



Multi-omic molecular profiling of lung cancer in COPD

Brian J. Sandri^{1,7}, Adam Kaplan^{2,7}, Shane W. Hodgson³, Mark Peterson¹, Svetlana Avdulov¹, LeeAnn Higgins⁴, Todd Markowski⁴, Ping Yang⁵, Andrew H. Limper⁶, Timothy J. Griffin⁴, Peter Bitterman¹, Eric F. Lock² and Chris H. Wendt^{1,3}

Affiliations: ¹Division of Pulmonary, Allergy, Critical Care, and Sleep Medicine, Dept of Medicine, University of Minnesota Medical School, Minneapolis, MN, USA. ²Division of Biostatistics, School of Public Health, University of Minnesota, Minneapolis, MN, USA. ³Pulmonary, Allergy, Critical Care, and Sleep Medicine, Veterans Affairs Medical Center, Minneapolis, MN, USA. ⁴Dept of Biochemistry, Molecular Biology and Biophysics, University of Minnesota, Minneapolis, MN, USA. ⁵Division of Epidemiology, Mayo Clinic, Rochester, MN, USA. ⁶Division of Pulmonary and Critical Care Medicine, Mayo Clinic, Rochester, MN, USA. ⁷Both authors contributed equally.

Correspondence: Chris H. Wendt, 1 Veterans Drive, Minneapolis, MN 55417, USA. E-mail: wendt005@umn.edu

 @ERSpublications

A multi-omic approach identified gene expression programmes distinguishing COPD lung stroma associated with lung cancer <http://ow.ly/j1Da30jWH1X>

Cite this article as: Sandri BJ, Kaplan A, Hodgson SW, *et al.* Multi-omic molecular profiling of lung cancer in COPD. *Eur Respir J* 2018; 52: 1702665 [<https://doi.org/10.1183/13993003.02665-2017>].

ABSTRACT Chronic obstructive pulmonary disease (COPD) is a known risk factor for developing lung cancer but the underlying mechanisms remain unknown. We hypothesise that the COPD stroma contains molecular mechanisms supporting tumourigenesis.

We conducted an unbiased multi-omic analysis to identify gene expression patterns that distinguish COPD stroma in patients with or without lung cancer. We obtained lung tissue from patients with COPD and lung cancer (tumour and adjacent non-malignant tissue) and those with COPD without lung cancer for profiling of proteomic and mRNA (both cytoplasmic and polyribosomal). We used the Joint and Individual Variation Explained (JIVE) method to integrate and analyse across the three datasets.

JIVE identified eight latent patterns that robustly distinguished and separated the three groups of tissue samples (tumour, adjacent and control). Predictive variables that associated with the tumour, compared to adjacent stroma, were mainly represented in the transcriptomic data, whereas predictive variables associated with adjacent tissue, compared to controls, were represented at the translational level. Pathway analysis revealed extracellular matrix and phosphatidylinositol-4,5-bisphosphate 3-kinase–protein kinase B signalling pathways as important signals in the tumour adjacent stroma.

The multi-omic approach distinguishes tumour adjacent stroma in lung cancer and reveals two stromal expression patterns associated with cancer.

This article has supplementary material available from erj.ersjournals.com

Received: Dec 20 2017 | Accepted after revision: May 06 2018

The content of this work is not subject to copyright. Design and branding are copyright ©ERS 2018

Introduction

Independent of smoking, chronic obstructive pulmonary disease (COPD) is a known risk factor for the development of lung cancer [1, 2]. This paradigm is further supported by the observation that tumours are more likely to arise in association with areas of emphysema [3, 4]. Central to the pathophysiology of airflow obstruction in COPD is remodelling of lung extracellular matrix (ECM) with loss of elasticity and increased resistance to airflow. This ECM remodelling is accompanied by small airway fibrosis and inflammation, a stromal milieu with the potential to promote tumourigenesis of epithelial cells that have previously incurred oncogenic mutations from exposure to tobacco or other oncogenic stressors.

It is well recognised that cancer initiation occurs as a progression of genetic mutations. However, it is becoming more apparent that tumours are integrally connected with their microenvironment [5, 6]. This microenvironment consists of ECM along with a stroma consisting of fibroblasts, endothelial cells and resident immune cells that can promote tumour progression. These stromal cells do not acquire genetic mutations, nevertheless, it is well accepted that cancer-associated fibroblasts, the primary stromal cell that produces ECM, are a major contributor in tumour progression. Recent evidence also places the fibroblast and ECM at the centre of tumour initiation. Senescent human fibroblasts have been shown to stimulate both malignant and premalignant epithelial cells to proliferate while “normal” epithelial cells remain unaffected [7]. In a murine model of breast cancer, fibroblasts devoid of the tumour suppressor gene *PTEN* dramatically accelerate cancer initiation, in a manner that is partially dependent on fibroblast secretory factors [8]. Our study sought to identify the specific alterations in the COPD-stroma that promote lung carcinogenesis, as these remain unknown.

The mechanisms creating a pro-malignant stromal environment are likely complex. Although proteins are the effector molecules, their regulation can occur at multiple levels including transcription, translation and post-translational modification. Early on, transcriptional regulation was the major focus of study; however, translation regulation has emerged as a significant mechanism in regards to carcinogenesis [9–11]. Therefore, genome-wide approaches are necessary to understand the complexity of carcinogenesis. In this regard, we sought to understand the role of the COPD lung stroma in lung cancer using a multi-omic approach.

Materials and methods

Study population and lung tissue samples

The University of Minnesota Institutional Review Board Human Subjects Committee approved this study. Surgical lung tissue samples were obtained from the Mayo Clinic Lung Specimen Registry from individuals with COPD who underwent surgical resection for Stage I or II adenocarcinoma (table 1). Tissue samples consisted of tumour and adjacent lung from 32 patients with adenocarcinoma and COPD. Controls were obtained from the National Institutes of Health Lung Tissue Research Consortium (NIH LTRC) and consisted of lung tissue from 32 individuals with COPD without lung cancer (table 1). All samples were obtained using identical procedures as outlined by the NIH LTRC. All subjects, cases and controls were current or former smokers. These lung tissue biopsies are referred to as “tumour, adjacent and control”. In total there were 32 tissue samples with shared transcriptome, translome and proteomic data (10 controls and 11 cases with one tumour and one adjacent tissue each; table 2). The number of samples that are available for each data type are shown in table 3.

