Protein kinase C isozymes; predictors of progression free survival in NSCLC patients

CURRENT STATUS: POSTED

Ann Rita Halvorsen  ann.rita.halvorsen@rr-research.no
Corresponding Author
ORCID: 0000-0003-1223-7948

Mads Haugland Haugen
Institute for cancer research, Oslo university hospital-Radiumhospitalaet

Åsa Kristina Öjlert
Institute for cancer research, Oslo university hospital-Radiumhospitalaet

Marius Lund-Iversen
Department of pathology, Oslo university hospital-Radiumhospitalaet

Lars Jørgensen
Department of cardiothoracic surgery, Oslo university hospital-Rikshospitalaet

Steinar Solberg
Department of cardiothoracic surgery, Oslo university hospital-Rikshospitalaet

Gunhild M. Mælandsmo
Insittute for cancer research, Oslo university hospital-Radiumhospitalaet

Odd Terje Brustugun
Institute for cancer research, Oslo university hospital-Radiumhospitalaet

Åslaug Helland
Institute for cancer research, Oslo university hospital-Radiumhospitalaet

DOI:
10.21203/rs.2.10276/v1

SUBJECT AREAS
Oncology

KEYWORDS
expression subtypes, PKC, NSCLC, adenocarcinoma, protein profiling, neuroendocrine markers, progression free survival
Abstract

Background

Protein expression is deregulated in cancer, and the proteomic changes observed in lung cancer may be a consequence of mutations in essential genes. The purpose of this study was to identify protein expression associated with prognosis in lung cancers stratified by smoking status, molecular subtypes, and EGFR-, TP53- and KRAS-mutations.

Methods

We performed profiling of 295 cancer-relevant phosphorylated and non-phosphorylated proteins, using reverse phase protein arrays. Biopsies from 80 patients with operable lung adenocarcinomas were analyzed for protein expression and association with progression free survival (PFS) were studied.

Results

Spearman rank correlation analysis identified 56 proteins with significant association to PFS (p<0.05). High expression of protein kinase C (PKC)-α and the phosphorylated state of PKC-α, PKC-β and PKC-δ, showed the strongest positive correlation to PFS, especially in the wild type samples. This was confirmed in gene expression data from 186 samples. Based on protein expression, unsupervised hierarchical clustering separated the samples into four subclusters enriched with the molecular subtypes TRU, PI or PP (p=0.0001). Subcluster 2 contained a smaller cluster (2a) enriched with samples of the subtype PP, low expression of the PKC isozymes, and associated with poor PFS (p=0.003) compared to the other samples. Subcluster 2a revealed increased expression of neuroendocrine markers, supporting the aggressive behavior. Low expression of the PKC isozymes in the subtype PP and a reduced relapse free survival was confirmed with the TCGA LUAD samples.

Conclusion

This study identified different proteins associated with PFS depending on molecular
subtype, smoking- and mutational-status, with PKC-α, PKC-β and PKC-δ showing the strongest correlation. Cluster analysis detected a subgroup of samples enriched for samples of the PP subtype and poor PFS, which may benefit from a more aggressive treatment regimen.

Introduction

Proteins are the functional players driving both normal and disease processes. Some of the most important types of mutations in lung cancer occur in \textit{EGFR}, \textit{TP53} and \textit{KRAS}. Mutations in these genes may lead to changes in many interacting pathways, leading to significantly altered protein expression. They are known to influence treatment response and regarded as essential for progression of lung cancer [1].

Some of the changes in protein expression observed in lung cancer are a consequence of mutations in essential driver genes, and targeted therapy is usually efficient for these subgroups of patients [2].

Patients with somatic genomic alterations in the \textit{EGFR} gene are routinely treated with \textit{EGFR} inhibitors, which have improved the outcome for this patient group [3].

Mutations in \textit{KRAS} leads to constitutively and persistent stimulus-independent activation of downstream pathways affecting tumor growth, proliferation and survival. Developing treatment targeting KRas has proved to be complicated, but ongoing studies
investigate inhibition of effector-molecules downstream of KRas, including ERK and MEK [4]. Nevertheless, MEK inhibitors are associated with early development of resistance due to crosstalk with other signaling pathways which make this approach challenging [5].

So far, no efficient therapy to re-establish the function of p53 is in clinical use, but studies with reactivation of p53 have been performed [6].

In order to improve outcome for lung cancer patients, stratification based on alterations in essential genes and the affected pathways may lead to better treatment strategies and increased response rate. To sub-classify non-small cell lung cancer (NSCLC), intrinsic molecular subtypes have been explored based on gene expression profiling. Three subtypes have been identified for the adenocarcinomas, namely the Bronchioid, Magnoid, and Squamoid subtype, later re-named to terminal respiratory unit (TRU), proximal proliferative (PP), and proximal inflammatory (PI), respectively [7-9].

The TRU subtype is most common among females and never-smokers, and often harbours EGFR mutations and a less invasive phenotype. Early stages of the TRU subtype have better prognosis. Gene expression profiles related to biological processes involved in
excretion, asthma, and surfactants are associated with the TRU subtype. The PP subtype is reported with a high frequency of \textit{KRAS} and \textit{TP53} mutations, over-expression of DNA repair genes and is often found in heavy smokers. The PI subtype is recognized with over-expression of defense response genes such as \textit{CXCL10} and is most common in high grade tumors [7]. At late stages, the PI subtype is associated with better survival compared to the other subtypes [8]. Chromosomal instability (CIN), copy number alterations, and genomewide DNA methylations are also reported to differ among the three subtypes, with the PP subtype having the highest CIN [8]. So far, microRNA expression and protein expression are reported to only partially correlate with the molecular subtypes of adenocarcinoma [9]. In order to treat lung cancer patients more efficiently, groups of patients who share common biological features such as mutations or pathway alterations should be identified and treated with drugs optimized for their subgroup. It has been known for a long time that never-smoking NSCLCs are recognized with a different underlying biology compared to ever-smokers. However, except from a handful of known genetic aberrations
such as EGFR mutations and ALK rearrangement, no subgroups of NSCLC are stratified for optimized treatment based on the molecular profile of the tumor.

