

SCIENTIFIC REPORTS



OPEN

Lipidomic Profiling of Lung Pleural Effusion Identifies Unique Metabotype for EGFR Mutants in Non-Small Cell Lung Cancer

Received: 31 May 2016
Accepted: 26 September 2016
Published: 14 October 2016

Ying Swan Ho^{1,*}, Lian Yee Yip^{1,*}, Nurhidayah Basri¹, Vivian Su Hui Chong¹, Chin Chye Teo¹, Eddy Tan¹, Kah Ling Lim², Gek San Tan², Xulei Yang³, Si Yong Yeo³, Mariko Si Yue Koh⁴, Anantham Devanand⁴, Angela Takano⁴, Eng Huat Tan⁵, Daniel Shao Weng Tan^{5,6} & Tony Kiat Hon Lim^{2,6}

Cytology and histology forms the cornerstone for the diagnosis of non-small cell lung cancer (NSCLC) but obtaining sufficient tumour cells or tissue biopsies for these tests remains a challenge. We investigate the lipidome of lung pleural effusion (PE) for unique metabolic signatures to discriminate benign versus malignant PE and EGFR versus non-EGFR malignant subgroups to identify novel diagnostic markers that is independent of tumour cell availability. Using liquid chromatography mass spectrometry, we profiled the lipidomes of the PE of 30 benign and 41 malignant cases with or without EGFR mutation. Unsupervised principal component analysis revealed distinctive differences between the lipidomes of benign and malignant PE as well as between EGFR mutants and non-EGFR mutants. Docosapentaenoic acid and Docosahexaenoic acid gave superior sensitivity and specificity for detecting NSCLC when used singly. Additionally, several 20- and 22- carbon polyunsaturated fatty acids and phospholipid species were significantly elevated in the EGFR mutants compared to non-EGFR mutants. A 7-lipid panel showed great promise in the stratification of EGFR from non-EGFR malignant PE. Our data revealed novel lipid candidate markers in the non-cellular fraction of PE that holds potential to aid the diagnosis of benign, EGFR mutation positive and negative NSCLC.

Lung cancer is the leading cause of cancer-related mortality worldwide, with non-small cell lung cancer (NSCLC) being the predominant form of the disease, accounting for ~85–90% of all cases^{1,2}. Currently, small biopsy and cytological examination of malignant cells forms a cornerstone in the diagnosis of lung cancer³. Beyond establishing a definitive diagnosis of NSCLC, the American College of Chest Physicians recommends such diagnostic workup to classify the subtypes of lung cancer (e.g. adenocarcinoma versus squamous) and the molecular status based on specific gene mutations as that would facilitate the prescription of personalized therapy for individual patients by clinicians^{3,4}. In particular, specific mutations in the epidermal growth factor receptor (EGFR) have been reported to be one of the top driver oncogenes in NSCLC, with a prevalence of 9–23%⁵. NSCLC patients harbouring EGFR mutations are known to be more responsive to tyrosine kinase inhibitors (TKIs) such as gefitinib and erlotinib, making such medications the first-line therapy instead of standard chemotherapy⁶.

Adequate tumour cell and tissue acquisition is paramount for diagnosis, histologic and molecular characterization of NSCLC. However, the limited availability of cells and tissues for these diagnostic workups is an on-going

¹Bioprocessing Technology Institute Agency for Science, Technology and Research (A*STAR), 20 Biopolis Way, #06-01 Centros, Singapore 138668, Singapore. ²Department of Pathology Singapore General Hospital 20 College Road, Academia, Level 10, Singapore 169856, Singapore. ³Institute of High Performance Computing, Agency for Science, Technology and Research (A*STAR), 1 Fusionopolis way, #16-16 Connexis, Singapore 138632, Singapore. ⁴Department of Respiratory and Critical Care Medicine, Singapore General Hospital, Outram Rd, Singapore 169608. ⁵National Cancer Centre, 11 Hospital Drive, Singapore 169610, Singapore. ⁶Genome Institute of Singapore, Agency for Science, Technology and Research (A*STAR), 20 Biopolis Way, #06-01 Centros, Singapore 138668, Singapore. *These authors contributed equally to this work. Correspondence and requests for materials should be addressed to Y.S.H. (email: ho_ying_swan@bti.a-star.edu.sg) or D.S.W.T. (email: daniel.tan.s.w@singhealth.com.sg) or T.K.H.L. (email: lim.kiat.hon@singhealth.com.sg)

challenge for pathologists. The presence of pleural effusion (PE) is commonly observed in the initial diagnosis of lung cancer with an occurrence of 7 to 30% of all lung cancer cases⁷. Cytology of malignant cells from PE facilitates diagnosis of NSCLC. Lung PE, however, can also be a manifestation of benign inflammatory conditions including pneumonia, tuberculosis and pulmonary disorders. While cytological examination of PE aids in diagnosis of NSCLC, it is noteworthy that the diagnostic performances of cytology is dependent on the tumour type, tumour burden in the pleural space and the cytologist's expertise⁸. In PE samples with low cell yields, diagnostic accuracy can be compromised, resulting in false-negative rates of more than 30%^{9,10}. Consequently, the identification of molecular biomarkers independent of tumour cell and tissue is pertinent to complement and circumvent the challenge in the diagnosis of NSCLC.

To date, several efforts to identify alternative strategies using PE supernatant to complement the diagnostic workup for NSCLC have been reported. Several of these studies examine the diagnostic potential of protein-based molecules (e.g. carcinoembryonic antigen (CEA) and α -fetoprotein) and carbohydrate antigens (e.g. CA125 and CA 9-9), which are commonly present in other cancer types^{11–13}. However, the outcomes of using abovementioned protein molecules as candidate malignancy markers are mixed as studies revealed large variations in their sensitivities for diagnosis. At present, molecular characterization of EGFR mutations is performed using DNA extracts from tumour tissue specimens. While DNA extracted from malignant PE supernatant has been suggested as an alternative, the large variation in the quantity and quality of the DNA present in such samples can compromise the sensitivity of this approach. As observed in a recent study performed by Liu and co-workers, the malignant PE supernatant was reported to have a low sensitivity of 63.6% in comparison to the tumour tissue¹⁴. Henceforth, the authors recommended that DNA extracted from malignant PE supernatant should not be used for mutation testing if tumour tissue is available. Nevertheless, it is noteworthy that other studies reported favourable clinical usefulness of cell-free DNA from blood for the detection of EGFR mutations as supported by promising diagnostic performance (sensitivity up to ~80% and specificity of 100%)^{15,16}. Collectively, these studies demonstrated some of the efforts to identify surrogate markers in liquid biopsy that are indicative of the mutation subtype so as to overcome the issue of limited availability of tumour tissue for genotyping.

More recently, metabolic reprogramming involving deregulated cellular energetics of tumour is enshrined as one of the hallmarks of cancer^{17,18}. In the context of NSCLC, metabolic profiling using NMR detected elevated levels of branched chain amino acids, lactate and alanine and suppressed levels of glucose, trimethylamine-N-oxide and creatinine in malignant PE¹⁹. Using an untargeted LC-MS approach, Lam *et al.* observed significant elevation of fatty acids FA(16:0), FA(18:0), FA(18:1) and FA(18:2), as well as a decrease in a ceramide species Cer(d18:1/16:0) in malignant PE²⁰. Oleic acid (FA(18:1)) was found to be the best individual differentiator between benign tuberculosis PE samples and malignant lung adenocarcinoma PE, with an area-under-curve (AUC) value of 0.962 from receiver operating characteristic (ROC) analysis.

While these metabolic studies illuminated the potential of using PE-derived metabolites as markers for malignancy in NSCLC, the alterations in the metabolic phenotype (metabotype) in PE and their utility in the diagnosis of NSCLC is not well studied. The identification of fatty acid marker metabolites in the study by Lam *et al.* coincide with the lipid reprogramming phenomenon in cancer biology that is gaining increasing recognition in recent years^{20,21}. Collectively, this underscored the value in scrutinizing the lipidome of PE in NSCLC diagnosis. It is noteworthy that while interesting lipid candidate markers were identified in the Lam *et al.* study, their extraction method (using 1:1:1 v/v/v of acetone/ethanol/methanol) is not well optimized to interrogate the PE lipidome. Additionally, there has been no other follow up studies to evaluate the validity of the key findings in additional patient cohorts. Therefore, the primary objective of this study is to compare the lipidomes of the non-cellular fraction of lung PE derived from benign and malignant patients. To our knowledge, there is no study to date that investigates the utility of small molecules in biofluid as potential surrogate markers for the stratification of mutational subtypes in lung cancer. Hence, our secondary objective aims to identify lipid species with the ability to distinguish malignant PE from NSCLC patients with and without EGFR mutations.

Results

Pleural effusion lipidomes. We undertook an in-depth LC-MS based lipidomic analysis of 71 PE samples (30 benign, 41 malignant—19 EGFR mutant, 17 non-EGFR mutant and 5 with unknown EGFR status) (Table 1). The enrolled subjects comprising of 61 Chinese, 5 Malay, 4 Indian and 1 other ethnicity, generally reflecting the predominant distribution of the major ethnic group in Singapore. No statistically significant differences were identified in terms of age (p-value = 0.09), gender (p-value = 0.15), smoking status (p-value = 0.46) and ethnicity (p-value = 0.25) between the benign and malignant patients.

From the lipidomic analysis of benign and malignant PE samples, we detected a diverse range of lipid species in the human PE (Table S1, Figure S1). These species were identified based on authentic standards or by comparing the characteristic MS² spectra of the respective lipid classes with online mass spectral databases²². The list of identified lipids includes long chain fatty acids, sphingolipids, phospholipids and triacylglycerols.

