

Multiple mutations of lung squamous cell carcinoma shared common mechanisms

Qianping Li^{1,*}, Junyi Hou^{2,*}, Zhaoyan Hu³, Biao Gu⁴, Yan Shi⁵

¹Department of Cardiothoracic Surgery, Shanghai University of Medicine & Health Sciences Shanghai Sixth People's Hospital East Campus, Shanghai, PR China

²Department of Gastroenterology, Shanghai University of Medicine & Health Sciences Shanghai Sixth People's Hospital East Campus, Shanghai, PR China

³Hangzhou Cancer Institute, Hangzhou cancer hospital, Hangzhou City, PR China

⁴Department of Thoracic Surgery, Huai'an First People's Hospital, Nanjing Medical University, Huai'an, Jiangsu, China

⁵Department of Emergency, The Affiliated Huai'an Hospital of Xuzhou Medical University and The Second People's Hospital of Huai'an, Huai'an, China

* Co-first authors

Correspondence to: Biao Gu, **email:** prof_zh@sina.com
Yan Shi, **email:** jfmsy@163.com

Keywords: lung squamous cell carcinoma, mutation, the cancer genome atlas, mRNA

Received: July 04, 2016

Accepted: October 19, 2016

Published: November 07, 2016

ABSTRACT

Lung squamous cell carcinoma (LUSC) is a subtype of non-small cell lung cancers which is the cause of 80% of all lung cancer deaths. The genes that highly mutated in patients with LUSC and their roles played in the tumorigenesis remains unknown. Data of patients with Lung squamous cell carcinoma (LUSC) were retrieved from The Cancer Genome Atlas (TCGA). Differentially expressed genes were identified between control and cancer samples. Patients and controls can be separated by mRNA expression level showing that the between-group variance and totally 1265 genes were differentially expressed between controls and patients. Top genes whose mutations highly occurred in patients with LUSC were identified, most of these genes were shown to be related with tumorigenesis in previous studies. All of the genes mostly mutated were independently correlated with expression levels of all genes. These mutations did not show the trend of co-occurrence. However, the influenced gene of these mutations had overlaps. After studying the intersection of these genes, a group of shared genes were identified. The shared pathways enriched which played critical role in LUSC were identified based on these shared genes. Different mutations had contribution to the progression of LUSC. Though these genes involved different specific mechanisms, most of them may share a common mechanism which is critical for LUSC. The results may suggest a neglected mechanism and also indicate a potential target for therapies.

INTRODUCTION

Lung cancer is the most common cause of deaths that related to cancer in the world. On the meantime, non-small cell lung cancers, causing about 80% of all lung cancer deaths in the United States, is the most frequent form of lung cancer [1]. Lung squamous cell carcinoma (LUSC) is one of the primary subtypes of non-small cell lung cancers.

Even when potentially curative surgery were carried out, about 40% of patients with LUSC will relapse within 5 years [2]. While most cancers have a steady increase in survival in the past decades, lung cancers have slow advance in this aspect, whose currently 5-year relative survival is only 18% and 7%, respectively [3]. These low rates are, to some extent, associated with the fact that more than 50% of patients with LUSC are diagnosed at a distant stage [3].

A lot of studies have been carried out to improve patients' prognosis. Wilkerson, Yin [4] categorized LUSC into four subtypes using mRNA expression, which could be used to hint survival outcomes. Cancer Genome Atlas Research [5] identified potential therapeutic targets like pathways included NFE2L2 and KEAP1, squamous differentiation genes, phosphatidylinositol-3-OH kinase pathway genes and so forth. However, considering the high prevalence and the poor prognosis of LUSC, it is worthwhile to study more about it. This paper aims to figure out the important pathways and mechanisms that lead to the LUSC.

RESULTS

Differentially expressed genes between controls and patients

Differentially expressed genes between controls and patients were identified using R package EBSeq [6]. Totally 1265 genes were differentially expressed.