Proteomics and RNA sequencing

Lung tissue was processed for isobaric tag for relative and absolute quantitation (iTRAQ) labelling following manufacturers’ recommendations and as previously published [12]. Lung tissue was processed for total and polysomal mRNA isolation and sequencing as previously described [13]. Detailed experimental design information is available in the supplementary material. Protein mass spectrometry data is available at the PRoteomics IDentifications (PRIDE) database (www.ebi.ac.uk/pride/archive/; accession number PXD007919). RNA sequencing data is available at the Gene Expression Omnibus (GEO) data repository (www.ncbi.nlm.nih.gov/geo/; accession number GSE106899).

TABLE 1 Patient demographics for proteomic data

Sample type	Age years	Males/females	FEV ₁ L	FEV ₁ % predicted	FVC L	FEV ₁ /FVC
Lung cancer (n=32)	68±5 (59–79)	14/18	1.70±0.74 (0.71–3.60)	53±19 (31–99)	2.97±1.07 (1.14–5.46)	56±11 (39–71)
Control (n=32)	56±7 (45–67)	18/14	0.56±0.19 (0.3–1.1)	18±4 (12–28)	2.03±0.68 (0.7–3.9)	29±9 (19–57)

Data are presented as n/n or mean±SD (range). FEV₁: forced expiratory volume in 1 s; FVC: forced vital capacity.

TABLE 2 Patient demographics for shared transcriptomic, translomic and proteomic data

Sample type	Age years	Males/females	FEV ₁ L	FEV ₁ % predicted	FVC L	FEV ₁ /FVC
Lung cancer (n=11)	64±4 (62–75)	3/8	1.61±0.84 (0.71–2.18)	50±16 (31–81)	2.77±1.08 (1.31–5.46)	53±12 (39–71)
Control (n=10)	53±6 (45–64)	7/3	0.51±0.14 (0.4–0.8)	15±4 (13–27)	2.14±0.82 (1.2–3.9)	24±6 (18–36)

Data are presented as n/n or mean±SD (range). FEV₁: forced expiratory volume in 1 s; FVC: forced vital capacity.

Statistical analysis

All data analysis was performed using the R Environment for Statistical Computing (www.r-project.org/; R version 3.3.2). Prior to data integration, data underwent pre-processing and independent screening as described in the supplementary material [14–16]. After pre-processing there were 1486 proteomic, 5989 transcriptome and 8014 translome genes or proteins available. Next, we determined an integrative composite biomarker for sample status (tumour, adjacent and control) from the three different omic data types (proteomic, transcriptome and translome) using a multi-step approach described previously [17].

For each data type, we first conducted a univariate linear modelling step for each gene or protein to filter out variables with little predictive power. After filtering, we performed an integrative dimension reduction of the remaining variables using the Joint and Individual Variation Explained (JIVE) method [18, 19]. JIVE is an extension of principal components analysis for multi-omic data and it identifies latent structures shared among and expressed within multiple high-throughput data types. “Joint components” define sample patterns that explain substantial variability across multiple data types and “individual components” are sample patterns that explain substantial variability in one data type but not others [17].

After performing JIVE we used the lung-cancer sample types (tumour, adjacent and control) as our outcomes in a multinomial logistic regression using the JIVE components as predictors. The components were either included or excluded from the model by a stepwise (forward/backward) selection algorithm to minimise the Akaike Information Criteria (AIC) [20]. This gave a composite model to predict the odds of a tissue sample being cancerous, tumour-adjacent, or non-cancerous COPD from all three omics datasets.

To assess the accuracy of the fitted model we performed N-fold (leave one out) cross-validation where a single tissue sample was removed as a test case, the model parameters were estimated using the remaining (n=31) samples as a training set and the prediction of the resulting model was assessed using the test case. This procedure was repeated for each sample as the test case.

To determine which genes or proteins were most influential, the JIVE components were mapped back to the original variable space to give meta-loadings for each variable [17]. These could be interpreted as giving the linear weight (*i.e.* coefficient) of each variable in the fitted model for each comparison. We performed a pathway analysis using the meta-loadings to determine which biological functions were most over-represented in the fitted model. Those genes which 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. The WEB-based GENE SeT AnaLysis Toolkit (WebGestalt; www.webgestalt.org/) [21] was used to perform an over-representation enrichment analysis using KEGG pathways as described in the supplementary material.

TABLE 3 Sample availability by data type and the overlap of all three data types

Availability	Data type			Overlap
	Proteome	Transcriptome	Translatome	
Sample type				
Tumour	32	11	11	11
Matched adjacent	32	11	11	11
Control	32	11	11	10
Total samples	96	33	33	32
Total patients	64	22	22	21

Results

Samples and data acquisition

We compared lung tumour and non-cancerous adjacent lung tissue from patients with COPD to tissue from patients with COPD but without lung cancer (figure 1). Proteomic analysis was available for 32 patients with COPD and lung cancer and 32 controls. Lung function was more severe in the non-cancerous control group with an average % predicted forced expiratory volume in 1 s (FEV1) of 18% compared to 53% in those with lung cancer (table 1). Proteomic analysis consisted of iTRAQ-based quantitative mass spectrometry across 16 runs, obtaining an average of 27 432 spectra matched to 12 457 distinct peptides and 2105 inferred proteins at a 1% global false discovery rate (FDR). Twenty-one subjects (11 cases and 10 controls) had samples adequate for both proteomic analysis and mRNA sequencing, both total and polyribosomal (table 2), with an average of 24 484 971 reads and a median read quality of 38. This led to a read mapping percentage of 52.6, identifying an average of 5602 genes per sample from mRNA sequencing of total mRNA (transcriptome) and polysome-associated mRNA (transcripts bound by more than three ribosomes). The latter identifies efficiently translated mRNA, hereafter referred to as the translatoome. The lower mapping was attributed to the relatively small sample mass requiring additional amplification.

Independent screening

Under univariate screening, the translatoome data did not yield significant independent biomarker associations with cancer status under the FDR threshold of 0.1. The transcriptome data attained five significant independent biomarker associations with cancer status and the proteomic data attained 1116 significant independent biomarker associations. The larger number of associations detected for the proteomic data is due in part to its much larger sample size ($n=96$ for proteome *versus* $n=33$ for mRNA), which yields more statistical power to detect independent associations under the rigorous multiple testing criteria used here.