The purpose of this study of lung adenocarcinoma, was to identify differences in protein expression levels associated with prognosis, stratified on EGFR, TP53 and KRAS mutations status, and smoking status.

Material And Methods

Patients diagnosed with operable NSCLC adenocarcinoma from 2006 to 2011, were included and analyzed for protein expression (n=80) in the Oslo cohort. The patients underwent curatively intended surgical resection at Rikshospitalet, Oslo University Hospital, Norway. Tumor samples were snap frozen in liquid nitrogen and stored at \(-80^\circ\text{C}\) until protein extraction. Pathological stage, mutation status and smoking status are outlined in Table 1. All the samples were classified as adenocarcinomas. Never-smokers are defined as those who have smoked less than 100 cigarettes.

The study was approved by the Regional Ethics Committee (S-05307), and written informed consent was obtained from all patients.

Table 1. Patient characteristics

Table 1. Mutation status, smoking history and stage for the patients analyzed for protein expression (n=80). Samples not mutated, wild type (WT), are shown in brackets.

Reverse phase protein arrays

We have performed profiling of 295 cancer relevant proteins of which 60 were in a phosphorylated state (S1 Table and supplementary material and methods) on the Oslo cohort, using the reverse phase protein array (RPPA) core facility at MD Anderson Cancer Center (Houston, TX). RPPA data (n=131 proteins) and phenotypes from the LUAD cohort were extracted from the cancer genome atlas (TCGA) data generated by the TCGA
Research Network:

http://cancergenome.nih.gov/. Samples with no expression subtype assigned and without registered relapse free survival were filtered out. The remaining 181 LUAD samples were utilized as validation set. Time to relapse was extracted for survival analyses. Of note, the median follow-up time for alive patients was 23.9 months for the LUAD cohort, compared to 60 months for the Oslo cohort.

**EGFR, TP53 and KRAS analyses**

Mutation analyses of *EGFR* exons 18–21 were performed using the TheraScreen EGFR mutation kit (DxS, Manchester, UK,) designed to detect 28 specific mutations in the *EGFR* gene. Assays were carried out according to the manufacturer’s protocol and with the use of the Roche LightCycler 480 real-time PCR system. Some of the results were previously published by Helland et al [10].

The *TP53* gene was analyzed by the Sanger Sequencing method in all the tumor samples. The procedure was performed on an Applied Biosystems 3730 DNA analyser according to the supplier’s handbook, Applied Biosystem 3730/3730X/DNA Analysers Part 4331467 Rev.B, as previously described [11]. More details are provided in supplementary material and methods.

We used the wobble-enhanced ARMS (WE-ARMS) method for detecting *KRAS* mutations in the lung adenocarcinoma samples. This mutation assay detects the seven most commonly
reported mutations in the KRAS gene—KRAS g.34G>C (p.G12R), g.34G>A (p.G12S),
g.34G>T (p.G12C), g.35G>A (p.G12D), g.35G>C (p.G12A), g.35G>T (p.G12V) and g.38G>A
(p.G13D)—by real-time PCR [12].

Gene expression and subtyping of adenocarcinoma samples

Gene expression was performed on 186 adenocarcinomas (including 79 of those with
protein expression) from the same cohort, using hybridization arrays (SurePrint G3
Human, 8x60K, Agilent Technologies). More details are provided in supplementary
material and methods. The adenocarcinoma samples were assigned a gene expression
subtype using the previously described 506 gene centroid classifier and Pearson
correlation [7,8]. Four of the samples had a negative correlation with all three subtypes and were
not assigned to any subtype. The raw data and normalized data are deposited in
ArrayExpress database:

www.ebi.ac.uk/arrayexpress under accession number: E-MTAB-7708.

Immunohistochemistry (IHC)

IHC staining for the neuroendocrine markers CD56, synaptophysin, chromogranin A and
NSE were performed using the antibodies as outlined in S2 Table and supplementary material and methods.

Statistics

All the statistical analysis were performed in R (v 3.3.2) [13] using RStudio (v 1.1.447). Hierarchical clustering was performed to visualize the total protein expression for all the samples with mutation status, stage, subtype, smoking history and progression free survival (PFS) data included. Hierarchical clustering was performed using the publicly available R-package Clustermap (Lingjærde and Steen, submitted). In brief, median-centered and log2-transformed RPPA-data were clustered using Pearson distance and average linkage, and with N=200 iterations in the PART method used to estimate the number of clusters, as described previously [14].

