



Hidden treasures in “ancient” microarrays: gene-expression portrays biology and potential resistance pathways of major lung cancer subtypes and normal tissue

Konstantinos Kerkentzes^{1,2}, Vincenzo Lagani², Ioannis Tsamardinos^{1,2}, Mogens Vyberg³ and Oluf Dimitri Røe^{4,5,6}*

¹ Department of Computer Science, University of Crete, Heraklion, Greece

² Institute of Computer Science, Foundation of Research and Technology – Hellas, Heraklion, Greece

³ Institute of Pathology, Aalborg University Hospital, Aalborg, Denmark

⁴ Department of Cancer Research and Molecular Medicine, Norwegian University of Science and Technology, Trondheim, Norway

⁵ Department of Oncology, Clinical Cancer Research Center, Aalborg University Hospital, Aalborg, Denmark

⁶ Cancer Clinic, Levanger Hospital, Nord-Trøndelag Health Trust, Levanger, Norway

Edited by:

Taofeek K. Owonikoko, Emory University, USA

Reviewed by:

Bo H. Chao, The Ohio State University, USA

Jeffrey Switchenko, Emory University, USA

*Correspondence:

Oluf Dimitri Røe, Department of Cancer Research and Molecular Medicine, Norwegian University of Science and Technology, Erling Skjalgssons gt. 1, Trondheim 7491, Norway
e-mail: oluf.roe@ntnu.no

Objective: Novel statistical methods and increasingly more accurate gene annotations can transform “old” biological data into a renewed source of knowledge with potential clinical relevance. Here, we provide an *in silico* proof-of-concept by extracting novel information from a high-quality mRNA expression dataset, originally published in 2001, using state-of-the-art bioinformatics approaches.

Methods: The dataset consists of histologically defined cases of lung adenocarcinoma (AD), squamous (SQ) cell carcinoma, small-cell lung cancer, carcinoid, metastasis (breast and colon AD), and normal lung specimens (203 samples in total). A battery of statistical tests was used for identifying differential gene expressions, diagnostic and prognostic genes, enriched gene ontologies, and signaling pathways.

Results: Our results showed that gene expressions faithfully recapitulate immunohistochemical subtype markers, as chromogranin A in carcinoids, cytokeratin 5, p63 in SQ, and TTF1 in non-squamous types. Moreover, biological information with putative clinical relevance was revealed as potentially novel diagnostic genes for each subtype with specificity 93–100% (AUC = 0.93–1.00). Cancer subtypes were characterized by (a) differential expression of treatment target genes as *TYMS*, *HER2*, and *HER3* and (b) overrepresentation of treatment-related pathways like cell cycle, DNA repair, and ERBB pathways. The vascular smooth muscle contraction, leukocyte trans-endothelial migration, and actin cytoskeleton pathways were overexpressed in normal tissue.

Conclusion: Reanalysis of this public dataset displayed the known biological features of lung cancer subtypes and revealed novel pathways of potentially clinical importance. The findings also support our hypothesis that even old omics data of high quality can be a source of significant biological information when appropriate bioinformatics methods are used.

Keywords: bioinformatics, carcinoid, lung adenocarcinoma, mesothelioma, microarray, small cell, squamous

INTRODUCTION

Lung cancer is the cancer entity that takes most lives worldwide and overall 5-year survival is still only 15%, despite recent progress in targeted therapy in small subsets of patients (1). The need to identify subgroups with different biology and treatment resistance patterns has been an underlying theme in research, but translating this to the clinic has been very difficult due to the complexity and heterogeneity of the disease (2). The major histological lung cancer subgroups include adenocarcinoma (AD), squamous cell carcinoma (SQ), large cell carcinoma (LC), the carcinoids (COID), and small-cell lung cancer (SCLC) (3). There is an obvious lack of molecular knowledge on lung cancer (4, 5) and an urgent need to dissect the biological differences between/within types both for

improving diagnosis, for detecting prognostic subgroups defined by novel molecular signatures and novel targets and strategies for personalized treatment.

More than a decade ago, genome-wide profiling was believed to allow detection of cancer subgroups and to be able to provide an answer to all these questions. In 2001, Bhattacharjee et al. proposed that the lung AD should be divided into four subgroups based on gene-expression profiles (6). These subgroup descriptions did not materialize into practice-changing information. However, several bioinformatics tools and resources were not yet developed. Particularly, gene annotation was not as complete and we did not have the same level of pathway insight. Could a fresh look into this material reveal new information? In this

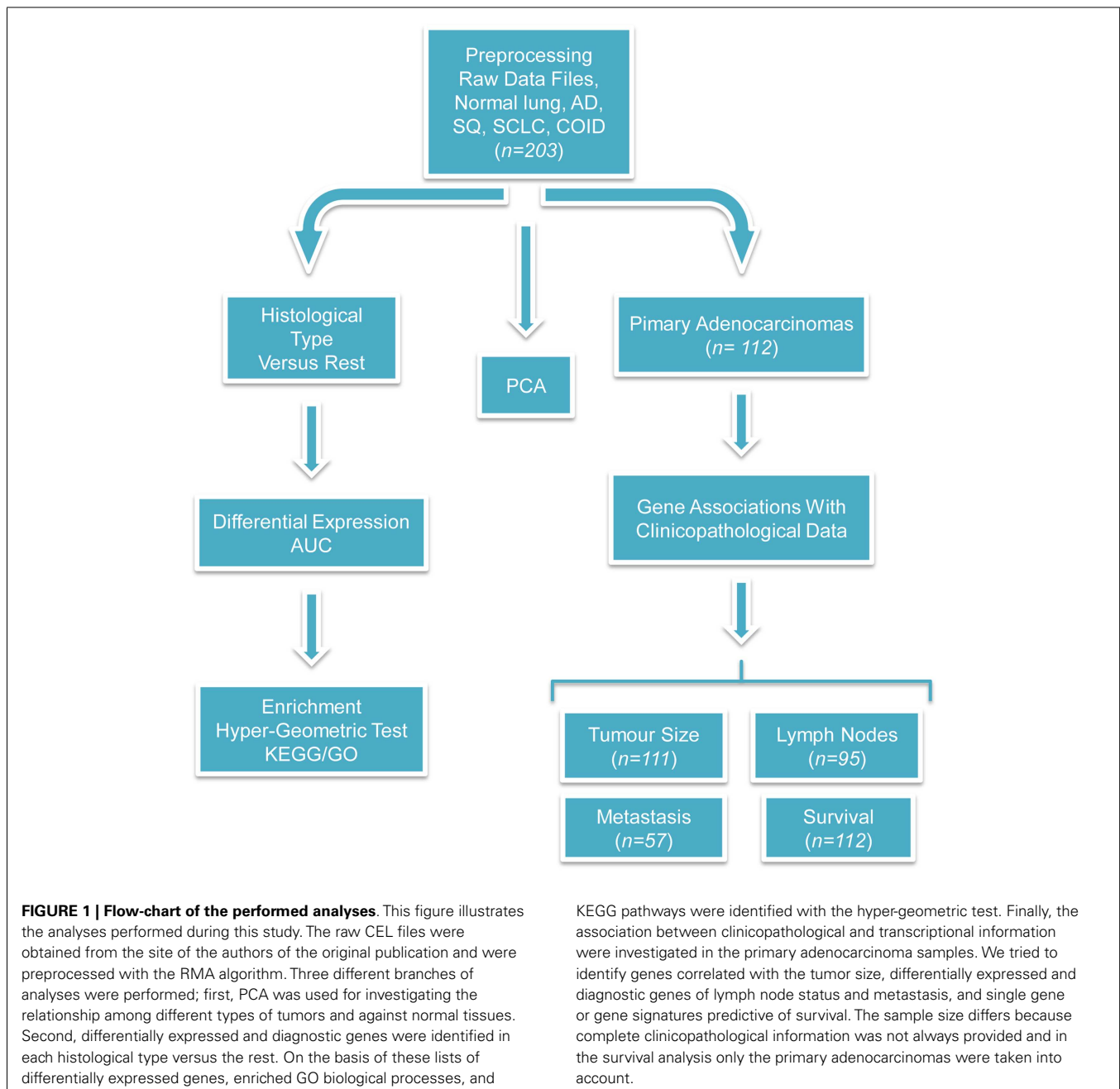
study, we reanalyzed the Bhattacharjee dataset *in silico* – still one of the largest, publicly available, high-quality microarray datasets on normal lung tissue and lung cancer histological subgroups (6). The aim was to characterize the classical histological subgroups in-depth based on differentially expressed genes, gene ontologies, and pathways, and comparing this with our own published mesothelioma dataset (7). In the AD group, we correlated gene expression to clinicopathological information and survival.

MATERIALS AND METHODS

DATA AND PREPROCESSING

Publicly available raw gene-expression profiling data (6) of 203 snap-frozen samples from lung tumors and normal tissue were

obtained from <http://www.broadinstitute.org/mpr/lung/>. These data have not been submitted in any repository (e.g., ArrayExpress, Gene Expression Omnibus, DNA databank of Japan) probably because it was published as early as 2001. These included histologically defined AD ($n = 126$), SQ ($n = 21$), COID ($n = 20$), SCLC ($n = 6$) cases, metastasis from other ADs ($n = 13$, suspected to be extrapulmonary metastases), and normal lung specimens ($n = 17$). Of these, 125 AD samples were associated with clinical data and with histological slides from adjacent sections (6). The total RNA extracted from samples was used to generate cRNA target, subsequently hybridized to human U95A v2 oligonucleotide probe arrays of 12,600 transcripts/8655 unique genes (Affymetrix, Santa Clara, CA, USA) according to standard protocols (8). We



calculated expression values from the raw data (CEL files) using the robust multi-array average (RMA) normalization approach as implemented in the Bioconductor’s “affy” package (9) and any probe with insufficient annotation was removed. When multiple probes were referring to the same gene, only the probe with the higher variation was retained. Principal components analysis (PCA) was performed and either the first two (2D) or three (3D) principal components were plotted in order to visualize a significant portion of the information on the data. A schematic representation of the different analyses performed on the lung cancer dataset in this study is provided (Figure 1).

IDENTIFICATION OF DIFFERENTIALLY EXPRESSED AND DIAGNOSTIC GENES

For each diagnostic group (normal/AD/SQ/SCLC/COID), differentially expressed and diagnostic genes were analyzed in a “one vs. all the rest” fashion (e.g., normal versus the union of AD/SQ/SCLC/COID, SQ versus the union of normal/AD/SCLC/COID, etc.). For differential expression, linear models and empirical Bayes statistics from Bioconductor’s “limma” package (10, 11) were used and any gene with false discovery rate (FDR) adjusted p -value below 0.05 was considered significant. For identifying diagnostic genes, the receiver operating characteristic (ROC) – area under the curve (AUC) (12) for each gene was determined using its expression as a ranking criterion, and genes with high AUC were deemed as the most diagnostic.

PATHWAY OVER/UNDERREPRESENTATION ANALYSIS

GO biological processes and KEGG pathways that were over-represented or underrepresented for each diagnostic group were identified via a hyper-geometric statistical test as implemented in the hyperGTest R function (13). In our analyses, the hyper-geometric test was used for contrasting specific GO terms/KEGG pathways against sets of genes “of interest.” These sets were defined, in turn, as all differentially expressed genes, the upregulated genes, and the downregulated genes of each diagnostic group.

COMPARISON WITH MESOTHELIOMA

Our previous study on microarray of mesothelioma versus parietal pleura has a gene list of differentially expressed genes as well as GO and KEGG pathways (7). We compared these findings with the lung cancer profiles to see possible similarities and differences. The mesothelioma versus normal pleura expression profile has been submitted to ArrayExpress registered with accession number E-MTAB-47.

CORRELATION WITH PHENOTYPE CHARACTERISTICS AND SURVIVAL

We used the supplementary information available for the ADs samples ($n = 126$; www.pnas.org/content/suppl/2001/11/13/191502998.DC1/SampleData.xls), for investigating possible associations among gene expressions and different phenotype. We correlated gene expression with tumor size by performing Spearman’s rank correlation test. The resulted p -values were corrected through Storey’s q -value method (14) (as implemented in Bioconductor’s “ q -value” package), where a cutoff threshold of 0.05 was used to control the FDR.

Differentially expressed and diagnostic genes for the metastatic vs. primary AD tumors and for each lymph node status were

identified as described in Section “Identification of Differentially Expressed and Diagnostic Genes.”

A complex analysis protocol was applied in order to identify gene signatures able to predict survival in AD. All the available clinicopathological information as sex, age, and smoking were included in the analysis, and missing values were replaced with the mean of their respective variables. The experimentation protocol from Lagani et al. (15) was applied. Briefly, this protocol consists of several multivariate regression and feature selection algorithms applied with a nested N -fold cross-validation procedure. The N -fold cross validation consists in subdividing the available sample in N -folds; in turn, each fold is held out for testing purposes while the remaining data (training set) is employed for model selection and fitting. The nested-cross validation is a generalization of the standard cross validation, where an inner cross validation is applied in the training set in order to select the best model. The performance estimation provided by nested-cross validation is known to be more accurate than the ones provided by simple cross validation (16).

VALIDATION ON EXTERNAL DATASETS

The validation on the robustness and reproducibility of our methods was tested by analysis of two additional, more recent lung cancer datasets (17, 18), as detailed in Text S1 in Supplementary Material. In brief, three ordered gene lists were produced for each dataset, one based on FDR adjusted p -value, one based on log-fold change, and one based on AUC. A statistical methods based on differentially weighting overlaps among lists depending by whether they occur at the extremes of in the middle of the lists [Bioconductor’s package “OrderedList,” (19)] was employed to test and measure the similarity between the lists of each ordering scheme.

