Pathol. Oncol. Res. (2014) 20:197–202
DOI 10.1007/s12253-013-9685-2

RESEARCH

Pathways Enrichment Analysis for Differentially Expressed Genes in Squamous Lung Cancer

Liqiang Qian · Qingquan Luo · Xiaojing Zhao · Jia Huang

Received: 6 June 2013 / Accepted: 30 July 2013 / Published online: 10 October 2013

Abstract Squamous lung cancer (SQLC) is a common type of lung cancer, but its oncogenesis mechanism is not so clear. The aim of this study was to screen the potential pathways changed in SQLC and elucidate the mechanism of it. Published microarray data of GSE3268 series was downloaded from Gene Expression Omnibus (GEO). Significance analysis of microarrays was performed using software R, and differentially expressed genes (DEGs) were harvested. The functions and pathways of DEGs were mapped in Gene Otology and KEGG pathway database, respectively. A total of 2961 genes were filtered as DEGs between normal and SQLC cells. Cell cycle and metabolism were the mainly changed functions of SQLC cells. Meanwhile genes such as MCM, RFC, FEN1, and POLD may induce SQLC through DNA replication pathway, and genes such as PTTG1, CCNB1, CDC6, and PCNA may be involved in SQLC through cell cycle pathway. It is demonstrated that pathway analysis is useful in the identification of target genes in SQLC.

Keywords Squamous lung cancer · Oncogenesis mechanism · Potential pathways · Target genes

Introduction

Lung cancer is the most common primary malignant tumors of the lung, the vast majority of which are originated in the bronchial epithelium, and squamous lung cancer (SQLC) is a common type of lung cancer [1]. Lung cancers including squamous cell carcinomas are a leading cause of cancer related death and carry a poor prognosis with the majority of patients presenting at a stage when curative treatments are no longer feasible [2]. In the recent 50 years, the incidence and mortality lung cancer are rising rapidly in all over the world, especially in developed countries; and men died of lung cancer ranks first in all died in cancer. However, there are so few drugs that are both active and tolerable in SQLC patients [3], and target agents have not yet shown to be successful in SQLC because of the lack of the knowledge of genomic alternate that drive SQLC [4]. Therefore, the research and treatment of SQLC is of great importance for human health.

In recent years, some researches of lung cancer using bioinformatics are undertaken. For example, a genome-wide study is conducted to search for methylation genes in SQLC patients, whose findings emphasizes the impact of methylation on the pathogenesis of SQLC [4]. Differentially expressed genes (DEGs) are identified between lung adenocarcinoma tissue and adjacent nonmalignant lung tissue, of which ERGIC3 may be an active gene in the development and progression of lung cancer [5]. Genetic network and gene set enrichment of lung cancer are conducted, and top 3 pathways of cell cycle, DNA replication, and RNA transport involved in lung cancer are identified [6]. However, few studies had reported the related genes and pathways of SQLC.

In this paper, biochip array technology analysis was used to compare the expression difference between SQLC and normal cells. Significant DEGs were tested by Limma package in R language. Function and pathway enrichment analysis was conducted to detect metabolic pathways changed inner and outside of cells, helping us to elucidate the mechanism of SQLC.

Liqiang Qian and Xiaojing Zhao should be regarded as co-first authors.

L. Qian · Q. Luo (✉) · X. Zhao · J. Huang
Shanghai Lung Cancer Center, Shanghai Chest Hospital,
Shanghai JiaoTong University,
NO.241 Huaihai Road, Shanghai 200030, China
e-mail: luoqingquanlqq@hotmail.com