For heatmap visualization of the data, values are normalized to the range [-1.2, 1.2] by application of a nonlinear sigmoid transformation \( f(x) = 1.2 \tanh(x/1.2) \). This limits the visual dominance of outlier values while maintaining the order of the values, since \( f \) is strictly increasing. PFS was calculated from the date of surgery until the date of event, defined as relapse, metastases or death from lung cancer. Correlation between protein expression and PFS was calculated using two-sided Spearman test. Student t-test was performed to identify differentially expressed proteins between the groups of interests. Survival analysis was performed using cox regression analysis and Kaplan Meyer survival plot. The significance threshold was set to \( p<0.05 \).
Results

We performed RPPA on biopsies from 80 lung adenocarcinoma biopsies. The tumor samples were tested for mutations in the EGFR-, TP53-, and KRAS genes, where 12.5% had EGFR mutations, 39% had KRAS mutations and 39% had mutations in the TP53 gene. In the group of patients with EGFR mutated tumors, four (representing 44%) were never smokers (Table 1). Out of 80 patients, 36 (45%) patients were registered with PFS > 60 months. Among these, three (8%) of the tumors were EGFR mutated, 15 (42%) were TP53 mutated, 13 (36%) were KRAS mutated and three patients were never smokers.

Hierarchical clustering analysis

Unsupervised hierarchical clustering analysis performed on protein expression grouped the samples into four subclusters as shown in the dendrogram (Fig 1). The parameters EGFR-, TP53-, KRAS- mutations, stage, event and smoking status were evenly distributed between the sub-clusters (p > 0.05 in Fisher Exact test). The molecular subtypes were significantly associated with the four subclusters (p = 0.0001), of which subcluster 1 (green branches) was mainly associated with the PI subtype, subcluster 2 (red branches) enriched with the PP subtype, and subcluster 3 and 4 (turquoise and blue branches) were associated with the TRU subtype. Subcluster 2 contained four smaller clusters (2a, 2b, 2c and 2d visualized in S1 Fig) of which 2a (marked with red box in Fig 1 and S1 Fig) was recognized with early relapse (median PFS = 12.5 months). We used multivariate cox regression analysis, where the samples in subcluster 2b, 2c and 2d were analyzed as one group and stage was included as a covariate in the model, confirming a significantly shorter PFS for samples within subcluster 2a compared to the other subclusters (p=0.003). This association was visualized with a Kaplan Meier survival plot (Fig 2). The heatmap further demonstrated that cluster 2a had very low expression of several proteins.
recognized as members in the MAPK pathway and mTOR pathway. In addition, protein kinase C (PKC)-α and phosphorylated PKC-α, PKC-β, and PKC-δ were significantly lower expressed in subcluster 2a compared to the other samples (wilcoxon rank test, FDR <0.05, S3 Table). To identify the proteins with the highest correlation to PFS within this group, Spearman Rank correlation was utilized. Eight proteins (myosin11, CD26, YAP, DJ1, YAP_pS127, CD171, HSP27_ps82 and hexokinaseII) were significantly associated (p< 0.05) with PFS in subcluster 2a (S4 Table). To check if subcluster 2a contained neuroendocrine-like features, the expression of the genes NCAM1, CHGA, SYP encoding for the neuroendocrine markers CD56, chromogranin A and synaptophysin respectively, were analyzed. All three markers revealed a higher gene expression in samples from subcluster 2a compared to all the other samples, but only SYP was significant (p=0.03). IHC staining for the proteins CD56, chromogranin A, synaptophysin and neuron-specific enolase (NSE) confirmed an expression of synaptophysin, CD56, chromogranin A, and NSE in nine, two, one, and nine of the 11 samples, respectively (S5 Table). Serum levels of the neuroendocrine markers proGRP, chromogranin A and NSE were elevated in one of the 11 samples (data not shown).

Unsupervised hierarchical clustering of mRNA expression data of the genes coding for the analyzed proteins (n=226) were performed on a cohort of 186 adenocarcinoma samples, including the 79 with RPPA data (S2 Fig). The clustering confirmed a different gene expression pattern for the three molecular subtypes (p = 1.3*E-16 in Fisher exact test). The subtype PI was scattered between several subclusters, but the majority of the TRU samples were found in the largest subcluster, whereas most of the PP subtype samples clustered together. Focusing on the 11 samples from subcluster 2a identified with protein expression, only the samples with the PP subtype (n=7) clustered together with samples recognized with poor PFS and PP subtype, when using mRNA data. Time to progression,
TP53 - and EGFR- mutation status were significantly differentially distributed between the subclusters (p < 0.05).

Protein expression associated with prognosis

Using Spearman rank correlation, we identified 56 proteins significantly associated with PFS (p<0.05), of which nine were in a phosphorylated state (S4 Table). High expression of PKC-α, and phosphorylated PKC-α, PKC-βII, and PKC-δ were all highly associated with increased PFS (Table 2). The proteins c-ABL, MIF, PAI-1, Jak2, vimentin and the phosphorylated state of JNK were also among the top twenty proteins significantly associated with PFS (p<0.05).

Table 2. Proteins associated with PFS

Table 2. Top twenty proteins associated with PFS (p<0.05) in the total group of samples (n=80).

Spearman Rank correlation performed pair-wise for 79 samples on the proteins and phosphorylated proteins, and the corresponding gene expression data (S3 Fig, S6 Table) revealed a positive correlation (>0.3) for 110 of the proteins, and a negative correlation for 42 of the proteins.

Next, we included all the samples with mRNA data (n=186) and performed a Spearman Rank correlation between genes and proteins associated with prognosis, and found 10 proteins/genes significantly associated with PFS in both analysis (Table 3 and S3 Fig).

Table 3. The pair-wise spearman Rank correlation between mRNA data and protein data

Table 3. The pair-wise spearman Rank correlation was performed on 79 samples, while the PFS analysis performed on mRNA data contained 186 samples. The overlap between PFS associated mRNA data and protein data are based on a threshold of p-value < 0.05.