IMMUNOHISTOCHEMISTRY

The immunohistochemistry images are from our institution, Institute of Pathology, Aalborg University Hospital and are representative for each diagnostic antibody. The microphotos are presented in accordance with the rules of the Review Board at our institution.

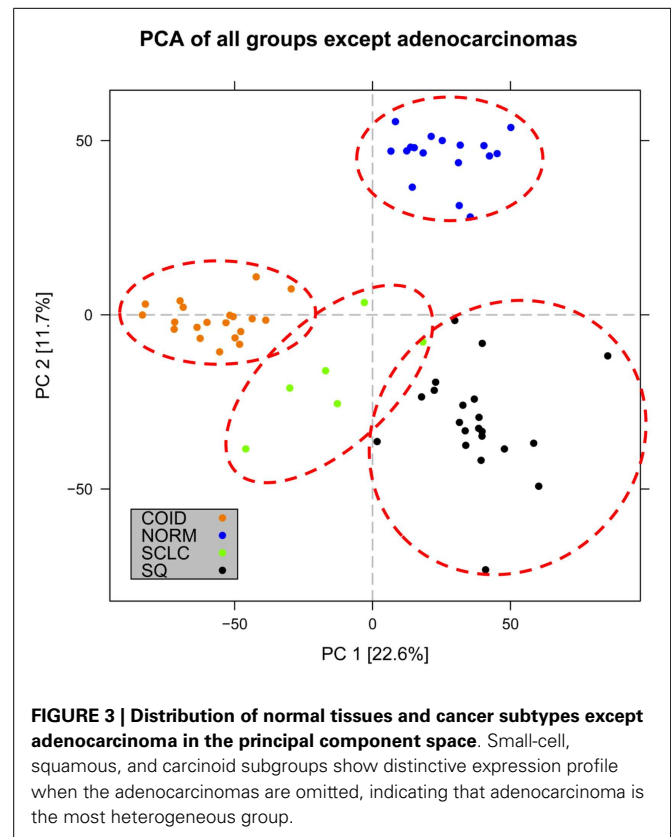
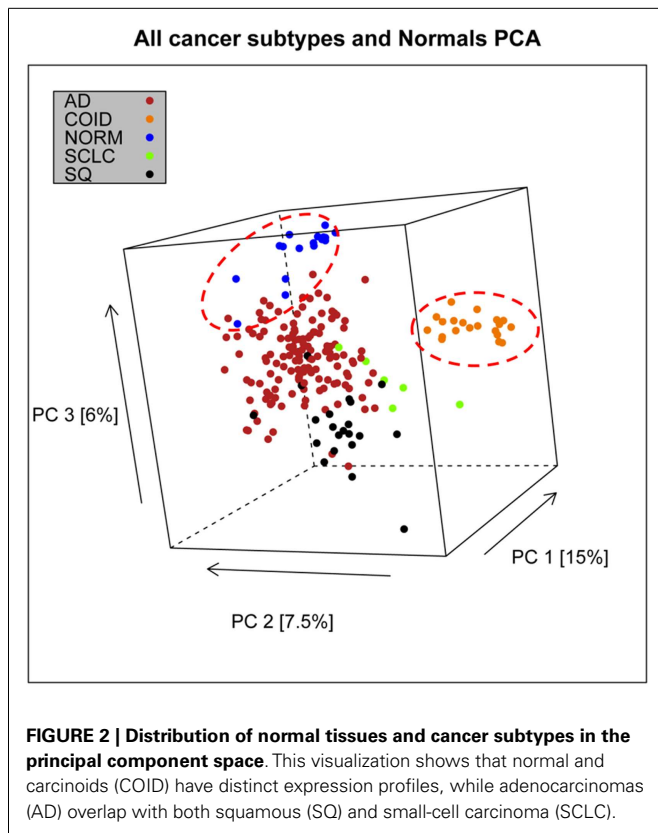
RESULTS

In the first analysis, differential gene expression of each tissue type versus all the other types was determined. Data showed that each of the tumor subtypes and normal tissue had >1000 differentially expressed genes, except the SCLC with 842 genes (Table 1; File S1 in Supplementary Material). Likewise each tissue type had tens of genes with an AUC greater than 0.8, indicating a strong

Table 1 | Differentially regulated genes between each diagnostic group versus the rest ($p < 0.05$).

Genes	Normal	Lung AD	SQ	COID	SCLC
Up	1338	1527	1611	3247	493
Down	1680	1975	2064	2130	349

Up, upregulated; Down, downregulated; AD, adenocarcinoma; SQ, squamous cell carcinoma; COID, carcinoids; SCLC, small-cell lung cancer.



discriminatory or diagnostic value of these genes (File S2 in Supplementary Material). Moreover, there were several significantly over and underrepresented ontologies and pathways in GO and KEGG (Files S3 and S4 in Supplementary Material). There was a distinctive expression profile of the normal tissue, SCLC, SQ, and COID as visualized by principal component analysis (PCA) in 2D and 3D. In contrast, AD showed high degree of heterogeneity and overlapped the expressions of SQ and SCLC (Figures 2 and 3). In the PCA, the expression of the AD metastases to the lung from other locations did not vary significantly from the primary lung ADs (Figure 4). However, between the primary AD and metastatic groups there were 267 differentially expressed genes with FDR adjusted p -value <0.05 . Below the main results of each tissue type is described.

NORMAL LUNG

The top five diagnostic genes were all overexpressed, *PECAM1* (AUC >0.99), *TGFBR2*, *CDH5*, *AGER*, and *TCF21* (File S2 in Supplementary Material). These results were similar with the results from the original paper that found all these genes overexpressed in the normal samples as well (6). The differentially expressed genes of normal lung tissue versus all types of lung tumors were overrepresented in several GO processes as actin cytoskeleton organization and regulation of developmental process and underrepresented in such as mRNA metabolic process and G1/S transition checkpoint (File S3 in Supplementary Material). Similar KEGG pathways were overrepresented, such as regulation of actin cytoskeleton, vascular smooth muscle contraction, and leukocyte trans-endothelial

migration (File S4 in Supplementary Material). Interestingly, these pathways were downregulated in tumors.

NON-SMALL CELL LUNG CANCER

Squamous cell lung cancer

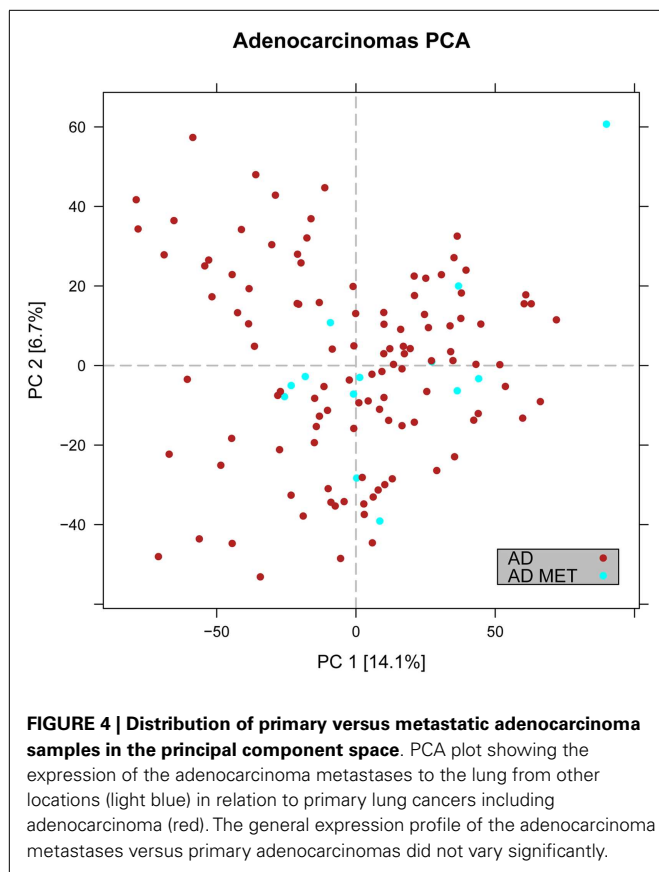
The top five genes with the highest predictive value were *KRT5* (AUC >0.98), *ATP11A*, *DSP*, *UBXN7*, and *RHOBTB2* from which *ATP11A* and *RHOBTB2* were underexpressed and the rest were overexpressed. The GO of the differentially expressed genes of SQ included M/G1 transition and DNA repair, while in KEGG several DNA repair mechanisms were overexpressed. Among the underrepresented was cellular response to cytokine stimulus and similarly in KEGG, cytokine–cytokine receptor interaction.

Adenocarcinoma

Among the top five genes with the highest diagnostic value *ERBB2* (*HER2*) (AUC >0.93), *GALE*, *PRSS8*, and *PYCR1* were overexpressed and *PGAP1* was downregulated. The GO of the differentially expressed genes of the primary lung AD showed antigen procession and presentation, identical as in KEGG, while response to bacterium and cell cycle was underrepresented in GO and KEGG, respectively (Files S3 and S4 in Supplementary Material).

SMALL CELL LUNG CANCER

The top five genes were all overexpressed, namely the *ISL1* (AUC = 1.00), *HMG2*, *CDKN2C*, *STMN1*, and *ACY1*. The GO of the differentially expressed genes of SCLC included mitotic cell cycle and DNA repair, while regulation of several metabolic



processes was underrepresented. In KEGG, cell cycle, several DNA repair mechanisms and one carbon pool by folate was overrepresented while cytokine–cytokine receptor interaction and proteasome were among the underrepresented pathways (Files S3 and S4 in Supplementary Material).

CARCINOIDS

The lung COID was the only tumor entity that separated completely from the rest of the tumors and normal tissue by PCA and AUC (Figures 2 and 3). There were seven genes with a diagnostic AUC = 1.00, namely *VAMP2*, *MYT1L*, *GPRASP1*, *TSPYL2*, *CHGA*, *MAPRE3*, and *SNAP91*, all overexpressed. Differentially expressed genes of COID in GO showed overrepresentation of regulation of angiogenesis and Hippo signaling cascade, and upregulation of synaptic transmission and neurological system process. In KEGG pathways, the lysosome, the SCLC, and phagosome pathways were overrepresented, and neuroactive-ligand receptor interaction neurological disease pathways such as Parkinson's, Huntington's, and Alzheimer's were significantly upregulated. Underrepresentation of GO included response to stress and immune-related ontologies while in KEGG the ribosome, the ECM-receptor interaction, phagosome, and the leukocyte trans-endothelial migration was downregulated (Files S3 and S4 in Supplementary Material).

PRIMARY LUNG ADENOCARCINOMA VERSUS LUNG METASTASES

In the PCA, we observed that the AD metastases of the lung from other sites had an overlapping and overall similar gene-expression

Table 2 | Genes significantly correlated with primary lung adenocarcinoma tumor size.

Symbol	Correlation	Adjusted <i>p</i> -value <0.05
<i>NR2C1</i>	0.4527	0.00564
<i>DLGAP1</i>	−0.4255	0.01395
<i>ZNF259P1</i>	−0.3952	0.0369
<i>CD22</i>	−0.3946	0.0369

Positive correlation implies higher expression in larger tumors.

profile with the primary lung AD (Figure 4). However, there were 267 differentially expressed genes, where 207 were downregulated in the primary lung AD (File S1 in Supplementary Material). The overexpressed gene with highest discriminatory power was the *SLC26A3* with AUC >0.88. In KEGG pathways, key genes of the PI3K-Akt and Jak-Stat, e.g., the oncogenes *MYC* and *MYB* and in the mTOR signaling pathways the *AKT1*, *MTOR*, *OSMR*, and *PKN1* were upregulated.

CLINICOPATHOLOGICAL CORRELATION AND SURVIVAL IN ADENOCARCINOMAS

We performed a Spearman rank test using only the primary ADs of the lung where clinical data were available (*n* = 112) to examine the correlation of gene expression and tumor size in centimeters. There were four genes with a significant *p*-value, and only one of them, the *NR2C1* was positively correlated with tumor size (Table 2). Analysis of N-status revealed only in the N0 cases versus N1–N3 two differentially expressed genes (*P* < 0.05) with an AUC of >0.74, *FLNB* and *ILVBL*, both downregulated. There were no differentially expressed gene related to metastasis, but there were very few cases in the M1 group (*n* = 4), most where MX (undefined M status).

The survival analysis did not achieve any notable results as no single gene neither gene signature could predict survival in this cohort. Particularly, we measured predictive performances in terms of concordance index (CI) and integrated brier score (IBS) (20). The first metric has an interpretation similar to the AUC and provides an estimation of the probability of correctly ranking two randomly chosen subjects according to their respective risk of experiencing the event of interest. The IBS evaluates the calibration of the predicted survival curves. In all our analyses, the predictive performances of any single gene and gene signature (as measured through nested-cross validation, see “Materials and Methods”) were around 0.5 (CI) and 0.25 (IBS), which corresponds to random guessing.

DISCUSSION

New insight in molecular subtyping is of increasing importance in the era of personalized medicine, as advances in high-throughput technology have revealed distinct molecular differences between and within tumor types and even a high degree of intratumoral heterogeneity. Here, we provide a proof-of-concept of how a well curated set of expression profiles on normal lung tissues and lung cancer subtypes can be used to make biologically and clinically relevant assumptions on the phenotype based on the transcription profiles.

