BMC Cancer

RESEARCH

Open Access

Gene expression in tumor and adjacent normal tissues in lung adenocarcinoma subtypes

Olga Y. Gorlova[1†], Ivan P. Gorlov[1*†], R. Taylor Ripley[1], Chao Cheng[1], Yafang Li[2], Bo Peng[1], Yanhong Liu[1], Hee-Jin Jang[1], Sung Wook Kang[1], Claire Lee[1], Priyanka Ranchod[1], Bryan M. Burt[3], Hyun-Sung Lee[1] and Christopher I. Amos[1,2*]

Abstract

Background Lung adenocarcinoma (LUAD) has several histologically distinct subtypes that differ by a number of clinical features including patient survival. Molecular mechanisms underlying histological and clinical differences between subtypes remain poorly understood.

Methods We conducted a comparative analyses of gene expression in acinar, lepidic, papillary and solid subtypes, as well as mucinous adenocarcinoma. We used a novel, more efficient approach to identify subtype-specific genes. We compared the mean gene expression level separately for tumors and adjacent normal tissue with pure or a highly represented (≥ 75%) subtype of interest to the mean expression in tumors where the subtype of interest was not present. We also performed tumor to adjacent normal tissue comparisons and identified genes differentially expressed between tumor and adjacent normal tissues for each subtype.

Results The number of subtype-specific genes varied from 1 for the acinar to 482 for the papillary subtype. Comparative analysis of gene expression in adjacent normal tissues also identified subtype-specific genes, 38 in total. Gene set enrichment analysis identified oxidative phosphorylation as a biological function associated with papillary, and immune response - with solid subtype. Using data on differential expression between tumor and adjacent normal tissue among the subtype-specific genes and existing evidence for association with lung carcinogenesis, we have identified several candidate subtype-specific driver genes.

Conclusio n We identified subtype-specific genes, biological functions, and potential drivers of subtype-specific carcinogenesis for LUAD subtypes. The study showed importance of gene expression in adjacent normal tissue for subtype-specific tumorigenesis.

Keywords Lung adenocarcinoma (LUAD), Predominant subtypes, Gene expression

†Olga Y. Gorlova and Ivan P. Gorlov contributed equally to this work.

*Correspondence:
Ivan P. Gorlov
ivan.gorlov@bcm.edu
Christopher I. Amos
CIAmos@salud.unm.edu
[1]Baylor College of Medicine, Houston, USA
[2]University of New Mexico, Albuquerque, USA
[3]University of California, Los Angeles, USA

Table 1 Basic demographic characteristics of lung adenocarcinoma patients

Characteristic	Statistics*
Age,mean (SD)	66.7 (9.1)
Sex	
Male	31 (47%)
Female	32 (53%)
Smoking Status	
Non-smokers	11 (16.7%)
Smokers	46 (74.2%)
Unknown	6 (9.1%)
Stage	
I	46 (72.6%)
II	10 (16.7%)
IIIA	7 (10.7%)

Table 2 Histological characteristics of the LUAD patients

LUAD type	LUAD subtype	*N* predominant	*N* pure or nearly pure
Non-mucinous			
	Acinar	39	28
	Lepidic	9	9
	Papillary	3	3
	Micropapillary	1	1
	Solid	5	2
Mucinous			7

Background

Lung adenocarcinoma (LUAD) comprises nearly half of all lung cancer cases. LUAD can be mucinous or non-mucinous, and among the latter, five major histological growth patterns (subtypes) are recognized: acinar, lepidic, papillary, micropapillary, and solid. In up to 90% of lung adenocarcinomas, several histologic patterns are present. Tumors with more than one histological pattern are classified as mixed type, with the pattern of the highest presence denoted as predominant [1]. Lepidic predominant subtype tends to be a low-grade subtype with a good prognosis [2]. Acinar and papillary are intermediate grades while predominant solid or micropapillary are high-grade subtypes with a poor prognosis and high chances to metastasize [3, 4]. The subtypes also differ by response to treatment [5].

Since adenocarcinoma subtypes differ by their morphology, one can expect that they will differ at the molecular level. Such differences indeed have been found [6–11]. However, a direct comparison of gene expression profiles between major subtypes of lung adenocarcinoma has not been conducted.

A better understanding of subtypes at the molecular level will lead to better prognostic differences in clinical outcomes and potentially better predictive response to subtype-specific therapeutic approaches. In this study, we used whole transcriptome RNA sequencing for gene expression profiling of the lung adenocarcinoma subtypes and mucinous lung adenocarcinoma. We have identified subtype-specific genes and biological functions associated with subtypes and demonstrated an importance of gene expression in adjacent normal tissue in the context of subtype tumorigenesis.

Material and methods

Samples

This study was performed in accordance with Institutional Review Board protocol at Baylor College of Medicine (H-35782). Informed consent was obtained for the collection of clinical data and biospecimens. A prospectively maintained, single-institution database was retrospectively queried. Eligible were patients with histologically confirmed lung adenocarcinoma who underwent complete surgical resection from 2017 to 2021. We used samples from tumor and adjacent normal tissues acquired from the patients diagnosed with stage I-IIIA lung adenocarcinoma and surgically treated at Baylor College of Medicine. A total of 63 patients with qualified paired tumor and adjacent non-neoplastic samples were selected for the study. Tables 1 and 2 and Supplementary Table S1 provide patient characteristics.