Venn diagrams demonstrate the number of significant proteins (FDR <0.1) for each comparison (figure 2). Within the proteomic independent association results, the majority of biomarker associations were

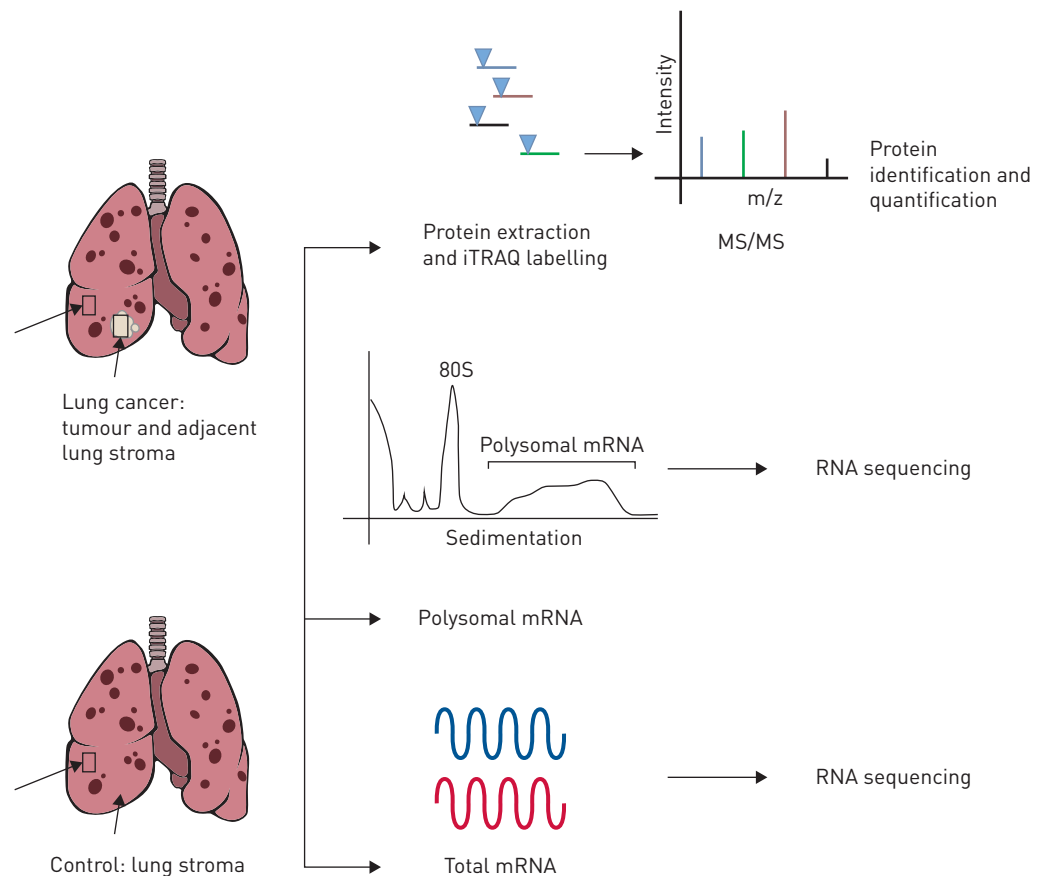


FIGURE 1 The multi-omic profiling approach for lung tumour, adjacent non-cancerous lung tissue and lung tissue from chronic obstructive pulmonary disease controls. iTRAQ: isobaric tag for relative and absolute quantitation; MS/MS: tandem mass spectrometry.

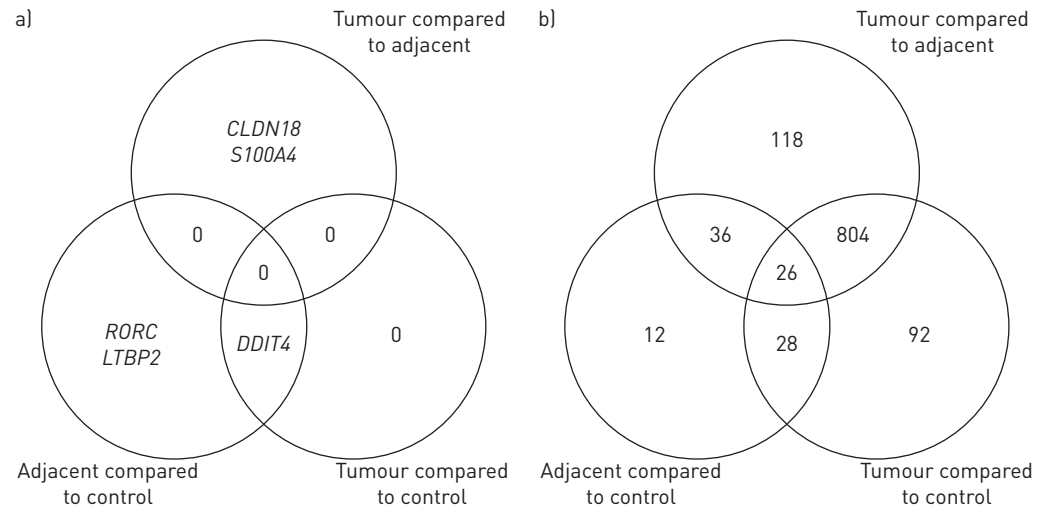


FIGURE 2 Venn diagrams for a) shared transcripts and b) proteins between sample comparisons using pairwise t-tests (false discovery rate (FDR) <0.1).

significant for comparing tumour samples to paired adjacent samples and also for comparing tumour samples to controls (figure 2b). A smaller number of proteins distinguished the adjacent samples from controls. For the transcriptomic data, one gene (*DDIT4*) was a shared significant independent biomarker between the cancer status associations of tumour and control as well as adjacent and control (figure 2a). Using a more stringent threshold, FDR <0.05, no transcriptomic associations remained and 314 total proteomic associations remain significant.

For the transcriptome data, 35 genes had a significant correlation with % predicted FEV₁ within controls, while for the translome data 18 genes had a significant correlation with % predicted FEV₁ within controls (FDR <0.1). These genes had no overlap with those genes that correlate with cancer status in figure 2. No other significant associations were detected with % predicted FEV₁ or age for the transcriptome, translome or proteomic data, either within controls or within cancer patients.

Data integration and predictive modelling identifies eight latent data structures

To integrate these three data sets into a useful predictive model we used the JIVE algorithm for multi-source data [18, 19]. JIVE is an extension of principal components analysis for multi-omic data and identifies latent structures (*i.e.* components) shared among and expressed within multiple high-throughput data types. Joint components define sample patterns that explain substantial variability across multiple data types and individual components are sample patterns that explain substantial variability in one data type but not others [17].