Lipidomic analysis highlighted key differences in benign and malignant PE. To compare the composition of the lipidomes between the benign and malignant patients, we performed PCA analysis to identify any intrinsic clustering pattern of the PE samples. The PCA analysis of the PE lipidomes revealed distinctive clustering of the benign and malignant cases, indicating the existence of metabolic differences between these two groups (Fig. 1A). Interestingly, closer scrutiny of the PCA scores plot for the malignant PE revealed further compositional differences in the lipidomes between the EGFR mutant and non-EGFR mutant groups within the malignant class (Fig. 1B).

In view of the heterogeneity in the malignant lipidomes between non-EGFR and EGFR mutants, we performed two separate pairwise supervised multivariate analyses using OPLS-DA subsequently to build models to identify potential lipid differentiators with VIP > 1 between: (i) benign vs EGFR mutant and (ii) benign vs

Characteristic	All Patients (n = 71)	Benign (n = 30)	Malignant (n = 41)
Gender			
Male, %	35 (49.3)	18 (60.0)	17 (41.5)
Female, %	36 (50.7)	12 (40.0)	24 (58.5)
Age at sample collection, years			
Mean \pm standard deviation	67 \pm 14	63 \pm 16	69 \pm 12
Smoking status			
Smoker (Current/Ex)	25	9	16
Non smoker	46	21	25
Ethnic group			
Chinese, %	61 (86.0)	25 (83.3)	36 (87.8)
Malay, %	5 (7.0)	1 (3.3)	4 (9.8)
Indian, %	4 (5.6)	3 (10.0)	1 (2.4)
Others, %	1 (1.4)	1 (3.3)	0 (0.0)
Histology			
Non-small cell lung adenocarcinoma	39 (54.9)		
Squamous-cell carcinoma	1 (1.4)		
Lymphoepithelioma-like lung carcinoma	1 (1.4)		
Non-malignant	30 (42.3)		
Cytology assessment			
Positive for malignancy	32 (45.0)		
Negative for malignancy	30 (42.3)		
Suspicious/atypical confirmed as malignant based on histology	9 (12.7)		
Mutation subtypes for malignant cases			
EGFR +			19 (46.3)
EGFR –		N.A.	17 (41.5)
Unknown			5 (12.2)
Subtypes for non-malignant cases			
Pneumonia		22 (73.3)	
Cardiopulmonary congestion		5 (16.7)	N.A.
Tuberculosis		3 (10.0)	

Table 1. Clinical characteristics of benign subjects (n = 30) and malignant non-small cell lung cancer patients (n = 41). n, Number of cases; EGFR, Epidermal growth factor receptor.

non-EGFR mutant cases. Additionally, each pairwise multivariate analysis was supplemented by the use of univariate statistical tools (Mann-Whitney U test, fold change analysis). From these analysis, we identified 45 lipid species satisfying the following criteria: VIP > 1²³, p-value < 0.05 and an average ratio \geq 1.5, in at least one of the pair-wise comparison analysis (Fig. 1C).

Unsaturated fatty acids, phospholipids and sphingolipids constituted some of the major lipid classes discriminating between the malignant from benign PE samples (Figs 1C, and S2). Within the benign patients, we observed no clear differentiation in the abundance of these lipid markers between the benign infective (pneumonia and tuberculosis) and benign non-infective (cardiopulmonary congestion) PE samples for markers indicated in the heat map. These lipid species, however, were significantly elevated in the malignant PE of NSCLC patients compared with benign PE. Consistent with the earlier observation of a heterogeneous malignant lipidomic profile, the heat map analysis further illustrated the metabolic differences present in malignant lung PE associated with their genotype (presence/absence of EGFR mutation) (Fig. 1C). For instance, ether-linked phospholipids such as PC(P-36:5) and PEtn(P-38:5) were found to be statistically different between benign and EGFR mutant PE samples but not between benign and non-EGFR mutant cases. Conversely, glycosylated ceramide species including Gb3(42:2) and Gb3(34:1) were found to be significantly elevated in non-EGFR mutant cases, but not in PE cases with EGFR mutation.

To account for the unique metabolite dependent on the driver mutation for NSCLC, we screened for only lipid species satisfying both the abovementioned criteria for “benign vs EGFR mutant” and “benign vs non-EGFR mutants” comparisons as candidate malignancy markers (Table 2). These identified candidates include a large number of unsaturated fatty acids, in addition to specific ceramide and sphingomyelin species. Subsequently, we assessed the diagnostic performance of this panel of 12 malignancy lipid species using ROC curve analysis (Fig. 2A,B, Table 2). Each of these lipid malignancy markers was able to discriminate between the malignant and benign groups with AUC values ranging from 0.66–0.87, sensitivity (SN) of 63.3–82.9%, specificity (SP) of 60.0–83.3% and accuracy (ACC) of 64.8–83.1%. Individual ROC analysis performed on each candidate indicated that the polyunsaturated fatty acids (PUFAs) (e.g. FA(22:5), FA(22:6)) gave the best performance as malignancy markers and should be further evaluated in larger scale studies. Combining an optimal number of four malignancy markers comprising of FA(22:6), FA(22:5), FA(23:0) and Gb3(42:2) derived from SVM modelling into a

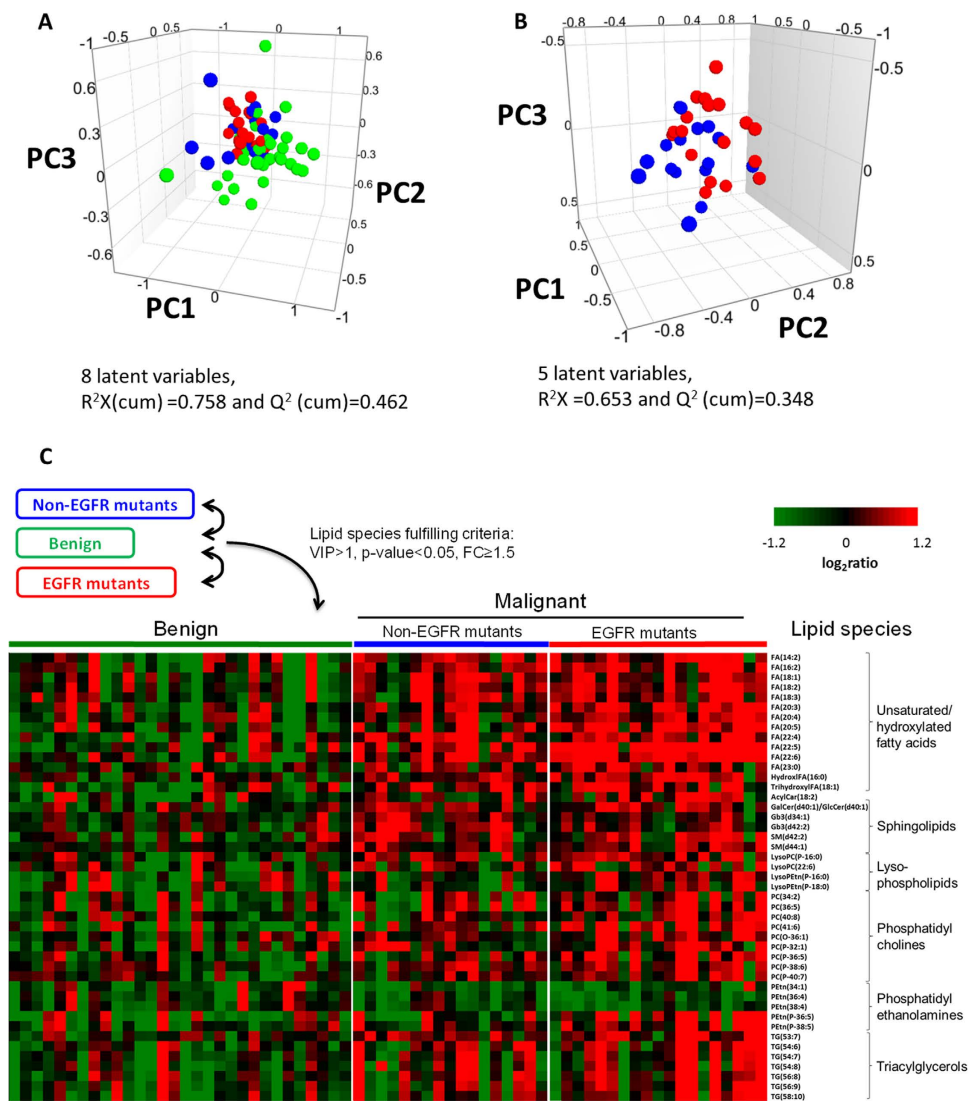


Figure 1. Principal component analysis scores plots for (A) benign ($n = 30$) and malignant lung PE ($n = 36$), and the PE of (B) EGFR mutant ($n = 19$) and non-EGFR mutant ($n = 17$) collected from the malignant patients. Green, blue and red circles represent the benign, malignant non-EGFR mutant and malignant EGFR mutant PE respectively. (C) Heat map of differential lipid metabolites derived from individual pairwise comparisons between benign, non-EGFR mutant and EGFR mutant PE samples. All represented species are statistically significant ($VIP > 1$, $p\text{-value} < 0.05$, fold change (FC) ≥ 1.5) for at least one of the pairwise comparisons. Metabolites are grouped according to their lipid classes.

single panel gave a comparable performance than when using the biomarkers alone ($AUC = 0.94$; $SN = 82.9\%$; $SP = 90.0\%$; $ACC = 85.9\%$) (Fig. 2C).

EGFR mutants exhibited greater metabolic derangement compared to non-EGFR mutants.