Protein expression with impact on prognosis differ depending on
mutational and smoking status

A significant association between PFS and expression of 10 proteins (S4 Table), including MEK-1, LCA3B, Collagen IV, CD49b and phosphorylated PRAS40_pT246 was seen in never smokers. No proteins were simultaneously significantly associated with PFS when comparing never-smokers and smokers (S4 Fig).

High expression of B7-H3 was associated with poor survival in the total group, but only found to be significant in KRAS and TP53 wild type tumors, smokers, and EGFR mutated samples after stratification (p<0.05, S4 Table).

When comparing similarities in PFS-associated proteins between the mutated samples and the wild type samples two, three and four overlapping proteins were identified in the EGFR, TP53 and KRAS subgroups, respectively. The proteins PKC-α and phosphorylated PKC-α, PKC-βII, and PKC-δ were the most significantly PFS-associated proteins in smokers, TP53wt, EGFRwt, and KRASwt. Among the mutated samples, fewer proteins were associated with PFS, and as expected, no overlapping proteins were found between EGFR and KRAS mutated samples (S4 Table and S4 Fig). The same analyses were applied to the 186 samples with gene expression data, where fewer genes were associated with PFS in samples harboring a mutation compared to wild type samples. In the groups stratified on mutational status and smoking status, the level of PRKCA, encoding PKC-α, were significantly associated with PFS in smokers, EGFR wt, KRAS wt, and TP53 wt samples. The level of PRKCD, encoding PKC-δ, was significantly associated with PFS among smokers, EGFR mut, EGFR wt, and KRAS wt, which confirmed the findings in protein data.

A tendency towards worse relapse free survival in patients with low expression of PKC-δ was also seen in the TCGA data (p=0.06, Fig 3). The PP subtype showed a significantly decreased expression of PKC-α, and phosphorylated PKC-α and PKC-δ when compared to PI and TRUE subtype. Further, the PP subtype also showed a reduced relapse free survival
when compared to non-PP samples (p=0.085, Fig 3).

**Discussion**

In this study we examined the expression of 235 proteins and 60 phospho-proteins in tumors from 80 patients diagnosed with lung adenocarcinomas. We identified 56 proteins and phospho-proteins likely to impact the patient outcome. When stratifying the samples according to mutations in the genes TP53, EGFR or KRAS, more proteins were associated with PFS in samples without mutations. This was also seen when analyzing the corresponding mRNA data. When we compared proteins associated with progression in never-smokers and former/current smokers, no overlapping proteins were found, confirming the differences between these groups of lung cancer. Cluster analysis identified a small subcluster containing 11 patients enriched with tumors of the PP subtype, recognized with early relapse, low expression of PKC isozymes, and increased expression of neuroendocrine markers.

**Cluster analysis**

Based on the expression of 295 proteins, unsupervised hierarchical clustering separated the samples into four subclusters. However, pathological stage, smoking status, or mutations in the genes TP53, KRAS or EGFR did not seem to impact the clustering. Interestingly, the four subclusters were significantly correlated with the molecular subtypes TRU, PP and Pl. Previous work using protein expression identified six subgroups of adenocarcinoma, where the subgroups partially overlapped with the three mRNA-derived subtypes. The PP subtype was further divided into two groups [
Subcluster 2a, enriched with PP subtype samples, was recognized with lower expression of members of the mTOR pathway and the MAPK pathway. This is in line with previous finding where TRU samples are associated with higher expression of these proteins when compared to PP samples [15].

Interestingly, cluster analysis performed on mRNAs corresponding to the proteins, grouped the samples with PP subtype from the protein-derived subcluster 2a together, but not the non-PP subtypes. However, the distinct pattern with very low expressed proteins within subcluster 2a, was not reflected using mRNA data. This can be explained with a lower correlation between phosphorylated proteins and the corresponding genes. Nine of the 11 samples within subcluster 2a showed a positive staining for synaptophysin which are considered as a neuroendocrine marker and can be used to confirm the diagnosis [16].

In total, all the 11 samples showed positive staining for at least one of the neuroendocrine marker or NSE. In a recent study on NSCLCs, a molecular subgroup enriched with PP subtype and shorter survival was identified. These samples had a mixed histology predominantly with adenocarcinomas with molecular expression pattern associated with neuroendocrine tumors [15].

LCNEC (large cell neuroendocrine carcinoma) can share some of the same pathological features as adenocarcinomas, and LCNEC with areas of adenomatous differentiations (mixed LCNEC) is described. Both pure LCNEC and mixed LCNEC tumors
exhibit an aggressive behavior and are associated with poor survival [17] as seen within subcluster 2a in our analyses. The hierarchical clustering performed on 186 mRNA samples resulted in a significantly different distribution of the molecular subtypes between the clusters, although the PI samples did not cluster together. Of note, the overlap between the genes used for the original sub-typing [8] and our clustering based on protein expression was sparse (n=18), indicating that the proteins included in our analysis seem to be important for the subtypes. Both the TRU- and PP-subtype clustered based on our mRNA data, and we suggest that the TRU- and PP-subtypes are more distinct subtypes compared to PI. In a large meta-study, the TRU subtype was identified as the most prognostically important subtype compared to the non-TRU subtype, arguing for the need to identify additional classifiers [18].