Through more than four decades, patients with lung cancer were generally treated uniformly in two large groups, the non-small cell lung cancers (NSCLC) and the SCLC. Current progress include more efficient chemotherapy compounds for so-called non-squamous carcinomas, including the multifolate inhibitor pemetrexed (21), and smaller subgroups with molecular changes such as EGFR mutations and ALK rearrangements that significantly predict the effect of biological targeted therapies (22, 23). However, the SQ, LC, SCLC, or COID and the majority of the AD still have no molecular prognostic or predictive markers in clinical use (4, 5) and most treatment modalities are stratified according to clinical staging as TNM (Tumor, Node, Metastasis), Stage I–IV and Karnofsky or WHO performance status.

In lung cancer, mRNA expression profiling showed promising results in defining some subgroups more than a decade ago, but surprisingly this information has still not led to any radical path to a personalized and effective treatment. Performing high-quality microarray analysis of human tissue is not easy, and several factors need to be optimal at sample collection, processing, and analysis, that may account for large variations of the results. Moreover, they are both time consuming and costly. However, currently there are thousands of public datasets on human material, and as both gene function knowledge and bioinformatics methods have evolved, these datasets may be a valuable source of new knowledge.

SELECTION OF DATASET AND IDEA OF ANALYSIS

The dataset used was from 2001, but was elected due to the unusual high number of participants where all the large subgroups (except the large cell type) were represented, the completeness of the clinical data, as well as the quality of the research center. The main focus of the original publication was to present several new molecular subgroups of lung AD and pinned out one subtype with less favorable prognosis. We did not try to recapitulate their findings, but we observed that some of the differences between the classical subgroups in their analysis were similar to our findings [Figure 1 in Ref. (6)]. The main idea of our analysis was to identify the differential and diagnostic genes for each classical tissue type versus all the other groups, hereby trying to define the most specific molecular characteristics of each tissue type. The results showed a high degree of coherence between our findings and previous knowledge, but also revealed novel potential biomarker and target genes and pathways.

VALIDATION ON IMMUNOHISTOCHEMICAL MARKERS

As histopathology and immunohistochemistry is the gold standard of cancer diagnosis, we validated our results by comparing the diagnostic gene results (File S2 in Supplementary Material) with the most common negative and positive immunohistochemical markers (proteins) for each subtype used for diagnosis in the clinic. These clinical diagnostic immunohistochemical markers are already validated through several studies, typically with sensitivity >80–90% (24–30). This comparison revealed a quite impressive overlap where genes encoding the diagnostic proteins were found in the top of their respective AUC list (Table 3; Figure 5). These included chromogranin A (*CHGA*) and synaptophysin (*SYN*) for COID, cytokeratin 5 (*KRT5*) and p63/p40 (*TP63*) for SQ, thyroid transcription factor 1 (*TTF1*, synonymous with *NKX2-1*) for AD,

and its homolog (*NKX2-2*) for SCLC (24–28). Interestingly, we also detected that the *NKX2-1* encoding *TTF1*, the most specific marker for cancers originating from the lung, was downregulated in SQ, in line with recent knowledge of SQ immunomarkers (28). In line with our findings, a large immunohistochemical study verified that a two-marker panel (*TTF1*/p63) is sufficient for subtyping of the majority of tumors as AD or SQ (29). Moreover, another obviously very expensive and time-consuming study on lung cancer subtypes testing >1000 cases with 108 antibodies got similar results, showing a five protein signature that could also separate cases that were undefined after *TTF1*/p63 analysis. This signature included the SQ positive markers *KRT5* (CK5), *TRIM29*, and *SLC7A5* (AUC 0.98, 0.93, and 0.83, respectively, in our analysis) and the AD positive markers *MUC1* and *CEACAM5* (AUC 0.81 and 0.78, respectively, in our analysis) (30).

However, several of the top genes diagnosing each subgroup have not been described in the literature for lung cancer, as the neuroendocrine marker *ISL1* (insulin gene enhancer protein ISL-1) for the SCLC, the *ATP11A* (ATPase, class VI, type 11A) for SQ, *GALE* (UDP-galactose-4-epimerase) for the AD, and *VAMP2* (vesicle-associated membrane protein 2/synaptobrevin 2) for the COID (see Results and Supplementary Material). Currently, some of these genes are validated in tumors in an ongoing collaboration project.

VALIDATION ON EXTERNAL DATASETS

In order to validate the robustness of our methods, two more datasets were retrieved, one published in 2005 (17) and one in 2010 (18), where the second dataset was obtained through the ExpressionBlast tool (31). Unfortunately, we were not able to identify any dataset with the exact cancer subtypes as in the Bhattacharjee et al. study, and thus a head-to-head comparison of our results was not possible. Due to this limitation, in each dataset we only contrasted the AD versus the healthy tissue samples. Moreover, since the measurements in each study were performed on the Affymetrix platform but different chip types, HGU-95Av2, HGU-133A, and HGU-133plus2, only the common genes of all three datasets were used ($n = 8556$). The genes of each study were ranked according to different criteria (FDR adjusted p -value for differential expression, log-fold change, and AUC). These ordered lists were compared across studies with a statistical method devised for evaluating similarity in gene ranks (19). In each pairwise comparison, the statistical tests rejected the null-hypothesis of ranks being different, thus indicating that results obtained on the Bhattacharjee et al. study is reproducible on other datasets (see Text S1 in Supplementary Material).

GENE ONTOLOGY AND PATHWAY ANALYSIS

Among the abundant gene and pathway information acquired through this quantitative analysis, several genes, and pathways recapitulating known phenotypes were detected, such as resistance to known chemotherapy regimens, targeted treatments, and tumor aggressiveness. Some of the gene ontologies/pathways with highest differential overrepresentation of genes will be discussed below. We observed that there were some similarities between these results and some of the finding of our previous study (7) on mesothelioma versus parietal pleura (normal mesothelial tissue). Thus, we

Table 3 | Immunomarker occurrence expressed as percentage of cases in lung adenocarcinoma, squamous cell carcinoma, small-cell carcinoma, and carcinoid.

Lung cancer	CGA (%)	CHGA, AUC	CK5 (%)	KRT5, AUC	p63/p40 (%)	TP63, AUC	SNP (%)	SYP, AUC	TTF1 (%)	NKX2-1/TTF1, AUC
AD	2	0.59D	5–30 ^{a,b}	0.71D	5–30 ^{a,b}	0.72D	10	0.56D	60–90	0.76U
SQ	0	ND	90–100	0.98U	90–100	0.89U	0	0.78D	0–20 ^a	0.89D
SCLC	70–90	ND	0–5	ND	0–5	ND	70–90	ND	70–100	0.85U (NKX2-2)
COID	95	1,0U	0	ND	0	ND	100	0.99U	50–70	0.72D

^aPredominantly focal reactions

^bp40 virtually negative in adenocarcinoma.

Values are based on recent large studies [Ref. (24–30)]. For each immunomarker protein, the diagnostic power of the respective gene (expressed in terms of AUC) is reported in blue.

CGA, chromogranin A; CHGA, CGA gene; CK5, cytokeratin 5; KRT5, CK5 gene; p63/p40, tumor protein p63/isoform of p63; TP63, p63 gene; SNP, synaptophysin; TTF1, thyroid transcription factor 1; NKX2-1, TTF1 encoding gene; NKX2-2, homolog of NKX2; D, downregulated; U, upregulated; ND, not differentially expressed; COID, carcinoids; AD, adenocarcinoma; SCLC, small-cell lung cancer; SQ, squamous cell lung cancer.

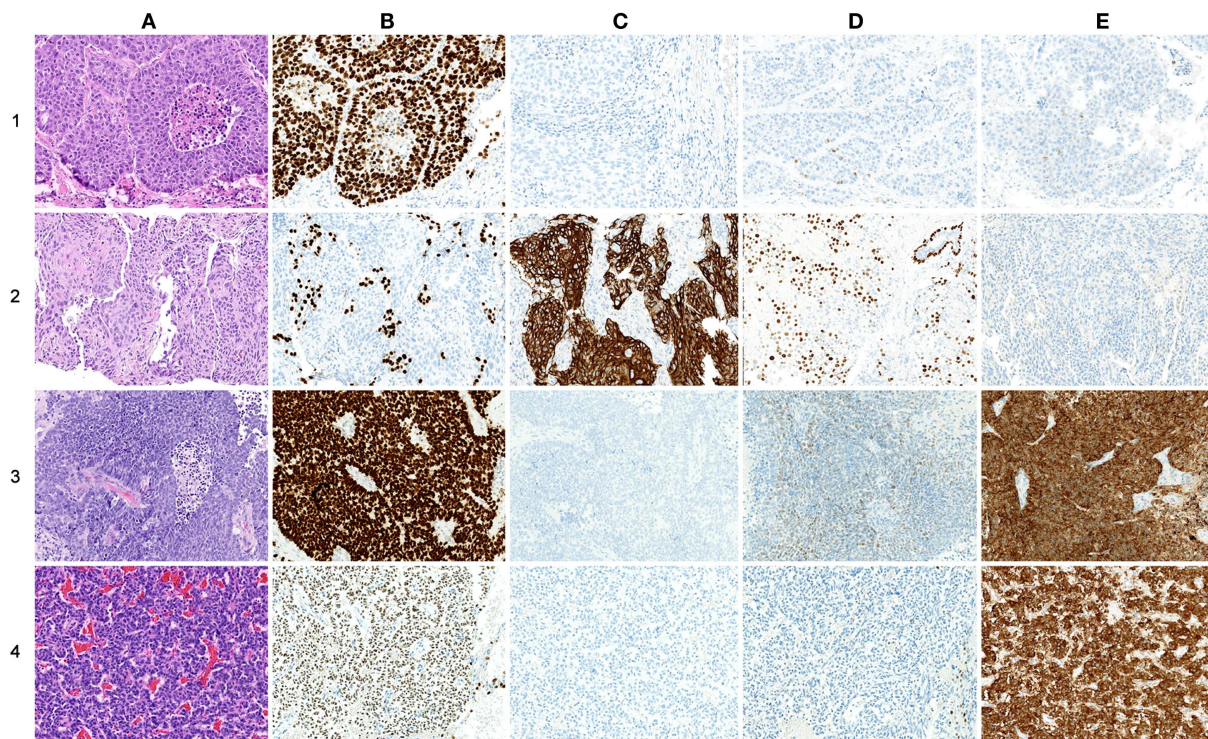


FIGURE 5 | Comparison of protein and gene-expression patterns in primary lung neoplasias. (A) hematoxylin and eosin (H&E), **(B)** thyroid transcription factor-1 (TTF1), **(C)** cytokeratin 5 (CK5), **(D)** p63, and **(E)** synaptophysin (SNP). All photos $\times 200$. (1) Adenocarcinoma, **(A)** H&E showing low differentiation with solid growth, **(B)** TTF1 showing strong nuclear staining of all tumor cells, **(C)** CK5 showing no cytoplasmic staining of tumor cells, **(D)** p63 showing nuclear staining of a few scattered tumor cells, and **(E)** SNP showing cytoplasmic staining of a few scattered tumor cells. (2) Squamous cell carcinoma, **(A)** H&E showing solid growth with slight squamous maturation to the left, **(B)** TTF1 showing no nuclear staining of tumor cells, scattered entrapped alveolar cells with normal TTF1 expression is seen, **(C)** CK5 showing strong cytoplasmic staining of most tumor cells,

(E) p63 showing widespread nuclear staining of tumor cells, in the upper right corner an entrapped bronchiolus with normal p63 expression of basal cells is seen, and **(E)** SNP showing no cytoplasmic staining of tumor cells. (3) Small-cell carcinoma, **(A)** H&E showing sheet-like growth, **(B)** TTF1 showing very strong nuclear staining of tumor cells, **(C)** CK5 showing no staining of tumor cells, **(D)** p63 showing widespread but weak nuclear staining of tumor cells, and **(E)** SNP showing strong cytoplasmic staining of tumor cells. (4) Carcinoid, **(A)** H&E showing trabecular growth, **(B)** TTF1 showing moderate nuclear staining of most tumor cells, **(C)** CK5 showing no cytoplasmic staining of tumor cells, **(D)** p63 showing no nuclear staining of tumor cells; a few entrapped alveolar cells are positive, and **(E)** SNP showing strong cytoplasmic staining of tumor cells.

also contrasted our current results against the key pathways that are differentially expressed in this treatment-refractory thoracic cancer.

Cell cycle

Analysis of cell cycle pathway in KEGG of each differential list of genes showed that in normal lung tissue only tumor suppressor genes (*CDKN1A/p21*, *CDKN1C/p57*, *RBL2/p130*) and the cyclin D family were overexpressed, reflecting a normal phenotype. Tumor suppressors are crucial factors in tumorigenesis, often deleted or downregulated in tumors. Cyclins are often upregulated in cancers (32); however, cyclin D1 has not been shown to be a negative prognostic factor in cancer (32), cyclin D2 is often methylated and thus downregulated in lung cancer (33), and cyclin D3 was suggested not to have a profound role in tumorigenesis (34). Opposite, in all the tumors, including mesothelioma, these genes were suppressed, while oncogenes, cyclins, cyclin dependent kinases, and other cell cycle driving genes were overexpressed (Figure 6). The SQ, SCLC but also mesothelioma displayed overexpressed Mini-Chromosome Maintenance complex genes, while not differentially expressed in AD and downregulated in COID. In the SQ, 56 of 63 genes in the cell cycle were overexpressed, including *CCNE/CDK2*, *CCNA/CDK2*, *CCNA/CDK1*, *CCNB/CDK1* complexes, *BUB1*, *BUB1B*, *BUB3*, and *MYC* oncogenes, the damage response genes as *ATR*, *CHEK1/2*, and five genes of the 14-3-3 family, where several have been linked to tumorigenesis including the hypomethylated SFN or Stratifin (35). Again, some of these findings are not surprising, but recapitulate prior knowledge. Moreover, the *TGFB11* was overexpressed in normal lung but less so in tumors, as seen in the Human Protein Atlas (36). Recently, a gene-expression study on aggressive versus more indolent lung COID showed that several of these cell cycle genes, including *BUB1*, were overexpressed in the aggressive forms (37).