After filtering by univariate multinomial logistic regression we had 612, 1108 and 574 genes for the proteomic, transcriptome and translome data, respectively. The number of proteins or genes at each stage of filtering and the overlap between transcriptome and translome at each stage is given in figure 3. Permutation testing identified two joint JIVE components, zero significant additional individual components for proteomic data, three individual components for the transcriptome and three individual components for the translome for a total of eight underlying latent patterns. The three sample groups (tumour, adjacent and control) are distinguished by the two joint components, as seen in the top-left panel of figure 4.

Next, we fit a multinomial logistic regression model to predict sample group using the eight identified JIVE components as predictors. That is, the log odds ratio that a sample belongs to a given group is modelled by a weighted linear combination of the two joint components, three individual components for the transcriptome and three individual components for the translome. The predicted group probabilities for each sample under N-fold cross-validation are shown in figure 5. The model robustly distinguishes all three sample groups. Using a class prediction probability of 0.5 as a threshold, just three samples were misclassified (90.6% accuracy): a tumour sample was classified as tumour adjacent (10 out of 11 correctly classified), an adjacent sample was misclassified as a control (10 out of 11 correctly classified) and a control sample was misclassified as a tumour (9 out of 10 correctly classified). The log predicted probabilities did not have a significant correlation with % predicted FEV₁ or age within control patients or within cancer patients ($p > 0.05$).

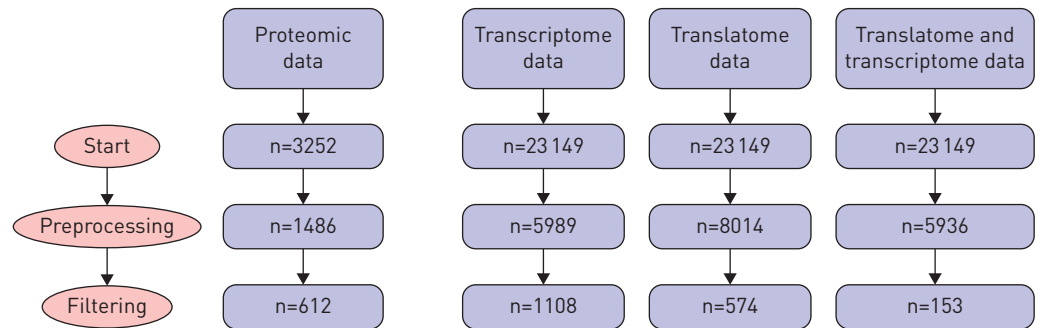


FIGURE 3 Diagram of the number of genes or proteins at different stages of preprocessing and filtering prior to Joint and Individual Variation Explained (JIVE) analysis. Preprocessing details are available in the supplementary material. The right-hand column gives the number of genes shared by the transcriptome and translatome data after each step.

Multinomial logistic regression model (supplementary figure S1) coefficients for the JIVE scores are summarised in supplementary table S1. The resulting gene and protein meta-loadings are shown for each comparison (tumour *versus* control, adjacent *versus* control and tumour *versus* adjacent) in figure 6. These plots show meta-loading against variable (gene or protein) index, revealing the size and direction of each variable's contribution to the predictive model for each comparison. It is apparent that predictive variables for tumours compared to control tissue were evenly distributed across proteomic, transcriptomic and translatomic data. Predictive variables of COPD stroma associated with cancer *versus* adjacent COPD stroma not associated with cancer were primarily found in translatomic data (polyribosomal-associated mRNA), while predictive variables of tumour *versus* adjacent were primarily found in transcriptomic data (total mRNA). This pattern suggests that lung tissue adjacent to lung cancers, but not containing lung cancer, display unique signatures that manifest at different levels of regulation. These datasets indicate that a thorough interrogation of not only the proteome, but also the transcriptome and translatome, provide molecular insight that would otherwise be discarded.

KEGG pathway analysis

Pathway enrichment analysis of the meta-loadings obtained from the JIVE predictive analysis revealed several significant pathways for each comparison across the three data types (table 4). For the proteome data, six pathways were significant for the tumour *versus* adjacent comparison and five were significant for the tumour *versus* control comparison. However, for the translatome data, six pathways were significant for the adjacent *versus* control comparison. For all other analyses all pathways were insignificant after correction for multiple comparisons (FDR >0.1). These findings are consistent with the translatome being most relevant to the distinction of adjacent *versus* control samples, while the proteomic data is more relevant for the other two comparisons. However, several pathways are shared among the different comparisons and data types. ECM-receptor interaction, focal adhesion and phosphatidylinositol-4,5-bisphosphate 3-kinase–protein kinase B (PI3K–Akt) signalling were highly significant for each of the three analyses that yielded significant results.

An analogous pathway analysis based on univariate t-statistics rather than JIVE meta-loadings for each dataset and comparison was also performed. The pathways achieving significance were again found only for the three analyses above (tumour *versus* adjacent for proteomic data, tumour *versus* control for proteomic data and adjacent *versus* control for translatome data). The results closely resembled those in table 3 (see supplementary table S2).

Discussion

The relationship between COPD and lung cancer was described over three decades ago [1, 2]; however, the molecular underpinnings of this observation remain elusive. In this study, we utilised a multi-omic approach on lung tissue that included tumour, adjacent non-malignant lung tissue and COPD lung tissue in subjects without lung cancer to uncover pre-malignant signals in lung stroma. This multi-omic approach included quantitative mass spectrometry proteomics along with RNA sequencing (RNA-Seq) on total cytoplasmic mRNA (transcriptomics) and polysomal-associated mRNA (translatomics). To integrate our multi-omic datasets, we utilised the joint and individual variation explained (JIVE) method [18, 19] that simultaneously models components that are shared across all datasets. From these we constructed a multi-omic predictive model that gave statistically robust separation of the three tissue types (*i.e.* tumour, non-tumour adjacent tissue and control COPD tissue).