Protein kinase C levels associated with survival

The protein kinase C is a group of enzymes known to be involved in diverse cellular functions, including cell proliferation, apoptosis and cell migration, and has been regarded as an onco-protein. The members of the PKC-family are encoded from nine different genes which have several known splice variants. Recent work has demonstrated that these proteins may have a more complex role than first assumed, which is supported by the

16
many failed clinical trials for cancer using PKC-inhibitors. In addition, mutational studies have revealed that most cancers have loss of function (LOF) mutations in genes belonging to the PKC-family, suggesting a tumor suppressor role for the proteins [19, 20]. A meta-study on the use of PKC-inhibitors combined with chemotherapy in lung cancer patients reported decreased response rate and disease control, compared to chemotherapy alone [21]. In our study, low expression of PKC-α, and phosphorylated PKC-α, PKC-βII and PKC-δ were associated with poor PFS. In addition, the levels of the PKC isozymes were strikingly lower in subcluster 2a which also contained the samples with the overall poorest PFS. An association with low expression of PKC-δ and decreased relapse free survival was confirmed in the LUAD TCGA samples. Low levels of PKC-α, and phosphorylated PKC-α and PKC-δ in subtype PP were also found in the TCGA samples. This support the findings in the Oslo cohort, with poor PFS and low expression of the isozymes of PKC in subgroup 2a containing mainly PP samples. Unfortunately, sparse information on mutational status, reduced number of proteins analyzed, short follow-up time, limited further validation on the TCGA samples. Thus, with regard to the classical role of the PKC-family, our results suggest a general tendency towards a tumor suppressor role for PKC-α, PKC-βII and PKC-δ in NSCLC.
In the *EGFR* mutated samples, PKC-δ levels positively correlated to patient survival. It has been reported that an activation of PKC-δ can be promoted by an activated EGFR [22]. Further, activation of PKC-δ may induce apoptosis and growth arrest resulting in reduced tumorigenesis [23]. As a response to DNA damage, it has been shown that over-expression of p53 increases the transcription of PKC-δ resulting in apoptosis. This may explain the significant correlation we discovered between expression of PKC-δ and PFS in *TP53* wild type samples, but not in the *TP53* mutated samples. These results were also reflected by gene expression data. It’s been demonstrated that PKC can phosphorylate many oncoproteins to suppress their activity, including KRas, PI3K and several tyrosine kinase receptors [24]. The oncoprotein KRas, recognized with activating mutations in cancer, can be suppressed by activated PKC, which is proposed as a novel approach to target KRas [25]. A negative correlation between KRas and PKC-α was recently described in colorectal cancer. Further, low expression of PKC-α was also associated with poor prognosis [26].
This is in concordance with our results, where high levels of PKC-α was associated with better PFS in KRAS mutated samples, whereas PKC-δ showed a higher correlation to PFS in those with KRAS wild type. Interestingly, no isozymes of PKC did influence on survival in never-smoking patients, further supporting this group as a distinct lung cancer disease driven by other mechanisms. This leads to the hypothesis that PKC may also have an essential role keeping oncoproteins in check.

Based on our results, we suggest that the association to PFS for the different PKC-isozymes is connected to mutational and smoking status. Results from gene expression analysis performed on 186 NSCLC samples strengthen these observations.

Proteins associated with PFS in subgroups of NSCLC

Interestingly, high expression of B7-H3, a molecule involved in immune checkpoint signaling, was correlated to poor outcome in our study, especially in smokers, those without any detected mutations in KRAS or TP53, and in those harboring an EGFR mutation. B7-H3 is a molecule known to inhibit T-cell activation in an immune suppressive manner. This protein has been shown to be linked to poor survival in cancer, and have been suggested as a new immune checkpoint target.

In a recent study of lung cancer patients, expression of B7-H3 was associated with overall survival only in smokers.
This indicates that future anti-B7-H3 therapy may have higher success rate among ever smoking lung cancer patients with KRAS or TP53 wild type tumors, or an EGFR mutation.

Within subcluster 2a, eight proteins were significantly associated with PFS, where low expression of myosin II showed the highest correlation with better PFS. This is also supported by a protein study on early stage lung cancer where myosin IIa was reported to be upregulated in stage Ia/Ib lung cancer patients with early relapse [29].

Interestingly, low expression of YAP, phosphorylated YAP(s127) and phosphorylated HSP27(s82) were associated with increased PFS in subcluster 2a. It has been shown that high expression of HSP27 leads to less phosphorylated YAP(s127). Further, phosphorylation of YAP on S127 decreased the activity of YAP since this prevent its translocation to the nucleus [30].

This also means that un-phosporylated YAP promotes tumor aggressiveness and is related to poor prognosis which is in line with our study. These finding highlight the central role HSP27 has in several pathways, including the Hippo pathway.

Correlation analysis between mRNA and protein expression

Spearman Rank correlation revealed a high correlation (R >0.3) between expression levels of almost half of the proteins and mRNAs. Previous studies have reported that much of the variation in mean-level protein expression can be explained by variation in mRNA expression [...
However, the variance in the proteomes across different tissue types can poorly be explained by the mRNA levels, highlighting a tissue-specific posttranscriptional regulation of gene expression. In a study of lung cancer, proteins involved in metabolic and translational pathways were highly correlated with mRNA expression, whereas proteins involved in extracellular matrix and adhesion, were not correlated or anti-correlated.

In a study of breast cancer, 35% of the proteins correlated significantly (R > 0.3) with mRNA expression. The proteins, Cyclin B1, cyclin E1, 4E-BP1, PKC-α and RAB 25 were highly correlated with mRNA expression in breast cancer.