All tumor and adjacent normal samples were obtained from completely resected tumors form lung cancer patients who underwent lobectomy. For tumor samples, pathologists first identified the tumor regions from the gross fresh specimens. H&E review determined the subtype composition. A tumor region adjacent to the region used for H&E examination was frozen for the subsequent use in RNA sequencing studies. Normal lung tissues were sampled from areas outside the tumor boundaries and most distant from the tumor.

RNA sequencing and data processing

Total RNA was extracted from lung tumors and adjacent normal tissues using the RNeasy Plus Kit (Qiagen, Cat# 74134) following the manufacturer's protocol. The quality of each RNA sample was thoroughly assessed to ensure high purity and integrity. These RNA samples were then used to prepare sequencing libraries, which were sequenced on the NovaSeq 6000 system (Illumina, San Diego) at the Human Genome Sequencing Core, targeting a minimum of 80 million reads per sample. The sequencing data were processed using the STAR aligner [12] for alignment and the RSEM quantification tool [13] for transcript quantification. For constructing the index, we used the reference genome and corresponding annotation files (FASTA and GTF formats) obtained from the GENCODE database (https://www.gencodegenes.org/human/). Gene expression levels were quantified in tra nscripts per million (TPM) and fragments per kilobase

of transcript per million mapped reads (FPKM) units. These results were further integrated with gene annotation information, including gene names, chromosome coordinates, and other relevant genomic features. To ensure accurate and comparable results across samples, we applied quantile normalization using the scikit-learn library.

Using pure subtypes or tumors with >= 75% of a subtype of interest to identify subtype- specific genes

Subtype-specific genes are typically identified through the comparison of the gene expression in a given predominant subtype versus all others, for example genes specific for lepidic subtype are identified by comparison of the gene expression levels in predominant lepidic with gene expression level in other predominant subtypes [14]. An obvious drawback of this approach is that in a tumor of mixed histology the predominant subtype comprises only a fraction of tumor tissue. Using "all other" group as a reference is not perfect because tumors in the reference group often include the subtype of interest, further complicating detection and interpretation of the result of the comparative gene expression analysis. To address these issues, to identify subtype-specific genes for each subtype, we contrasted tumors representing pure or predominant (> =75%) subtype of interest with all other tumors that did not have the subtype of interest. For example, to identify acinar-specific genes we compared gene expression in 100% acinar adenocarcinoma with all other tumors that did not contain acinar subtype. We found that this approach identifies more differentially expressed genes with a higher level of statistical significance compared to the traditional mixed histology approach to identify subtype-specific genes. An exception was made for the micropapillary subtype, for which we used tumors with predominant micropapillary subtype (even if it represented less than 75%), since pure micropapillary subtype was not present in our set. Mucinous LUAD tumors were compared with non-mucinous tumors taken together. Separately, we analyzed gene expression in adjacent normal tissues, contrasting each subtype with others as above.

Differentially expressed genes and gene enrichment analysis

We also identified genes differentially expressed between tumor and adjacent normal tissues for each subtype. For this we used all tumors in which a subtype of interest was present and applied paired t-test weighted on the percentage of the given subtype in the tumor.

We used the Database for Annotation, Visualization and Integrated Discovery (DAVID) for gene enrichment analysis [15].

Results

Nonrandom co-occurrence of histological subtypes in tumors

We used two characteristics to estimate the correlations between subtypes: (i) a simple presence of a given subtype in a tumor: presence was coded as one and absence as zero, and (ii) percentage of a given histological subtype in a tumor. The correlations were similar for the two characteristics we have used. Figure 1 shows a color matrix of pairwise correlations for percentage of a subtype in a tumor. We found that the percentage of the lepidic subtype is negatively correlated with the percentage of the acinar, micropapillary, and solid subtypes. We found a significant positive correlation between the percentages of the acinar and solid subtypes. All other pairwise correlations were not statistically significant.

Subtype-specific genes identified based on the comparative analysis of the gene expression in tumor tissue

Table 3 shows the numbers of subtype-specific genes that remained statistically significant after adjustment

	Acinar	Lepidic	Papillary	Micropapillary	Solid
Acinar	1.00	**0.062**	0.025	0.052	**0.098**
Lepidic	**-0.25**	1.00	0.016	**0.064**	**0.082**
Papillary	0.16	0.13	1.00	0.039	0.002
Micropapillary	0.23	**-0.25**	0.20	1.00	0.001
Solid	**0.31**	**-0.29**	-0.04	0.02	1.00
r≥	-1	-0.8	-0.6	-0.4	-0.2	0	0.2	0.4	0.6	0.8	1

Fig. 1 Pairwise correlations of subtypes across tumors. Statistically significant correlations are shown in black boxes. Color coding is shown on the bottom. The numbers below the diagonal are correlation coefficients and the numbers above the diagonal are R^2

Table 3 Number of subtype-specific genes identified for each subtype based on the analysis of the gene expression in tumor tissue. Number of subtype-specific genes FDR < 0.05

Subtype	Number of subtype-specific genes FDR < 0.05
Acinar	1
Lepidic	10
Micropapillary	18
Mucinous	48
Papillary	482
Solid	59

Table 4 Gene enrichment analysis for subtype-specific genes

Subtype	Category	Term	FDR
Acinar	None detected		
Lepidic	KEGG pathway	mTOR signaling pathway	0.048
Micropapillary	None detected		
Papillary	Cellular component	Mitochondrion	1.2×10^{-13}
Solid	Biological process	Immune response	1.1×10^{-5}
	Biological process	B cell activation	2.2×10^{-5}
	Biological process	Adaptive immune response	6.2×10^{-3}

for multiple testing with false discovery rate (FDR) < 0.05. Subtype-specific genes were identified by the comparisons of the mean expression in tumor tissues of pure/predominant subtype versus other tumors with zero content of the subtype histology. The numbers of subtype-specific genes varied from 1 for the acinar subtype to 482 for papillary subtype. The full lists of differentially expressed genes for the five subtypes of non-mucinous LUAD and for mucinous LUAD can be found in Supplementary Table S2. The complete list of the genes with *p*-values for comparison of the mean expression in predominant subtype compared to all other tumors can be found in Supplementary Material.