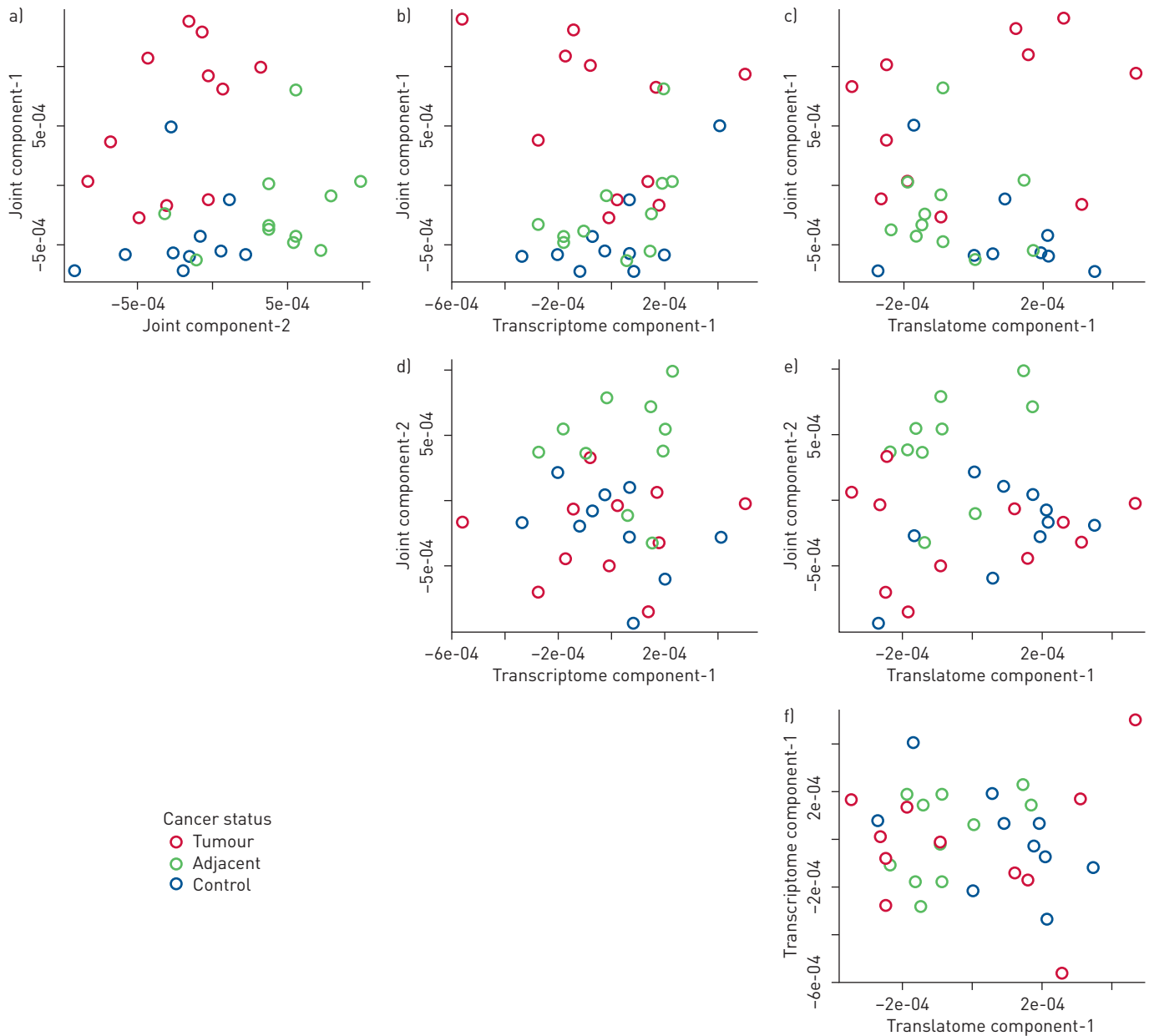


FIGURE 4 Scatterplots of joint and individual variation explained (JIVE) scores for two joint components and the first individual components for the transcriptome data and translome data. Each point represents one lung tissue sample and colors give cancer status. The first joint component shared by the three data types (a-c) shows a separation between non-cancer (control and adjacent) and cancer samples. The second joint component (d, e) shows a distinction between the adjacent samples and the rest of the samples. The individual structures (transcriptome and translome, respectively) did not show strong differentiation of cancer status (f).

Among our datasets the proteomic dataset had the largest sample size due to limitations in the original sample mass. We observed the largest differences in expression when comparing tumour and non-tumour tissues, either adjacent tissue or that from COPD controls. This is not surprising considering cancers replace normal lung structure creating a unique molecular background. Interestingly, we found significant changes between stroma proteome from tissue adjacent to the lung cancer compared to control tissue, *i.e.* that not associated with lung cancer.

To explore and integrate the three multi-omic datasets we utilised the JIVE algorithm after filtering by univariate multinomial logistic regression. The JIVE algorithm reduced the large dataset variables to eight latent patterns. The three sample groups (tumour, adjacent and control) were illustrated most by two joint components. Six additional individual components or patterns were identified for the transcriptome and translome (three each). Using these components, we generated a model to predict sample group identity, which robustly distinguished and separated all three sample groups. Next, we determined the size and