This is in line with our study, where these proteins showed a high correlation value (Rho > 0.6). Interestingly, HER2 was highly correlated in the breast cancer study across all subtypes, but this protein was poorly correlated with mRNA expression in our lung study (Rho =0.17). On the other hand, EGFR revealed Rho=0.72 in our study, while in breast cancer a correlation between R=0.15 - 0.3 was found. This indicates that genes known to be deregulated in a specific cancer type may be regulated by other mechanisms. Proteins such as p53, CDKN1B, and MAPK14 showed very low correlation with the mRNA
expression both in our study on lung cancer and in the breast cancer study [33]. Lack of synergy between the level of proteins and mRNAs measured in the cells can have several explanations including copy number aberrations, miRNA expression and methylation.

**Conclusion**

These results demonstrate that essential mutations in lung carcinomas affect several proteins associated with outcome. Based on our results, expression of PKCα and phosphorylated PKCα, PKCβ, and PKCδ seem to be positively associated with PFS, with different isozymes linked to smoking and mutational status of *EGFR, KRAS* and *TP53*. These results illustrate the need to better understand the biological context in order to further improve targeted therapy in cancer. This study supports that a therapy restoring the level of specific isozymes of PKC activity may be beneficial for subgroups of lung cancer patients based on the genetic background. We identified a subgroup of samples enriched with the molecular subtype PP, recognized with early relapse, increased expression of neuroendocrine markers, and a distinct protein expression pattern, including low levels of PKC isozymes. These patients may benefit from a more aggressive treatment regimen. Proteins associated with PFS among never smokers were strikingly different compared to the other investigated subgroups. This is not surprising, but underscores the need for a more stratified therapy in order to improve clinical outcome.

**Declarations**

**Acknowledgment**

We want to thank Ingjerd Solvoll for help with collecting tissue samples and clinical
information. We thank Nils Bolstad and the Department of Medical Biochemistry, Oslo University Hospital- Radiumhospitalen for the analyses of neuroendocrine markers in serum. The authors gratefully acknowledge the contribution from the patients participating in this study. We thank the support provided by the Sister Institution Network Fund (SINF) in FY14 to Prof. Gordon Mills and the RPPA core facility at the MD Anderson Cancer Center for RPPA analysis (MDACC CCSG grant P30 CA016672 and NCI # CA16672). This work was supported by grants from Norwegian Cancer Society, and the regional health authorities.

References

1. Helland A, Brustugun OT, Nakken S, Halvorsen AR, Donnem T, Bremnes R, Busund LT, Sun J, Lorenz S, Solberg SK et al: High number of kinome-mutations in non-small cell lung cancer is associated with reduced immune response and poor relapse-free survival. *Int J Cancer* 2017, 141(1):184-190.

2. Pakkala S, Ramalingam SS: Personalized therapy for lung cancer: Striking a moving target. *JCI insight* 2018, 3(15).

3. Zhou C, WuYL, Chen G, Feng J, Liu XQ, Wang C, Zhang S, Wang J, Zhou S, Ren S et al: Erlotinib versus chemotherapy as first-line treatment for patients with advanced EGFR mutation-positive non-small-cell lung cancer (OPTIMAL, CTONG-0802): a multicentre, open-label, randomised, phase 3 study. *The Lancet Oncology* 2011, 12(8):735-742.

4. Tomasini P, Walia P, Labbe C, Jao K, Leighl NB: Targeting the KRAS Pathway in Non-Small Cell Lung Cancer. *Oncologist* 2016, 21(12):1450-1460.

5. Bernabe R, Patrao A, Carter L, Blackhall F, Dean E: Selumetinib in the treatment of non-small-cell lung cancer. *Future oncology (London, England)* 2016, 12(22):2545-2560.
6. Duffy MJ, Synnott NC, Crown J: Mutant p53 as a target for cancer treatment. *Eur J Cancer* 2017, 83:258-265.

7. Hayes DN, Monti S, Parmigiani G, Gilks CB, Naoki K, Bhattacharjee A, Socinski MA, Perou C, Meyerson M: Gene expression profiling reveals reproducible human lung adenocarcinoma subtypes in multiple independent patient cohorts. *Journal of clinical oncology: official journal of the American Society of Clinical Oncology* 2006, 24(31):5079-5090.

8. Wilkerson MD, Yin X, Walter V, Zhao N, Cabanski CR, Hayward MC, Miller CR, Socinski MA, Parsons AM, Thorne LB et al: Differential pathogenesis of lung adenocarcinoma subtypes involving sequence mutations, copy number, chromosomal instability, and methylation. *PloS one* 2012, 7(5):e36530.

9. Network TCGAR: Comprehensive molecular profiling of lung adenocarcinoma. *Nature* 2014, 511(7511):543-550.

10. Helland A, Skaug HM, Kleinberg L, Iversen ML, Rud AK, Fleischer T, Sagerup C, Solberg S, Jorgensen L, Ariansen S et al: EGFR gene alterations in a Norwegian cohort of lung cancer patients selected for surgery. *Journal of thoracic oncology: official publication of the International Association for the Study of Lung Cancer* 2011, 6(5):947-950.

11. Halvorsen AR, Silwal-Pandit L, Meza-Zepeda LA, Vodak D, Vu P, Sagerup C, Hovig E, Myklebost O, Borresen-Dale AL, Brustugun OT et al: TP53 Mutation Spectrum in Smokers and Never Smoking Lung Cancer Patients. *Frontiers in genetics* 2016, 7:85.