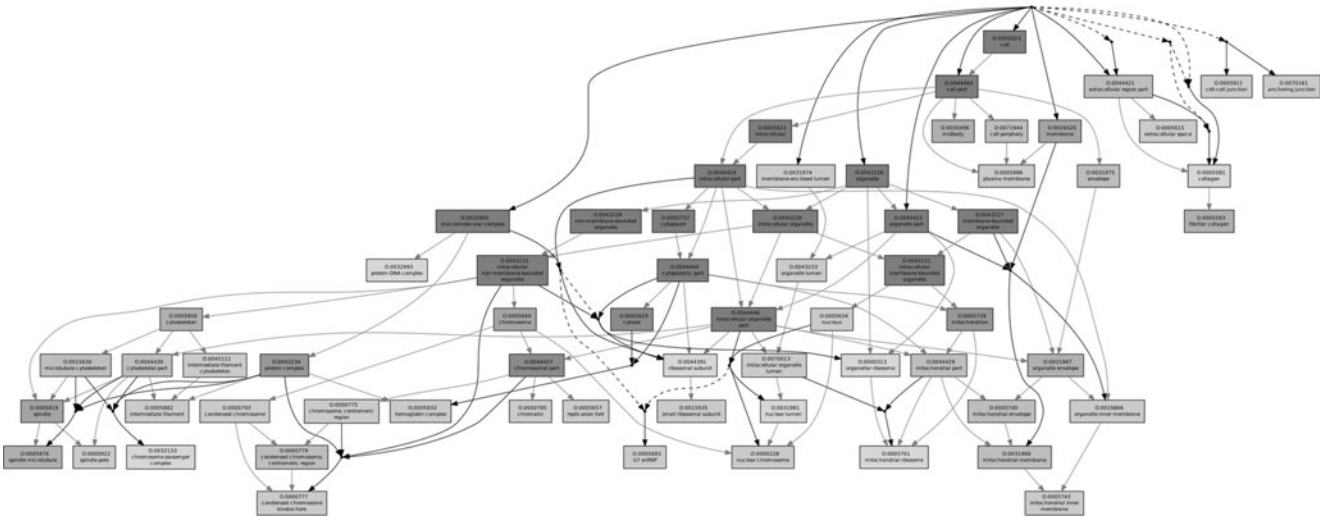

Fig. 1 The enriched Gene Ontology (GO) terms of cellular component of the differentially expressed genes (DEGs). The *colored entries* are the significantly enriched (FDR<0.05) ones, and the deeper the color is the more significance is

Methods and Materials

Expression Profile Microarrays

Gene expression profiles data GSE3268 [7] of SQLC cells were downloaded from the National Center For Biotechnology Information Gene Expression Omnibus (GEO) data repository (http://www.ncbi.nlm.nih.gov/geo/). Only five pairs of samples are available, and each pair of samples represents a single patient with squamous lung cancer, of which one is derived from the cancer cells, and the other is from the normal cells (platform: GPL96 [HG-U133A] Affymetrix Human Genome U133A Array).