Principal component analysis (PCA) was carried out on the mRNA expression level among all genes and only the differentially genes. Even with the whole gene group, mRNA expression level can separate patients with control group (Figure 1A), which demonstrated that the biggest variance among all the samples was whether the sample had LUSC. In this case, principal component 1 and 2 accounted for 9.6% and 4.8% of the variance, respectively. With only the differentially expressed genes (Figure 1B), though the variance represented by PC1 and PC2 decreased, the within-group variance also decreased.

The GO terms and pathways related with these genes were summarized in Table 1. These genes shared only a few pathway or COG terms. This suggested that a lot of mechanisms resulted in the difference from patients to control. These mechanisms masked other pathway or terms. When considering the enriched pathway, the olfactory transduction was significantly enriched which can be explained by the involvement of lung in the olfactory transduction. Also, neuroactive ligand-receptor interaction pathway was usually related with cancer progression [7]. Interestingly, systemic lupus erythematosus was enriched, which seemed unrelated to LUSC. However, in most patients with systemic lupus erythematosus, lung involvement is a known complication. Also lupus patients experienced an elevated risk of different cancers including lung cancer.

Mutations occurred in patients with LUSC

Top genes whose mutations highly occurred in patients with LUSC were identified with the somatic mutations datasets from TCGA. The overall results were shown in Table 2. Most of these genes were shown to be related with tumorigenesis in previous studies

mathematically or biologically. Here, abParts stood for parts of antibodies which are mostly variable regions. Due to the fact it was not a gene, it was excluded in the following analysis. On the meantime, it is quite common for a variable regions to mutate. *PCDHGC5*, *PCDHAC2*, *SPTA1*, *XIRP2* and *FLG* seemed unrelated with cancer based the reference study.

The other genes were all shown to have connection with cancer progression. However, some of them were only related to cancer without evidences whether they would lead to cancer progression, or they were mutated led by cancer or other genes. These genes were marked as red in the Table 2. Our attention was mainly put on the genes that had proved to have contribution to tumorigenesis. These genes are possibly important driver genes leading to LUSC, instead of just passenger genes.

Another interesting point is that average of mutations of a certain gene in was different. For example, TTN was mutated in 129 patients and every patient had 2.61 TTN mutations in average. On the other hand, every patient had 1.04 mutations of TP53 in average. The ratio of non-synonymous mutations and patients might suggest that whether one mutation in this gene is critical to tumorigenesis. This result also showed that any mutation on TP53, together mutations on other genes, might lead to cancer. In this study, MLL2 also had a relative low ratio of non-synonymous mutations and patients.

Mutation patterns and global gene expression

Strikingly, all of the genes mostly mutated were independently correlated with expression levels of all genes (Figure 2) and 1261 differentially expressed genes (data not shown). However, Gerstung, Pellagatti [8] found that the principal component values of target genes varied widely across the different mutations or indels in patients with myelodysplastic syndromes. It suggested the different mutation patterns between LUSC and chronic blood cancer.

Figure 3 showed the pairwise heatmap between top genes. The upper triangle showed that whether mutations of two different genes co-occurred and the lower triangle showed that whether the expression levels of two different genes were correlated. TTN and SYNE1, CSMD3 and SPTA1, MUC16 and ZFH4, ZFH4 and SI, and FAM135B and COL11A1 had the highest possibilities that their mutations occur simultaneously. But mainly, these mutations did not occur on the same time. Combining the expression levels of these genes, it suggested that there was no relation between mutation and expression. The mutation on one gene may not influence the expression of it.