Thymidylate synthase

The one carbon pool by folate pathway is the target of several antineoplastic treatments, where thymidylate synthase (TS) is a key enzyme. Pemetrexed is a multifolate antagonist that inhibits replication through inhibition of folate-dependent enzymes as TS, GARFT, and DHFR, where TS is the main target with many orders of magnitude higher affinity for pemetrexed (38). The TS encoded by *TYMS*, is a key protein that catalyzes the methylation of deoxyuridylate (dUMP) to deoxythymidylate (dTMP) that maintains the dTMP pool critical for DNA replication and repair and has shown a strong correlation to the effect of pemetrexed. Low TS expression increases the pemetrexed response *in vitro* (39, 40) and a recent meta-analysis on clinical data showed that both *TYMS* and TS expression were inversely correlated to the effect of pemetrexed in lung cancer (41). This is in line with our findings where *TYMS* was not overexpressed in AD where pemetrexed has a main role in treatment, but significantly overexpressed in SCLC and SQ where it has no proven effect, and is not used in the clinic (42–44). Interestingly, while in the normal tissue, no genes in this one carbon pool by folate pathway were overexpressed, several genes were upregulated in the cancers (Figure S1 in Supplementary Material). The overexpression of *TYMS* in our mesothelioma

cohort, where pemetrexed is a key treatment, was probably due to the fact that most of the mesothelioma cases included were generally resistant to pemetrexed, and had overexpressed TS (45, 46). In the COID, the *TYMS* was not overexpressed, indicating that COID could be a potential tumor group for multifolate inhibitors. To our knowledge, this has not been tested yet.

ERBB signaling pathway

The genes of the epidermal growth factor receptor (ERBB) signaling pathway were differentially expressed in each tissue type, including the *ERBB2 (HER2)*, *ERBB3 (HER3)*, and *ERBB4*. These are type I transmembrane growth factor receptors that activate intracellular signaling pathways in response to extracellular signals and activate numerous downstream pathways involved in the regulation of differentiation, migration, proliferation, and survival (47). Interestingly, *EGFR (HER1)*, encoding EGFR, the primary target of the tyrosine kinases gefitinib and erlotinib, was not overexpressed in any type. This may have a simple explanation, either that only a minority has activating mutations of EGFR (17 and 2% of the non-squamous cancers in Caucasian and Afro-American population, respectively), and thus did not show a statistical difference as a group, or the array did not detect mutant EGFR (48). Currently, the HER2 receptor is a very important drug target in HER2 overexpressing breast and gastric cancers. It is known that the HER2 protein is overexpressed in 20% of lung cancer cases, but treatment with trastuzumab in this group has not been beneficial except probably in 1–2% of cases with a *HER2* gene driver mutation (49). More surprising was the finding that *HER3* was overexpressed in AD. Recently afatinib, a selective, orally bioavailable ERBB family blocker that irreversibly blocks signaling from EGFR, HER2, and ERBB4 showed an increased progression-free survival of *EGFR* mutated ADs of the lung in a phase III study versus chemotherapy (50). The *ERBB4* was overexpressed in the COID, recapitulating immunohistochemistry showing that 100% of lung COID expresses *ERBB4* (51) and thus indicating a potential novel candidate target for COID (Figure S1 in Supplementary Material).

DNA repair

The backbone of cytotoxic treatment of non-EGFR mutated lung cancers as well as mesothelioma includes platinum derivatives cisplatin and carboplatin. They are often designated cross-linking agents due to their capacity of inducing lethal DNA damage in cells through DNA inter- and intra-strand crosslinks. However, the response rates in the first line treatment of advanced cases are approximately 35%, meaning that the intrinsic resistance is high (52). Resistance to the cross-linking agents requires a functioning DNA repair system, especially the pathways that take care of crosslinks and adducts. In that respect, this analysis showed a marked difference between tumors and normal lung tissue. Tumors showed overrepresentation and overexpression of genes of all the DNA repair pathways, including the NER, BER, HR, MMR, and NHEJ, as well as the Fanconi anemia, all known to play a role in platinum resistance, while in normal lung only one DNA repair gene was overexpressed, the *XPC* in NER (Figure 7) (7, 53). Importantly, the tumor types with the highest intrinsic resistance to platinum, the COID and SQ also had the highest number of

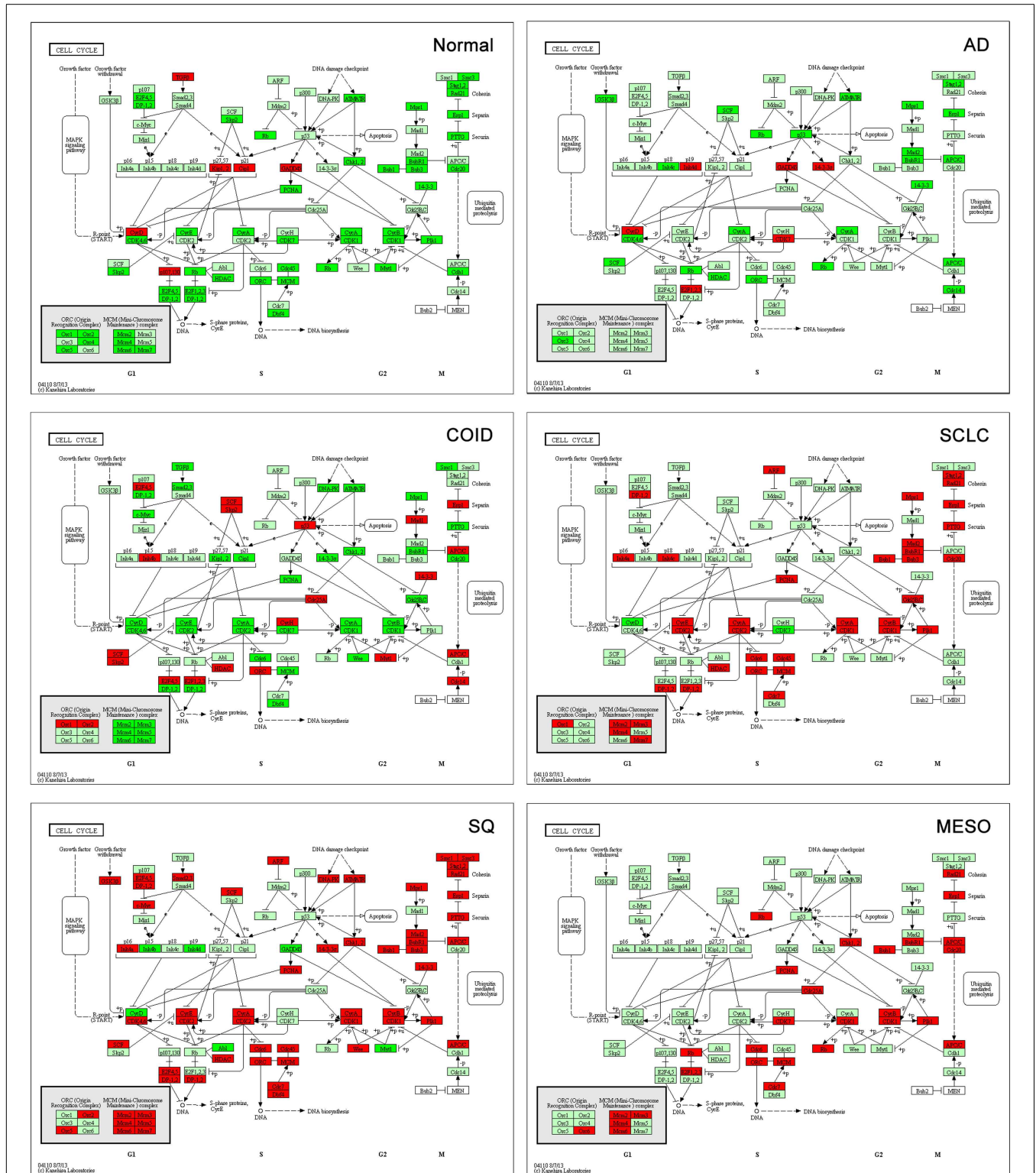


FIGURE 6 | Cell cycle pathway. Tumor suppressor genes (CDKN1A/p21, CDKN1C/p57, RBL2/p130) and the cyclin D family were overexpressed in normal lung tissue while there was type-specific overexpression of oncogenes, cyclins, cyclin dependent kinases, and other cell cycle driving genes in tumors. Genes belonging to the mini-chromosome maintenance complex were overexpressed in squamous and small-cell lung cancer, as well as in mesothelioma. The same genes were not differentially expressed in AD

and downregulated in COID. Fifty-six out of 63 genes in the Cell Cycle pathway were overexpressed in SQ, including CCNE/CDK2, CCNA/CDK2, CCNA/CDK1, CCNB/CDK1 complexes, BUB1, BUB1B, BUB3, and MYC oncogenes, damage response genes (ATR, CHEK1/2) and five genes of the 14-3-3 family. Red is overexpressed, dark green downregulated. SQ, squamous type; AD, adenocarcinoma; SCLC, small-cell lung cancer; COID, carcinoid; MESO, mesothelioma.

repair genes overexpressed (Table 4). The NER pathway is the principal and most analyzed pathway related to cisplatin resistance, and here, there was an array of significantly overexpressed NER genes in the COID, SCLC, SQ, and MESO tumors. Interestingly, the most heterogeneous tumor group, both genetically and regarding treatment response, the AD, had the least NER genes overexpressed (Figure 7). In the SQ, 25 NER genes were differentially expressed, and only four were downregulated, the *CUL4*, *XPD*, *ERCC2*, and *ERCC8*. Key genes of the whole NER repair process through

damage recognition, DNA unwinding, incision, DNA synthesis, and ligation were overexpressed (Figure 7). There are also several similarities with mesothelioma in this respect where overexpression of two or more of *CDK7*, *GTF2H2*, *PCNA*, *RFC4*, and the *RFC5* were shared by the lung tumors (Table 4). Importantly, a recent study showed that DNA repair proteins of the NER pathway can predict effect of adjuvant cisplatin in SQ but not in AD, showing the relevance of more knowledge related to these pathways (54).

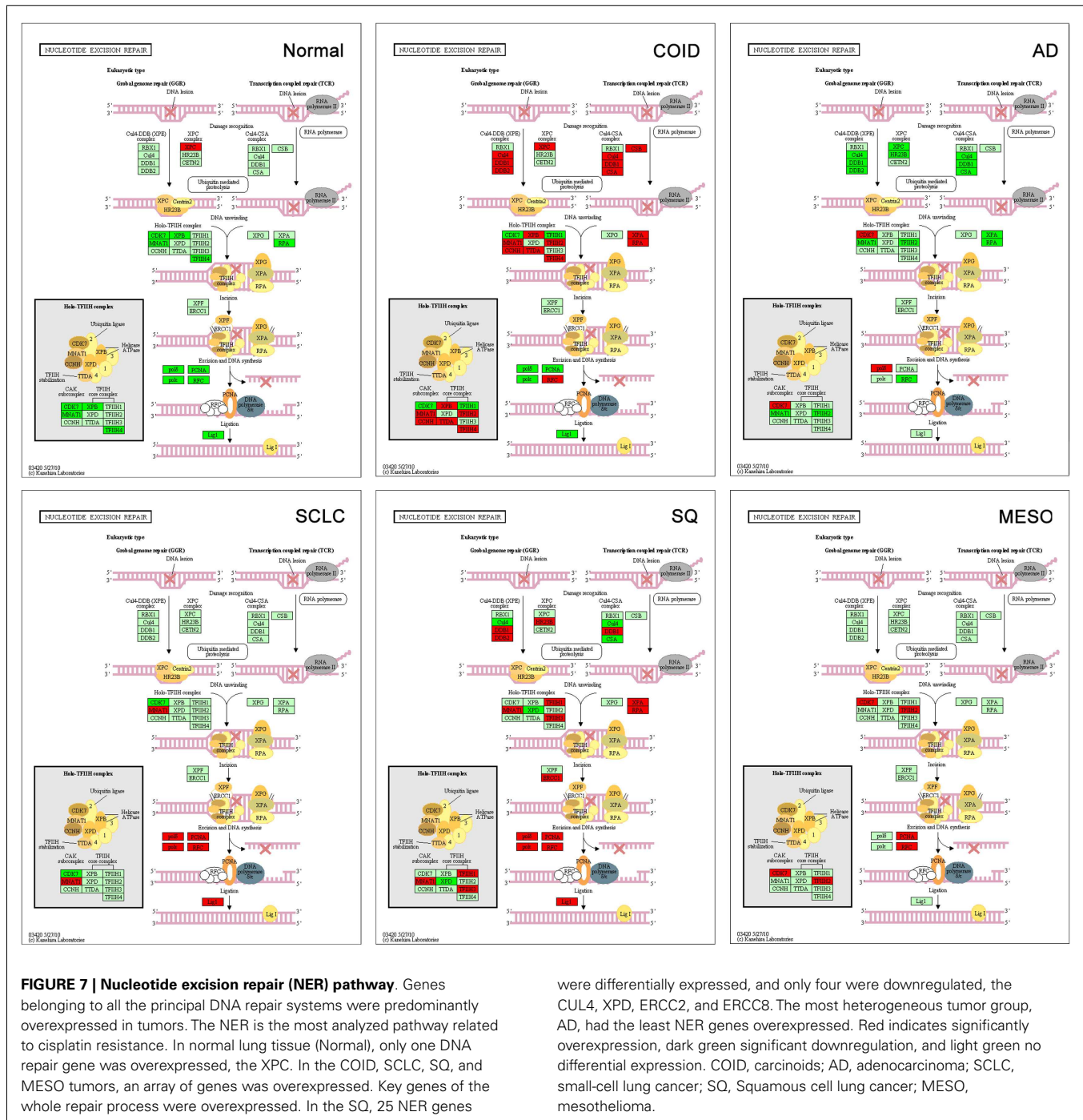


Table 4 | Differential expression for each tumor type versus rest in KEGG pathways related to DNA repair.