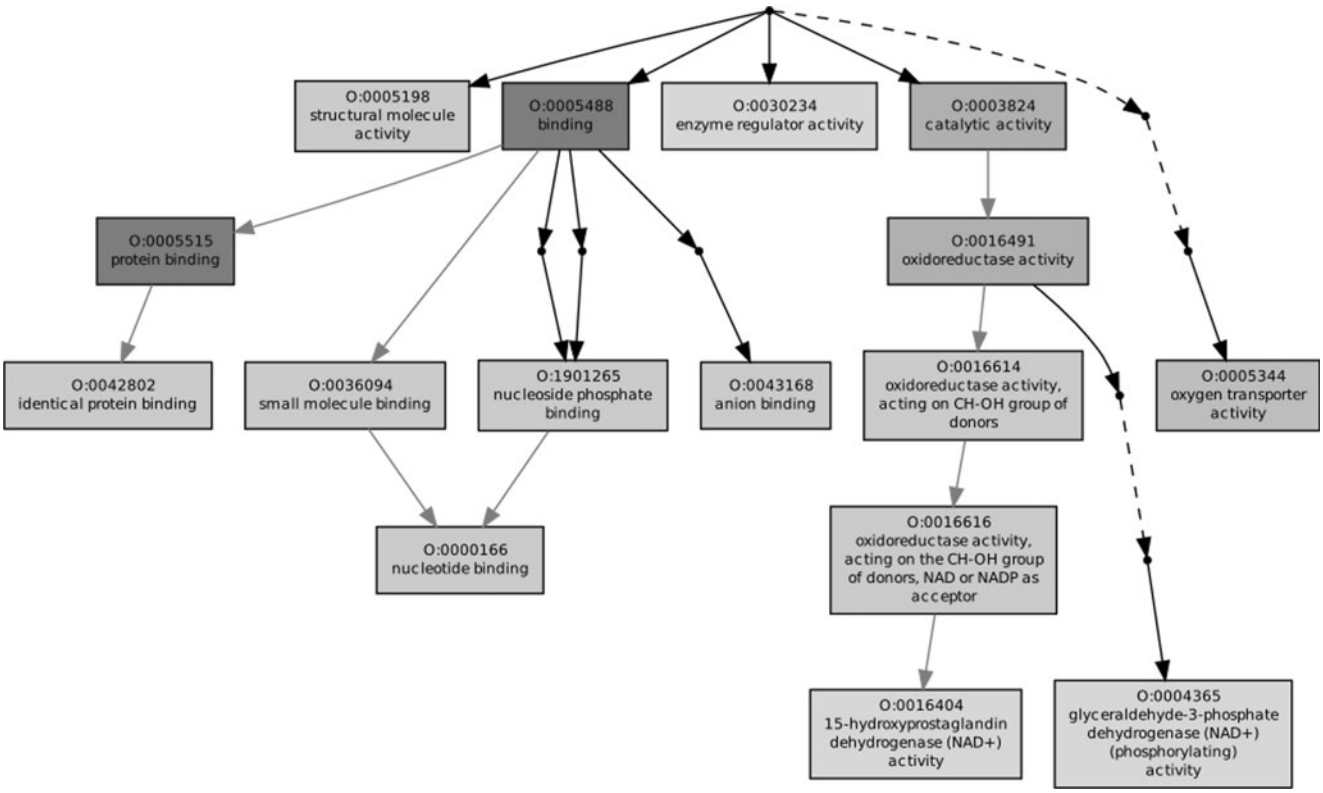

Fig. 2 The enriched Gene Ontology (GO) terms of molecular function of the differentially expressed genes (DEGs). The *colored entries* are the significantly enriched (FDR<0.05) ones, and the deeper the color is the more significance is

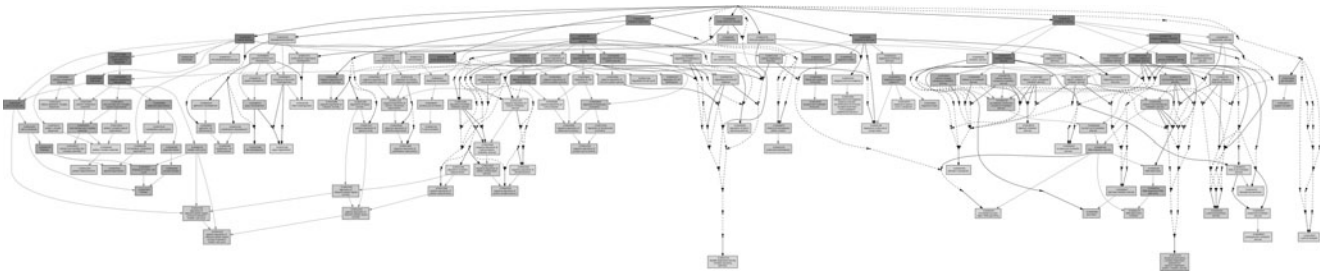

Fig. 3 The enriched Gene Ontology (GO) terms of biological progress of the differentially expressed genes (DEGs). The *colored entries* are the significantly enriched ones (FDR<0.05) ones, and the *deeper the color* is the more significance is

DEGs Analysis

The original data of the profiles were firstly analyzed by R language software (v.2.13.0) [8], and then by Geoquery [9] and Limma package. Geoquery can quickly access the expression profiling data on the GEO, while Limma is the most popular method in analyzing DEGs [10, 11]. Preprocessed data were extracted and log2 transformed by Geoquery, and then was divided into two groups: normal and cancer group. Next, the significance of DEGs was tested by linear regression model software package Limma, and corrected with Benjamin and Hochberg (BH) test [12]. Genes with FDR (false discovery rate) <0.05 were selected as DEGs.

Function Annotation

Gene Otology (GO) analysis is a commonly used approach in functional studies [13]. GOEAST (Gene Ontology Enrichment Analysis Software Tollkit) is an easy-to-use web-based toolkit that identifies statistically overrepresented GO terms within given gene sets [14]. In order to trace cell changes and functions of the DEGs, GOEAST was used to identify over-represented GO categories in cellular component, molecular function and biological processes based on the hypergeometric distribution, with FDR<0.05.

Pathway Annotation

Kyoto Encyclopedia of Gens And Genomes (KEGG) pathway is used as a reference knowledge base for understanding signal transduction, cellular process and biological pathways [15]. All metabolic and non-metabolic pathways were downloaded from the open KEGG pathway database, and cluster analysis of DEGs was conducted by Gene Set Analysis Toolkit V2 platform based on hypergeometric distribution. The count number larger than 2 and FDR<0.05 were chosen as cut-off criterion.