Effects of mutations on expression

After figuring out all genes that were differentially expressed between groups where a certain gene was

Table 1: The pathway and COG enrichment of genes differentially expressed genes between patients and controls

Category	Term	P-Value	Benjamini
Pathway	Neuroactive ligand-receptor interaction	1.40E-14	1.70E-12
	Olfactory transduction	1.80E-10	1.10E-08
	Systemic lupus erythematosus	4.60E-04	1.90E-02
	Secondary metabolites biosynthesis, transport, and catabolism	1.50E-03	1.90E-02
COG	Signal transduction mechanisms / Cytoskeleton / Cell division and chromosome partitioning / General function prediction only	2.70E-02	1.70E-01
	Amino acid transport and metabolism	4.20E-02	1.70E-01
	Lipid metabolism	9.30E-02	2.70E-01

mutated or not, the summary of the logarithm fold changes was plotted as a barplot (Figure 4). There seemed no significant difference between different genes.

After studying the pathway involved, it should that different genes had some common pathway, including neuroactive ligand-receptor interaction, retinol metabolism, drug metabolism, steroid hormone biosynthesis and so forth. The fully list is not shown here. It is quite interesting because there seemed no common patterns among these genes. They were independently correlated with expression levels of all genes and their mutation did not show co-occurrence.

Functional analysis of mutation-related expression change

After studying the intersection of these differentially genes of different genes, a group of shared gene were identified. Only the genes that were shared by at least four genes were extracted here. Pathway enrichment was carried out on these group of genes (Table 3).

The shared pathway enrichment results of each genes were close to this table as expected. This table showed a shared mechanisms and pathways in patients with different gene mutations and suggested the important roles that these pathways played in LUSC.

DISCUSSION

This study identified genes which were highly mutated in patients with LUSC. Among all of the genes that had the most mutations, some of them lack evidence to link them with the cancer. One reason may be that there are high possibility that mutations occur on these genes, even though they themselves did not have impact on the progression of LUSC. For example, abParts was identified, which is the most variable regions in antibodies. It, certainly, had higher mutation rate. On the other hand, the length of the gene has an impact. The longer the gene is, the higher possibility that a mutation occurs is. Proteins of TTN and MUC16 are extremely long, respectively 34350 and 22152. This may induce bias in the identification of

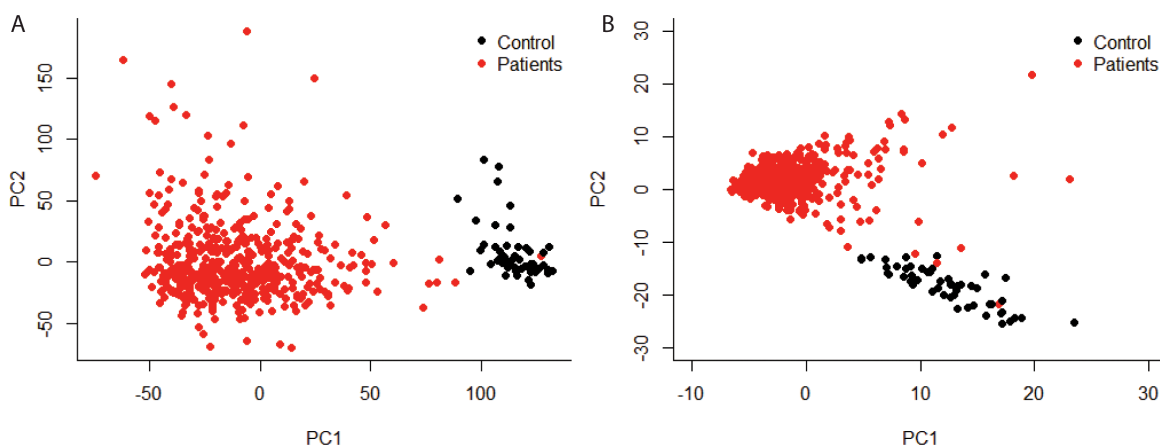


Figure 1: Principal component analysis of gene expression level. Only principal component 1 and 2 were shown. (A) all the genes were used; (B) only genes that were differentially expressed between patients and control were used.