Gene set enrichment analysis of the genes identified by the analysis of the gene expression in tumor tissue

We used DAVID to identify biological functions associated with subtypes. To ensure a robust analysis we used the most significant 500 genes for each subtype. Based on our experience this number is optimal number for gene set enrichment analysis. Table 4 shows the results of the analysis. No significant enrichment was detected for the acinar and micropapillary subtypes. For the papillary subtype we found an extremely significant enrichment for the genes associated with mitochondrion. Interestingly, the absolute majority of mitochondrion associated genes are upregulated in the papillary subtype compared to the other subtypes, suggesting that increased energy production is a distinct biological feature of papillary adenocarcinoma. Five hundred genes with most significant difference in mean expression between predominantly solid subtype versus all other tumors showed a strong enrichment for adaptive, B-cell mediated immune response. A borderline significant enrichment for mTOR signaling pathway was detected for the lepidic subtype-associated genes.

Table 5 The number of subtype-specific genes identified for each subtype based on the analysis of the gene expression in adjacent normal tissue

Subtype	Number of subtype-specific genes FDR < 0.05
Acinar	0
Lepidic	20
Micropapillary	2
Mucinous	0
Papillary	7
Solid	9

Subtype-specific genes detected by comparative analysis of the gene expression in adjacent normal tissue

Since gene expression level in normal lung tissue shows strong inter-individual variation [16, 17] we tested the possibility that gene expression in adjacent normal tissue can be subtype specific. Our working hypothesis was that some gene expression profiles in normal tissue, for example due to tobacco exposure, may enhance development of some subtypes and suppress the development of other subtypes. For detection of subtype specific genes based on the analysis of the gene expression in adjacent normal tissue we used the same approach for this analysis as for analysis of the gene expression in tumor tissues. Table 5 shows numbers of subtype-specific genes for adjacent normal tissue that remained statistically significant after adjustment for multiple testing.

The total number of subtype-specific genes identified by the comparative analysis of adjacent normal tissue is much lower compared to that identified by analysis of the gene expression in tumors: 38 versus 618 genes. The largest number of the normal tissue subtype-specific genes was found for the lepidic subtype: 20 genes, or 53% of all detected genes among the lung tumors. No acinar-specific genes or genes specific to mucinous LUAD were detected by the analysis of gene expression in adjacent normal tissue. The full list of subtype-specific genes identified by the analysis of the gene expression in adjacent normal tissues can be found in Supplementary Table S3.

Gene set enrichment analysis of the genes identified by the analysis of the gene expression in adjacent normal tissue

Table 6 shows the results of the gene-set enrichment analysis of the top genes detected by the analysis of adjacent normal tissues. Genes associated with solid subtype show a significant enrichment by the genes involved in inflammatory response (FDR < 0.03). Genes associated with mucinous LUAD are enriched by the genes involved

Table 6 Gene enrichment analysis for the top subtype-specific differentially expressed genes identified by analysis of the gene expression in adjacent normal tissue

Subtype	Category	Term	FDR
Mucinous	Biological process	Positive regulation of tumor necrosis factor production	0.022
	Biological process	Inflammatory response	0.036
Acinar	Biological process	Mitochondrial translation	0.015
Lepidic	Biological process	Positive regulation of angiogenesis	0.00017
	Biological process	Positive regulation of cell migration	0.0054
Micropapillary	None detected		
Papillary	Biological process	Fibroblast growth factor receptor activity	0.0069
	Molecular function	Macrophage colony-stimulating factor signaling pathway	0.05
Solid	Biological process	Inflammatory response	0.0011

in regulation of tumor necrosis (FDR < 0.02). Gene enrichment analysis of the acinar-specific genes found statistically significant enrichment for the genes associated with oxidative phosphorylation – "mitochondrial translation" (FDR < 0.01).

Overlap of subtype-specific genes with genes differentially expressed between normal and tumor tissue in a given subtype

We estimated the overlap between subtype-specific genes both detected by analysis of the gene expression in tumor and adjacent normal tissues and genes differentially expressed (DE) between normal and tumor tissue in a given subtype. A paired t-test weighted on the percentage of a given subtype in the tumor was used for all tumors. The observed number of overlapping genes was compared with the number expected under the assumption that these two sets of genes are independent. The expected numbers were computed as a product of the proportion of subtype-specific genes among all genes analyzed and the proportion of the genes DE between normal and tumor tissue among all genes analyzed, multiplied by the total number of the genes in the analysis – 19,509. Table 7 shows that the observed and expected number of subtype-specific genes that also were differentially expressed between normal and tumor tissues are in good agreement in most cases except for the subtype-specific genes identified using adjacent normal tissue of lepidic and micropapillary tumors, which showed a significant co-occurrence.

