SUPPLEMENT: METHODS

Proteomics

Tissue Processing: Lung tissue samples (100 mg) were washed in 1X PBS solution until clear of blood. Washed samples were placed in Eppendorf SafeLock tubes (Eppendorf North America, Hauppauge, NY) containing 75 µL 2.0 mm Zirconium Oxide beads, 50 µL 1.0 mm Zirconium Oxide beads (Next Advance, Averill Park, NY), 50 µL 0.5 mm Zirconium Oxide beads with 6X sample-volume lysis buffer consisting of 0.4 M TEAB, 7 M urea, 2 M thiourea, 20% methanol and 4 mM TCEP (Sigma-Aldrich, St. Louis, MO). The Bullet Blender Storm bead mill homogenizer (Next Advance, Averill Park, NY) was utilized to thoroughly homogenize the sample at 4^{o} C and maximum agitation for 10 minutes (Next Advance, Averill Park, NY). Thorough lysis membrane-bound protein release, and complete cellular membrane disruption was achieved using a Barocycler NEP2320 (Pressure Biosciences Inc., South Easton, MA). The samples underwent thirty cycles at 36^{o} C, 35,000 psi for 30 seconds and ~0 psi for 10 seconds in pressure cycling technology tubes (Pressure Biosciences Inc., South Easton, MA) with 150 µL sample buffer. Methyl methanethiosulfonate (MMTS, Sigma-Aldrich, St. Louis, MO) was added to achieve a final concentration of 8 mM. The samples were transferred to a 1.6 mL microfuge Protein LoBind tube (Eppendorf North America, Hauppauge, NY) and incubated 15 minutes at room temperature. The samples were microfuged at 4^{o} C, 12,000 x g for 10 minutes and the supernatant transferred to a new Protein LoBind microfuge tube (Eppendorf North America, Hauppauge, NY).

Enzymatic (Tryptic) Digestion of Protein: Total protein concentrations were determined using the Bradford assay (Bio-Rad, Hercules, CA). A 60 µg aliquot of each sample was added to a 1.6 mL tube and all samples were brought to equal volume with sample lysis buffer without TCEP. Samples were diluted four-fold with mass spec grade water (Fisher Scientific W5SK-1, Pittsburgh, PA). Trypsin Gold (Promega, Madison, WI) was added at a trypsin to total protein ratio of 1:30. Samples were incubated for 16 hours at 37^{o} C, frozen at -80^{o} C for 0.5 hours and dried in a speed vacuum centrifuge (Thermo Scientific, Waltham, MA).

Peptide Purification: Dried peptides were resuspended in 1 mL of reverse phase solvent [98% water, 2% acetonitrile, 0.1% Trifluoroacetic acid (TFA)] (Sigma-Aldrich, St. Louis, MO). Resprep 3 cc C18 cartridges (Restek, Bellefonte, PA) were conditioned with 1 mL 80% acetonitrile, 0.1% TFA, (Sigma-Aldrich, St. Louis, MO) followed by 3 mL of washing solvent. Samples were then added to the cartridge with a flow rate of approximately one drop per second (~1.7psi) followed by 3 mL of washing solvent. Peptides were eluted in 1 mL of 70% acetonitrile, 0.1% TFA (Sigma-Aldrich, St. Louis, MO) and vacuum dried. Samples were then resuspended in 0.5 M TEAB (2 µg/µL) and a 1.5 µL aliquot was analysed on a Linear Trap Quadrupole (LTQ) mass spectrometer (Thermo Scientific, Waltham, MA) to check for proper peptide digestion and sample integrity.

Quantitative Mass Spectrometry (iTRAQ): Labelling and Detection: A 20 µg aliquot of each processed sample was labelled with iTRAQ 8-plex reagents according to the manufacturer's protocol (ABSciex, Framingham, MA). Following the labelling reaction, all samples and controls were multiplexed, vacuum dried and purified using a Resprep 3 cc MCX cartridge (Restek, Bellefonte, PA).

Protein Fractionation and Data Acquisition: Each sample was resuspended in Buffer A (10mM ammonium formate, pH 10 in 98:2 water:acetonitrile) and fractionated offline by high pH C18 reversed-phase (RP) chromatography followed by fraction concatenation for 2D proteomic analysis. A MAGIC 2002 HPLC (Michrom BioResources, Inc., Auburn, CA) was used with a C18 Gemini-NX column, 150 mm x 2 mm internal diameter, 5 µm particle, 110 Å pore size (Phenomenex, Torrence, CA). The flow rate was 150 µL/minute with a gradient from 0-35% Buffer B (10 mM ammonium formate, pH 10 in 10:90 water:acetonitrile) over 60 minutes, followed by 35-60% over 5 minutes. Fractions were collected every 2 minutes and UV absorbance was monitored at 215 and 280 nm. Peptide containing fractions were divided into two equal numbered groups, "early" and "late". The first "early" fraction was concatenated with the first "late" fraction, and so on. Concatenated samples were vacuum dried and resuspended in load solvent (98:2:0.01, water:acetonitrile:formic acid). Digested peptide mixtures were desalted with C18 resin according to the stop and go procedure. Aliquots of 1-1.5 µg of total peptide were dissolved in 5.5 µL of load solvent A (98:2:0.01, water:acetonitrile:formic acid) and loaded directly