Results

Differentially Expressed Genes

T-test was used to identify genes differentially expressed between normal cells and cancer cells, with BH test corrections. At a FDR value of 0.05, a total of 3730 probes were identified to be differentially expressed in lung cancer samples compared with normal samples, which corresponded to 2961 DEGs.

Gene Otology Analysis

To determine the function of DEGs, the DEGs were mapped to the GO database by GOEAST. Cellular components in which the DEGs were most located, such as microtubule, centromere, chromosome, cell-cell junction, and mitochondrion are shown in Fig. 1 (Fig. 1). Figure 2 displays the molecular function of DEGs, for instance, binding of proteins, oxygen transmission and NAD + metabolism on the respiratory chain (Fig. 2). Meanwhile, Fig. 3 shows the biological

Table 1 The biological pathways enriched in KEGG of lung cancer cells

KEGG pathway	Genes	*p* value	FDR value
DNA replication	RFC2/3/4/5, MCM2-7, FEN1,POLD1/2, POLE2	2.84E-08	5.17E-06
Cell cycle	PTTG1, CCNB1, CDC6, CHEK1/2, PCNA	3.97E-07	3.61E-05
Mismatch repair	RFC2/3/4/5, POLD1/2, MSH2/6	9.20E-5	0.056
Proteasome	PSMB1/4, PSMA1/2, SHFM1	0.0002	0.0091
Systemic lupus erythematosus	H2AFZ, HLA-DRB1, HIST1H2BD	0.0003	0.109
Metabolic pathways	ACACB, ACSL1, ADH1B, AOX1	0.0008	0.0243
Glutathione metabolism	SRM, PRM1, GSTM3	0.0010	0.0260