The complete list of subtype-specific genes together with the fold change and corresponding p-value of the difference of the mean expression value between normal and tumor tissue is shown in Supplementary Table **S4**. Genes that are (1) subtype-specific and (2) differentially expressed between normal and tumor tissues are the top candidates for subtype-specific drivers of lung carcinogenesis.

Discussion

The highest number of specific genes were identified for the papillary subtype – 482 genes. Fifty-seven, or 12% of the total number of papillary-specific genes are genes related to oxidative phosphorylation in mitochondrion. The enrichment is extremely significant with the *p*-value adjusted for multiple testing $= 1.2 \times 10^{-13}$. Almost all mitochondrion-related genes (52 out of 57) are upregulated in the papillary subtype compared to other subtypes. Genes upregulated in the papillary subtype include those from both mitochondrial (20) and nuclear (37)

Table 7 The observed and expected numbers of subtype specific (SS) genes among the genes differentially expressed (DE) between normal (N) and tumor (T) tissues

Type/Subtype	Tissue	DE* NvsT* for a given subtype				
		DE NvsT	SS* genes	SS among DE NvsT Obs*	SS among DE NvsT Exp*.	*p*-value**
Mucinous LUAD	Normal	1763	0	**0**	**0.0**	1
Mucinous LUAD	Tumor	1763	96	**5**	**8.9**	0.222
Acinar	Normal	5145	0	**0**	**0.0**	1
Acinar	Tumor	5145	1	**0**	**0.3**	1
Lepidic	Normal	3926	55	**40**	**21.4**	**$< 10^{-5}$**
Lepidic	Tumor	3926	13	**7**	**2.7**	0.03
Micropapillary	Normal	662	2	**1**	**0.1**	**0.0035**
Micropapillary	Tumor	662	34	**1**	**1.2**	0.887
Papillary	Normal	3193	16	**3**	**2.7**	0.828
Papillary	Tumor	3193	714	**144**	**120.0**	0.0195
Solid	Normal	3095	11	**0**	**1.8**	0.380
Solid	Tumor	3095	69	**13**	**11.2**	0.568

DE Differentially expressed, *NvsT* Normal versus tumor, *SS* Subtype specific, *Obs*. Observed, *Exp*. Expected

**Computed using χ^2 test or Fisher exact test as appropriate; bolded are *p*-values significant after adjustment for multiple testing

genomes. Therefore, the results of the analysis suggest that upregulation of energy production is the biological mechanism driving papillary adenocarcinoma development. There is abundant evidence that oxidative phosphorylation plays an important role in different types of cancer, including lung cancer [18–20]. The results of this analysis indicate that oxidative phosphorylation can be most important for the papillary subtype.

Three biological functions, namely, "Immune response", "B cell activation", and "Adaptive immune response" were associated with the solid subtype indicating that immune response may be important in solid subtype adenocarcinomas. Solid predominant subtype has been shown to have an increased expression of PD-L1 and a high proportion of dual positive PD-L1– and tumor-infiltrating lymphocytes [21]. We did not identify any biological processes or functions associated with the acinar or micropapillary subtypes. We found some evidence for the role of mTOR signaling pathway in lepidic subtype, which is consistent with the reports that activation of the Akt/mTOR pathway may play an important role in lung tumorigenesis [22]. The most significant findings were for the lepidic subtype, which is enriched by the genes associated with positive regulation of angiogenesis suggesting that anti-angiogenesis drugs like bevacizumab may be the appropriate therapeutic for this subtype.

Lung adenocarcinoma subtypes exhibit profound differences in tumor morphology including immune cell infiltration, tumor microenvironment and angiogenesis [23] making it challenging to identify subtype-specific drivers of tumorigenesis. As a first step in this direction, we identified genes differentially expressed between normal and tumor tissues among subtype-specific genes. Genes differentially expressed between normal and tumor tissues are more likely to be cancer drivers compared to the genes with the same level of expression in tumor and adjacent normal tissue [24]. We identified 110 papillary-specific genes that are differentially expressed between normal and tumor tissue. Twenty genes specific to mucinous LUAD were also differentially expressed between normal and tumor tissue in mucinous LUAD, while for the lepidic subtype there were seven and for the solid subtype, six DE genes. Only one micropapillary-specific gene – *ALX1* was differentially expressed between normal and tumor tissue. To further narrow down the list of potential subtype-specific candidate drivers among the genes DE between normal and tumor tissue we checked if they were reported to be important for lung carcinogenesis. We used the list of known lung cancer genes from the recent review by Liu et al. [25]. The following genes were considered candidates for adenocarcinoma subtype-specific driver genes. Two candidate driver genes were identified for the lepidic subtype: *TMEM63A* and *NFATC4*. For the micropapillary subtype ALX1 was identified as the only candidate driver. Four genes: *SCGB2A2*, *AGR2*, *FUT2*, and *TMPRSS4* were identified as potential drivers for mucinous LUAD. *TTN* was the only candidate gene identified in the solid subtype. The largest number of candidates, 29 genes, were identified for the papillary subtype: *TSTD1*, *LSM12*, *NUDT16L1*, *TMEM54*, *LRG1*, *GSS*, *ARSD*, *TRIM2*, *GOLPH3L*, *TSEN34*, *EPCAM*, *ATIC*, *ARMC7*, *EIF6*, *TOR4A*, *TSPYL5*, *TUBA1C*, *TMEM181*, *SERPIND1*, *PDCD6*, *TIMM17A*, *TACO1*, *TMTC4*, *TAX1BP3*, *TMCO1*, *CXADR*, *POLB*, *TST*, and *CDH1*. No candidate genes were identified for the acinar subtype. Many of the identified candidates have been shown to be important for cancer development including lung adenocarcinoma. For example, the transmembrane proteins (TMEM54) plays role in cell proliferation, migration, invasion, and epithelial-mesenchymal transition [26]. LRG1 influences tumorigenesis by inducing vascularization in solid tumors including lung tumors [27]. TRIM2 is highly expressed in lung adenocarcinomas and influences tumor cell proliferation, migration, and invasion in lung adenocarcinoma [28]. Overexpression of GOLPH3 is associated with poor survival in non-small-cell lung cancer [29]. EPCAM expression in lung tumors is modulated by methylation and is associated with distant metastases [30]. 5-Aminoimidazole-4-carboxamide ribonucleotide formyltransferase/IMP cyclohydrolase (ATIC) facilitates cell growth and migration in lung adenocarcinoma [31]. EIF6 is overexpressed in non–small cell lung cancer and is associated with worse overall survival in lung adenocarcinoma [32]. It has been demonstrated that TSPYL5 is involved in cell growth and the resistance to radiation in lung adenocarcinoma cells through the regulation of PTEN/AKT pathway [33]. TUBA1C is a prognostic biomarker for lung adenocarcinoma modulating immune cell infiltration in the tumor microenvironment [34]. It has been demonstrated that high SERPINH1 expression predicts poor prognosis in lung adenocarcinoma [35]. Genetic polymorphisms in PDCD6 genes are associated with the risk of different cancers including lung adenocarcinoma [36]. Recent study found that TIMM17A overexpression in lung adenocarcinoma and its association with prognosis [37].