12. Hamfjord J, Stangeland AM, Skrede ML, Tveit KM, Ikdahl T, Kure EH: Wobble-enhanced ARMS method for detection of KRAS and BRAF mutations. *Diagnostic molecular pathology: the American journal of surgical pathology, part B* 2011, 20(3):158-165.

13. Team RDC: R: language and environment for statistical computing. In., vol. ver.3.1.2. URL
14. Nilsen G, Borgan O, Liestol K, Lingjaerde OC: Identifying clusters in genomics data by recursive partitioning. *Stat Appl Genet Mol Biol* 2013, 12(5):637-652.

15. Chen F, Zhang Y, Parra E, Rodriguez J, Behrens C, Akbani R, Lu Y, Kurie JM, Gibbons DL, Mills GB et al: Multiplatform-based molecular subtypes of non-small-cell lung cancer. *Oncogene* 2017, 36(10):1384-1393.

16. Thunnissen E, Borczuk AC, Flieder DB, Witte B, Beasley MB, Chung JH, Dacic S, Lantuejoul S, Russell PA, den Bakker M et al: The Use of Immunohistochemistry Improves the Diagnosis of Small Cell Lung Cancer and Its Differential Diagnosis. An International Reproducibility Study in a Demanding Set of Cases. *Journal of thoracic oncology : official publication of the International Association for the Study of Lung Cancer* 2017, 12(2):334-346.

17. Fasano M, Della Corte CM, Papaccio F, Ciardiello F, Morgillo F: Pulmonary Large-Cell Neuroendocrine Carcinoma: From Epidemiology to Therapy. *Journal of thoracic oncology : official publication of the International Association for the Study of Lung Cancer* 2015, 10(8):1133-1141.

18. Ringner M, Jonsson G, Staaf J: Prognostic and Chemotherapy Predictive Value of Gene-Expression Phenotypes in Primary Lung Adenocarcinoma. *Clin Cancer Res* 2016, 22(1):218-229.

19. Cooke M, Magimaidas A, Casado-Medrano V, Kazanietz MG: Protein kinase C in cancer: The top five unanswered questions. *Mol Carcinog* 2017, 56(6):1531-1542.

20. Newton AC: Protein kinase C as a tumor suppressor. *Semin Cancer Biol* 2018, 48:18-26.

21. Zhang LL, Cao FF, Wang Y, Meng FL, Zhang Y, Zhong DS, Zhou QH: The protein kinase
C (PKC) inhibitors combined with chemotherapy in the treatment of advanced non-small cell lung cancer: meta-analysis of randomized controlled trials. *Clinical & translational oncology : official publication of the Federation of Spanish Oncology Societies and of the National Cancer Institute of Mexico* 2015, 17(5):371-377.

22. Denning MF, Dlugosz AA, Threadgill DW, Magnuson T, Yuspa SH: Activation of the epidermal growth factor receptor signal transduction pathway stimulates tyrosine phosphorylation of protein kinase C delta. *J Biol Chem* 1996, 271(10):5325-5331.

23. Mascia F, Denning M, Kopan R, Yuspa SH: The black box illuminated: signals and signaling. *The Journal of investigative dermatology* 2012, 132(3 Pt 2):811-819.

24. Newton AC: Protein kinase C: perfectly balanced. *Crit Rev Biochem Mol Biol* 2018, 53(2):208-230.

25. Wang MT, Holderfield M, Galeas J, Delrosario R, To MD, Balmain A, McCormick F: K-Ras Promotes Tumorigenicity through Suppression of Non-canonical Wnt Signaling. *Cell* 2015, 163(5):1237-1251.

26. Chen S, Wang Y, Zhang Y, Wan Y: Low expression of PKCalpha and high expression of KRAS predict poor prognosis in patients with colorectal cancer. *Oncology letters* 2016, 12(3):1655-1660.

27. Lee YH, Martin-Orozco N, Zheng P, Li J, Zhang P, Tan H, Park HJ, Jeong M, Chang SH, Kim BS et al: Inhibition of the B7-H3 immune checkpoint limits tumor growth by enhancing cytotoxic lymphocyte function. *Cell Res* 2017, 27(8):1034-1045.

28. Inamura K, Yokouchi Y, Kobayashi M, Sakakibara R, Ninomiya H, Subat S, Nagano H, Nomura K, Okumura S, Shibutani T et al: Tumor B7-H3 (CD276) expression and smoking history in relation to lung adenocarcinoma prognosis. *Lung cancer (Amsterdam, Netherlands)* 2017, 103:44-51.

29. Maeda J, Hirano T, Ogiwara A, Akimoto S, Kawakami T, Fukui Y, Oka T, Gong Y, Guo R,
Inada H et al: Proteomic analysis of stage I primary lung adenocarcinoma aimed at individualisation of postoperative therapy. Br J Cancer 2008, 98(3):596-603.

30. Vahid S, Thaper D, Gibson KF, Bishop JL, Zoubeidi A: Molecular chaperone Hsp27 regulates the Hippo tumor suppressor pathway in cancer. Scientific reports 2016, 6:31842.

31. Edfors F, Danielsson F, Hallstrom BM, Kall L, Lundberg E, Ponten F, Forsstrom B, Uhlen M: Gene-specific correlation of RNA and protein levels in human cells and tissues. Mol Syst Biol 2016, 12(10):883.

32. Franks A, Airoldi E, Slavov N: Post-transcriptional regulation across human tissues. PLoS Comput Biol 2017, 13(5):e1005535.