FDR false discovery data

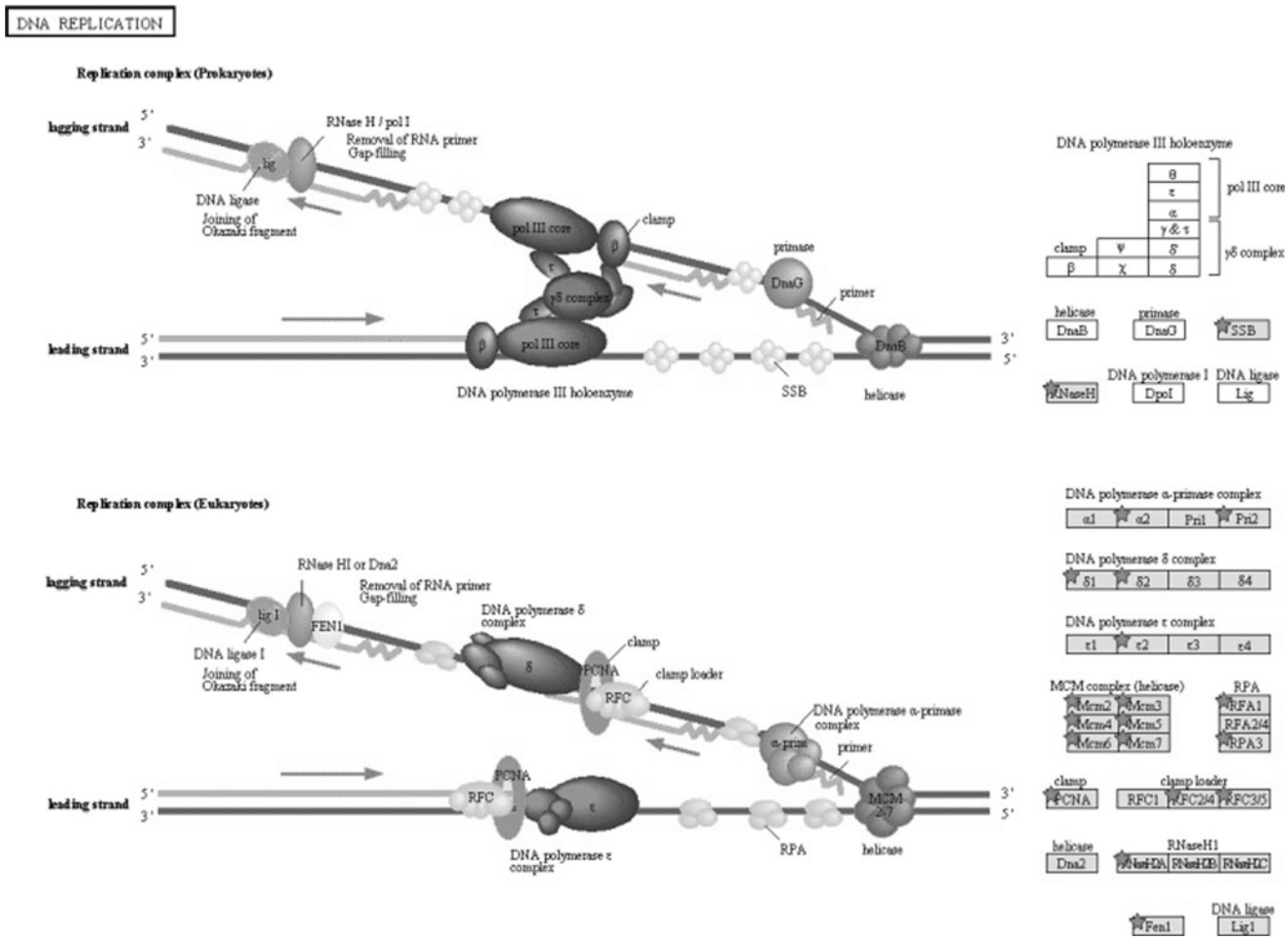

Fig. 4 The enriched DNA replication pathway of the differentially expressed genes (DEGs). The *stars* represent the DEGs changed in this pathway

process of DEGs, such as cell division, cell cycle, development and regeneration of organs, differentiation and maturation of lymphocyte, signal transduction, and the metabolism of sugar, protein and fat (Fig. 3).

Pathway Analysis

KEGG pathway enrichment analysis was used to identify the significant biological pathway related with the DEGs. Total 7 pathways were identified and listed in Table 1 (Table 1). The most significant pathway was DNA replication (FDR=5.17E-06), and the genes such as replication factor C 3 (RFC3), minichromosome maintenance complex component 4 (MCM4), flap endonuclease 1 (FEN1), and polymerase (DNA directed) delta (POLD) were enriched in this pathway (Fig. 4). Cell cycle was the second significant pathway (FDR=3.61E-05), and genes such as pituitary tumor transforming gene 1 (PTTG1), cyclin B1, cell division cycle 6 homolog, and proliferating cell nuclear antigen (PCNA) were enriched in this pathway (Fig. 5).

Discussion

Although outcomes for patients with lung adenocarcinoma have been improved, SQLC currently lacks therapeutically exploitable genetic alterations [16]. Therefore, further studies are needed to detect more genetic alternations. In the present study, 2961 DEGs of SQLC and normal cells were screened, and then the functions and pathway enrichment analysis of the DEGs found that the main functional changes of SQLC cells were cell cycle and metabolism, which is consistent with the significantly enriched pathways by KEGG. The top 2 pathways: DNA replication and cell cycle confirmed the active division ability of lung cancer cells.

MCM2-7, RFC2, RFC3, RFC4, RFC5, FEN1, and POLD are the mainly genes close related with DNA replication.