An unusual and surprising finding of the study was the observation that gene expression in adjacent normal tissue is also associated with histological subtype of lung adenocarcinoma. We identified 38 subtype-specific genes using comparative analysis of gene expression in adjacent normal tissue only. The majority of them were detected for the lepidic subtype. Interestingly, the number of lepidic-specific genes detected by the analysis of gene expression in normal tissue was higher than that identified in tumor tissue: 20 versus 10 genes. This difference suggests that gene expression in adjacent normal tissue may be may be as important as the gene expression in

tumor tissue for the lepidic subtype. We have identified 9 solid-specific, 7 papillary-specific, and 2 micropapillary-specific genes based on the analysis of the gene expression in adjacent normal lung tissue. Subtype-specific genes identified by the analysis of gene expression in adjacent normal tissue are different from subtype specific genes identified based on the analysis of the gene expression in tumor tissues. Only five subtype-specific genes identified using normal tissue analysis were among genes detected by the analysis of gene expression in tumor tissues.

In observational studies like ours it is impossible to say if the specific tumor histology influences gene expression in adjacent normal tissue, or the adjacent normal tissue influences gene expression in tumor. The results of our analysis suggest the importance of normal tissue when we are looking for the genes associated with a specific adenocarcinoma subtype.

We found that the lepidic subtype-specific genes identified by analysis of adjacent normal tissue significantly overlap with the genes differentially expressed between normal and tumor tissues of the lepidic subtype (Table 7). Interestingly, the majority of identified subtype-specific genes (50 out 55 genes) are upregulated in the lepidic subtype compared to other subtypes and all the 40 differentially expressed genes are upregulated in tumor tissue compared to the adjacent normal. One plausible explanation of the co-occurrence can be that gene expression in the pre-disease lung might be critical for the development of the lepidic subtype. We hypothesize that some patients develop a "pro-lepidic" gene expression profile as a result of tobacco smoking (as most LUAD patients, including those with the lepidic subtype, are current or former smokers), environmental exposures, and/or the genetic make-up, which is in line with the fact that the for the adjacent normal tissue, the highest number of subtype-specific genes was identified for this subtype. This "pro-lepidic" profile initiates cancer development and can be further enhanced by somatic mutations, CNVs or methylation of promotor regions that can significantly modify gene expression [38–40]. The top lepidic subtype-specific genes upregulated in tumor tissue of lepidic LUAD include *MS4A15*, *CAVIN2*, *CDO1*, *TNNC1*, and *CASP12*. All these genes are reported to be cancer-related [41].

We also found a statistically significant co-occurrence between the micropapillary subtype-specific genes and the genes differentially expressed between normal and tumor tissue in the micropapillary subtype; however, as this result is based on a single gene, we are not sure that this finding is real.

The results of the gene enrichment analysis differ for the adjacent normal tissue- and tumor tissue-detected subtype-specific genes. In the analysis of the adjacent normal tissue, the most significant finding was for the lepidic subtype, enriched by "positive regulation of angiogenesis" with the p-value = 0.0002 after adjustment for multiple testing. From the statistical point of view the result is more significant compared to the enrichment analysis based on the subtype-specific genes detected by the analysis of gene expression in tumor tissue, which showed enrichment by "mTOR signaling pathway", p = 0.04. Acinar-specific genes detected by the analysis of gene expression in normal tissue identified "mitochondrial translation" as top biological process, while no biological function was identified for this subtype based on subtype-specific genes identified in tumor. The top biological function for the papillary subtype was "Positive regulation of tumor necrosis factor production". For the solid subtype we identified "Inflammatory response" as the top function. Thus, gene enrichment analysis for subtype-specific genes detected by the analysis of gene expression in adjacent normal tissue identified a different, complementary set of biological functions associated with individual subtypes, further supporting the hypothesis that detection of subtype-specific genes based on the analysis of the gene expression in adjacent normal tissue can be useful for the understanding of molecular mechanisms underlying different subtypes of lung adenocarcinoma.