Genes	NER	BER	NHEJ	HR	MMR
NOR	XPC	None	None	None	None
COID	<i>CCNH, CUL4B, DDB1, DDB2, ERCC3, ERCC6, ERCC8, GTF2H2, GTF2H4, MNAT1, RFC1, RFC2, RFC4, RPA2, XPA, XPC</i>	<i>LIG3, MUTYH, OGG1, XRCC1</i>	<i>DNL4, LIG4, RAD50, XRCC6 (KU70)</i>	<i>RAD51C, RAD51D, XRCC2</i>	<i>MLH1, MLH3, MSH3, RFC1, RFC2, RFC4, RPA2</i>
AD	CDK7, POLD4	<i>MPG, NTHL1, POLD4, PARP3, PARP4</i>	<i>MRE11A, XRCC4, XRCC6 (KU70)</i>	<i>MRE11A, POLD4</i>	POLD4
SCLC	LIG1, MNAT1, PCNA, POLD3, POLD4, POLE, POLE2, POLE3, RFC3, RFC4, RFC5	<i>FEN1, LIG1, MUTYH, PARP1, PARP2, PARP3, PCNA, POLD3, POLD4, POLE, POLE2TDG, UNG</i>	FEN1	<i>BLM, BRCA2, POLD3, POLD4, RAD54L, RAD51C</i>	<i>EXO1, LIG1, MSH2, MSH6, POLD3, POLD4, PCNA, RFC3, RFC4, RFC5</i>
SQ	<i>DDB1, DDB2, ERCC1, GTF2H1, GTF2H3, LIG1, MNAT1, PCNA, POLD2, POLD4, POLE, POLE2, POLE3, RAD23A, RAD23B, RFC2, RFC3, RFC4, RFC5, RPA1, RPA2, XPA</i>	<i>APEX2, FEN1, HMBG1, LIG1, MBD4, PCNA, POLD2, POLD4, POLE, POLE2, POLE3, POLB, PARP1, PARP2, PARP3, TDG, UNG</i>	<i>LIG4, FEN1, PRKDC, XRCC5 (KU80)</i>	<i>BLM, BRCA2, POLD2, POLD4, RPA1, RPA2, RAD51C, RAD51D, RAD54L, SHFM1</i>	<i>LIG1, MSH2, MSH6, POLD2, POLD4, PCNA, RFC2, RFC3, RFC4, RFC5, RPA1, RPA2</i>
MESO	CDK7, GTF2H2, PCNA, RFC4, RFC5	FEN1, PCNA	FEN1, RAD50, XRCC4	BRCA2, DSS1, EME1, RAD50, SHFM1, RAD54L, SSBP1	MSH6, PCNA, RFC4, RFC5, SSBP1

For each pathway, only the overexpressed genes are shown. Genes that are common between several lung cancer subtypes are in bold. The mesothelioma results are based on mesothelioma versus parietal pleura (see Materials and Methods).

BER, base excision repair; HR, homologous recombination; MMR, mismatch repair; NER, nucleotide excision repair; NHEJ, non-homologous end joining; NOR, normal lung tissue; COID, carcinoids; AD, adenocarcinoma; SCLC, small-cell lung cancer; SQ, squamous cell lung cancer; MESO, mesothelioma.

Other important platinum resistance repair pathways were highly overexpressed, including the HR pathway. The HR is the principal DNA double-strand break (DSB), non-error-prone repair mechanism that takes place in the late S-G2 phase of the cell cycle and involves generation of a single-stranded region of DNA, strand invasion, formation of a Holliday junction, DNA synthesis using the intact strand as a template, branch migration, and resolution. Our results showed that SQ and SCLC had overexpressed genes throughout the pathway with 10 of 11 genes overexpressed including the very important *BRCA2*, *RAD51* paralogs and *BLM* (Bloom syndrome mutated) (Figure 8) (55). In normal lung tissue, none of the HR genes were overexpressed.

One indispensable part of cisplatin induced DNA inter-strand crosslink repair is the Fanconi anemia pathway (56, 57). The crosslinks are recognized by *FANCM* and associated proteins, recruiting the FA core complex. The *FANCD2* and *FANCI* are monoubiquitinated by the FA core complex. The monoubiquitinated *FANCD2/FANCI* becomes an active form and interacts with a series of DNA repair proteins and facilitates downstream repair pathways. In the normal tissue, all *FANCD2/FANCI* genes were downregulated, while in all tumors several were overexpressed (File S1 in Supplementary Material).

In SCLC, the standard treatment consists of etoposide combined with cisplatin, which gives a high response rate but an almost 100% recurrence rate. Etoposide is a DNA topoisomerase II poison that prevents the re-ligation of topoisomerase II-induced

single-strand breaks. However, the DNA repair profile was very similar to the SQ, so that other genes also interfere both for the higher response, and the high relapse rate. A recent study on deciphering the resistance genes of etoposide, the polymerase beta (*POLB*), and the homolog of *TTF1*, *NKX2-2* were identified. Overexpression of *NKX2-2* has been shown to predict a more dismal survival (58). In our data, *NKX2-2* was overexpressed only in the SCLC with a diagnostic AUC of 0.85 (Table 3).

PRIMARY VERSUS METASTATIC ADENOCARCINOMA

The distribution of the gene expressions of metastatic AD from colon ($n = 8$), breast cancer ($n = 2$), and unspecified ($n = 3$) was in the same range as the primary lung AD (Figure 4), so the overall expression profiles were similar. However, differential expression of 267 genes was detected in metastatic versus primary tumors. The gene with highest discriminatory power was the *SLC26A3*, synonymous to downregulated in adenoma (DRA) that was overexpressed in the metastatic AD. This gene encodes a transmembrane glycoprotein that transports chloride ions across the cell membrane in exchange for bicarbonate ions and is specifically localized to the normal mucosa of the lower intestinal tract. Downregulation of this gene was not only inversely correlated to adenomas and ADs of the colon but also to the clinical stage of the tumors (59, 60). However, this gene/protein is not expressed in normal lung, or in lung cancer (36), so this finding probably reflected the relative overweight of colon cancer metastases in this material, and how

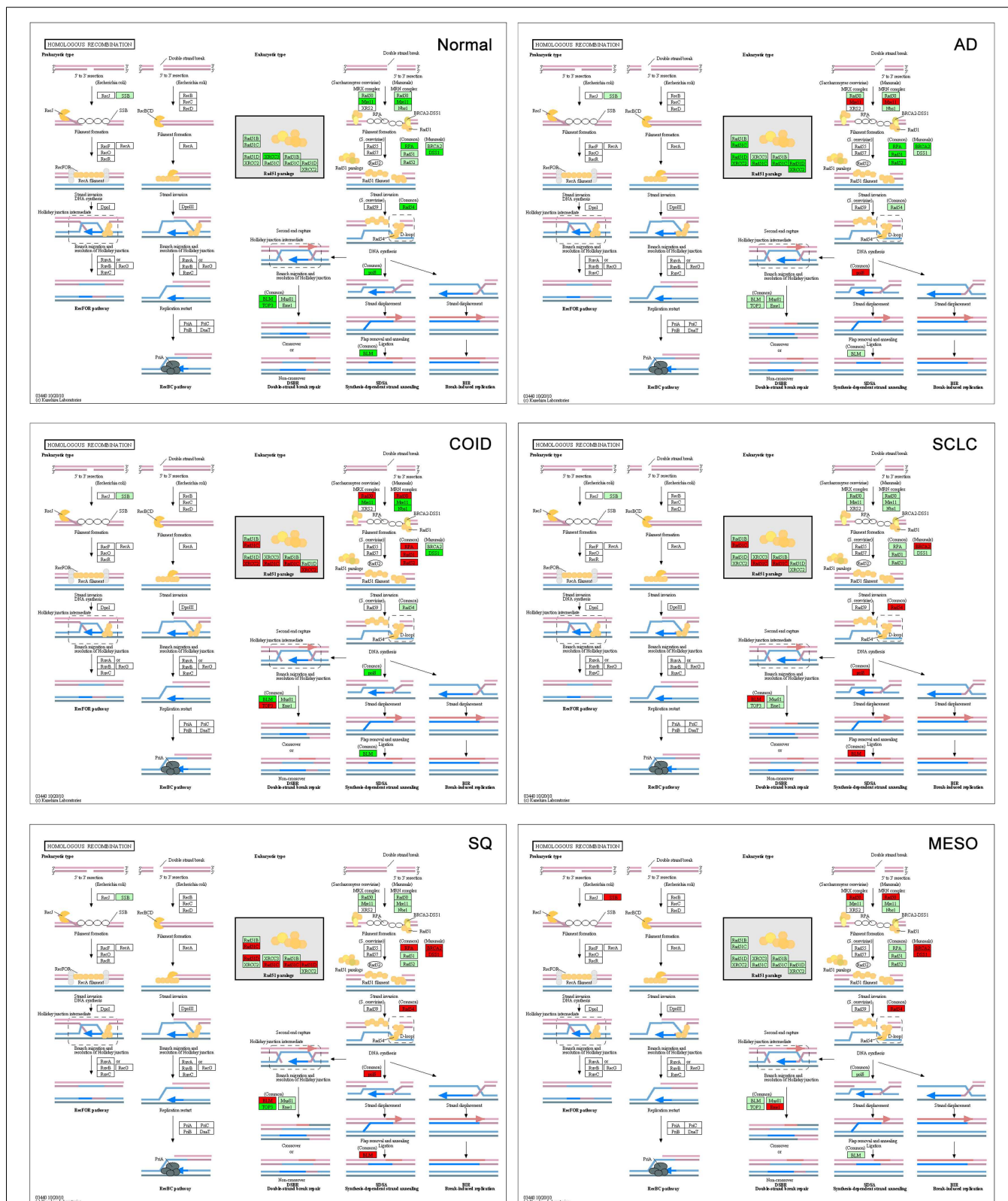


FIGURE 8 | Homologous recombination (HR) pathway. The HR is the principal, non-error-prone, repair mechanism for DNA double-strand breaks. SQ and SCLC have 10 of 11 overexpressed genes throughout the pathway, including the very important BRCA2, RAD51 paralogs, and BLM (Blom

syndrome mutated). Red indicates significantly overexpression, dark green significant downregulation, and light green no differential expression. COID, carcinoids; AD, adenocarcinoma; SCLC, small-cell lung cancer; SQ, Squamous cell lung cancer; MESO, mesothelioma.

such an analysis could identify putative markers for colon cancer metastasis to the lung.

NORMAL LUNG TISSUE

There are few published accounts of the gene profile of normal lung tissue, as it usually is used as a control to describe tumor tissue. However, in this analysis, we have also tried to describe the normal tissue as related to the tumors. Actually the observed, almost non-existent overexpression, and even downregulation of DNA repair systems pathways was a novel finding (Figure S2 in Supplementary Material). That does not mean that the normal tissue do not have DNA repair, only that the levels are much lower in relation to the cancers. In this respect, the finding that some other crucial mechanisms for cell maintenance and survival were overexpressed in normal tissue while being downregulated or indifferent in the tumors, was surprising. Below, we will discuss some of the pathways that clearly were most upregulated in the normal tissue and downregulated in the cancers.

Vascular smooth muscle contraction

Contraction of vasculature is important in controlling the blood flow and influx of oxygen and nutrients to tissues. Vascular smooth muscle contraction pathway genes were significantly overrepresented and upregulated in normal tissue (15/36 genes, $p < 0.0001$, File S4 in Supplementary Material) opposed to SQ, SCLC, and mesothelioma, were none of the genes of the smooth muscle cell membrane were overexpressed in the cancers (Figure 9). Recently, a bioinformatics study on gene expression of lung cancer of unspecified type versus normal showed that vascular smooth muscle contraction pathway was negatively regulated in tumor, in line with our findings (61). The significantly overexpressed genes in the normal lung included potassium channel genes (*KCNMB1*), the vaso-constricting angiotensin receptor II genes (*AGTR1*), endothelin receptor type A (*EDNRA*), and vasodilating genes as the adenosine (*ADORA2B*), calcitonin (*CRLCR*, *RAMPI*, *RAMP2*, and *RAMP3*), and natriuretic peptide receptor (*NPR1*). How can this be explained? Normal vasculature needs a well-functioning vasoconstrictor and dilator system, while some of these functions may be defective in pathological tumor vessels formed by neo-angiogenesis (62). Tumors often have a higher microvessel density than normal tissues; however, these pathological blood vessels are often described as less elastic and with different physiology than normal vessels. As an example, the endothelium of the microvessels has been incriminated in promoting tumor growth by autocrine loops (63).