33. Myhre S, Lingjaerde OC, Hennessy BT, Aure MR, Carey MS, Alsner J, Tramm T, Overgaard J, Mills GB, Borresen-Dale AL et al: Influence of DNA copy number and mRNA levels on the expression of breast cancer related proteins. Molecular oncology 2013, 7(3):704-718.

Tables

Table 1. Patient characteristics

| Condition      | Number |
|----------------|--------|
| **EGFR**       |        |
| Mutated (WT)   | 9 (71) |
| **KRAS**       |        |
| Mutated (WT)(na) | 31 (48)(1) |
| **TP53**       |        |
| Mutated (WT)   | 31 (49) |
| Smoking history|        |
| Never (current or former) | 9 (71) |
| Stage          |        |
| Ia/ib          | 44     |
| IIa/IIb        | 17     |
| IIIa           | 17     |
| na             | 2      |

Table 1. Mutation status, smoking history and stage for the patients analyzed for protein
expression (n=80). Samples not mutated, wild type (WT), are shown in brackets.

Table 2. Proteins associated with PFS

| Protein            | Estimate | P-value |
|--------------------|----------|---------|
| c-Abl              | -0.386   | 0.0004  |
| PKC-βII_pS660      | 0.361    | 0.0010  |
| MIF                | -0.359   | 0.0011  |
| PAI-1              | -0.348   | 0.0016  |
| PKC-α_pS657        | 0.337    | 0.0023  |
| Jak2               | 0.318    | 0.0041  |
| JNK_pT183_Y185     | -0.311   | 0.0051  |
| PKC-α              | 0.310    | 0.0052  |
| PKC-δ_pS664        | 0.310    | 0.0052  |
| Vimentin           | 0.307    | 0.0056  |
| eEF2K              | 0.300    | 0.0069  |
| RPA32              | 0.296    | 0.0077  |
| Caveolin1          | 0.296    | 0.0077  |
| LC3A-B             | -0.293   | 0.0083  |
| PI3K-p85           | 0.289    | 0.0093  |
| B-Raf              | 0.285    | 0.0105  |
| Pdcd-1L1           | 0.283    | 0.0110  |
| Fibronectin        | -0.280   | 0.0117  |
| Rictor             | 0.279    | 0.0123  |
| B7-H3              | -0.273   | 0.0145  |

Table 2. Top twenty proteins associated with PFS (p<0.05) in the total group of samples (n=80).

Table 3. The pair-wise spearman Rank correlation between mRNA data and protein data
Table 3. The pair-wise spearman Rank correlation was performed on 79 samples, while the PFS analysis performed on mRNA data contained 186 samples. The overlap between PFS associated mRNA data and protein data are based on a threshold of p-value < 0.05.

Supplemental File Legend

S1 Fig. Hierarchical clustering of subcluster 2

S2 Fig. Hierarchical clustering of genes corresponding to the proteins included in RPPA

S3 Fig. Proteins associated with PFS in groups stratified on mutational status and smoking status

S4 Fig. Waterfall plot shows the correlation between protein expression (included proteins in phosphorylated state) and mRNA expression. Y-axis displays the Spearman’s rho coefficient. The proteins/genes are distributed on x-axis ordered after degree of correlation. The 10 overlapping PFS associated proteins/genes are displayed with black spikes.

S1 Table. List of proteins and the types of antibodies used in the RPPA analysis
S2 Table. List of antibodies used for IHC

S3 Table. Proteins significantly differentially expressed (FDR<0.05) in samples from subcluster 2a compared to the other subclusters

S4 Table. Proteins significantly associated with PFS in the total group, in subgroup 2a, and after stratification for smoking- and mutational status

S5 Table. IHC staining of neuroendocrine markers in samples from subcluster 2a

S6 Table. Spearman Rank correlation between gene expression and expression of proteins/phosphoproteins

Figures
Unsupervised hierarchical cluster analysis based on protein expression from 80 NSCLC samples. Clinical variables such as smoking status, mutations status of the gene TP53, KRAS and EGFR, PFS (ranging from one month = light green to dark green = 60 months) and event, were included to see if these features were enriched within the clusters. Events were divided into four categories; no event, relapse (which also includes metastasis), dead of lung cancer and dead of other reasons. The samples clustered into four subclusters marked with green, red, turquoise, and blue branches. Beneath subcluster two (red branch) a red box indicates a smaller cluster named 2a, and a green box is drawn beneath subcluster 2b, 2c and 2d. Subcluster 2a is recognized with poor PFS and enriched with subtype PP.
Kaplan Meier survival plot shows that subcluster 2a had worse PFS compared to the other subclusters (p=0.003). The different subclusters are generated from the hierarchical clustering in Fig 1. Subcluster 2b contains 2b, 2c and 2d.
Figure 3

A: Kaplan Meyer survival plot demonstrate reduced relapse free survival in patients with PP subtype. B: Kaplan Meyer survival plot shows that low levels of phosphorylated PKCd (p= 0.0683, log rank test) are associated with reduced relapse free survival. The molecular subtypes TRU, PI and PP display different levels of C: phosphorylated PKCδ, D: phosphorylated PKCα and E: PKCα. These results are calculated from LUAD samples in TCGA.

Supplementary Files

This is a list of supplementary files associated with the primary manuscript. Click to download.

Fig_S3_waterfall.eps
Fig_S2_hclust_mRNAmod.eps
supplementary material and methods.docx
Supplementary Tables.docx
Fig_S1_hclust.eps
Fig_S4_PFSprot overlap.eps