Table 2: The summary of top genes that are frequently mutated in patients with LUSC

Symbol	Non-synonymous Mutations	Patients	Experimental Evidences
TTN	337	129	Mathematically based prediction without biological evidence [9, 10] [%]
abParts ^s	221	110	
TP53	147	141	Tumor protein
MUC16	142	77	Tumor antigen
CSMD3	124	81	Contributing to lung tumorigenesis [13]
RYR2	120	76	Related with apoptosis and carcinogenesis [14, 15]
LRP1B	103	69	Related with lung cancer [%]
PCDHGC5	103	69	
PCDHAC2	100	53	
USH2A	93	67	Mutations are frequent in cancer [9] [%]
ZFHX4	90	65	Related with glioblastoma [16]
SYNE1	70	52	Methylated in lung cancer [17] [%]
RYR3	54	41	Regulator of breast cancer cell [18]
MLL2	52	43	Associated with gain-of-function p53 mutations [19]
SPTA1	51	40	
FAM135B	49	37	Promote malignancy of oesophageal squamous cell carcinoma [20]
XIRP2	48	33	
COL11A1	46	35	Promote tumor progression in ovarian cancer [21]
SI	46	36	Related with cancer development [%]
FLG	45	36	

^sabParsts meant parts of antibodies which are mostly variable regions.

[%]Reports showed that mutations and/or mRNA expression of these genes were related to tumors but no experiments showed that knockout or mutations would lead to cancer progression.

Table 3: Pathway enrichment for shared differentially expressed genes

Term	P-Value	Benjamini
Neuroactive ligand-receptor interaction	3.20E-14	4.10E-12
Drug metabolism	6.80E-09	4.30E-07
Retinol metabolism	1.40E-07	4.40E-06
Metabolism of xenobiotics by cytochrome P450	5.20E-07	1.30E-05
Steroid hormone biosynthesis	1.10E-05	2.30E-04
Androgen and estrogen metabolism	8.90E-05	1.60E-03
Tyrosine metabolism	3.20E-04	5.00E-03
Ascorbate and aldarate metabolism	4.30E-04	6.00E-03
Pentose and glucuronate interconversions	5.80E-04	7.30E-03
Starch and sucrose metabolism	1.30E-03	1.50E-02
Maturity onset diabetes of the young	2.80E-03	2.90E-02

mutations. However, basically, the identified genes that are highly mutated were reported to be related to cancer or lung cancer in the previous studies and also some of

them may play roles as driver genes that result in LUSC. Here TTN and SYNE1 were used as an example. Kim, Hong [9] found that TTN had dominant frequencies in

eleven different tumors and Greenman, Stephens [10] stated that its functions were compatible with a role in oncogenesis. SYNE1 expressed in skeletal and smooth muscle and localizes to the nuclear membrane, but a lot of studies reported that missense mutations, silent mutations, nonsense mutations, and frameshift deletions on SYNE1 were observed in colon cancer, stomach cancer, breast cancer and so forth [11, 12].

These genes showed independently pattern with the expression level of all genes. Also, there seemed no co-occurrence between these genes. However, the differentially expressed genes between patients with or without a certain gene have an intersection. 741 out of 3022 genes, which were differentially expressed in any gene, appeared in at least four groups. It suggested that the mechanisms that the mutation of these genes led to LUSC may have an overlap. The pathways enriched using the shared genes may be extremely important.

Among these pathways, only neuroactive ligand-receptor interaction was enriched using differentially expressed genes between patients and controls. Interestingly, most of these pathways are related to metabolism, including metabolism of retinol, xenobiotics, androgen and estrogen, tyrosine, ascorbate and aldarate, and starch and sucrose. It suggests that any mutation on the driver genes may lead to the different patterns of metabolism, while LUSC itself has a smaller impact on the metabolism.