From the earlier analyses, we noticed that the candidate malignancy markers were predominantly in higher abundance in the PE of the EGFR mutants compared to the non-EGFR mutants (Fig. 1C). The most prominent observations were the increase in levels of PUFAs comprising of 20 and 22 carbons, as well as ether-linked phosphatidylethanolamine (PEtn) and its corresponding lysoPEtn species in EGFR mutants. These trends are exemplified in Fig. 3A,B for FA(22:6) and FA(20:5), illustrating that the EGFR mutants exhibit greater derangement in lipid metabolism compared to non-EGFR mutants. To further understand the influence of NSCLC driver mutations (EGFR/non-EGFR) on the lipidomic profile, we performed a combination of Mann-Whitney U test, OPLS-DA analysis and ROC analyses to identify lipid species that discriminated the EGFR mutants from the non-EGFR mutants (Table 3). Interestingly, we observed that PUFAs comprising of 20 and 22 carbons and PEtn species were significantly elevated in EGFR mutants by 1.6 to 2.6-fold compared to non-EGFR mutant cases (Table 3). Each of these lipid species was able to discriminate between the EGFR mutants and non-EGFR mutants with AUC ranging from 0.67–0.78, SN of 63.2–89.5%, SP of 58.8–82.4% and ACC of 66.7–77.8% (Table 3, Figs 3C,D). A combination of an optimal number of 7 lipid species comprising of FA(20:5), FA(22:4), FA(22:5),

Lipid name ^a	Benign vs non-EGFR mutants	Benign vs EGFR mutants	Diagnostic performance as malignancy marker			
	Ratio ^b	Ratio ^b	AUC ^c	SN (%)	SP (%)	ACC (%)
<i>Unsaturated/hydroxylated fatty acids</i>						
Hydroxyl FA(16:0) [(R)-2-Hydroxyhexadecanoic acid]	1.56	1.83	0.74 (0.62–0.85)	73.17	63.33	69.01
FA(14:2) [5,8-Tetradecadienoic acid]	1.64	2.54	0.77 (0.65–0.88)	78.05	70.00	74.65
FA(18:1) [Oleic acid]	1.80	1.71	0.76 (0.65–0.88)	70.73	73.33	71.83
FA(18:2) [Linoleic acid]	1.65	1.95	0.81 (0.70–0.91)	80.49	76.67	78.87
FA(18:3) [Linolenic acid]	1.53	1.81	0.74 (0.63–0.86)	70.73	70.00	70.42
FA(20:5) [Eicosapentaenoic acid]	1.82	4.43	0.79 (0.68–0.90)	75.61	73.33	74.65
FA(22:4) [Adrenic acid]	2.20	2.33	0.80 (0.69–0.90)	75.61	73.33	74.65
FA(22:5) [Docosapentaenoic acid]	2.46	6.11	0.87 (0.79–0.96)	82.93	83.33	83.10
FA(22:6) [Docosahexaenoic acid]	1.89	3.17	0.87 (0.79–0.95)	82.93	83.33	83.10
<i>Sphingolipids</i>						
GalCer(d40:1)/GlcCer(d40:1)	1.82	1.51	0.72 (0.60–0.85)	80.49	60.00	71.83
SM(d44:1)	1.53	1.53	0.73 (0.62–0.85)	63.33	69.01	64.79
SM(d42:2)	1.64	1.74	0.66 (0.54–0.79)	63.33	64.79	69.01

Table 2. Potential lipid malignancy markers for differentiating the PE of benign and malignant patients with NSCLC and their diagnostic performance. ^aIndividual abbreviated lipid names are provided based on the following convention: Lipid class (total number of carbons: total number of double bonds). ^bRatio calculated relative to benign. ^cAUC value obtained based on receiver operating characteristic (ROC) analysis with 95% confidence interval range provided in parentheses. VIP, variable importance for projection value; AUC, area under curve for ROC analysis; SN, sensitivity; SP, specificity; ACC, accuracy; FA, fatty acid; GalCer/GlcCer, galactosylceramide/glucosylceramide; SM, sphingomyelin. Listed markers satisfy statistical threshold (ratio > 1.5, p-value < 0.05, VIP > 1) in both pair-wise comparisons between EGFR (n = 19) and non-EGFR mutant (n = 17) cases with benign PE (n = 30).

FA(23:0), PC(41:6), PEtn(38:4), Gb3(42:2) based on SVM modelling, resulted in a more superior performance compared to using the biomarkers alone (AUC = 0.86; SN = 84.2%; SP = 82.4%; ACC = 83.3%) (Fig. 3E).

Discussion

The motivation for this study stemmed from challenges in the diagnosis of NSCLC, which is highly dependent on the availability of tissue biopsy or cells for the standard testing of malignancy and mutation status (e.g. EGFR, ALK mutations). The limited tumour tissue and cell availability, the low and variable sensitivity of cytology (ranging from 43–91%) and the significant false-negative rate in the event of insufficient cell numbers from PE provided the impetus to investigate novel, complementary diagnostic markers^{4,10,24–26}. The utilisation of cell-free DNA from blood samples as a surrogate for tumour biopsy to determine EGFR mutation status represents one such promising alternative that is currently evaluated clinically to assist the diagnosis of NSCLC^{15,16}. Here, we adopted a lipidomic-based approach to screen the PE profile so as to identify lipid differentiators suggestive of malignancy.

The combination of multivariate and univariate analyses revealed clear differences in the lipidomes between benign and malignant PE (Fig. 1). More importantly, malignant PE cases with a known EGFR mutation exhibited a more distinctive metabolic phenotype in comparison to non-EGFR mutant cases, with higher abundance of several lipid classes relative to benign PE. This observation has important implications in the identification of reliable malignancy markers that will apply to both EGFR and non-EGFR mutant cases; the commonly used strategy of biomarker identification does not take into consideration the potential heterogeneity of metabolic phenotypes between malignant cases harbouring different driver mutations (e.g. EGFR mutant vs non-EGFR mutant). As such, to account for such heterogeneity and to ensure that reliable indicators of malignancy can be selected, separate pair-wise comparisons (benign vs EGFR mutants, benign vs non-EGFR mutants) were performed. Subsequently, only lipid species which satisfied the statistical selection criteria for both sets of pair-wise comparisons were identified as candidate markers for malignancy. This strategy appears to be effective - as shown in Table 2, individual ROC analysis yielded AUC values of 0.70 and above for the majority of these lipid species, indicating that each feature had good diagnostic performance in distinguishing between benign and malignant PE samples regardless of molecular subtypes.

A large group of unsaturated or hydroxylated fatty acids were found to be elevated in our malignant PE samples. In particular, these elevations of FA(18:1) and FA(18:2) in malignancy recapitulated Lam *et al.*²⁰ observations. Although our AUC values (ROC analysis) were comparatively lower to theirs (0.76–0.81 as compared to 0.962²⁰), we suggest that this may be due to the different underlying medical conditions of the benign controls. In our study, benign PE samples were attributed to multiple non-malignant causes (pneumonia, tuberculosis, cardiopulmonary etc.), whereas Lam *et al.* recruited solely tuberculosis subjects as benign controls. We believe that the selection of a diversified benign patient cohort (control) would render a more robust selection of PE-derived malignancy markers as PE can be attributed to various benign causes.

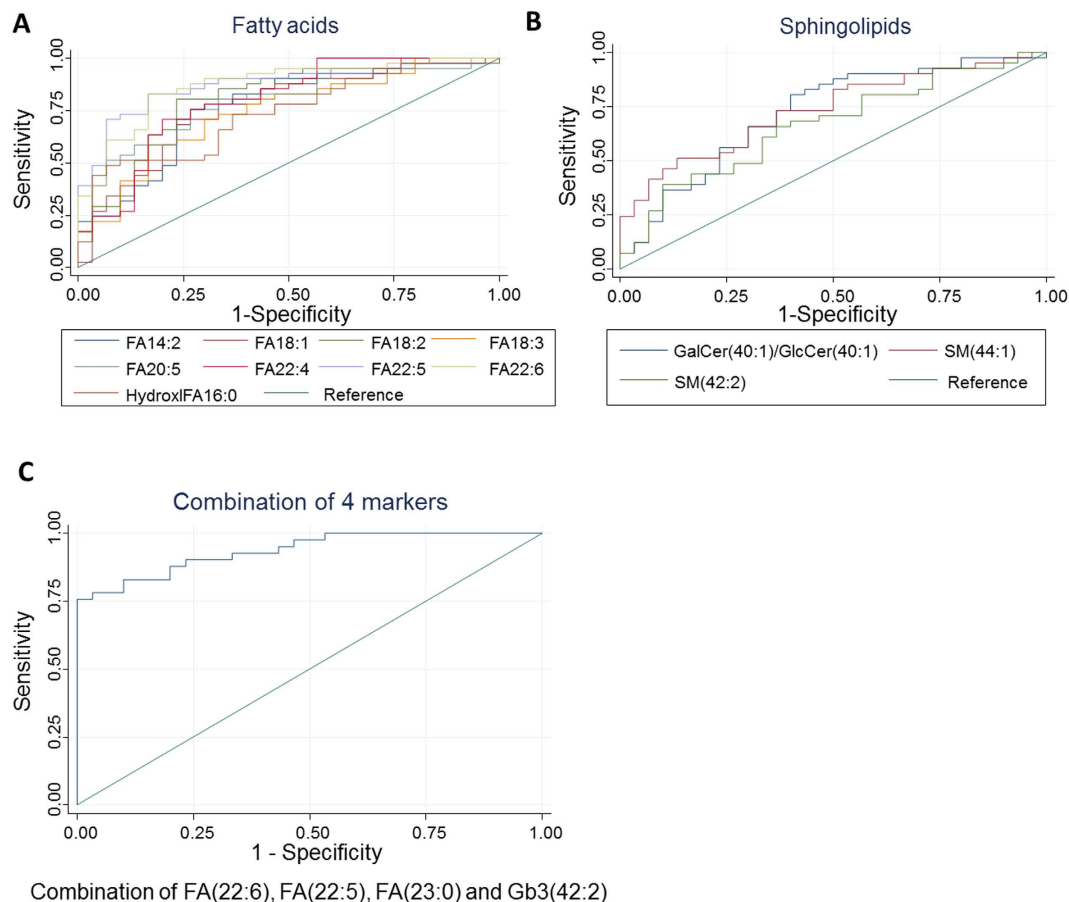


Figure 2. Diagnostic performance of lipid markers in discriminating pleural effusion from malignant NSCLC patients (n = 41) from benign subjects (n = 30). ROC curves of malignant versus benign subjects for individual PE lipid markers in the class of (A) fatty acids (B) sphingolipids. (C) ROC curves of malignant versus benign subjects for an optimal combination of 4 lipid malignancy markers from SVM modelling.