An intriguing finding was the overexpression of the *ILK* gene (integrin-linked kinase) in normal lung ($p = 3.71 \times 10^{-7}$) and its downregulation in SQ ($p = 0.0037$). Loss of this integrin-linked kinase induces failure in the formation of a unitary layer in ILK-deficient vascular smooth muscle cells and also induces abnormal contractility, through the activation of *RHOA/ROCK2* that was also overexpressed in normal lung (64). In the literature, we were not able to identify accounts of this gene profile in lung tumors and normal lung tissue.

Leukocyte trans-endothelial migration

Leukocyte trans-endothelial migration pathway genes were highly overrepresented and upregulated in normal tissue versus the rest

(16/35 genes, $p < 0.0001$, File S4 in Supplementary Material), and thus downregulated in the tumors. The migration of cells through the vasculature is an important mechanism that is complex, active, and requires the presence and activity of several genes. This mechanism is very important for the flux of neutrophils in case of inflammation but also lymphocytes for cytotoxic T-lymphocytes and killer T-cells. Leukocyte trans-endothelial migration downregulation was found to correlate with recurrence after operation of stage I lung cancer in four large datasets, and thus seem to play a role in tumor aggressiveness (65). We have previously shown that this pathway was downregulated in mesothelioma versus normal parietal pleura and that it was even more downregulated after acquired pemetrexed and platinum resistance (66). Here, we show that it is a consistent trait in all the lung tumors (Figure S3 in Supplementary Material). Among the consistently downregulated genes in tumors were the *PECAM1*, *JAM2*, and *CDH5*. The *PECAM1/CD31* (platelet/endothelial cell adhesion molecule) is a member of the immunoglobulin superfamily that is expressed on the surface of platelets, monocytes, neutrophils, and T-cell subsets and is also a major constituent of the endothelial cell intercellular junction. It modulates multiple functions besides trans-endothelial migration, integrin-mediated cell adhesion, angiogenesis, apoptosis, cell migration, and negative regulation of immune cell signaling (67). Inhibition of the *JAM2/JAMB* (junctional adhesion molecule) decreases leukocyte infiltration (68). Vascular endothelial cadherin or *CDH5* is a key protein controlling the endothelial barrier and its disruption by specific antibody both amplifies metastasis in normal mice and overcomes the genetic resistance in mice (69). Moreover, *CDH5* is a candidate tumor suppressor and low expression strongly correlated to decreased survival in neuroblastoma (70). One is tempted to speculate that this is a trait of aggressive and treatment resistant cancers as well, where the antitumor immune system cannot reach the tumor cell due to hindering the entrance of immunological cells in tumor resulting in a barrier to immunologic tumor rejection and thus to immunological therapies. Interestingly, this coincides with results from an immunotherapy trial showing good responses in mesothelioma, but only in small tumors (71).

Actin cytoskeleton

Regulation of actin cytoskeleton genes was the top overrepresented and upregulated pathway in normal tissue versus the rest (29/58 genes, $p < 0.0001$, File S4 in Supplementary Material) (Figure S3 in Supplementary Material). Actin cytoskeleton is larger than any organelle, and is comprised of actin filaments and microtubules, that make cell locomotion and cell division possible, as well as keeping cell polarity and morphology. Within this pathway some genes were uniquely overexpressed in normal tissue including *IQGAP1* and *IQGAP2* (IQ motif containing GTPase activating protein). These are members of a family of scaffold proteins where most studied *IQGAP1* modulates several cellular functions, including cell–cell adhesion, transcription, cytoskeletal architecture, migration, and selected signaling pathways, reviewed in Ref. (72). Overexpression of this gene has been associated to several cancer types, in contrast to our findings, where it was downregulated or indifferent in the tumors. However, *IQGAP2* overexpression is protective against tumorigenesis, as the loss or

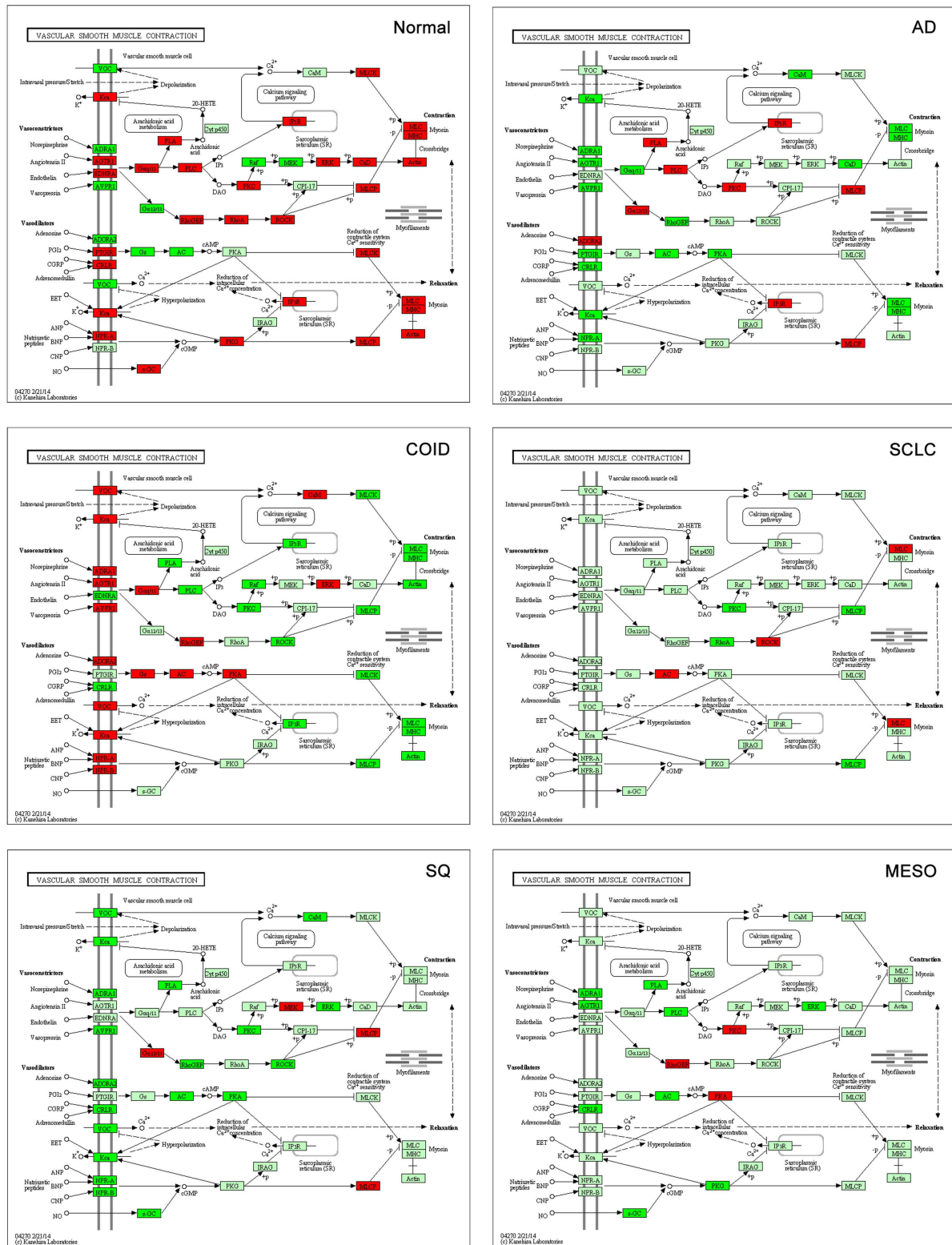


FIGURE 9 | Vascular smooth muscle contraction pathway. Genes belonging to the vascular smooth muscle contraction pathway were significantly overrepresented and upregulated in normal tissue (15/36 genes, $p < 0.0001$). None of the genes of the smooth muscle cell membrane (according to KEGG pathways) were overexpressed in SQ,

SCLC, and mesothelioma. Red indicates significant overexpression, dark green significant downregulation, and light green no differential expression. COID, carcinoids; AD, adenocarcinoma; SCLC, small-cell lung cancer; SQ, squamous cell lung cancer; MESO, mesothelioma.

downregulation of *IQGAP2* has been shown to play a role in development of hepatocellular, gastric, and prostate cancer (73). Another uniquely overexpressed gene in normal lung tissue was the *RHOA* (ras homolog family member A). The *RHOA* is involved in a multitude of processes as controlling epithelial cell junctions as well as regulating microtubule function and cytokinesis (74). It stimulates the formation of active non-muscle myosin filaments and long, unbranched F-actin, and is crucial for cell polarity, of which is disturbed in epithelial–mesenchymal transition (*EMT*) during metastasis. Cytokinesis dysregulation also leads to the formation of polyploid and aneuploid cells that are prone to tumor formation. Recently, it was shown that the increased degradation of *RHOA* encoded protein RhoA by altered autophagy induced pathological cytokinesis resulting in multinucleation and aneuploidy (75). In the Human Protein Atlas, the RhoA is expressed in all normal respiratory tissues while only in 34% of lung cancers, supporting our findings (36).

GENES CORRELATED TO TUMOR SIZE BUT NOT TO SURVIVAL

The original dataset provided survival data only for the AD patients, and all were treated surgically. Only type of operation performed and TNM stage was shown, but no information on adjuvant or systemic treatment in general. Eligible for survival analysis were 31 cases in Stage IA, 40 patients with Stage IB, four with Stage IIA, 20 with Stage IIB, 7 with Stage IIIA, 3 with Stage IIIB, 3 with Stage IV, in total 108 cases. In the original paper, an AD subgroup called C2 or neuroendocrine expression type, was claimed to have a less favorable survival compared to the rest of the AD and that one gene, kallikrein 11 (*KLK11*) could separate this group from the rest of the tumors (6). However, the C2 cases were only nine, making valid conclusions on survival difficult, and the kallikrein 11 was upregulated in the other groups as well (original publication, **Figure 4**). In our analysis of only the primary lung AD (**Figure 1**), no single gene or gene signature could predict survival. This is not surprising as the survival is an endpoint influenced by many factors, and this population had a very heterogenous stage spectrum, probably heterogenous systemic treatment that was not reported and gene-expression heterogeneity as we reported. Probably, a clinically more homogenous group would provide a better basis for discovery of prognostic and predictive signatures.

There was, however, significant correlation of four genes with tumor size, where one gene, *NR2C1* (nuclear receptor subfamily 2, group C, member 1), showed a positive correlation that is increased expression with larger tumor (**Table 2**). Interestingly, this not very well-characterized gene seems to play an important role in early embryonic development by regulating key genes involved in stem cell self-renewal, commitment, and differentiation (76). Finally, there were two genes significantly downregulated in N0 versus N1–3 cases. One was the *FLNB*, one of the mammalian filamins, large actin-binding protein that is important for migration of cells (77).

NOVEL BIOINFORMATICS APPLICATIONS VERSUS THE METHODS USED IN THE ORIGINAL PAPER

Several groundbreaking advancements have been introduced in the field of bioinformatics in the 13 years that separate the Bhat-tacharjee study and this analysis. Some of the statistical methods

used in this work, for example, quantile normalization (78), were either not available or not widely known at the time of the first study, while they are nowadays widely accepted and used.

Importantly, the last 10 years have witnessed a dramatic effort by the bioinformatics community for creating and maintaining comprehensive on-line resources able to store in a structured way the massive mole of biological knowledge that is steadily generated. The KEGG pathways database (79), the Gene Expression Omnibus (80), and the REACTOME database (81, 82) are just a few examples of such resources.

Particularly, most of the conclusions drawn in the present work were obtained by integrating data-driven results (set of differentially expressed genes) with the pathway-oriented information provided by the KEGG and GO database. Such type of study would have been severely limited back in 2001, when the information stored in these repositories was much scarcer.

STUDY DRAWBACKS

The main weakness of this study is that it is based on transcriptome data obtained in the early days of microarray technology, when mRNA extraction procedure and microarray technical solution were not as mature, optimized, and reliable as today. As a first consequence, the number of gene-expression levels measured in the original study was inferior to the number of genes that can now be quantified. There was no other dataset analyzing identical subgroups, so a direct comparison with our study could not be performed. However, the results are aligned with other studies in several respects indirectly showing the high quality of the data, and comparative analysis performed on two separate datasets showed a high concordance of differentially expressed genes, ranked gene lists and AUC (see Text S1 in Supplementary Material). Another problem is the scarcity of SCLC cases, which can be attributed to the fact that they are seldom operated and fine needle cytology often gain too few cells for analysis. Ideally, large cell or other rare subgroups of NSCLC should have been analyzed, but these were not included in this dataset. Based on our findings, we would recommend to use old microarray datasets for novel data mining, given that the study design and the sample treatment are of provably adequate quality.

CONCLUSION

The lung cancer gene-expression profiles provided through this analysis, recapitulated some of the known features of classical histological subtypes, indicated novel candidate diagnostic markers and outlined new information on differences between normal lung tissue and tumor. The overrepresentation and overexpression of DNA repair genes was remarkably consistent in all tumors, and recapitulated known treatment resistance patterns. Novel interesting features of upregulated pathways in normal lung may shed some light on the biology of the normal lung as well as in tumors, deserving further study. This study provides an example of how the integrative *in silico* analysis of “old data” in conjunction with novel biological knowledge and computational techniques can provide much information not deducible at the time when the data were produced.