MATERIALS AND METHODS

TCGA lung squamous cell carcinoma dataset

Clinical information, level-3 data of microarray and mutation information from patients with lung squamous cell carcinoma were retrieved from The Cancer Genome

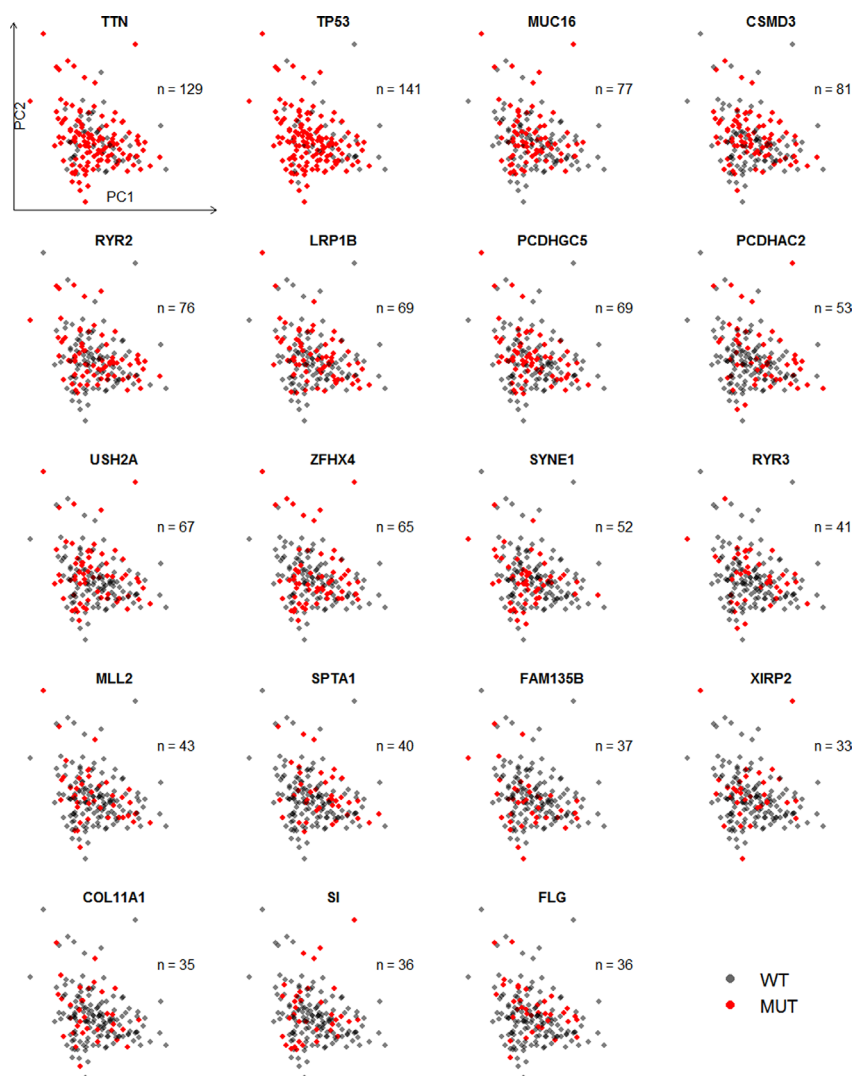


Figure 2: Scatter plot of the first two principal components of expression levels of all available genes with mutation status of 19 genes. This principal component analysis involved all 501 patients with mRNA expression level and 51 controls, though controls and patients without mutations' information were excluded in the plot.

Atlas (TCGA). This data set contains 504 patients, within which 501 and 497 patients, respectively, had mRNA and mutations information. On the meantime, mRNA expression levels of another 51 samples from normal solid tissue were used as control.