In the biological context of cancer metabolism, the phenomenon of increased fatty acids and phospholipids in malignant samples may be attributed to the induction of the key lipogenic enzymes, their nuclear receptors and transcriptional regulators. Elevated abundance of fatty acids is consistent with previous reports on prostate and breast cancers^{27,28}, and has been associated with the overexpression of fatty acid synthase (FAS). FAS overexpression is known to facilitate the *de novo* synthesis of fatty acids for the production of membrane phospholipids and energy production via beta-oxidation, and FAS is increasingly being recognised as a key trait which confers tumour growth and survival advantages^{29,30}. More specifically, the elevation of unsaturated fatty acids in our study also suggested an association with stearoyl-CoA desaturase (SCD) overexpression, which converts saturated fatty acids into unsaturated fatty acids. The observation corroborates with the reported increases in SCD-1 overexpression for both lung carcinoma cell lines³¹ and lung tumour-initiating cells³² in recent studies. In addition, liver X receptors (LXR) are a set of oxysterol-activated nuclear receptors involved in regulating cholesterol, fatty acid and glucose homeostasis^{33–38}. Sterol regulatory element binding protein 1 (SREBP-1) belongs to a family of membrane-bound transcription factors that can directly activate expression of genes implicated in fatty acids, triglycerides and phospholipid metabolism^{39,40}. Both LXR and SREBP-1 expression have been reported in human lung tissues and cell lines^{41–43}. LXRs can regulate lipogenesis through regulating SREBP-1 or by directly targeting genes such as FAS and SCD-1 downstream of SREBP-1C pathway^{33,34,38}. At the functional level, upregulation and activation of LXRs may explain the elevations of fatty acids and phospholipids. However, it is to be noted that there have been other reports of some PUFAs such as eicosapentaenoic acid (FA20:5) and docosahexaenoic acid (FA22:6) functioning as LXR antagonists^{39,44}. As such, further mechanistic studies will likely be required to determine the detailed associations between LXR and PUFAs in NSCLC. In a recent study, Dai *et al.*⁴⁵ found massive lung lipid accumulation, M1 macrophage-predominant lung inflammation that eventually proceed to lesions resembling peripheral squamous cell lung cancer when both LXR α and LXR β were inactivated in LXR α ; LXR β double null mutant mice models. Together with studies illustrating the anti-proliferative effect of LXR agonists in NSCLC^{45–47}, the findings underpin the importance of LXR signalling in lung cancer. Emerging evidence also suggest SREBP-1 is a critical link between oncogenic signalling and tumour metabolism⁴⁸. Aberrant activation of SREBP and induction of expression of its target genes has been found in several cancer types (e.g. breast, ovarian, prostate cancer), promoting cancer proliferation, migration and invasion⁴⁹. In certain subtypes of glioblastoma multiforme that express an activated mutant form of EGFR, high levels of nuclear SREBP-1 was observed⁵⁰. Considering the regulatory

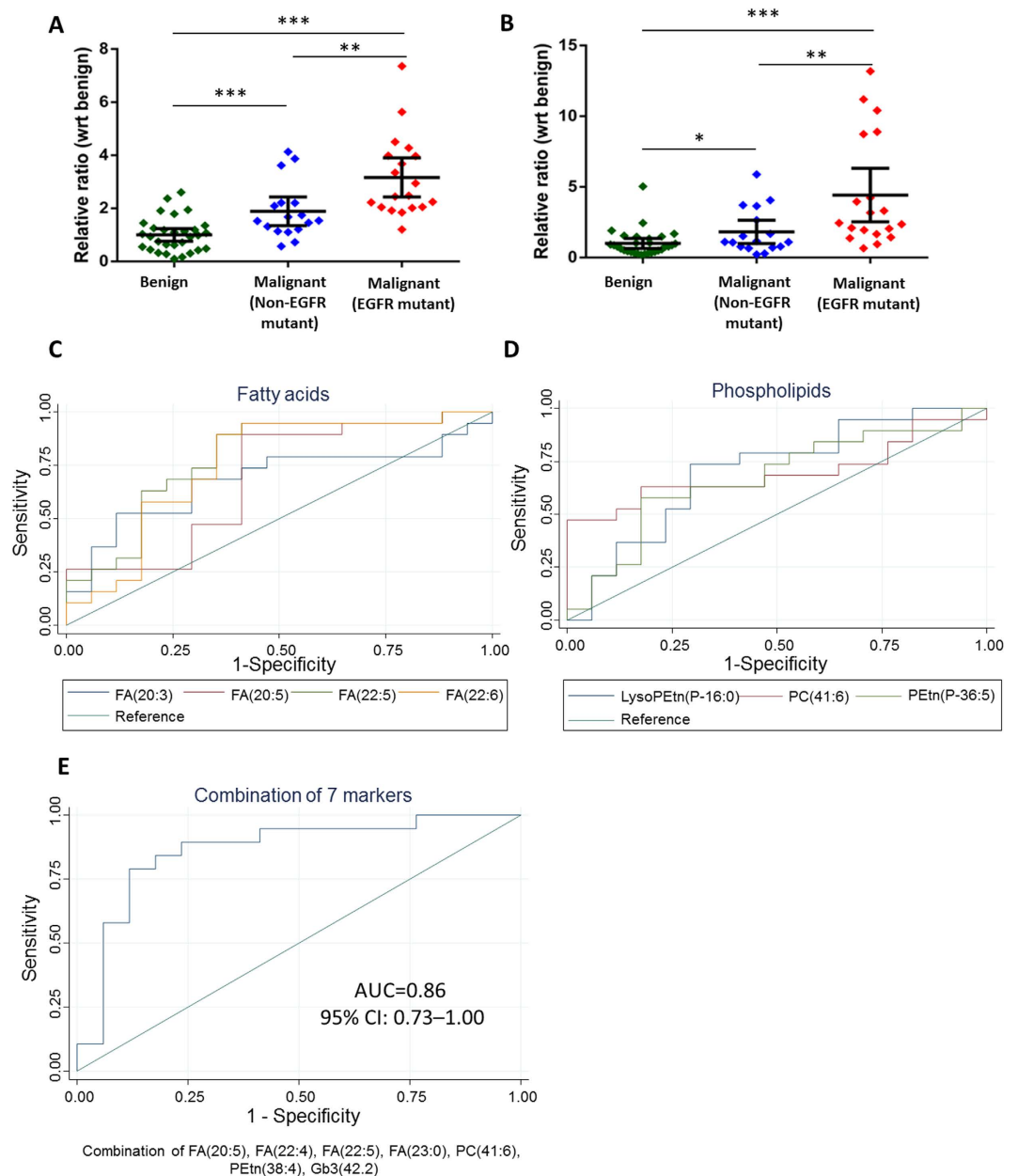


Figure 3. Dot plots describing the relative levels of (A) FA(22:6), (B) FA(20:5) in benign (green), non-EGFR mutant (blue) and EGFR mutant (red) PE samples. p-value calculated based on Mann-Whitney U test, where *denotes $p < 0.05$, **denotes $p < 0.01$, ***denotes $p < 0.001$. ROC curves of non-EGFR mutant versus EGFR mutant subjects for individual PE lipid markers in the class of (C) fatty acids and (D) phospholipids. (E) ROC curves of non-EGFR mutant versus EGFR mutant subjects for an optimal combination of 7 lipid malignancy markers from SVM modelling.

role of EGFR-PI3K-Akt signalling to activate SREBP-1^{48,50} and the inherent function of SREBP-1 to regulate lipid metabolism^{39,40}, it will be interesting to examine its association with the increase of PUFAs and phospholipids in NSCLC patients, particularly those observed in EGFR mutants in subsequent studies.