AUTHOR CONTRIBUTIONS

Conceived and designed the experiments: Oluf Dimitri Røe, Ioannis Tsamardinos, Vincenzo Lagani, and Konstantinos Kerkentzes. Performed the experiments: Konstantinos Kerkentzes, Vincenzo Lagani, and Ioannis Tsamardinos. Analyzed the data: Oluf Dimitri Røe, Ioannis Tsamardinos, Vincenzo Lagani, and Konstantinos Kerkentzes. Contributed reagents/materials/analysis tools: Ioannis Tsamardinos, Mogens Vyberg, and Oluf Dimitri Røe. Wrote the paper: Konstantinos Kerkentzes, Oluf Dimitri Røe, Ioannis Tsamardinos, Vincenzo Lagani, and Mogens Vyberg.

ACKNOWLEDGMENTS

We thank Prof. Hans E. Krokan for critical reading of the manuscript. Dr. Oluf Dimitri Røe is funded by the Liaison Committee between the Central Norway Regional Health Authority (RHA) and the Norwegian University of Science and Technology (NTNU), grant no. 46056639. Ioannis Tsamardinos and Konstantinos Kerkentzes were partially funded by STATegra EU FP7 project, No. 306000 and EPILOGEAS GSRT ARISTEIA II project, No. 3446.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at <http://www.frontiersin.org/Journal/10.3389/fonc.2014.00251/abstract>

Figure S1 | Clinically relevant differentially expressed pathways in each tissue type versus the rest. The pathway maps are arranged according to the scheme in the left-top corner. Red indicates overexpression in genes, dark green downregulation. In the Cell Cycle pathway, tumor suppressor genes were overexpressed in normal lung tissue while oncogenes and tumor driving cyclins were overexpressed in cancers. The gene thymidylate synthase (TYMS or TS, belonging to the pathway “one carbon pool by folate”) is relevant for tumor growth and is also a treatment target. Notably, TYMS was overexpressed only in the tumors that are generally refractory to the drug pemetrexed, as the squamous and the small-cell lung cancer. TYMS was not overexpressed in mesothelioma, but it is known that TYMS expression is highly variable in this cancer. In the ERBB pathway, the ERBB2/HER2 and ERBB3/HER3 were overexpressed in adenocarcinoma, while the ERBB4 was overexpressed in the carcinoids.

Figure S2 | Genes in overrepresented DNA repair pathways are predominantly overexpressed in cancer. The figure reports the base excision repair (BER), homologous recombination (HR), and mismatch repair (MMR) pathways. Red indicates significantly overexpressed genes, dark green significantly downregulated, and light green not differentially expressed.

Figure S3 | Pathways predominantly overexpressed in the normal lung tissue. The regulation of KEGG pathways actin cytoskeleton, leukocyte trans-endothelial migration, and smooth muscle vasculature contraction. Red indicates significantly overexpressed genes, dark green significantly downregulated, and light green not differentially expressed.

Text S1 | Detailed report of the validation procedure and results.

File S1 | Differentially expressed genes lists. A zip file containing the differentially expressed genes of each histological type versus the rest and of the metastatic versus the primary adenocarcinomas.

File S2 | AUC lists. A zip file containing the AUC of the differentially expressed genes of each histological type versus the rest and of the metastatic versus the primary adenocarcinomas.

File S3 | GO analysis. A zip file containing the enriched GO biological processes resulting from the hyper-geometric test for each histological type versus the rest.

File S4 | KEGG analysis. A zip file containing the enriched KEGG pathways processes resulting from the hyper-geometric test for each histological type versus the rest.

REFERENCES

- Alberg AJ, Brock MV, Ford JG, Samet JM, Spivack SD. Epidemiology of lung cancer: diagnosis and management of lung cancer, 3rd ed: American College of Chest Physicians evidence-based clinical practice guidelines. *Chest* (2013) **143**:1S–29S. doi:10.1378/chest.12-2345
- Mirsadraee S, Oswal D, Alizadeh Y, Caulo A, van Beek E Jr. The 7th lung cancer TNM classification and staging system: review of the changes and implications. *World J Radiol* (2012) **4**:128–34. doi:10.4329/wjr.v4.i4.128
- Travis WD, Brambilla E, Riely GJ. New pathologic classification of lung cancer: relevance for clinical practice and clinical trials. *J Clin Oncol* (2013) **31**:992–1001. doi:10.1200/JCO.2012.46.9270
- Smit EF, Socinski MA, Mullaney BP, Myrand SP, Scagliotti GV, Lorigan P, et al. Biomarker analysis in a phase III study of pemetrexed-carboplatin versus etoposide-carboplatin in chemo-naïve patients with extensive-stage small-cell lung cancer. *Ann Oncol* (2012) **23**:1723–9. doi:10.1093/annonc/mdr563
- Subramanian J, Govindan R. Chemotherapy: continued lack of progress in SCLC. *Nat Rev Clin Oncol* (2010) **7**:77–8. doi:10.1038/nrclinonc.2009.223
- Bhattacharjee A, Richards WG, Staunton J, Li C, Monti S, Vasa P, et al. Classification of human lung carcinomas by mRNA expression profiling reveals distinct adenocarcinoma subclasses. *Proc Natl Acad Sci U S A* (2001) **98**:13790–5. doi:10.1073/pnas.191502998
- Røe OD, Anderssen E, Helge E, Pettersen CH, Olsen KS, Sandeck H, et al. Genome-wide profile of pleural mesothelioma versus parietal and visceral pleura: the emerging gene portrait of the mesothelioma phenotype. *PLoS One* (2009) **4**:e6554. doi:10.1371/journal.pone.0006554
- Golub TR, Slonim DK, Tamayo P, Huard C, Gaasenbeek M, Mesirov JP, et al. Molecular classification of cancer: class discovery and class prediction by gene expression monitoring. *Science* (1999) **286**:531–7. doi:10.1126/science.286.5439.531
- Gautier L, Cope L, Bolstad BM, Irizarry RA. Affy – analysis of Affymetrix GeneChip data at the probe level. *Bioinformatics* (2004) **20**:307–15. doi:10.1093/bioinformatics/btg405
- Smyth GK. Linear models and empirical bayes methods for assessing differential expression in microarray experiments. *Stat Appl Genet Mol Biol* (2004) **3**:Article3. doi:10.2202/1544-6115.1027
- Gentleman RC, Carey VJ, Bates DM, Bolstad B, Dettling M, Dudoit S, et al. Bioconductor: open software development for computational biology and bioinformatics. *Genome Biol* (2004) **5**:R80. doi:10.1186/gb-2004-5-10-r80
- Fawcett T. An introduction to ROC analysis. *Pattern Recognit Lett* (2006) **27**:861–74. doi:10.1016/j.patrec.2005.10.010
- Falcon S, Gentleman R. Using GStats to test gene lists for GO term association. *Bioinformatics* (2007) **23**:257–8. doi:10.1093/bioinformatics/btl567
- Storey JD. A direct approach to false discovery rates. *J R Stat Soc Series B Stat Methodol* (2002) **64**:479–98. doi:10.1111/1467-9868.00346
- Lagani V, Tsamardinos I. Structure-based variable selection for survival data. *Bioinformatics* (2010) **26**:1887–94. doi:10.1093/bioinformatics/btq261
- Statnikov A, Aliferis CE, Tsamardinos I, Hardin D, Levy S. A comprehensive evaluation of multicategory classification methods for microarray gene expression cancer diagnosis. *Bioinformatics* (2005) **21**:631–43. doi:10.1093/bioinformatics/bti033
- Yap YL, Lam DC, Luc G, Zhang XW, Hernandez D, Gras R, et al. Conserved transcription factor binding sites of cancer markers derived from primary lung adenocarcinoma microarrays. *Nucleic Acids Res* (2005) **33**:409–21. doi:10.1093/nar/gki188
- Hou J, Aerts J, den Hamer B, van Ijcken W, den Bakker M, Riegman P, et al. Gene expression-based classification of non-small cell lung carcinomas and survival prediction. *PLoS One* (2010) **5**:e10312. doi:10.1371/journal.pone.0010312
- Lottaz C, Yang X, Scheid S, Spang R. OrderedList – a bioconductor package for detecting similarity in ordered gene lists. *Bioinformatics* (2006) **22**:2315–6. doi:10.1093/bioinformatics/btl385