Differentially expressed genes

Differentially expressed genes were found using R packages EBSeq [6] between control and cancer samples. Biological functions were summarized using Clusters

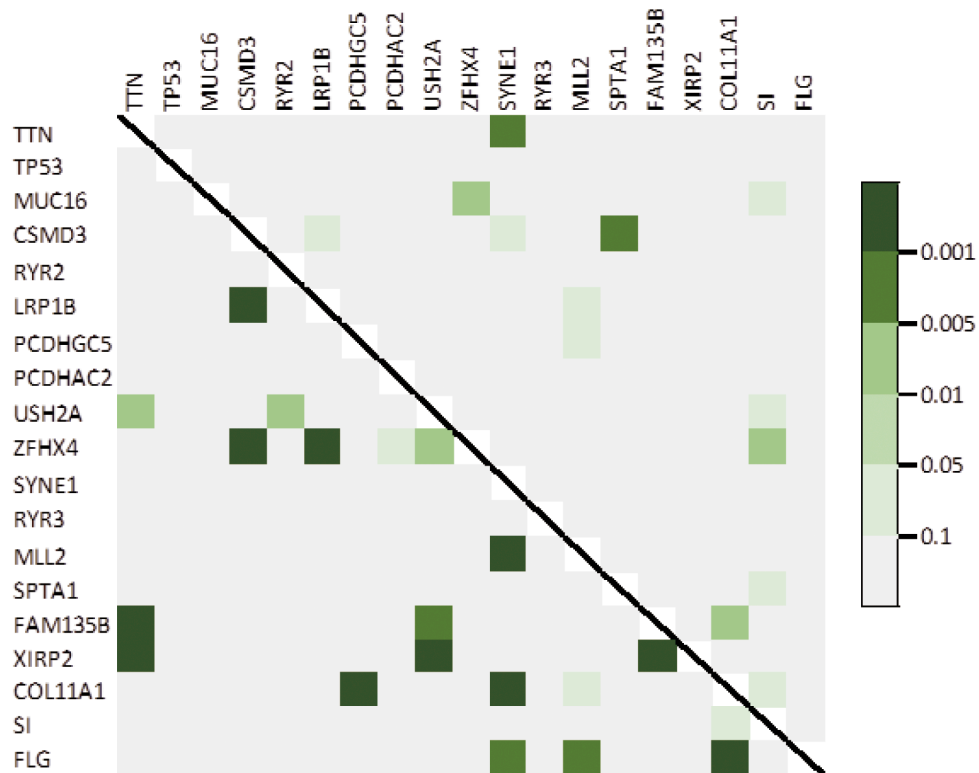


Figure 3: Heatmap of observed pairwise mutation patterns (upper triangle) and the pairwise correlation of expression level (lower triangle).

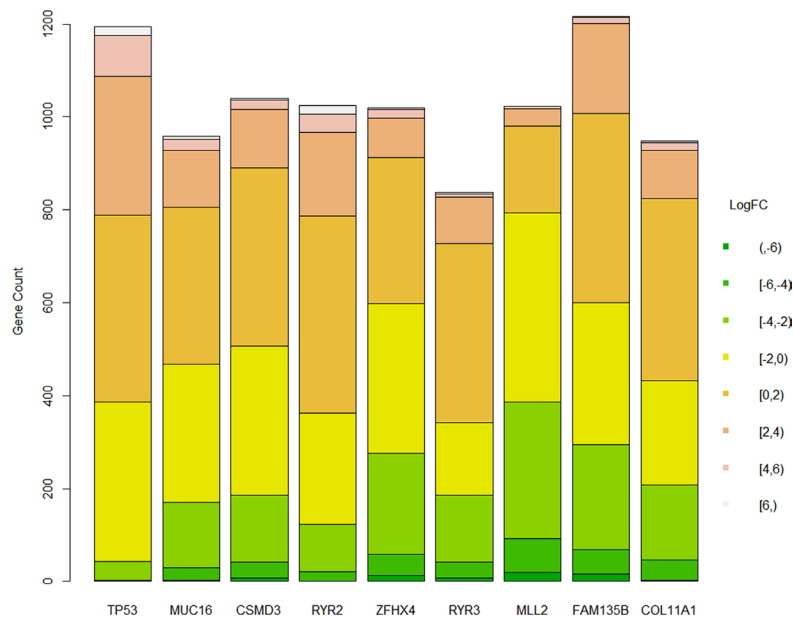


Figure 4: Difference of logarithm fold changes on genes that were differentially expressed whether a gene was mutated.

of Orthologous Groups (COG) terms [22] and pathway enrichment using DAVID Bioinformatics Resources 6.7 [23]. Principal component analysis (PCA) was carried out on the mRNA expression level among all genes, as well as among only the differentially genes.