Dysregulation of sphingolipid metabolism also has important implications to cancer pathogenesis owing to their diverse structural and signalling roles in regulating cell proliferations, differentiation and apoptosis⁵¹. In this study, we observed elevated glycosylated ceramide and sphingomyelin species in malignancy. Ceramide, a central molecule in sphingolipid biosynthesis, is generally regarded as a powerful tumour suppressor for its ability to induce apoptosis and anti-proliferative response⁵¹. Conversely, glycosylation of ceramide by glucosylceramide synthase (GCS) into glucosylceramide and glycosphingolipids is known to stimulate tumour cell proliferation, inhibit apoptosis and promotes tumour cell survival⁵¹. Further, increase GCS activities have been implicated in the development of chemotherapy resistance in several types of cancer^{52,53}. Additionally, ceramide levels can also be attenuated through the activity of sphingomyelin synthase (SMS) that transfers phosphocholine headgroup

Non-EGFR vs EGFR mutants					
Lipid name ^a	Ratio (relative to non-EGFR)	AUC ^b	SN (%)	SP (%)	ACC (%)
<i>Polyunsaturated fatty acids</i>					
FA(20:3) [<i>Eicosatrienoic acid</i>]	1.67	0.68 (0.50–0.87)	68.42	70.59	69.44
FA(20:5) [<i>Eicosapentaenoic acid</i>]	2.43	0.68 (0.50–0.87)	89.47	58.82	75.00
FA(22:5) [<i>Docosapentaenoic acid</i>]	2.49	0.78 (0.62–0.94)	73.68	70.59	72.22
FA(22:6) [<i>Docosahexaenoic acid</i>]	1.68	0.75 (0.58–0.93)	89.47	64.71	77.78
<i>Phospholipids</i>					
LysoPEtn(P-16:0)	1.57	0.70 (0.52–0.88)	73.68	70.59	72.22
PC(41:6)	2.60	0.70 (0.51–0.88)	63.16	82.35	72.22
PEtn(P-36:5)	2.30	0.67 (0.48–0.85)	63.16	70.59	66.67

Table 3. Lipid candidates capable of distinguishing EGFR mutation status in the PE of NSCLC patients and their diagnostic performance. ^aIndividual abbreviated lipid names are provided based on the following convention: Lipid class (total number of carbons: total number of double bonds). ^bAUC value obtained based on receiver operating characteristic (ROC) analysis with 95% confidence interval range provided in parentheses. VIP, variable importance for projection value; AUC, area under curve for ROC analysis; SN, sensitivity; SP, specificity; ACC, accuracy; FA, fatty acid; LysoPEtn(P-), ether-linked lysophosphatidylethanolamine; PC, phosphatidylcholine; PEtn(P-), ether-linked phosphatidylethanolamine. Listed markers satisfy statistical threshold (ratio > 1.5, p-value < 0.05, VIP > 1) in both pair-wise comparisons of EGFR mutant cases with benign PE and non-EGFR mutant cases.

to ceramide forming sphingomyelin. The functions of SMS may however go beyond the regulation of ceramide levels. For instance, sphingomyelin isolated from shed membrane vesicles of tumour cells has been demonstrated to promote tumour growth, metastases, invasiveness, endothelial cell migration and angiogenesis⁵⁴. Interestingly, a comparison of human lung tumour tissues compared to their matched non-tumour tissues revealed lower SMS-1 expression⁵⁵ and sphingomyelin abundance in lung tumour tissues⁵⁶. SMS activation by anti-tumour 2-hydroxyoleic acid to A549 human lung adenocarcinoma cells has been found to increase SM at plasma membrane, modify lipid raft properties, enhance localization of protein involved in cell apoptosis or differentiation and trigger tumour cell death⁵⁷. On the other hand, ceramide-enriched rafts, generated from the displacement of cholesterol from the raft by membrane ceramide released from sphingomyelin, has been proposed to lead to re-localization of EGFR in the ceramide-enriched rafts, thereby stabilizing aberrant EGFR signalling^{58–63}. As such, the intricate balance between these sphingolipids is crucial in determining the fate of cancer cells due to their biologically active roles. Our identification of elevated glycosylated ceramide and sphingomyelin species in the malignant PE may hence reflect the lung cancer cells' attempt to regulate sphingolipid metabolism to neutralize the downstream cell death signals initiated by ceramide.

In our study, we observed a more distinctive metabolic phenotype for the EGFR mutants, characterised by trends of higher relative fold change and higher AUC values for the pair-wise comparison between benign and EGFR mutant cases in comparison to non-EGFR mutants. As such, we further scrutinize for differential lipids between EGFR mutants and non-EGFR mutants [satisfying the selection criteria for both pairwise comparisons (EGFR mutant vs benign cases, EGFR mutant vs non-EGFR mutant cases)] (Table 3). These lipid species belonged to two major lipid classes, including PUFAs and phospholipids. PUFAs have been previously implicated in the activation and phosphorylation of EGFR in human endothelial cells⁶⁴. More significantly, the same study reported that the EGFR-phosphorylating activity of fatty acids was associated with chain length and degree of saturation. PUFAs with longer chain lengths such as FA(20:5) and FA(22:6) were more effective in activating EGFR, while short chain, saturated fatty acids had no effect on the receptor. These results corroborated with the fatty acid species we have identified, which were comprised solely of 20 and 22 chain length PUFAs. Currently, the protective roles of PUFAs in cancer development remain controversial^{65–67}. Our observation revealed a positive association of higher PUFAs levels in PE of EGFR mutants. As it is possible that PUFAs in EGFR mutant cells might play a role in sustaining the higher EGFR activation rate, the mechanistic function of PUFAs in EGFR and non-EGFR driven NSCLC should be further elucidated to determine its effect on the growth inhibition or promotion of EGFR mutant tumours^{64,68}. Nonetheless, the biological relevance of these fatty acids substantiated their roles as biomarkers for diagnosis and patient stratification as illustrated by their diagnostic performance (e.g. AUC and ACC).

Long chain and ether-linked PEtn species, including lysoPEtn(P-16:0), were another lipid class in differential abundance between EGFR mutants and non-EGFR mutants. Ether-linked species, also known as plasmalogens, are reported to play pertinent physiological roles in intracellular signalling and confer protective roles during oxidant-induced stress⁶⁹. Elevated levels of plasmalogens have previously been found in various cancer tumours⁷⁰, as well as breast and melanoma cancer cell line models⁷¹. In the same study, the authors reported the heightened expression of alkylglyceronephosphate synthase (AGPS), a critical enzyme for the synthesis of ether-linked lipids, in both the cell lines and primary breast tumours. Subsequent knockdown of AGPS resulted in the impairment of migration, invasion, cell proliferation and survival in both cell lines. In our study, plasmalogen species appear to be generally higher in EGFR mutants when compared to both non-EGFR mutants and benign PE samples (Fig. 1C). This lipid signature suggests that a more active ether-linked lipid synthesis pathway might be present

in EGFR mutants and the plausible role of AGPS in the pathogenesis of EGFR mutation-driven NSCLC remains to be determined.

Based on our ROC analysis of the lipid markers identified in Tables 2 and 3, the lipid species displayed good performance diagnostics (e.g. AUC, SN, SP and ACC) when used singly. Using SVM, we selected an optimal number of lipid markers for use in combination from the list of 45 candidate malignancy markers in the heat map in Fig. 1C. By combining the lipid markers, we were able to achieve improvement in the diagnostic performance to stratify patients based on their PE lipid composition into the following groups: (i) benign vs malignant and (ii) EGFR vs non-EGFR mutants. Some of the lipid species such as FA(22:5), FA(22:6), FA(20:5) were identified as candidate diagnostic markers by both the pairwise comparison strategy and the SVM strategy, to be used either singly or in combination. This underscored their importance in NSCLC and/or the underlying driver mutation. Interestingly, the SVM modelling also unveiled other lipids species (Figs 2C and 3E) that contributed towards the stratification of the patients not identified using our criteria (Tables 2 and 3) which deserves further investigations.

There are specific strengths and limitations that have to be considered for this study. Using a global lipidomics approach, we were able to detect a broad range of structurally diverse lipid species and elucidate interesting and important biomarkers for the diagnosis of NSCLC that is currently unexplored. As PE can occur due to a variety of non-malignant causes, we recruited subjects admitted due to a wide range of underlying pathology (e.g. pneumonia, tuberculosis and cardiopulmonary congestion) as “benign” controls so as to identify biomarkers that can be used in a wide clinical context. In consideration of the unique metabotype of NSCLC patients with different molecular subtypes (EGFR- and non-EGFR mutants), we adopted a pairwise comparison approach to screen for lipid malignancy markers with good diagnostic performance instead of treating them as a homogenous group. A limitation, however, was that the small sample size for other driver mutations did not allow for the derivation of additional unique lipidomic signatures which correlated to their respective genotypes to be performed. The semi-quantitative nature associated with such untargeted studies also calls for follow-up targeted studies involving an independent validation cohort comprising of a larger number of patients, to verify the predictive accuracy of these novel candidate markers and their ability to complement cytological assessment of PE samples. Additionally, mechanistic validation studies may also be performed on EGFR mutant and non-EGFR mutant cell lines to elucidate the biological roles of these candidate malignancy lipids, as this will provide interesting mechanistic insights to substantiate their relevance in NSCLC.

In conclusion, the study yielded a series of lipid molecules, many of which are novel, that show promise as potential malignancy and/or EGFR mutant specific markers. In particular, 20 and 22 carbon chain length PUFAs had the highest potential in distinguishing between benign, EGFR mutant and non-EGFR mutant PE samples. Such biomarkers will not only facilitate diagnosis, but also assist in patient stratification for personalised therapy (e.g. administration of TKIs to patients with EGFR mutation). The use of SVM modelling was found to facilitate the selection of an optimal combination of biomarkers that showed improved diagnostic performance compared to using single biomarkers. Additionally, our work highlights the importance of taking different phenotypic behaviour of different genotypes into consideration in the selection of suitable malignancy candidate markers. Finally, information on the perturbation of lipid pathways in NSCLC gleaned from lipidomic profiling has shed new and established insights that further substantiate the importance of lipid metabolic reprogramming in cancer biology.

Patients and Methods

Study Patient and Sample Collection. Lung PE samples were obtained from 71 patients admitted to Singapore General Hospital and National Cancer Centre Singapore between December 2012 and December 2014. These PE samples collected via thoracocentesis were centrifuged at 805 g at 4 °C for 10 mins upon collection. The NSCLC malignant cases were assessed based on clinical diagnosis, cytology of the cells isolated from PE and histology of tumour tissues. Patients whose histological examinations did not show any malignancy were classified as benign subjects. The patient demographics and characteristics, including age, gender, cytology and histology are provided in Table 1. In our study, there were 30 benign and 41 malignant cases. Of the malignant cases, 19 cases were confirmed to harbour EGFR mutations (EGFR mutant), 17 cases did not possess EGFR mutations (non-EGFR mutant), while the molecular status of the remaining 5 malignant cases were unknown due to insufficient samples for mutation testing following cytology and histology examination. All the PE supernatants were stored at –80 °C prior to LC-MS-based lipidomic analysis. All patients recruited gave informed consent and this study was approved by and carried out in accordance with the guidelines of the Singapore Health Services centralized institutional review board (CIRB 2010/516/B).