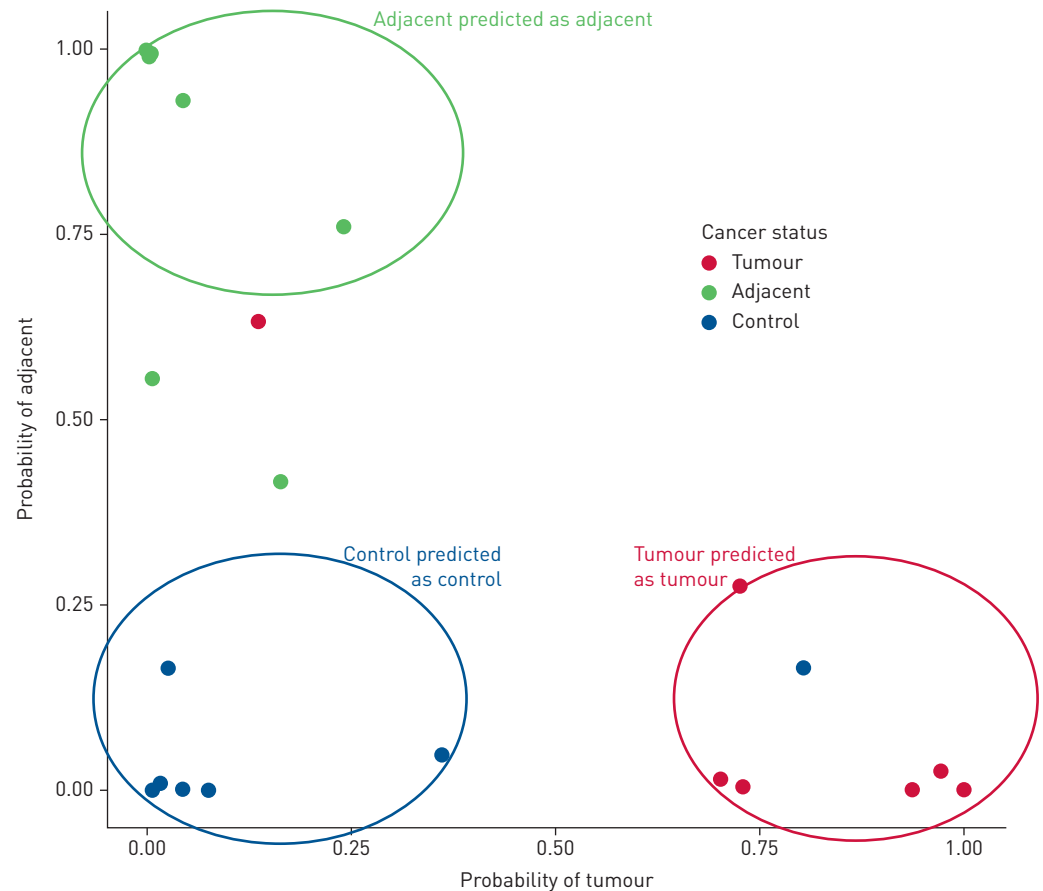


FIGURE 5 Predicted group probabilities for multinomial logistic regression using N-fold cross-validation (n=32).

direction of each variable's contribution to the model. Predictive variables of tumour compared to adjacent lung stroma were found primarily in the transcriptomic data set. This is not surprising since nonsmall cell lung cancer (NSCLC) consists of epithelial cells that have undergone genetic alterations often manifesting in alterations in transcription [22]. However, the predictive variables of non-malignant stroma, *i.e.* tumour adjacent stroma *versus* control stroma, were manifested in the polyribosomal mRNA dataset or translome [23], suggesting regulation occurs at the level of translation [24, 25]. These findings are consistent with other reports that gene expression in cancer initiation and progression is controlled at the level of translational [26–28].

Pathway analysis of the genes demonstrating changes in the adjacent tissue translome identified ECM and PI3K–Akt signalling as potential pre-malignant pathways. Both ECM and PI3K–Akt over-representation were common in the tumour and adjacent tissue, but differed from controls. Changes in the proteome were also related to the ECM and the PI3K–Akt signalling pathways, but these differences indicated over-representation in the tumour tissue relative to adjacent tissue and controls. This suggests that abnormal ECM and PI3K–Akt play a role as pre-malignant signals.

Translation is a tightly regulated process and dysregulation of this process is associated with the development and progression of many cancers, including lung cancer [26]. The eukaryotic translation initiation factor 4E (eIF4E) forms a complex that enables translation. Under homeostatic conditions, eIF4E is sequestered, however, stimulation *via* the PI3K–Akt–mammalian target of rapamycin (mTOR) pathway leads to assembly of the eIF4E complex and increased translation [29]. Elevated levels of eIF4E are associated with both experimental and human cancers, including lung cancer [26, 29]. Our data was consistent with these findings, as the tumour samples demonstrated increased representation of the PI3K–Akt signalling pathway compared to both adjacent tissue and controls. However, the role of eIF4E in pre-malignant lesions is less defined. Li *et al.* [28] identified eIF4E positive cells in adjacent, non-malignant tissues in association with oral squamous cell cancer. We identified the PI3K–Akt signalling pathway to be significantly over-represented in tissue adjacent to the tumour compared to control stroma and this pattern was expressed primarily in the translome or polyribosomal fraction. Thus, lung stroma that is histologically free of lung cancer has pre-malignant signals of increased translation *via* the PI3K–Akt signalling pathway.

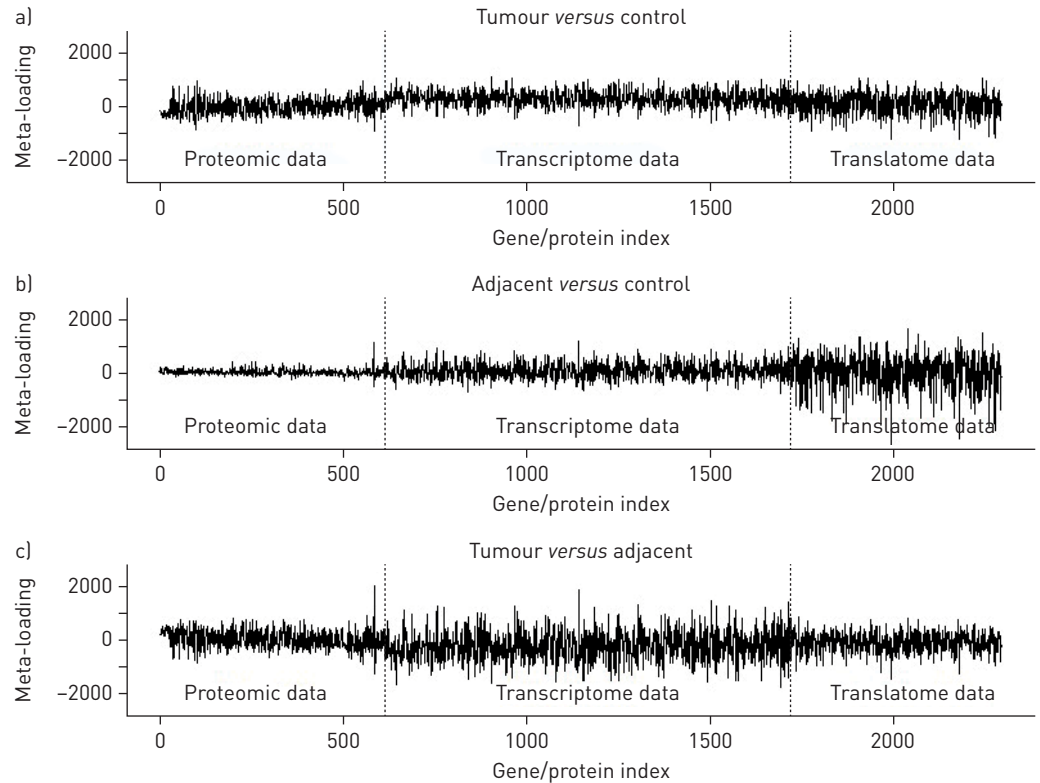


FIGURE 6 Meta-loadings for variables in each of the three data types (proteome, transcriptome and translome) for each of the sample comparisons considered: a) tumour *versus* control, b) adjacent *versus* control and c) tumour *versus* adjacent.