20. Graf E, Schmoor C, Sauerbrei W, Schumacher M. Assessment and comparison of prognostic classification schemes for survival data. *Stat Med* (1999) **18**:2529–45. doi:10.1002/(SICI)1097-0258(19990915/30)18:17/18<2529:AID-SIM274>3.0.CO;2-5
21. Paz-Ares LG, de Marinis F, Dediu M, Thomas M, Pujol JL, Bidoli P, et al. PARAMOUNT: final overall survival results of the phase III study of maintenance pemetrexed versus placebo immediately after induction treatment with pemetrexed plus cisplatin for advanced nonsquamous non-small-cell lung cancer. *J Clin Oncol* (2013) **31**(23):2895–902. doi:10.1200/JCO.2012.47.1102
22. Chen X, Liu Y, Røe OD, Qian Y, Guo R, Zhu L, et al. Gefitinib or erlotinib as maintenance therapy in patients with advanced stage non-small cell lung cancer: a systematic review. *PLoS One* (2013) **8**:e59314. doi:10.1371/journal.pone.0059314
23. Shaw AT, Kim DW, Nakagawa K, Seto T, Crinó L, Ahn MJ, et al. Crizotinib versus chemotherapy in advanced ALK-positive lung cancer. *N Engl J Med* (2013) **368**:2385–94. doi:10.1056/NEJMoa1214886
24. Bishop JA, Teruya-Feldstein J, Westra WH, Pelosi G, Travis WD, Rekhtman N. p40 (DeltaNp63) is superior to p63 for the diagnosis of pulmonary squamous cell carcinoma. *Mod Pathol* (2012) **25**:405–15. doi:10.1038/modpathol.2011.173
25. Brunnström H, Johansson L, Jirström K, Jönsson M, Jönsson P, Planck M. Immunohistochemistry in the differential diagnostics of primary lung cancer: an investigation within the Southern Swedish Lung Cancer Study. *Am J Clin Pathol* (2013) **140**:37–46. doi:10.1309/AJCP50RDXSCSBTB0
26. Jagirdar J. Application of immunohistochemistry to the diagnosis of primary and metastatic carcinoma to the lung. *Arch Pathol Lab Med* (2008) **132**:384–96. doi:10.1043/1543-2165(2008)132
27. Ordóñez NG. Value of thyroid transcription factor-1 immunostaining in tumor diagnosis: a review and update. *Appl Immunohistochem Mol Morphol* (2012) **20**:429–44. doi:10.1097/PAI.0b013e31825439bc
28. Ordóñez NG. Thyroid transcription factor-1 is not expressed in squamous cell carcinomas of the lung: an immunohistochemical study with review of the literature. *Appl Immunohistochem Mol Morphol* (2012) **20**:525–30. doi:10.1097/PAI.0b013e318251d8c1
29. Rekhtman N, Ang DC, Sima CS, Travis WD, Moreira AL. Immunohistochemical algorithm for differentiation of lung adenocarcinoma and squamous cell carcinoma based on large series of whole-tissue sections with validation in small specimens. *Mod Pathol* (2011) **24**:1348–59. doi:10.1038/modpathol.2011.92
30. Ring BZ, Seitz RS, Beck RA, Shasteen WJ, Soltermann A, Arbogast S, et al. A novel five-antibody immunohistochemical test for subclassification of lung carcinoma. *Mod Pathol* (2009) **22**:1032–43. doi:10.1038/modpathol.2009.60
31. Zinman GE, Naiman S, Kanfi Y, Cohen H, Bar-Joseph Z. ExpressionBlast: mining large, unstructured expression databases. *Nat Methods* (2013) **10**:925–6. doi:10.1038/nmeth.2630
32. Zhang LQ, Jiang F, Xu L, Wang J, Bai JL, Yin R, et al. The role of cyclin D1 expression and patient's survival in non-small-cell lung cancer: a systematic review with meta-analysis. *Clin Lung Cancer* (2012) **13**:188–95. doi:10.1016/j.clcc.2011.10.003
33. Castro M, Grau L, Puerta P, Gimenez L, Venditti J, Quadrelli S, et al. Multiplexed methylation profiles of tumor suppressor genes and clinical outcome in lung cancer. *J Transl Med* (2010) **8**:86. doi:10.1186/1479-5876-8-86
34. Usuda H, Naito M, Saito T, Emura I. Immunohistochemistry of cyclin D3 in pulmonary carcinomas. *Virchows Arch* (1996) **428**:159–63. doi:10.1007/BF00200658
35. Radhakrishnan VM, Jensen TJ, Cui H, Futscher BW, Martinez JD. Hypomethylation of the 14-3-3 sigma promoter leads to increased expression in non-small cell lung cancer. *Genes Chromosomes Cancer* (2011) **50**:830–6. doi:10.1002/gcc.20904
36. Berglund L, Björling E, Oksvold P, Fagerberg L, Asplund A, Szgyarto CA, et al. A gene-centric human protein atlas for expression profiles based on antibodies. *Mol Cell Proteomics* (2008) **7**(10):2019–27. doi:10.1074/mcp.R800013-MCP200
37. Swarts DR, Van Neste L, Henfling ME, Eijkenboom I, Eijk PP, van Velthuyzen ML, et al. An exploration of pathways involved in lung carcinoid progression using gene expression profiling. *Carcinogenesis* (2013) **34**(12):2726–37. doi:10.1093/carcin/bgt271
38. Zhao R, Zhang S, Hanscom M, Chattopadhyay S, Goldman ID. Loss of reduced folate carrier function and folate depletion result in enhanced pemetrexed inhibition of purine synthesis. *Clin Cancer Res* (2005) **11**:1294–301.
39. Schultz RM, Chen VJ, Bewley JR, Roberts EF, Shih C, Dempsey JA. Biological activity of the multitargeted antifolate, MTA (LY231514), in human cell lines with different resistance mechanisms to antifolate drugs. *Semin Oncol* (1999) **26**:68–73.
40. van der Wilt CL, Backus HH, Smid K, Comijn L, Veerman G, Wouters D, et al. Modulation of both endogenous folates and thymidine enhance the therapeutic efficacy of thymidylate synthase inhibitors. *Cancer Res* (2001) **61**:3675–81.
41. Wang T, Chuan Pan C, Rui Yu J, Long Y, Hong Cai X, De Yin X, et al. Association between TYMS expression and efficacy of pemetrexed-based chemotherapy in advanced non-small cell lung cancer: a meta-analysis. *PLoS One* (2013) **8**:e74284. doi:10.1371/journal.pone.0074284
42. Grønberg BH, Bremnes RM, Aasebø U, Brunsvig P, Flotten O, Amundsen T, et al. A prospective phase II study: high-dose pemetrexed as second-line chemotherapy in small-cell lung cancer. *Lung Cancer* (2009) **63**:88–93. doi:10.1016/j.lungcan.2008.04.003
43. Socinski MA, Smit EF, Lorigan P, Konduri K, Reck M, Szczesna A, et al. Phase III study of pemetrexed plus carboplatin compared with etoposide plus carboplatin in chemotherapy-naïve patients with extensive-stage small-cell lung cancer. *J Clin Oncol* (2009) **27**:4787–92. doi:10.1200/JCO.2009.23.1548
44. Scagliotti GV, Parikh P, von Pawel J, Biesma B, Vansteenkiste J, Manegold C, et al. Phase III study comparing cisplatin plus gemcitabine with cisplatin plus pemetrexed in chemotherapy-naïve patients with advanced-stage non-small-cell lung cancer. *J Clin Oncol* (2008) **26**:3543–51. doi:10.1200/JCO.2007.15.0375
45. Zucali PA, Giovannetti E, Destro A, Mencoboni M, Ceresoli GL, Gianoncelli L, et al. Thymidylate synthase and excision repair cross-complementing group-1 as predictors of responsiveness in mesothelioma patients treated with pemetrexed/carboplatin. *Clin Cancer Res* (2011) **17**:2581–90. doi:10.1158/1078-0432.CCR-10-2873
46. Wu MF, Hsiao YM, Huang CF, Huang YH, Yang WJ, Chan HW, et al. Genetic determinants of pemetrexed responsiveness and nonresponsiveness in non-small cell lung cancer cells. *J Thorac Oncol* (2010) **5**:1143–51. doi:10.1097/JTO.0b013e3181e0b954
47. Hsieh AC, Moasser MM. Targeting HER proteins in cancer therapy and the role of the non-target HER3. *Br J Cancer* (2007) **97**:453–7. doi:10.1038/sj.bjc.6603910
48. Leidner RS, Fu P, Clifford B, Hamdan A, Jin C, Eisenberg R, et al. Genetic abnormalities of the EGFR pathway in African American patients with non-small-cell lung cancer. *J Clin Oncol* (2009) **27**:5620–6. doi:10.1200/JCO.2009.23.1431
49. Mazières J, Peters S, Lepage B, Cortot AB, Barlesi F, Beau-Faller M, et al. Lung cancer that harbors an HER2 mutation: epidemiologic characteristics and therapeutic perspectives. *J Clin Oncol* (2013) **31**:1997–2003. doi:10.1200/JCO.2012.45.6095
50. Sequist LV, Yang JC, Yamamoto N, O'Byrne K, Hirsh V, Mok T, et al. Phase III study of afatinib or cisplatin plus pemetrexed in patients with metastatic lung adenocarcinoma with EGFR mutations. *J Clin Oncol* (2013) **31**:3327–34. doi:10.1200/JCO.2012.44.2806
51. Rickman OB, Vohra PK, Sanyal B, Vrana JA, Aubry MC, Wigle DA, et al. Analysis of ErbB receptors in pulmonary carcinoid tumors. *Clin Cancer Res* (2009) **15**:3315–24. doi:10.1158/1078-0432.CCR-08-2549
52. Vansteenkiste J. Improving patient management in metastatic non-small cell lung cancer. *Lung Cancer* (2007) **57**(Suppl 2):S12–7. doi:10.1016/S0169-5002(07)70422-1
53. Wang S, Wu X, Chen Y, Zhang J, Ding J, Zhou Y, et al. Prognostic and predictive role of JWA and XRCC1 expressions in gastric cancer. *Clin Cancer Res* (2012) **18**:2987–96. doi:10.1158/1078-0432.CCR-11-2863
54. Pierceall WE, Olausson KA, Rousseau V, Brambilla E, Sprott KM, Andre F, et al. Cisplatin benefit is predicted by immunohistochemical analysis of DNA repair proteins in squamous cell carcinoma but not adenocarcinoma: theranostic modeling by NSCLC constituent histological subclasses. *Ann Oncol* (2012) **23**:2245–52. doi:10.1093/annonc/mdr624
55. Roe OD, Andersson E, Sandeck H, Christensen T, Larsson E, Lundgren S. Malignant pleural mesothelioma: genome-wide expression patterns reflecting general resistance mechanisms and a proposal of novel targets. *Lung Cancer* (2010) **67**:57–68. doi:10.1016/j.lungcan.2009.03.016
56. Liu T, Ghosal G, Yuan J, Chen J, Huang J. FAN1 acts with FANCI-FANCD2 to promote DNA interstrand cross-link repair. *Science* (2010) **329**:693–6. doi:10.1126/science.1192656

57. Chirnomas D, Taniguchi T, de laVega M, Vaidya AP, Vasserman M, Hartman AR, et al. Chemosensitization to cisplatin by inhibitors of the Fanconi anemia/BRCA pathway. *Mol Cancer Ther* (2006) 5:952–61. doi:10.1158/1535-7163.MCT-05-0493
58. Lawson MH, Cummings NM, Rassl DM, Russell R, Brenton JD, Rintoul RC, et al. Two novel determinants of etoposide resistance in small cell lung cancer. *Cancer Res* (2011) 71:4877–87. doi:10.1158/0008-5472.CAN-11-0080
59. Schweinfest CW, Henderson KW, Suster S, Kondoh N, Papas TS. Identification of a colon mucosa gene that is down-regulated in colon adenomas and adenocarcinomas. *Proc Natl Acad Sci U S A* (1993) 90:4166–70. doi:10.1073/pnas.90.9.4166
60. Antalis TM, Reeder JA, Gotley DC, Byeon MK, Walsh MD, Henderson KW, et al. Down-regulation of the down-regulated in adenoma (DRA) gene correlates with colon tumor progression. *Clin Cancer Res* (1998) 4:1857–63.
61. Fang X, Netzer M, Baumgartner C, Bai C, Wang X. Genetic network and gene set enrichment analysis to identify biomarkers related to cigarette smoking and lung cancer. *Cancer Treat Rev* (2013) 39:77–88. doi:10.1016/j.ctrv.2012.06.001
62. Hoang MV, Whelan MC, Senger DR. Rho activity critically and selectively regulates endothelial cell organization during angiogenesis. *Proc Natl Acad Sci U S A* (2004) 101:1874–9. doi:10.1073/pnas.0308525100
63. Williams B. Mechanical influences on vascular smooth muscle cell function. *J Hypertens* (1998) 16:1921–9. doi:10.1097/00004872-199816121-00011
64. Kogata N, Tribe RM, Fässler R, Way M, Adams RH. Integrin-linked kinase controls vascular wall formation by negatively regulating Rho/ROCK-mediated vascular smooth muscle cell contraction. *Genes Dev* (2009) 23:2278–83. doi:10.1101/gad.535409
65. Lu Y, Wang L, Liu P, Yang P, You M. Gene-expression signature predicts post-operative recurrence in stage I non-small cell lung cancer patients. *PLoS One* (2012) 7:e30880. doi:10.1371/journal.pone.0030880
66. Røe OD, Szulkin A, Anderssen E, Flatberg A, Sandeck H, Amundsen T, et al. Molecular resistance fingerprint of pemetrexed and platinum in a long-term survivor of mesothelioma. *PLoS One* (2012) 7:e40521. doi:10.1371/journal.pone.0040521
67. Gao Q, Zhou J, Huang X, Chen G, Ye F, Lu Y, et al. Selective targeting of checkpoint kinase 1 in tumor cells with a novel potent oncolytic adenovirus. *Mol Ther* (2006) 13:928–37. doi:10.1016/j.ymthe.2005.12.009
68. Gery S, Komatsu N, Kawamata N, Miller CW, Desmond J, Virk RK, et al. Epigenetic silencing of the candidate tumor suppressor gene Per1 in non-small cell lung cancer. *Clin Cancer Res* (2007) 13:1399–404. doi:10.1158/1078-0432.CCR-06-1730
69. Weis S, Cui J, Barnes L, Cheresch D. Endothelial barrier disruption by VEGF-mediated Src activity potentiates tumor cell extravasation and metastasis. *J Cell Biol* (2004) 167:223–9. doi:10.1083/jcb.200408130
70. Fujita T, Igarashi J, Okawa ER, Gotoh T, Manne J, Kolla V, et al. CHD5, a tumor suppressor gene deleted from 1p36.31 in neuroblastomas. *J Natl Cancer Inst* (2008) 100:940–9. doi:10.1093/jnci/djn176
71. Giacinti C, Giordano A. RB and cell cycle progression. *Oncogene* (2006) 25:5220–7. doi:10.1038/sj.onc.1209615
72. White CD, Brown MD, Sacks DB. IQGAPs in cancer: a family of scaffold proteins underlying tumorigenesis. *FEBS Lett* (2009) 583:1817–24. doi:10.1016/j.febslet.2009.05.007
73. Xie Y, Yan J, Cutz JC, Rybak AP, He L, Wei F, et al. IQGAP2, a candidate tumour suppressor of prostate tumorigenesis. *Biochim Biophys Acta* (2012) 1822:875–84. doi:10.1016/j.bbdis.2012.02.019
74. Akhshi TK, Wernike D, Piekny A. Microtubules and actin work together to polarize cells for division and migration. *Cytoskeleton (Hoboken)* (2013) 71(1):1–23. doi:10.1002/cm.21150
75. Belaid A, Cerezo M, Chargui A, Corcelle-Termeau E, Pedoutour F, Giuliano S, et al. Autophagy plays a critical role in the degradation of active RHOA, the control of cell cytokinesis, and genomic stability. *Cancer Res* (2013) 73:4311–22. doi:10.1158/0008-5472.CAN-12-4142
76. Shyr CR, Kang HY, Tsai MY, Liu NC, Ku PY, Huang KE, et al. Roles of testicular orphan nuclear receptors 2 and 4 in early embryonic development and embryonic stem cells. *Endocrinology* (2009) 150:2454–62. doi:10.1210/en.2008-1165
77. Baldassarre M, Razinia Z, Burande CF, Lamsoul I, Lutz PG, Calderwood DA. Filamins regulate cell spreading and initiation of cell migration. *PLoS One* (2009) 4:e7830. doi:10.1371/journal.pone.0007830
78. Bolstad BM, Irizarry RA, Astrand M, Speed TP. A comparison of normalization methods for high density oligonucleotide array data based on variance and bias. *Bioinformatics* (2003) 19:185–93. doi:10.1093/bioinformatics/19.2.185
79. Kanehisa M, Goto S. KEGG: kyoto encyclopedia of genes and genomes. *Nucleic Acids Res* (2000) 28:27–30. doi:10.1093/nar/28.1.27
80. Edgar R, Domrachev M, Lash AE. Gene expression omnibus: NCBI gene expression and hybridization array data repository. *Nucleic Acids Res* (2002) 30:207–10. doi:10.1093/nar/30.1.207
81. D'Eustachio P. Pathway databases: making chemical and biological sense of the genomic data flood. *Chem Biol* (2013) 20:629–35. doi:10.1016/j.chembiol.2013.03.018
82. Croft D. Building models using reactome pathways as templates. *Methods Mol Biol* (2013) 1021:273–83. doi:10.1007/978-1-62703-450-0_14

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Received: 11 July 2014; accepted: 02 September 2014; published online: 29 September 2014.

Citation: Kerkentzes K, Lagani V, Tsamardinos I, Vyberg M and Røe OD (2014) Hidden treasures in "ancient" microarrays: gene-expression portrays biology and potential resistance pathways of major lung cancer subtypes and normal tissue. *Front. Oncol.* 4:251. doi: 10.3389/fonc.2014.00251

This article was submitted to *Thoracic Oncology*, a section of the journal *Frontiers in Oncology*.

Copyright © 2014 Kerkentzes, Lagani, Tsamardinos, Vyberg and Røe. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.