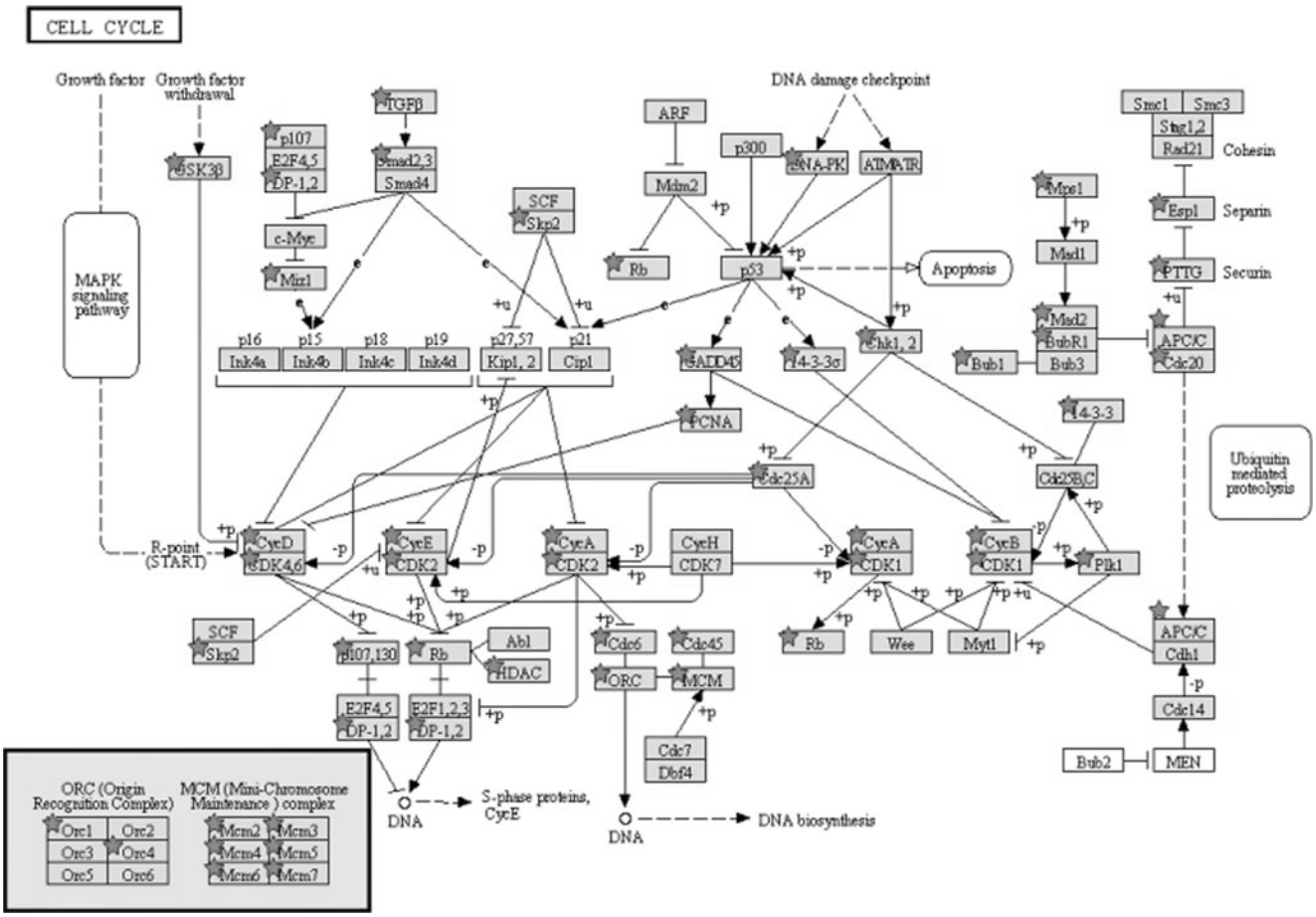

Fig. 5 The enriched cell cycle pathway of the differentially expressed genes (DEGs). The *stars* represent the DEGs changed in this pathway

MCM proteins are essential replication initiation and elongation factors, consisting of MCM2-7 [17]. MCM4 is highly expressed in non-small cell lung cancer (NSCLC), and plays an essential role in the proliferation of some NSCLC cells [18]. RFC is the clam loader in DNA replication. The deregulation of DNA repair and replication caused by RFC3 mutation and expression loss can cause the cancer pathogenesis in gastric and colorectal cancers [19]. FEN1 is a central component of cellular DNA metabolism, and FEN1 efficiency and specificity are critical to the maintenance of genome fidelity [20]. Kikuchi et al. [21] report that FEN1 is an important gene in human carcinogenesis and gene polymorphisms in FEN1 confers susceptibility to lung cancer. Allera-Moreau et al. [22] report that the expressions of the replicative DNA polymerases POLD and its processivity factor PCNA are slightly increased in NSCLC tumors.

In the cell cycle, PTTG1, CCNB1, CDC6, and PCNA are closely related genes. It is reported that PTTG1 can promote migration and invasion of human NSCLC [23]. The expression of CCNB1 and CDC6 are found to be in high levels in lung cancer [24], and up-regulated in SQLC [25]. The suppression of CCNB1 protein expression will lead to cell cycle arrest in G2 phase [26]. PCNA is up-regulated; and its expression level is correlated to the pathological type, lymph node metastasis, staging and differentiation grade in lung cancer [27]. PCNA works not just as a proliferation index but also can response to lung aggression [28].