Reagents and Materials. Reagents were obtained as follows: Optima grade methanol, isopropanol: Fisher Scientific (Loughborough, UK); tricine: Sigma-Aldrich (St Louis, MO); ammonia solution “AnalaR” 25%: VWR (Poole, UK); acetonitrile, chloroform, acetic acid, formic acid, ¹³C-labelled isotopic fatty acid standards (palmitoleic acid-¹³C₁₆, palmitic acid-1,2,3,4-¹³C₄, linoleic acid-¹³C₁₈, stearic acid-¹³C₁₈): Merck (Whitehouse Station, NJ); odd-chain lipid standards: phosphatidylcholine, PC(9:0/9:0); PC(17:0/17:0); PC(21:0/21:0); phosphatidylethanolamine, PEtn(15:0/15:0); PEtn(17:0/17:0): Avanti Polar Lipids (Alabaster, AL). The ¹³C-labelled isotopic and odd-chain lipid standards constitute the lipid reference standard mixture.

Lipid Extraction. 50 µL of each PE sample was aliquoted and 10 µL of lipid reference standard mixture was added to the sample as retention time reference prior to a two-phase modified Bligh and Dyer⁷² lipid extraction protocol. Each sample was extracted by sequential addition of methanol, chloroform and 3.8 mM tricine (1:1:0.5 v/v/v, total 2 mL), with sample vortexed for 1 min following each addition. The samples were then centrifuged at 12,000 g at 4 °C for 20 min, following which each sample separated into two fractions—the top methanolic layer contained the polar metabolites while the bottom chloroform layer contained the lipid species. The bottom

chloroform fraction enriched with lipid species was collected and stored at -80°C prior to analysis. Quality control (QC) samples were prepared by mixing equal amount of all the PE samples and each QC sample was extracted as described above. All samples were randomized for extraction.

Lipidomic Profiling using LC-MS. An ultra-high performance liquid chromatography (UHPLC) system (Ultimate 3000, ThermoFisher Scientific, MA) interfaced with a Q-Exactive mass spectrometer (ThermoFisher Scientific, MA) was utilized for lipidomic analysis. Prior to each analytical batch, instrument maintenance (source cleaning and mass calibration) was performed. A reversed phase LC column (Acquity CSH, 1.0×50 mm, $1.7\ \mu\text{m}$ particle size, Waters Corp) was used for separation with two solvents: 'A' comprising of acetonitrile, methanol and water (2:2:1) with 0.1% acetic acid and 0.1% ammonia solution, and 'B' comprising of isopropanol with 0.1% acetic acid and 0.1% ammonia solution. All samples were dried using a sample concentrator (Bio-technie, Minneapolis, MN), reconstituted in a 50:50 (v/v) mixture of solvents A and B. The UHPLC autosampler temperature was set at 4°C and the injection volume for each sample was $2\ \mu\text{L}$. All samples were processed in technical triplicates on the LC-MS system. The LC program is as follows: the column was first equilibrated for 1 min at 1% B with a flow rate of $0.1\ \text{ml}\ \text{min}^{-1}$. The gradient was increased from 1% B to 82.5% B over 9 min before B was increased to 99% for a 5 min wash at a flow rate of $0.15\ \text{ml}\ \text{min}^{-1}$. The column was re-equilibrated for 2.2 min at 1% B. Column temperature was maintained at 45°C . The eluent from the LC system was directed into the MS. Electrospray ionization (ESI) in the MS was conducted in both positive and negative modes in full scan with a mass range of 120 to 1800 m/z , resolution of 70,000, automatic gain control (AGC) target of 1×10^6 ions (ESI+) or 3×10^6 ions (ESI-) and maximum injection time (IT) of 100 ms (ESI+) or 200 ms (ESI-). The heated electrospray ionization (HESI) source used a spray voltage of 3.7 kV (ESI+) and 3.2 kV (ESI-), capillary temperature of 350°C (ESI+ and ESI-), sheath gas flow of 25 (ESI+ and ESI-) and auxiliary gas flow of 10 arbitrary units (ESI+ and ESI-).

QC samples were analysed at regular intervals throughout each batch analysis to monitor the reproducibility of the LC-MS. The extracted samples were re-randomized for LC-MS analysis such that the injection order was independent from the order of sample preparation to minimise systematic bias.

Data Pre-processing and Statistical Analysis. The raw LC-MS data obtained was then pre-processed and analysed using the XCMS peak finding algorithm⁷³. The spiked lipid reference standards had relative standard deviations of less than 20% across all samples, demonstrating the high reproducibility of our extraction and LC-MS method. The QC mixture was used for signal correction between and within each batch analysis. Mass peaks with poor repeatability within the QC samples (coefficient of variation more than 30%) were removed. Total area normalisation (based on ratio of area of each mass peak against sum of peak areas within each sample) was applied to the remaining features in the dataset to correct for minor variations in sample preparation and analysis. The normalised data were exported to SIMCA-P+ (version 13.0.3, Umetrics, Umea, Sweden) for multivariate data analysis to identify potential PE biomarkers.

Data were mean-centred and Pareto scaled in SIMCA-P+. Subsequently, an unsupervised principal component analysis (PCA) was utilised to determine the quality of LC-MS data obtained based on the tight clustering of the QC samples and to derive an overview of similarities and differences between individual PE samples. This was followed by supervised orthogonal projection to latent structures-discriminant analysis (OPLS-DA) where a model was built to identify individual lipid components that were distinctly different between (1) benign, (2) EGFR mutant and (3) non-EGFR mutant malignant cases in a pair-wise manner. These lipid components are identified based on their variable importance for projection (VIP) values. Lipid species with higher VIP ($\text{VIP} > 1$) made a greater contribution towards distinguishing the comparator groups in the OPLS-DA model and were considered as potential biomarkers. Univariate analysis was performed using the Mann-Whitney U test at p -value < 0.05 to verify the statistical significance of these potential biomarkers. Fold change was calculated by taking the ratio of the peak areas contributed by the lipid species of the two comparator groups.

Support Vector Machines Modelling. The support vector machines (SVMs) model, first proposed by Vapnik and his colleagues⁷⁴, is a widely used machine learning technique for pattern recognition. The SVMs construct a boundary that maximizes the distance between the designated class of each sample (e.g. whether the sample is "benign" or "malignant"). An optimal boundary separating the sample class is then defined. In this study, we used the popular libSVM package⁷⁵ with linear kernel function to perform the classification, where involved parameters are automatically selected by Bayesian Optimization⁷⁶. The recursive feature elimination (RFE) method, based on backward sequential selection strategy⁷⁷, was used to select the best features of the SVM classifier. Starting with a full candidate set of malignancy lipid markers, features (lipid markers) were removed sequentially such that the variation of separating boundary was minimized and until the desired number of features was reached. Different desired number of features was evaluated to determine the performance of the various feature combinations.

In the construction of a real pattern classification system, the data available are generally limited, such that there is a need for a validation technique to estimate how a classification system will perform in practice. In our study, the k -fold cross-validation⁷⁸ is used to estimate the classification performance. In a single round of k -fold cross-validation, the dataset is first randomly partitioned into k subsets (folds), which are of approximately equal size and are mutually exclusive. A SVM classifier is then trained and tested k times, and at each time, one of the subsets is set aside as the testing data while the remaining $k - 1$ subsets set as the training data. In our study, leave-one-out cross validation (i.e. $k = 71$) was used.

ROC analyses were then performed for the two optimal combinations of lipid markers capable of differentiating the PE between (i) the benign and malignant patients and (ii) non-EGFR and EGFR mutants. ROC analyses were also performed for the identified lipid species ($\text{VIP} > 1$, p -value < 0.05 , fold change ≥ 1.5). The ROC is plotted using Stata/MP 14.0 statistical package (Stata Corp, LP) based on the predicted real value of each sample from the trained SVM model.

Metabolite Identification. Mass peaks were first putatively identified based on mass comparison (less than 5 ppm error) with entries from the Kyoto Encyclopedia of Genes and Genome (www.genome.jp/kegg) and the Human Metabolome Database (www.hmdb.ca). Subsequently, the identities of lipid species of interest were verified by MS² spectral comparison with commercially available standards where possible, or by comparison to mass spectral databases available online.