The analysis of subtype co-occurrence demonstrated that the presence of the lepidic subtype in a tumor decreases chances to detect the micropapillary or solid subtype in the same tumor. The lepidic subtype is characterized by the largest number of subtype-specific genes identified by the analysis of adjacent normal tissue and these genes showed an enrichment for the positive regulation of angiogenesis. It is possible that the lepidic subtype is promoted by normal lung tissue background with increased angiogenesis. We find it challenging to explain a positive correlation between the presence of acinar and solid subtypes in the same tumor. For a solid subtype there is a strong activation of immune response based on results of the gene enrichment analysis. We found that acinar-specific normal tissue genes are enriched by the genes associated with "mitochondrial translation" function. It is possible that those two functions act synergistically helping both acinar and solid subtypes to propagate in the same tumor.

One of the limitations of this study is that we cannot tell if the differences in gene expression between subtypes are due to the differences in gene expression in tumor cells themselves, or they are the result of the differences between subtypes in the cellular composition that includes not only tumor cells but also infiltrating immune cells, fibroblasts, etc. Thus, the differences in the normalized RNA counts integrate (i) the differences in gene expression within tumor cells and (ii) the possible

differences in cellular composition. It is hard to quantify the relative contribution of each source. The effect of cellular composition is probably small because in the gene enrichment analysis of subtype-specific genes we did not see any significant enrichment for the genes specific for immune cell types.

Our study has limitations in that the overall sample size is limited and the number of tumors with a pure or predominant subtype is also low. Therefore, further independent studies are needed to verify the findings.

Conclusion

To conclude, we identified subtype-specific genes and biological functions with some of them being extremely statistically significant. Overlapping subtype specific genes with the genes differentially expressed between tumor and adjacent normal tissue in a given subtype we have identified candidate driver genes for each subtype. The list of subtype-specific driver genes was further refined by evidence that genes are important in lung tumorigenesis. Intriguing finding of the study is a strong evidence that gene expression in adjacent normal tissue is associated and may influence tumor histology.

Abbreviations

LUAD	Lung Adenocarcinoma
DAVID	Database for Annotation, Visualization and Integrated Discover
DE	Differentially Expressed

Supplementary Information

The online version contains supplementary material available at https://doi.org/10.1186/s12885-025-14496-z.

Supplementary Table S1. Characteristics of study participants.

Supplementary Table S2. Subtype specific genes identified by comparative analysis of gene expression in tumor tissue.

Supplementary Table S3. Subtype specific genes identified by comparative analysis of gene expression in adjacent normal tissue.

Supplementary Table S4. Subtype specific genes identified by analysis of tumor tissue "Tumor" and adjacent normal tissue "Normal" with p-values for subtype specific differences in gene expression between tumor (T) and normal (N) tissues.

Supplementary Material. An Excel file with multiple sheets showing p-values for the comparison of mean gene expression in predominant versus all other subtypes. Data for tumor and normal tissues are shown separately. Variation in degree of freedom (df) is due to missing data.

Acknowledgements

None.

Disclosure

None.

Authors' contributions

Olga Y. Gorlova: Conceptualization, Writing – original draft, Writing – review and editing, Data curation, Formal analysis. Ivan P. Gorlov: Conceptualization, Writing – original draft, Writing – review and editing, Investigation, Formal analysis. R. Taylor Ripley: Writing – review and editing, Investigation, Supervision, Resources. Chao Cheng: Writing – review and editing, Resources. Yafang Li: Writing – review and editing, Formal analysis. Bo Peng: Writing – review and editing, Investigation. Yanhong Liu: Writing – review & editing, Data curation. Hee-Jin Jang: Writing – review and editing, Investigation. Sung Wook Kang: Writing – review and editing, Investigation. Claire Lee: Writing – review and editing, Investigation. Priyanka Ranchod: Writing – review and editing, Investigation. Bryan M. Burt: Writing – review and editing, Supervision, Resources. Hyun-Sung Lee: Writing – review and editing, Supervision, Resources. Christopher I. Amos: Writing – review and editing, Supervision, Resources, Conceptualization.

Funding

This study was supported by Cancer Prevention Research Institute of Texas (CPRIT) RR170048 and RP200443 awards, National Institutes of Health (NIH) awards U19CA203654, R01CA285882, R03CA282953, R37 (R01) CA289419-01, U24 2U24OH009077-15-00, R01CA275762, R01CA243483, R21AI159379, R37 CA248478, a US Department of Defense Impact Award W81XWH-22-1-0657, French National Research Agency award ANR-23-IAHU-007, and the Helis Medical Research Foundation. The Human Tissue Acquisition and Pathology core supports tissue collection with funding from P30 Cancer Center Support Grant (NCI-CA125123).

Data availability

Research data supporting this publication are available from the corresponding author upon request. Gene expression data used in this study were submitted to GEO https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE283245.

Declarations

Ethics approval and consent to participate

Ethical approval was granted by the Institutional Ethics Committee of Baylor College of Medicine, Houston, TX and written informed consent was obtained for all study participants (H-35782). The study was conducted in accordance with the principles of the Declaration of Helsinki.

Competing interests

The authors declare no competing interests.