Mutations identification

Top 20 genes with most mutations in patients were identified using mutation information. References searching was carried out to figure out whether these genes were related to lung squamous cell carcinoma in the previous reports. The experimentally proved genes where had impact on the cancer progression were focused on the following analysis.

Mutation patterns with mRNA expression

Global analysis Scatter plot of the first two principal components was plotted [8] was carried out to find out whether mutations of specific genes occurred among some group of patients, whether certain mutations occurred simultaneously in patients, and whether the expression levels of these most mutated genes were correlated with other.

Effects of mutations on expression

Patients were separated into two groups that whether the patient had mutations of a certain gene [8] using EBseq package. In this study, except for abParts, which stood for the parts of antibodies which are mostly variable regions, the differentially expressed genes between two groups separated by another 19 genes were found. COG terms and pathway enrichment were carried out based on the genes identified by each mutation.

Shared differentially expressed genes

Genes that differentially expressed in groups whether a certain gene was mutated were shared by different groups. Using genes that shared by at least four group were extracted and GO term and pathway enrichment were carried out as well.

CONFLICTS OF INTEREST

None declared.

GRANT SUPPORT

This work was supported by the SUMHS Collaborative Innovation Focus (NO.HMCI-16-11-004).

REFERENCES

1. Raponi M, Dossey L, Jatko T, Wu X, Chen G, Fan H, Beer DG. MicroRNA classifiers for predicting prognosis of squamous cell lung cancer. *Cancer Res.* 2009; 69:5776–83.
2. Hoffman PC, Mauer AM, Vokes EE. Lung cancer. *The Lancet.* 2000; 355:479–485.
3. Siegel RL, Miller KD, Jemal A. Cancer statistics. 2015. *CA Cancer J Clin.* 2015; 65:5–29.
4. Wilkerson MD, Yin X, Hoadley KA, Liu Y, Hayward MC, Cabanski CR, Muldrew K, Miller CR, Randell SH, Socinski MA, Parsons AM, Funkhouser WK, Lee CB, et al. Lung squamous cell carcinoma mRNA expression subtypes are reproducible, clinically important, and correspond to normal cell types. *Clin Cancer Res.* 2010; 16:4864–75.
5. Cancer Genome Atlas Research, N. Comprehensive genomic characterization of squamous cell lung cancers. *Nature.* 2012; 489:519–25.
6. Leng N, Dawson JA, Thomson JA, Ruotti V, Rissman AI, Smits BM, Haag JD, Gould MN, Stewart RM, Kendziorski C. EBSeq: an empirical Bayes hierarchical model for inference in RNA-seq experiments. *Bioinformatics.* 2013; 29:1035–43.
7. Huan J, Wang L, Xing L, Qin X, Feng L, Pan X, Zhu L. Insights into significant pathways and gene interaction networks underlying breast cancer cell line MCF-7 treated with 17beta-estradiol (E2). *Gene.* 2014; 533:346–55.
8. Gerstung M, Pellagatti A, Malcovati L, Giagounidis A, Porta MG, Jädersten M, Dolatshad H, Verma A, Cross NC, Vyas P, Killick S, Hellström-Lindberg E, Cazzola M, et al. Combining gene mutation with gene expression data improves outcome prediction in myelodysplastic syndromes. *Nat Commun.* 2015; 6:5901.
9. Kim N, Hong Y, Kwon D, Yoon S. Somatic mutome profile in human cancer tissues. *Genomics Inform.* 2013; 11:239–44.
10. Greenman C, Stephens P, Smith R, Dalgliesh GL, Hunter C, Bignell G, Davies H, Teague J, Butler A, Stevens C, Edkins S, O'Meara S, Vastrik I, et al. Patterns of somatic mutation in human cancer genomes. *Nature.* 2007; 446:153–8.
11. Sjoblom T, Jones S, Wood LD, Parsons DW, Lin J, Barber TD, Mandelker D, Leary RJ, Ptak J, Silliman N, Szabo S, Buckhaults P, Farrell C, et al. The consensus coding sequences of human breast and colorectal cancers. *Science.* 2006; 314:268–74.
12. Cancer Genome Atlas Network. Comprehensive molecular characterization of human colon and rectal cancer. *Nature.* 2012; 487:330–7.
13. Liu P, Morrison C, Wang L, Xiong D, Vedell P, Cui P, Hua X, Ding F, Lu Y, James M, Ebben JD, Xu H, Adjei AA, et al. Identification of somatic mutations in non-small cell lung carcinomas using whole-exome sequencing. *Carcinogenesis.* 2012. 33:1270–6.