The role of AKT in COPD is less well known. In a study by HosGOOD *et al.* [30], a single nucleotide polymorphism (SNP) of the *PTEN* gene was identified that associated with COPD. *PTEN* is an important regulator of cell-cycle progression and cellular survival and also functions as a negative regulator of AKT. Loss of *PTEN* is associated with cancer initiation and therefore it is designated as a tumour suppressor

TABLE 4 Significantly over-represented pathways (false detection rate (FDR) <0.1) using the average absolute meta-loading threshold

Data	Comparison	Pathway	Genes n	FDR
Proteome	Tumour <i>versus</i> adjacent	ECM-receptor interaction	8	2.62×10^{-4}
Proteome	Tumour <i>versus</i> adjacent	Toxoplasmosis	8	2.62×10^{-4}
Proteome	Tumour <i>versus</i> adjacent	SCLC	7	2.66×10^{-4}
Proteome	Tumour <i>versus</i> adjacent	PI3K-Akt signalling pathway	8	4.25×10^{-4}
Proteome	Tumour <i>versus</i> adjacent	Glycolysis/gluconeogenesis	5	1.99×10^{-2}
Proteome	Tumour <i>versus</i> adjacent	Focal adhesion	11	2.81×10^{-2}
Proteome	Tumour <i>versus</i> control	SCLC	8	1.6×10^{-5}
Proteome	Tumour <i>versus</i> control	ECM-receptor interaction	9	1.6×10^{-5}
Proteome	Tumour <i>versus</i> control	Toxoplasmosis	9	1.6×10^{-5}
Proteome	Tumour <i>versus</i> control	PI3K-Akt signalling pathway	9	5.76×10^{-5}
Proteome	Tumour <i>versus</i> control	Focal adhesion	15	1.58×10^{-4}
Translatome	Adjacent <i>versus</i> control	ECM-receptor interaction	12	3.2×10^{-6}
Translatome	Adjacent <i>versus</i> control	Protein digestion and absorption	7	8.85×10^{-3}
Translatome	Adjacent <i>versus</i> control	Focal adhesion	14	1.48×10^{-2}
Translatome	Adjacent <i>versus</i> control	Bile secretion	5	4.38×10^{-2}
Translatome	Adjacent <i>versus</i> control	Dilated cardiomyopathy	7	4.38×10^{-2}
Translatome	Adjacent <i>versus</i> control	PI3K-Akt signalling pathway	15	4.6×10^{-2}

ECM: extracellular matrix; SCLC: small cell lung cancer; PI3K-Akt: phosphatidylinositol-4,5-bisphosphate 3-kinase-protein kinase B.

gene. In a murine model of breast cancer, loss of *PTEN* in stromal fibroblasts dramatically accelerated both cancer initiation and progression [8]. These findings suggest that AKT regulation may be the link between COPD and cancer initiation.

The tumour stroma consists of non-malignant cells, such as fibroblasts, mesenchymal cells and immune cells along with vasculature and ECM. In the last two decades, it has become apparent that the tumour stroma plays an active role in tumour cell recruitment, growth and spread [6, 31, 32]. However, the role of the ECM as a pre-malignant signal remains less clear. Abnormal ECM remains a hallmark of COPD since the discovery of elevated elastolytic and other related proteases that disrupt the ECM in COPD [33, 34]. These matrix-degrading proteases are produced by non-malignant cells, but once released they can be wielded by malignant or pre-malignant cells to aid micro-invasiveness or activation of signalling pathways [35]. As expected we found evidence of ECM differentiation between tumour and adjacent tissue, but also found differences when comparing tumour adjacent non-malignant stroma and COPD lung tissue not associated with cancer. These findings are tempered by the difference in COPD severity between controls and those with cancer, but highlights the important and somewhat covert role ECM likely has in both COPD and the pre-malignant stroma.

A limitation of our study is the difference in COPD severity in lung cancer associated COPD *versus* controls (*i.e.* controls had more severe COPD relative to those with lung cancer). Since our dataset was not large enough to control for COPD severity, we cannot determine with certainty whether some of the differences observed are related to COPD severity. However, the same pathways (*e.g.* ECM and PI3K–Akt) were represented in the distinction of tumour tissue from matched adjacent tissue, which suggests they are related to carcinogenesis and not COPD severity. Moreover, the identified genes and the predictions from the multinomial model did not correlate with COPD severity within the control or cancer cohorts. Although our adjacent non-malignant lung tissue did not contain histological evidence of malignancy, we cannot guarantee that the tumour influenced stromal phenotype at a distance through secreted factors. In addition, we cannot guarantee that the control group did not carry pre-malignant changes that would manifest at a later time point. However, if true this would introduce noise and weaken our statistical analysis, which is not suggested by the robust distinctions found in our study.

In conclusion, we set out to explore the COPD cancer-associated stroma for molecular signals supporting tumorigenesis. We found that COPD lung tissue adjacent to, but not containing, adenocarcinoma expresses distinct molecular signals compared to both tumour and non-malignant COPD tissue (the control). Specifically, molecular signals in tumour-associated tissue were manifested at the level of translation and were enriched for ECM and PI3K–Akt signalling. This suggests abnormal ECM and PI3K–Akt signalling as potential pre-malignant pathways.

Conflict of interest: None declared.

Support statement: This research was supported by the National Heart, Lung and Blood Institute (NHLBI) grant R01-HL107612. Funding information for this article has been deposited with the Crossref Funder Registry. This material is the result of work supported with resources and the use of facilities at the Minneapolis VAMC. The contents do not represent the views of the U.S. Department of Veterans Affairs or the United States Government.