Received: 16 March 2025 / Accepted: 11 June 2025
Published online: 14 July 2025

References

1. Nicholson AG, Tsao MS, Beasley MB, Borczuk AC, Brambilla E, Cooper WA, Dacic S, Jain D, Kerr KM, Lantuejoul S, et al. The 2021 WHO classification of lung tumors: impact of advances since 2015. J Thorac Oncol. 2022;17(3):362–87.
2. Butnor KJ. Controversies and challenges in the histologic subtyping of lung adenocarcinoma. Transl Lung Cancer Res. 2020;9(3):839–46.
3. Borczuk AC. Updates in grading and invasion assessment in lung adenocarcinoma. Mod Pathol. 2022;35(Suppl 1):28–35.
4. Li R, Li Z, Yang Z, Qiu B, Tan F, Xue Q, Gao S, He J. The presence of micropapillary and/or solid subtypes is an independent prognostic factor for patients undergoing curative resection for stage I lung adenocarcinoma with ground-glass opacity. Transl Lung Cancer Res. 2024;13(2):256–68.
5. Clay TD, Do H, Sundararajan V, Moore MM, Conron M, Wright GM, McLachlan SA, Dobrovic A, Russell PA. The clinical relevance of pathologic subtypes in metastatic lung adenocarcinoma. J Thorac Oncol. 2014;9(5):654–63.
6. Bhattacharjee A, Richards WG, Staunton J, Li C, Monti S, Vasa P, Ladd C, Beheshti J, Bueno R, Gillette M, et al. Classification of human lung carcinomas by mRNA expression profiling reveals distinct adenocarcinoma subclasses. Proc Natl Acad Sci U S A. 2001;98(24):13790–5.
7. Cancer Genome Atlas Research N. Author correction: comprehensive molecular profiling of lung adenocarcinoma. Nature. 2018;559(7715):E12.
8. Endoh H, Tomida S, Yatabe Y, Konishi H, Osada H, Tajima K, Kuwano H, Takahashi T, Mitsudomi T. Prognostic model of pulmonary adenocarcinoma by expression profiling of eight genes as determined by quantitative real-time reverse transcriptase polymerase chain reaction. J Clin Oncol. 2004;22(5):811–9.