14. Mariot P, Prevarskaya N, Roudbaraki MM, Le Bourhis X, Van Coppenolle F, Vanoverbergh K, Skryma R. Evidence of functional ryanodine receptor involved in apoptosis of prostate cancer (LNCaP) cells. *The Prostate*. 2000; 43:205–214.
15. Faryna M, Konermann C, Aulmann S, Bermejo JL, Brugger M, Diederichs S, Rom J, Weichenhan D, Claus R, Rehli M, Schirmacher P, Sinn HP, Plass C, et al. Genome-wide methylation screen in low-grade breast cancer identifies novel epigenetically altered genes as potential biomarkers for tumor diagnosis. *FASEB J*. 2012; 26:4937–50.
16. Chudnovsky Y, Kim D, Zheng S, Whyte WA, Bansal M, Bray MA, Gopal S, Theisen MA, Bilodeau S, Thiru P, Muffat J, Yilmaz OH, Mitalipova M, et al. ZFH4 interacts with the NuRD core member CHD4 and regulates the glioblastoma tumor-initiating cell state. *Cell Rep*. 2014; 6:13–24.
17. Tessema M, Belinsky SA. Mining the epigenome for methylated genes in lung cancer. *Proc Am Thorac Soc*. 2008. 5:806–10.
18. Zhang L, Liu Y, Song F, Zheng H, Hu L, Lu H, Liu P, Hao X, Zhang W, Chen K. Functional SNP in the microRNA-367 binding site in the 3'UTR of the calcium channel ryanodine receptor gene 3 (RYR3) affects breast cancer risk and calcification. *Proc Natl Acad Sci USA*. 2011; 108:13653–8.
19. Zhu J, Liu Y, Song F, Zheng H, Hu L, Lu H, Liu P, Hao X, Zhang W, Chen K. Gain-of-function p53 mutants co-opt chromatin pathways to drive cancer growth. *Nature*. 2015; 525:206–11.
20. Song Y, Li L, Ou Y, Gao Z, Li E, Li X, Zhang W, Wang J, Xu L, Zhou Y, Ma X, Liu L, Zhao Z et al. Identification of genomic alterations in oesophageal squamous cell cancer. *Nature*. 2014; 509:91–5.
21. Wu YH, Chang TH, Huang YF, Huang HD, Chou CY. COL11A1 promotes tumor progression and predicts poor clinical outcome in ovarian cancer. *Oncogene*. 2014; 33:3432–40.
22. Galperin MY, Makarova KS, Wolf YI, Koonin EV. Expanded microbial genome coverage and improved protein family annotation in the COG database. *Nucleic Acids Res*. 2015; 43:D261–9.
23. Huang da W, Sherman BT, Lempicki RA. Bioinformatics enrichment tools: paths toward the comprehensive functional analysis of large gene lists. *Nucleic Acids Res*. 2009; 37:1–13.