive controls, including SignalSlide positive control (HDLM-2, Cell Signaling Technology, Danvers, Massachusetts), human placenta (formalin fixed, paraffin embedded) and human tonsil (formalin fixed, paraffin embedded) were used as positive controls for PD-L1 staining. SignalSlide negative control (#13747, Cell Signaling Technology, Danvers, Massachusetts) and cell line HEK293 (fetal kidney epithelial cell line) were used as negative controls for PD-L1 staining. Human tonsil was employed as a positive control for most of the other staining with internal negative control depending on the cell compartment. The PD-L1 scoring was only performed on the tumor cells, while inflammatory cells, including macrophages and lymphocytes present in the tumor associated stroma were excluded from this study using histology pattern recognition software (GENIE, Aperio, Leica Biosystems). The PD-1 staining was scored only on the inflammatory cells. All IHC markers were validated and optimized using positive and negative controls. The lung cancer specimens were stained using the same IHC conditions in an automated stainer (Leica Bond Max), simultaneously with positive and negative controls. For the immune profiling analysis, cell density was measured as the number of positive cells per mm square in the core of the tumor at the same region for each marker, overlapping the different regions using digital image analysis (Aperio, Leica Biosystems). On each whole tumor section, five random areas were selected by the pathologist to confirm the presence of viable tumor cells and avoid areas with necrosis. The final cell density corresponds to the average of the cell densities evaluated in the five areas. This method has been well-established in previous studies ([20](#_ENREF_20), [21](#_ENREF_21)). Immunohistochemistry was independently scored by two pathologists blinded to EMT status and clinical features (done by E.R.P and J.C.R).

**Reverse phase protein array (RPPA) analysis**

For RPPA analysis of TCGA and PROSPECT specimens, the tumors were lysed in RPPA lysis buffer [1% Triton X-100, 50mM HEPES (pH 7.4), 150mM NaCl, 1.5mM MgCl2, 1mM EGTA, 100mM NaF, 10mM NaPPi, 10% glycerol supplemented with fresh PMSF (1mM final concentration), Na3VO4 (1mM final concentration), protease (complete) and phosphatase (phosSTOP) inhibitor cocktail (Roche Applied Science), as we have previously published([22](#_ENREF_22)). Five serial dilutions of each protein lysate were printed on nitrocellulose-coated slides using an Aushon Biosystems 2470 arrayer (Burlington, MA) and stained sequentially with primary and secondary antibodies in an autostainer (BioGenex), prior to signal detection using a signal amplification system and DAB-based colorimetric reaction. MicroVigene Software (VigeneTech) as well as an in-house R package was used to assess spot intensities and the SuperCurve method was applied to estimate protein levels in each sample. For comparisons, data were log transformed (to the base of 2) and median-centered across antibodies to correct for protein loading. Rabbit polyclonal antibody to PD-L1 (Cat: ab174838, Abcam, MA) at 1:250 dilution was used to detect PD-L1 expression by RPPA. Differences in protein expression between “epithelial” and “mesenchymal” lung adenocarcinoma tumor samples were compared by t-test. Pearson correlation between CD274 (PD-L1) mRNA expression and PD-L1 protein expression by RPPA were assessed, along with E-cadherin protein RPPA and PD-L1 RPPA expression. All statistical analyses were performed using R packages (version 2.10.0).