9. Hayes DN, Monti S, Parmigiani G, Gilks CB, Naoki K, Bhattacharjee A, Socinski MA, Perou C, Meyerson M. Gene expression profiling reveals reproducible human lung adenocarcinoma subtypes in multiple independent patient cohorts. J Clin Oncol. 2006;24(31):5079–90.
10. Parmigiani G, Garrett-Mayer ES, Anbazhagan R, Gabrielson E. A cross-study comparison of gene expression studies for the molecular classification of lung cancer. Clin Cancer Res. 2004;10(9):2922–7.
11. Wilkerson MD, Yin X, Walter V, Zhao N, Cabanski CR, Hayward MC, Miller CR, Socinski MA, Parsons AM, Thorne LB, et al. Differential pathogenesis of lung adenocarcinoma subtypes involving sequence mutations, copy number, chromosomal instability, and methylation. PLoS ONE. 2012;7(5):e36530.
12. Dobin A, Davis CA, Schlesinger F, Drenkow J, Zaleski C, Jha S, Batut P, Chaisson M, Gingeras TR. STAR: ultrafast universal RNA-seq aligner. Bioinformatics. 2013;29(1):15–21.
13. Li B, Dewey CN. RSEM: accurate transcript quantification from RNA-Seq data with or without a reference genome. BMC Bioinformatics. 2011;12:323.
14. Luo J, Ma K, Shi Y, Chen Z, Zhao M, Huang Y, Wang S, Xi J, Zhan C, Xu S, et al. Genetic analyses of differences between solid and nonsolid predominant lung adenocarcinomas. Thorac Cancer. 2018;9(12):1656–63.
15. Sherman BT, Hao M, Qiu J, Jiao X, Baseler MW, Lane HC, Imamichi T, Chang W. DAVID: a web server for functional enrichment analysis and functional annotation of gene lists (2021 update). Nucleic Acids Res. 2022;50(W1):W216–21.
16. Gruber MP, Coldren CD, Woolum MD, Cosgrove GP, Zeng C, Baron AE, Moore MD, Cool CD, Worthen GS, Brown KK, et al. Human lung project: evaluating variance of gene expression in the human lung. Am J Respir Cell Mol Biol. 2006;35(1):65–71.
17. Halloran JW, Zhu D, Qian DC, Byun J, Gorlova OY, Amos CI, Gorlov IP. Prediction of the gene expression in normal lung tissue by the gene expression in blood. BMC Med Genomics. 2015;8:77.
18. Rao S, Mondragon L, Pranjic B, Hanada T, Stoll G, Kocher T, Zhang P, Jais A, Lercher A, Bergthaler A, et al. AIF-regulated oxidative phosphorylation supports lung cancer development. Cell Res. 2019;29(7):579–91.
19. Tufail M, Jiang CH, Li N. Altered metabolism in cancer: insights into energy pathways and therapeutic targets. Mol Cancer. 2024;23(1):203.
20. Ashton TM, McKenna WG, Kunz-Schughart LA, Higgins GS. Oxidative phosphorylation as an emerging target in cancer therapy. Clin Cancer Res. 2018;24(11):2482–90.
21. Dong ZY, Zhang C, Li YF, Su J, Xie Z, Liu SY, Yan LX, Chen ZH, Yang XN, Lin JT, et al. Genetic and immune profiles of solid predominant lung adenocarcinoma reveal potential immunotherapeutic strategies. J Thorac Oncol. 2018;13(1):85–96.
22. Dobashi Y, Watanabe Y, Miwa C, Suzuki S, Koyama S. Mammalian target of rapamycin: a central node of complex signaling cascades. Int J Clin Exp Pathol. 2011;4(5):476–95.
23. Zha L, Matsu-Ura T, Sluka JP, Murakawa T, Tsuta K. Morphological basis of the lung adenocarcinoma subtypes. iScience. 2024;27(5): 109742.
24. Tang H, Wang S, Xiao G, Schiller J, Papadimitrakopoulou V, Minna J, Wistuba II, Xie Y. Comprehensive evaluation of published gene expression prognostic signatures for biomarker-based lung cancer clinical studies. Ann Oncol. 2017;28(4):733–40.
25. Liu Y, Zhao M, Qu H. A database of lung cancer-related genes for the identification of subtype-specific prognostic biomarkers. Biology. 2023;12(3): 357.
26. Herrera-Quiterio GA, Encarnacion-Guevara S. The transmembrane proteins (TMEM) and their role in cell proliferation, migration, invasion, and epithelial-mesenchymal transition in cancer. Front Oncol. 2023;13:1244740.
27. Sage WM, Hurst CR, Silverman JF, Bortz WM. 2nd: intensive care for the elderly: outcome of elective and nonelective admissions. J Am Geriatr Soc. 1987;35(4):312–8.
28. Lin Z, Lin X, Zhu L, Huang J, Huang Y. TRIM2 directly deubiquitinates and stabilizes Snail1 protein, mediating proliferation and metastasis of lung adenocarcinoma. Cancer Cell Int. 2020;20:228.
29. Tang W, Han M, Ruan B, Jin W, Lou J, Yuan X, Chen D, Chen Y, Shin VY, Jin H, et al. Overexpression of GOLPH3 is associated with poor survival in Non-small-cell lung cancer. Am J Transl Res. 2016;8(4):1756–62.
30. Cui Y, Li J, Liu X, Gu L, Lyu M, Zhou J, Zhang X, Liu Y, Zhu H, Zhang T, et al. Dynamic expression of EpCAM in primary and metastatic lung cancer is controlled by both genetic and epigenetic mechanisms. Cancers (Basel). 2022. https://doi.org/10.3390/cancers14174121.
31. Niu N, Zeng J, Ke X, Zheng W, Fu C, Lv S, Fu J, Yu Y. ATIC facilitates cell growth and migration by upregulating Myc expression in lung adenocarcinoma. Oncol Lett. 2022;23(4):131.
32. Gantenbein N, Bernhart E, Anders I, Golob-Schwarzl N, Krassnig S, Wodlej C, Brcic L, Lindenmann J, Fink-Neuboeck N, Gollowitsch F, et al. Influence of eukaryotic translation initiation factor 6 on non-small cell lung cancer development and progression. Eur J Cancer. 2018;101:165–80.
33. Kim EJ, Lee SY, Kim TR, Choi SI, Cho EW, Kim KC, Kim IG. TSPyl5 is involved in cell growth and the resistance to radiation in A549 cells via the regulation of p21(WAF1/Cip1) and PTEN/AKT pathway. Biochem Biophys Res Commun. 2010;392(3):448–53.
34. Bian T, Zheng M, Jiang D, Liu J, Sun H, Li X, Liu L, Zhang J, Liu Y. Prognostic biomarker TUBA1C is correlated to immune cell infiltration in the tumor microenvironment of lung adenocarcinoma. Cancer Cell Int. 2021;21(1):144.
35. Zhang H, Yan X, Gu H, Xue Q, Liu X. High SERPINH1 expression predicts poor prognosis in lung adenocarcinoma. J Thorac Dis. 2022;14(12):4785–802.
36. Hashemi M, Bahari G, Markowski J, Malecki A, Los MJ, Ghavami S. Association of PDCD6 polymorphisms with the risk of cancer: evidence from a meta-analysis. Oncotarget. 2018;9(37):24857–68.
37. Miao L, Wu D, Zhao H, Xie A. TIMM17A overexpression in lung adenocarcinoma and its association with prognosis. Sci Rep. 2024;14(1):8840.
38. Shao X, Lv N, Liao J, Long J, Xue R, Ai N, Xu D, Fan X. Copy number variation is highly correlated with differential gene expression: a pan-cancer study. BMC Med Genet. 2019;20(1):175.
39. Spainhour JC, Lim HS, Yi SV, Qiu P. Correlation patterns between DNA methylation and gene expression in the cancer genome atlas. Cancer Inform. 2019;18: 1176935119828776.
40. Ding J, McConechy MK, Horlings HM, Ha G, Chun Chan F, Funnell T, Mullaly SC, Reimand J, Bashashati A, Bader GD, et al. Systematic analysis of somatic mutations impacting gene expression in 12 tumour types. Nat Commun. 2015;6:8554.
41. Zhu K, Wang B, Li Y, Yu Y, Chen Z, Yue H, Meng Q, Tian D, Liu X, Shen W, et al. CAVIN2/SDPR functioned as a tumor suppressor in lung adenocarcinoma from systematic analysis of Caveolae-Related genes and experimental validation. J Cancer. 2023;14(11):2001–14.

Publisher's Note