onto a 12 cm X 75-µm internal diameter fused silica pulled-tip (New Objective Woburn, MA) capillary column packed in-house with MagicC18AQ resin (5 µm, 200 Å pore size; Michrom BioResources Auburn, CA) with load solvent at a flow rate of 800 nL/min using an Eksigent 1D+LC nanoflow system (Dublin, CA) and a MicroAS autosampler. Peptides were eluted using a gradient of 10–40% B Solvent over 55 at 320 nL/min. The column was mounted in a nanospray source directly in line with a Velos Orbitrap mass spectrometer (Thermo Scientific Inc., Waltham, MA).

iTRAQ Statistical Analysis: Raw files obtained directly from the Orbitrap Velos XL Mass Spectrometer were imported into GalaxyP where raw files were converted to mzML format using msconvert and then into .mgf files using MGF formatter. Protein Pilot 4.5 search was performed with a target-decoy version database generated from the Human UniProt database (12/1/2016) and ABSciex contaminant database. False discovery rate (FDR) analysis employed the Proteomics System Performance Evaluation Pipeline Software within the ProteinPilot suite of software.

Transcriptomics and Translatomics:

RNA isolation and polyribosome profiling: Frozen lung tissue samples ranging in weight from 25-60 mg were individually ground to a fine powder using a liquid nitrogen cooled ceramic mortar and pestle. For RNA quality control (QC), approximately 20% of each sample was processed with TRI Reagent (Sigma Aldrich, St Louis, MO). Once the initial RNA QC was completed, polyribosome

preparations were performed in random sample pairs.[35] For this purpose, the remainder of each frozen powdered sample was solubilized in 100 µL of lysis buffer (10 mM Tris-HCl at pH 8.0, 150 mM NaCl, 5 mM MgCl2, 1% Nonidet-P40, 1% sodium deoxycholate, 40 mM dithiothreitol, 500 U/mL RNAsin (Promega, Madison, WI), 40 mM VRC (vanadyl ribonucleoside complex), and 150 µg/ml cycloheximide) and mixed by pipetting. Nuclei and insoluble material were removed by centrifugation (12,000 x g, 10 seconds, at 4°C). Extraction buffer (50 µL; 0.2 M Tris-HCl at pH 7.5, 0.3 M NaCl, 10 mM PMSF) was added to the supernatant and the sample was mixed by pipetting. The sample was centrifuged (12,000 x g, 5 min, at 4° C) and 25% of the supernatant was processed with TRI Reagent to generate a cytosolic total RNA sample. The remaining supernatant was layered onto a 5 mL, linear sucrose gradient (0.5–1.5 M), 10 mM Tris-HCl pH 7.5, 140 mM NaCl, 3 mM $MgCl_2$, 10 mM DTT. Gradients were centrifuged (200,000 x g in a Beckman SW55Ti rotor for 80 minutes at 4^o C) and the gradients were fractionated into ten, 0.5 mL fractions utilizing an ISCO density gradient fractionator with absorbance monitored at 254 nm. Each fraction was collected into a tube containing 50 µL 1%SDS and 20 µL 0.5 M EDTA. The polyribosomal mRNA from fractions 7, 8, 9 and 10 were individually processed with 1 mL of TRI Reagent and 200 µL chloroform and the resultant RNA pellets from these four fractions were combined to create the polyribosomal mRNA sample corresponding to each tissue sample. Polyribosomal mRNA and cytosolic RNA were sent to the University of Minnesota Genomics Core for quality assessment by Caliper High Sensitivity Lab Chip. Those samples passing the

Caliper QC were processed into indexed library samples for analysis by RNA sequencing. This was accomplished using the SMART-Seq protocol.

RNA-Seq analysis: Sequencing was performed on a 51SR Dual indexed run on the HiSeq 2000 at the University of Minnesota Genomics Center.

Quality Assessment: Libraries were quantified using a fluorimetric PicoGreen assay. Library size was assessed using capillary electrophoresis on an Agilent BioAnalyzer 2100 and only libraries quantifying higher than 10 nM were analysed. Libraries were pooled and assessed using Kapa qPCR to ensure adapters were ligated to DNA fragments and functional.

Cluster generation and sequencing: Nextera libraries were hybridized to a single read flow cell and individual fragments were clonally amplified by bridge amplification on the Illumina cBot. Once clustering was complete, the flow cell was loaded on a HiSeq 2000 and sequenced using Illumina SBS chemistry. Upon completion of read 1, an 8-base pair index read for Index 1 was performed. The Index 1 product was removed and the template was re-annealed to the flow cell surface. The run proceeded with 8 chemistry-only cycles, followed by an 8-base pair index read to read Index 2.