References

- DeSantis, C. E. *et al.* Cancer treatment and survivorship statistics, 2014. *CA Cancer J Clin* **64**, 252–271, doi: 10.3322/caac.21235 (2014).
- Molina, J. R., Yang, P., Cassivi, S. D., Schild, S. E. & Adjei, A. A. Non-Small Cell Lung Cancer: Epidemiology, Risk Factors, Treatment, and Survivorship. *Mayo Clinic Proceedings* **83**, 584–594 (2008).
- Travis, W. D. *et al.* Diagnosis of lung cancer in small biopsies and cytology: implications of the 2011 International Association for the Study of Lung Cancer/American Thoracic Society/European Respiratory Society classification. *Arch Pathol Lab Med* **137**, 668–684, doi: 10.5858/arpa.2012-0263-RA (2013).
- Rivera, M. P., Mehta, A. C. & Wahidi, M. M. Establishing the diagnosis of lung cancer: Diagnosis and management of lung cancer, 3rd ed: american college of chest physicians evidence-based clinical practice guidelines. *Chest* **143**, e142S–e165S, doi: 10.1378/chest.12-2353 (2013).
- Siegelin, M. D. & Borczuk, A. C. Epidermal growth factor receptor mutations in lung adenocarcinoma. *Lab Invest* **94**, 129–137, doi: 10.1038/labinvest.2013.147 (2014).
- Zhou, C. *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. *Lancet Oncology* **12**, 735–742, doi: 10.1016/S1470-2045(11)70184-X10.1016/S14702045(11)70184-X (2011).
- Eleftheriou, D., Kallianos, A., Adrianopoulos, A. & Rapti, A. Evaluation of pleural effusion at first diagnosis of lung cancer. *J Thorac Dis* **4**, AB28 (2012).
- Porcel, J. M. & Light, R. W. Diagnostic approach to pleural effusion in adults. *Am Fam Physician* **73**, 1211–1220 (2006).
- Abouzgheib, W., Bartter, T., Dagher, H., Pratter, M. & Klump, W. A prospective study of the volume of pleural fluid required for accurate diagnosis of malignant pleural effusion. *Chest* **135**, 999–1001, doi: 10.1378/chest.08-2002 (2009).
- Motherby, H. *et al.* Diagnostic accuracy of effusion cytology. *Diagn Cytopathol* **20**, 350–357 (1999).
- Burgess, L. J. Biochemical analysis of pleural, peritoneal and pericardial effusions. *Clin Chim Acta* **343**, 61–84, doi: 10.1016/j.cccn.2004.02.002 (2004).
- Davies, H. E. *et al.* Clinical impact and reliability of pleural fluid mesothelin in undiagnosed pleural effusions. *Am J Respir Crit Care Med* **180**, 437–444, doi: 10.1164/rccm.200811-1729OC (2009).
- Gaspar, M. J., de Miguel, J., Garcia Diaz, J. D. & Diez, M. Clinical utility of a combination of tumour markers in the diagnosis of malignant pleural effusions. *Anticancer Res* **28**, 2947–2952 (2008).
- Liu, X. *et al.* The diagnostic accuracy of pleural effusion and plasma samples versus tumour tissue for detection of EGFR mutation in patients with advanced non-small cell lung cancer: comparison of methodologies. *J Clin Pathol* **66**, 1065–1069, doi: 10.1136/jclinpath-2013-201728 (2013).
- Sacher, A. G. *et al.* Prospective Validation of Rapid Plasma Genotyping for the Detection of EGFR and KRAS Mutations in Advanced Lung Cancer. *JAMA Oncol*, doi: 10.1001/jamaoncol.2016.0173 (2016).
- Karachaliou, N. *et al.* Association of EGFR L858R Mutation in Circulating Free DNA With Survival in the EURTAC Trial. *JAMA Oncol* **1**, 149–157, doi: 10.1001/jamaoncol.2014.257 (2015).
- Yoshida, G. J. Metabolic reprogramming: the emerging concept and associated therapeutic strategies. *Journal of Experimental & Clinical Cancer Research* **34**, 1–10, doi: 10.1186/s13046-015-0221-y (2015).
- Pavlova, Natalya N. & Thompson, Craig B. The Emerging Hallmarks of Cancer Metabolism. *Cell Metab* **23**, 27–47, doi: <http://dx.doi.org/10.1016/j.cmet.2015.12.006> (2016).
- Zhou, X. M. *et al.* Metabonomic classification and detection of small molecule biomarkers of malignant pleural effusions. *Anal Bioanal Chem* **404**, 3123–3133, doi: 10.1007/s00216-012-6432-6 (2012).
- Lam, C. W. & Law, C. Y. Untargeted mass spectrometry-based metabolomic profiling of pleural effusions: fatty acids as novel cancer biomarkers for malignant pleural effusions. *J Proteome Res* **13**, 4040–4046, doi: 10.1021/pr5003774 (2014).
- Beloribi-Djefallia, S., Vasseur, S. & Guillaumond, F. Lipid metabolic reprogramming in cancer cells. *Oncogenesis* **5**, e189, doi: 10.1038/oncsis.2015.49 (2016).
- Dunn, W. B. *et al.* Procedures for large-scale metabolic profiling of serum and plasma using gas chromatography and liquid chromatography coupled to mass spectrometry. *Nat. Protocols* **6**, 1060–1083, doi: <http://www.nature.com/nprot/journal/v6/n7/abs/nprot.2011.335.html#supplementary-information> (2011).
- Chong, I.-G. & Jun, C.-H. Performance of some variable selection methods when multicollinearity is present. *Chemometrics and Intelligent Laboratory Systems* **78**, 103–112, doi: 10.1016/j.chemlab.2004.12.011 (2005).
- Kremer, R., Best, L. A., Savulescu, D., Gavish, M. & Nagler, R. M. Pleural fluid analysis of lung cancer vs benign inflammatory disease patients. *British Journal of Cancer* **102**, 1180–1184, doi: 10.1038/sj.bjc.6605607 (2010).
- Martin Diaz, E., Arnau Obrer, A., Martorell Cebollada, M. & Canto Armengod, A. Thoracocentesis for the assessment of lung cancer with pleural effusion. *Arch Bronconeumol* **38**, 479–484 (2002).
- Nance, K. V., Shermer, R. W. & Askin, F. B. Diagnostic efficacy of pleural biopsy as compared with that of pleural fluid examination. *Mod Pathol* **4**, 320–324 (1991).
- Chua, M. E., Sio, M. C., Sorongon, M. C. & Morales, M. L. Jr. The relevance of serum levels of long chain omega-3 polyunsaturated fatty acids and prostate cancer risk: A meta-analysis. *Can Urol Assoc J* **7**, E333–343, doi: 10.5489/auaj.1056 (2013).
- Pouchieu, C. *et al.* Prospective associations between plasma saturated, monounsaturated and polyunsaturated fatty acids and overall and breast cancer risk - modulation by antioxidants: a nested case-control study. *PLOS One* **9**, e90442, doi: 10.1371/ (2014).
- Mashima, T., Seimiya, H. & Tsuruo, T. De novo fatty-acid synthesis and related pathways as molecular targets for cancer therapy. *Br J Cancer* **100**, 1369–1372, doi: 10.1038/sj.bjc.6605007 (2009).
- Menendez, J. A. & Lupu, R. Fatty acid synthase and the lipogenic phenotype in cancer pathogenesis. *Nat Rev Cancer* **7**, 763–777 (2007).
- Roongta, U. V. *et al.* Cancer cell dependence on unsaturated fatty acids implicates stearyl-CoA desaturase as a target for cancer therapy. *Mol Cancer Res* **9**, 1551–1561, doi: 10.1158/1541-7786.MCR-11-0126 (2011).
- Noto, A. *et al.* Stearyl-CoA desaturase-1 is a key factor for lung cancer-initiating cells. *Cell Death Dis* **4**, e947, doi: 10.1038/cddis.2013.444 (2013).
- Schultz, J. R. *et al.* Role of LXRs in control of lipogenesis. *Genes & Development* **14**, 2831–2838, doi: 10.1101/gad.850400 (2000).
- Hong, C. & Tontonoz, P. Liver X receptors in lipid metabolism: opportunities for drug discovery. *Nat Rev Drug Discov* **13**, 433–444, doi: 10.1038/nrd4280 (2014).
- Zhao, C. & Dahlman-Wright, K. Liver X receptor in cholesterol metabolism. *J Endocrinol* **204**, 233–240, doi: 10.1677/joe-09-0271 (2010).