References

- 1 Tockman MS, Anthonisen NR, Wright EC, *et al.* Airways obstruction and the risk for lung cancer. *Ann Intern Med* 1987; 106: 512–518.
- 2 Skillrud DM, Offord KP, Miller RD. Higher risk of lung cancer in chronic obstructive pulmonary disease. A prospective, matched, controlled study. *Ann Intern Med* 1986; 105: 503–507.
- 3 Wilson DO, Weissfeld JL, Balkan A, *et al.* Association of radiographic emphysema and airflow obstruction with lung cancer. *Am J Respir Crit Care Med* 2008; 178: 738–744.
- 4 Hohberger LA, Schroeder DR, Bartholmai BJ, *et al.* Correlation of regional emphysema and lung cancer: a lung tissue research consortium-based study. *J Thorac Oncol* 2014; 9: 639–645.
- 5 Radisky D, Hagios C, Bissell MJ. Tumors are unique organs defined by abnormal signaling and context. *Semin Cancer Biol* 2001; 11: 87–95.
- 6 Pupa SM, Menard S, Forti S, *et al.* New insights into the role of extracellular matrix during tumor onset and progression. *J Cell Physiol* 2002; 192: 259–267.
- 7 Krtolica A, Parrinello S, Lockett S, *et al.* Senescent fibroblasts promote epithelial cell growth and tumorigenesis: a link between cancer and aging. *Proc Natl Acad Sci USA* 2001; 98: 12072–12077.
- 8 Trimboli AJ, Cantemir-Stone CZ, Li F, *et al.* Pten in stromal fibroblasts suppresses mammary epithelial tumours. *Nature* 2009; 461: 1084–1091.
- 9 Jacobson BA, Alter MD, Kratzke MG, *et al.* Repression of cap-dependent translation attenuates the transformed phenotype in non-small cell lung cancer both in vitro and in vivo. *Cancer Res* 2006; 66: 4256–4262.
- 10 Avdulov S, Li S, Michalek V, *et al.* Activation of translation complex eIF4F is essential for the genesis and maintenance of the malignant phenotype in human mammary epithelial cells. *Cancer Cell* 2004; 5: 553–563.
- 11 Larsson O, Li S, Issaenko OA, *et al.* Eukaryotic translation initiation factor 4E induced progression of primary human mammary epithelial cells along the cancer pathway is associated with targeted translational deregulation of oncogenic drivers and inhibitors. *Cancer Res* 2007; 67: 6814–6824.

- 12 Bhargava M, Viken KJ, Dey S, *et al.* Proteome profiling in lung injury after hematopoietic stem cell transplantation. *Biol Blood Marrow Transplant* 2016; 22: 1383–1390.
- 13 Parker MW, Rossi D, Peterson M, *et al.* Fibrotic extracellular matrix activates a profibrotic positive feedback loop. *J Clin Invest* 2014; 124: 1622–1635.
- 14 Lindstrom J, Szpiro A, Sampson PD, *et al.* SpatioTemporal: spatio-temporal model estimation. <https://cran.r-project.org/package=SpatioTemporal> Date last accessed: June 07, 2018. Date last updated: August 12, 2013.
- 15 Johnson WE, Li C, Rabinovic A. Adjusting batch effects in microarray expression data using empirical Bayes methods. *Biostatistics* 2007; 8: 118–127.
- 16 Benjamini Y, Hochberg Y. Controlling the false discovery rate: a practical and powerful approach to multiple testing. *J R Stat Soc Ser B* 1995; 57: 289–300.
- 17 Kaplan A, Lock EF. Prediction with dimension reduction of multiple molecular data sources for patient survival. *Cancer Inform* 2017; 16: 1176935117718517.
- 18 Lock EF, Hoadley KA, Marron JS, *et al.* Joint and individual variation explained (Jive) for integrated analysis of multiple data types. *Ann Appl Stat* 2013; 7: 523–542.
- 19 O'Connell MJ, Lock EF. RJIVE for exploration of multi-source molecular data. *Bioinformatics* 2016; 32: 2877–2879.
- 20 Akaike H. A new look at the statistical model identification. *IEEE Trans Automatic Control* 1974; 19: 716–723.
- 21 Wang J, Duncan D, Shi Z, *et al.* WEB-based GENE SeT AnaLysis Toolkit (WebGestalt): update 2013. *Nucleic Acids Res* 2013; 41: W77–W83.
- 22 Fujii T, Dracheva T, Player A, *et al.* A preliminary transcriptome map of non-small cell lung cancer. *Cancer Res* 2002; 62: 3340–3346.
- 23 Masvidal L, Hulea L, Furic L, *et al.* mTOR-sensitive translation: cleared fog reveals more trees. *RNA Biol* 2017; 14: 1299–1305.
- 24 Liu Y, Beyer A, Aebersold R. On the dependency of cellular protein levels on mRNA abundance. *Cell* 2016; 165: 535–550.
- 25 Schwanhausser B, Busse D, Li N, *et al.* Global quantification of mammalian gene expression control. *Nature* 2011; 473: 337–342.
- 26 Bauer C, Diesinger I, Brass N, *et al.* Translation initiation factor eIF4G is immunogenic, overexpressed, and amplified in patients with squamous cell lung carcinoma. *Cancer* 2001; 92: 822–829.
- 27 Lazaris-Karatzas A, Montine KS, Sonenberg N. Malignant transformation by a eukaryotic initiation factor subunit that binds to mRNA 5' cap. *Nature* 1990; 345: 544–547.
- 28 Li Y, Li B, Xu B, *et al.* Expression of p53, p21(CIP1/WAF1) and eIF4E in the adjacent tissues of oral squamous cell carcinoma: establishing the molecular boundary and a cancer progression model. *Int J Oral Sci* 2015; 7: 161–168.
- 29 Mamane Y, Petroulakis E, LeBacquer O, *et al.* mTOR, translation initiation and cancer. *Oncogene* 2006; 25: 6416–6422.
- 30 Hosgood HD III, Menashe I, He X, *et al.* PTEN identified as important risk factor of chronic obstructive pulmonary disease. *Respir Med* 2009; 103: 1866–1870.
- 31 Bremnes RM, Donnem T, Al-Saad S, *et al.* The role of tumor stroma in cancer progression and prognosis: emphasis on carcinoma-associated fibroblasts and non-small cell lung cancer. *J Thorac Oncol* 2011; 6: 209–217.
- 32 Micke P, Ostman A. Tumour-stroma interaction: cancer-associated fibroblasts as novel targets in anti-cancer therapy? *Lung Cancer* 2004; 45: Suppl 2, S163–S175.
- 33 Burgess JK, Mauad T, Tjin G, *et al.* The extracellular matrix - the under-recognized element in lung disease? *J Pathol* 2016; 240: 397–409.
- 34 Kulkarni T, O'Reilly P, Antony VB, *et al.* Matrix remodeling in pulmonary fibrosis and emphysema. *Am J Respir Cell Mol Biol* 2016; 54: 751–760.
- 35 Hanahan D, Weinberg RA. The hallmarks of cancer. *Cell* 2000; 100: 57–70.