Primary analysis and de-multiplexing: Base call (.bcl) files for each cycle of sequencing were generated by Illumina Real Time Analysis (RTA) software. The base call files and run folders were exported to servers at the University of Minnesota Supercomputing Institute. Primary analysis and de-multiplexing were performed using Illumina CASAVA software 1.8.2. Quality control checks on raw

sequence data for each sample were performed with FastQC. Read mapping was performed via Bowtie (v2.2.4.0) using the UCSC human genome (hg19) as reference. Gene quantification was done via Cuffquant for FPKM values and Feature Counts for raw read counts. 50bp FastQ paired-end reads (n= ~ 50Million per sample) were trimmed using Trimmomatic (v 0.33) enabled with the optional "-q" option; 3bp sliding-window trimming from 3' end requiring minimum Q30.

Statistical Analysis

Data Pre-processing

We considered all genes from the transcriptome and translatome data with at least thirty-three percent non-zero entries. The read counts for transcriptome and translatome were $log(1+X_{ij})$ transformed for all subsequent analyses. We considered all proteomic variables identified by mass spectrometry with at least eighty percent non-missing entries. The remaining missing values in the proteomic data were imputed by singular value decomposition with the R package *SpatioTemporal.*[14] We adjusted the proteomic data for systematic effects within iTraq runs using the *ComBat* Non-Parametric Empirical Bayes adjustment.[15]

Independent Screening

Within each data-type, we assessed the relationships between lung cancer status separately for each gene or protein to identify powerful univariate biomarkers. We retained all of the 96 samples in the proteomic, and 33 samples in the transcriptome and translatome data, respectively, for these independent

screening analyses. We consider each pair-wise comparison between the three sample types. For the class comparisons between *tumour* and *control,* and *adjacent* and *control,* we used 2-sample t-tests. For the comparison between *tumour* and *adjacent* we used paired t-tests, because tumour and adjacent samples were collected from the same individuals. To adjust for multiple comparisons we applied a Benjimani-Hochberg *false-discovery rate* (FDR) adjustment, [16] and we consider those genes or proteins with FDR<0.1.

Pathway Analysis

The genes or proteins available after pre-processing for each data-type were used as the reference list, and those genes or proteins with a p-value less than 0.05 from the two-sample or paired t-tests used for independent screening were used for enrichment analysis. We performed three enrichment analyses for each data-type: one for genes that distinguish tumour *adjacent* samples from *control* samples, another for genes that distinguish *tumour* samples from *control* samples, and another for genes that distinguish *adjacent* from *control* samples.

We also performed a pathway analysis as described above using the meta-loadings from the JIVE multinomial logistic regression analysis, to determine which biological functions were most overrepresented in the fitted model. (Figure S1) Those genes that had an absolute meta-loading greater than or equal to the mean of the absolute meta-loading for a given comparison were used as the active gene set for each data-type.

	Log Odds (Tumour/ Control)	**Std. Error**	**Wald P-value**	**Log Odds (Adjacent/ Control)**	**Std. Error**	**Wald P-Value**
(Intercept)	0.23	0.99	0.82	-0.11	1.00	0.92
Joint - 1	5981.75	1755.18	**<0.01**	-415.45	1856.98	0.82
Joint - 2	-1463.66	1954.76	0.45	5044.09	1728.28	**<0.01**
Transcriptome -1	-3538.76	1797.90	**0.04**	2981.02	1159.36	**0.01**
Transcriptome - 2	1070.37	1168.52	0.36	-666.68	2829.41	0.81
Transcriptome - 3	-1408.66	722.75	0.05	1427.61	1639.51	0.38
Translatome - 1	27.20	3234.18	0.99	-3919.07	3369.21	0.24
Translatome - 2	2512.90	1498.80	0.09	5441.82	3008.44	0.07
Translatome -3	-252.07	221.20	0.25	-600.03	346.16	0.08

Table S1: Multinomial logistic regression coefficients

Data	Comparison	Pathway	# of Genes	FDR
Proteome	Tumour vs. Adjacent	Focal Adhesion	27	9.07E-04
Proteome	Tumour vs. Adjacent	ECM-receptor interaction	10	5.27E-02
Proteome	Tumour vs. Adjacent	Toxoplasmosis	10	5.27E-02
Proteome	Tumour vs. Adjacent	Amoebiasis	14	8.2E-02
Proteome	Tumour vs. Adjacent	Regulation of Actin Cytoskeleton	16	8.2E-02
Proteome	Tumour vs. Adjacent	Small Cell Lung Cancer	8	9.59E-02
Proteome	Tumour vs. Control	Focal Adhesion	25	1.51E-02
Proteome	Tumour vs. Control	Amoebiasis	15	1.51E-02
Proteome	Tumour vs. Control	ECM-receptor interaction	10	2.82E-02
Proteome	Tumour vs. Control	Toxoplasmosis	10	2.82E-02
Proteome	Tumour vs. Control	Small Cell Lung Cancer	8	8.79E-02
Translatome	Adjacent vs. Control	ECM-receptor interaction	10	2.72E-03

Table S2: Significantly overrepresented pathways (FDR<0.1), using a p-value threshold of p<0.05 under the t-test.