36. Chawla, A., Repa, J. J., Evans, R. M. & Mangelsdorf, D. J. Nuclear Receptors and Lipid Physiology: Opening the X-Files. *Science* **294**, 1866–1870 (2001).
37. Rong, X. *et al.* LXRs regulate ER stress and inflammation through dynamic modulation of membrane phospholipid composition. *Cell Metab* **18**, 685–697, doi: 10.1016/j.cmet.2013.10.002 (2013).
38. Calkin, A. C. & Tontonoz, P. Transcriptional integration of metabolism by the nuclear sterol-activated receptors LXR and FXR. *Nat Rev Mol Cell Biol* **13**, 213–224 (2012).
39. Sampath, H. & Ntambi, J. M. Polyunsaturated fatty acid regulation of genes of lipid metabolism. *Annu Rev Nutr* **25**, 317–340, doi: 10.1146/annurev.nutr.25.051804.101917 (2005).
40. Horton, J. D., Goldstein, J. L. & Brown, M. S. SREBPs: activators of the complete program of cholesterol and fatty acid synthesis in the liver. *The Journal of Clinical Investigation* **109**, 1125–1131, doi: 10.1172/jci15593 (2002).
41. Higham, A. *et al.* The role of the liver X receptor in chronic obstructive pulmonary disease. *Respir Res* **14**, 106, doi: 10.1186/1465-9921-14-106 (2013).
42. Gong, H. *et al.* Activation of the Liver X Receptor Prevents Lipopolysaccharide-induced Lung Injury. *Journal of Biological Chemistry* **284**, 30113–30121, doi: 10.1074/jbc.M109.047753 (2009).
43. Majem, M. & Remon, J. Tumor heterogeneity: evolution through space and time in EGFR mutant non small cell lung cancer patients. *Transl Lung Cancer Res* **2**, 226–237, doi: 10.3978/j.issn.2218-6751.2013.03.09 (2013).
44. Yoshikawa, T. *et al.* Polyunsaturated Fatty Acids Suppress Sterol Regulatory Element-binding Protein 1c Promoter Activity by Inhibition of Liver X Receptor (LXR) Binding to LXR Response Elements. *Journal of Biological Chemistry* **277**, 1705–1711, doi: 10.1074/jbc.M105711200 (2002).
45. Dai, Y.-b. *et al.* Ablation of Liver X receptors α and β leads to spontaneous peripheral squamous cell lung cancer in mice. *Proceedings of the National Academy of Sciences* **113**, 7614–7619, doi: 10.1073/pnas.1607590113 (2016).
46. CHUU, C.-P. & LIN, H.-P. Antiproliferative Effect of LXR Agonists T0901317 and 22(R)-Hydroxycholesterol on Multiple Human Cancer Cell Lines. *Anticancer Research* **30**, 3643–3648 (2010).
47. Wu, Y. *et al.* LXR ligands sensitize EGFR-TKI-resistant human lung cancer cells *in vitro* by inhibiting Akt activation. *Biochemical and Biophysical Research Communications* **467**, 900–905, doi: http://dx.doi.org/10.1016/j.bbrc.2015.10.047 (2015).
48. Guo, D., Bell, E. H., Mischel, P. & Chakravarti, A. Targeting SREBP-1-driven lipid metabolism to treat cancer. *Current pharmaceutical design* **20**, 2619–2626 (2014).
49. Baenke, F., Peck, B., Miess, H. & Schulze, A. Hooked on fat: the role of lipid synthesis in cancer metabolism and tumour development. *Disease Models and Mechanisms* **6**, 1353–1363 (2013).
50. Guo, D. *et al.* EGFR Signaling Through an Akt-SREBP-1-Dependent, Rapamycin-Resistant Pathway Sensitizes Glioblastomas to Antitumorigenic Therapy. *Science Signaling* **2**, ra82–ra82 (2009).
51. Dyatlovitskaya, E. V. & Kandyba, A. G. Role of biologically active sphingolipids in tumor growth. *Biochemistry (Moscow)* **71**, 10–17, doi: 10.1134/s0006297906010020 (2006).
52. Liu, Y.-Y., Han, T.-Y., Giuliano, A. E. & Cabot, M. C. Ceramide glycosylation potentiates cellular multidrug resistance. *The FASEB Journal* **15**, 719–730, doi: 10.1096/fj.00-0223com (2001).
53. Lucci, A. *et al.* Glucosylceramide: a marker for multiple-drug resistant cancers. *Anticancer Res* **18**, 475–480 (1998).
54. Kim, C. W. *et al.* Extracellular Membrane Vesicles from Tumor Cells Promote Angiogenesis via Sphingomyelin. *Cancer Res* **62**, 6312–6317 (2002).
55. Rozhkova, A. V. *et al.* Expression of sphingomyelin synthase 1 (SGMS1) gene varies in human lung and esophagus cancer. *Molecular Biology* **48**, 340–346, doi: 10.1134/s0026893314030170 (2014).
56. Marien, E. *et al.* Non-small cell lung cancer is characterized by dramatic changes in phospholipid profiles. *International Journal of Cancer* **137**, 1539–1548, doi: 10.1002/ijc.29517 (2015).
57. Barceló-Coblijn, G. *et al.* Sphingomyelin and sphingomyelin synthase (SMS) in the malignant transformation of glioma cells and in 2-hydroxyoleic acid therapy. *Proceedings of the National Academy of Sciences* **108**, 19569–19574, doi: 10.1073/pnas.1115484108 (2011).
58. Sot, J. *et al.* Cholesterol displacement by ceramide in sphingomyelin-containing liquid-ordered domains, and generation of gel regions in giant lipidic vesicles. *FEBS Lett* **582**, 3230–3236, doi: 10.1016/j.febslet.2008.08.016 (2008).
59. Gulbins, E. & Kolesnick, R. Raft ceramide in molecular medicine. *Oncogene* **22**, 7070–7077 (2003).
60. Bollinger, C. R., Teichgräber, V. & Gulbins, E. Ceramide-enriched membrane domains. *Biochimica et Biophysica Acta (BBA) - Molecular Cell Research* **1746**, 284–294, doi: http://dx.doi.org/10.1016/j.bbamcr.2005.09.001 (2005).
61. Grassme, H., Riethmüller, J. & Gulbins, E. Biological aspects of ceramide-enriched membrane domains. *Prog Lipid Res* **46**, 161–170, doi: 10.1016/j.plipres.2007.03.002 (2007).
62. Filosto, S. *et al.* EGF Receptor Exposed to Oxidative Stress Acquires Abnormal Phosphorylation and Aberrant Activated Conformation That Impairs Canonical Dimerization. *PLoS One* **6**, e23240, doi: 10.1371/journal.pone.0023240 (2011).
63. Goldkorn, T., Chung, S. & Filosto, S. Lung Cancer and Lung Injury: The Dual Role of Ceramide. *Handbook of experimental pharmacology*. 93–113, doi: 10.1007/978-3-7091-1511-4_5 (2013).
64. Vacaresse, N. *et al.* Activation of epithelial growth factor receptor pathway by unsaturated fatty acids. *Circ Res* **85**, 892–899 (1999).
65. Azrad, M., Turgeon, C. & Demark-Wahnefried, W. Current Evidence Linking Polyunsaturated Fatty Acids with Cancer Risk and Progression. *Frontiers in Oncology* **3**, 224, doi: 10.3389/fonc.2013.00224 (2013).
66. Liu, J. & Ma, D. W. L. The Role of n-3 Polyunsaturated Fatty Acids in the Prevention and Treatment of Breast Cancer. *Nutrients* **6**, 5184–5223, doi: 10.3390/nu6115184 (2014).
67. Zhang, Y.-F., Gao, H.-F., Hou, A.-J. & Zhou, Y.-H. Effect of omega-3 fatty acid supplementation on cancer incidence, non-vascular death, and total mortality: a meta-analysis of randomized controlled trials. *BMC Public Health* **14**, 1–12, doi: 10.1186/1471-2458-14-204 (2014).
68. Schley, P. D., Brindley, D. N. & Field, C. J. (n-3) PUFA Alter Raft Lipid Composition and Decrease Epidermal Growth Factor Receptor Levels in Lipid Rafts of Human Breast Cancer Cells. *The Journal of Nutrition* **137**, 548–553 (2007).
69. Nagan, N. & Zoeller, R. A. Plasmalogens: biosynthesis and functions. *Prog Lipid Res* **40**, 199–229 (2001).
70. Snyder, F. & Wood, R. Alkyl and Alk-1-enyl ethers of glycerol in lipids from normal and neoplastic human tissues. *Cancer Res* **29**, 251–257 (1969).
71. Benjamin, D. I. *et al.* Ether lipid generating enzyme AGPS alters the balance of structural and signaling lipids to fuel cancer pathogenicity. *Proc Natl Acad Sci USA* **110**, 14912–14917, doi: 10.1073/pnas.1310894110 (2013).
72. Bligh, E. G. & Dyer, W. J. A rapid method of total lipid extraction and purification. *Canadian Journal of Biochemistry and Physiology* **37**, 911–917, doi: 10.1139/o59-099 (1959).
73. Smith, C. A., Want, E. J., O’Maille, G., Abagyan, R. & Siuzdak, G. XCMS: Processing mass spectrometry data for metabolite profiling using nonlinear peak alignment, matching and identification. *Anal Chem* **78**, 779–787 (2006).
74. Boser, B. E., Guyon, I. M. & Vapnik, V. N. In *Proceedings of the fifth annual workshop on Computational learning theory* 144–152 (ACM, Pittsburgh, Pennsylvania, USA, 1992).
75. Chang, C.-C. & Lin, C.-J. LIBSVM: A library for support vector machines. *ACM Trans. Intell. Syst. Technol.* **2**, 1–27, doi: 10.1145/1961189.1961199 (2011).
76. Snoek, J., Larochelle, H. & Adams, R. P. In *Advances in neural information processing systems* 2951–2959.

77. Guyon, I., Weston, J., Barnhill, S. & Vapnik, V. Gene Selection for Cancer Classification using Support Vector Machines. *Machine Learning* **46**, 389–422, doi: 10.1023/a:1012487302797 (2002).
78. Efron, B. & Tibshirani, R. Improvements on Cross-Validation: The 632+ Bootstrap Method. *Journal of the American Statistical Association* **92**, 548–560, doi: 10.1080/01621459.1997.10474007 (1997).

Acknowledgements

The financial support for this study is provided by the Biomedical Research Council of the Agency for Science, Technology and Research (A*STAR), Singapore, under the POLARIS program (SPF 2012/001). We thank Li-Lian Kwok and Cindy Lim from the National Cancer Centre, Singapore for their help and advice on ROC analysis and Dr. Say Kong Ng from the Bioprocessing Technology Institute for his invaluable suggestions in the preparation of this manuscript.

Author Contributions

Conception and design: Y.S.H., D.S.W.T. and T.K.H.L. Patient recruitment and acquisition of clinical information: E.H.T., D.S.W.T., T.K.H.L., K.L.L., G.S.T., M.S.Y.K., A.D. and A.T. Acquisition of lipidomics data: L.Y.Y., N.B., V.S.H.C., C.C.T. and E.T. Analysis and interpretation of data (i.e. statistical analysis, biostatistics, computational analysis): Y.S.H., L.Y.Y., N.B., X.Y. and S.Y.Y. Writing and reviewing of manuscript: Y.S.H. and L.Y.Y. Study supervision: Y.S.H., E.H.T., D.S.W.T. and T.K.H.L.

Additional Information

Supplementary information accompanies this paper at <http://www.nature.com/srep>

Competing financial interests: The authors declare no competing financial interests.

How to cite this article: Ho, Y. S. *et al.* Lipidomic Profiling of Lung Pleural Effusion Identifies Unique Metabotype for EGFR Mutants in Non-Small Cell Lung Cancer. *Sci. Rep.* **6**, 35110; doi: 10.1038/srep35110 (2016).



This work is licensed under a Creative Commons Attribution 4.0 International License. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in the credit line; if the material is not included under the Creative Commons license, users will need to obtain permission from the license holder to reproduce the material. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>

© The Author(s) 2016