The changes of mismatch repair pathway suggested the ability of the DNA damage self-healing of cancer cells also changed. Moreover, the changes of proteasome, metabolic pathways, and glutathione metabolism further indicated that the metabolic ability of lung cancer cells had changed. The changes of proteasome may result in the degeneration of cancer suppressor protein p53 and cell cycle inhibiting factor, and then provide a base of cancerization [29].

Conclusions

By microarray analysis of DEGs, and the function and pathway enrichment of DEGs, the global and molecular level changes of lung cancer was studied. It is helpful to elucidate the mechanism of lung cancer.

References

1. Tseng C-Y, Huang Y-C, Su S-Y, Huang J-Y, Lai C-H, Lung C-C, Ho C-C, Liaw Y-P (2012) Cell type specificity of female lung cancer associated with sulfur dioxide from air pollutants in Taiwan: an ecological study. BMC Publ Health 12(1):4
2. Giangreco A, Lu L, Vickers C, Teixeira VH, Groot KR, Butler CR, Ilieva EV, George PJ, Nicholson AG, Sage EK (2012) β–Catenin determines upper airway progenitor cell fate and preinvasive squamous lung cancer progression by modulating epithelial–mesenchymal transition. J Pathol 226(4):575–587
3. Sequoia Ecosystem and Recreation Preserve Act of 1999 (1999) (trans: Rep. George E. Brown J). 106th Congress edn
4. Hammerman P (2012) How far away is targeted treatment for squamous cell lung cancer? Oncology Times UK 9(11):21
5. Wu M, Tu T, Huang Y, Cao Y (2013) Suppression subtractive hybridization identified differentially expressed genes in lung adenocarcinoma: ERGIC3 as a novel lung cancer-related gene. BMC Cancer 13(1):44
6. Fang X, Netzer M, Baumgartner C, Bai C, Wang X (2012) Genetic network and gene set enrichment analysis to identify biomarkers related to cigarette smoking and lung cancer. Cancer Treat Rev
7. Wachi S, Yoneda K, Wu R (2005) Interactome-transcriptome analysis reveals the high centrality of genes differentially expressed in lung cancer tissues. Bioinformatics 21(23):4205–4208
8. Team RC (2008) R: a language and environment for statistical computing. R Found Stat Comput
9. Davis S, Meltzer PS (2007) GEOquery: a bridge between the Gene Expression Omnibus (GEO) and BioConductor. Bioinformatics 23(14):1846–1847
10. Diboun I, Wernisch L, Orengo C, Koltzenburg M (2006) Microarray analysis after RNA amplification can detect pronounced differences in gene expression using limma. BMC Genomics 7(1):252
11. Smyth GK (2004) Linear models and empirical bayes methods for assessing differential expression in microarray experiments. Stat Appl Genet Mol Biol 3(1):3
12. Benjamini Y, Hochberg Y (1995) Controlling the false discovery rate: a practical and powerful approach to multiple testing. J R Stat Soc Series B (Methodological):289–300
13. Hulsegge I, Kommadath A, Smits MA (2009) Globaltest and GOEAST: two different approaches for Gene Ontology analysis. In: BMC proceedings. BioMed Central Ltd, p S10
14. Zheng Q, Wang X-J (2008) GOEAST: a web-based software toolkit for Gene Ontology enrichment analysis. Nucleic Acids Res 36(suppl 2):W358–W363
15. Zhang JD, Wiemann S (2009) KEGGgraph: a graph approach to KEGG PATHWAY in R and bioconductor. Bioinformatics 25(11): 1470–1471
16. Weiss J, Sos ML, Seidel D, Peifer M, Zander T, Heuckmann JM, Ullrich RT, Menon R, Maier S, Soltermann A (2010) Frequent and focal FGFR1 amplification associates with therapeutically tractable FGFR1 dependency in squamous cell lung cancer. Sci Transl Med 2(62):62ra93–62ra93
17. Tye BK (1999) MCM proteins in DNA replication. Annu Rev Biochem 68(1):649–686
18. Kikuchi J, Kinoshita I, Shimizu Y, Kikuchi E, Takeda K, Aburatani H, Oizumi S, Konishi J, Kaga K, Matsuno Y (2011) Minichromosome maintenance (MCM) protein 4 as a marker for proliferation and its clinical and clinicopathological significance in non-small cell lung cancer. Lung Cancer 72(2):229–237
19. Kim YR, Song SY, Kim SS, An CH, Lee SH, Yoo NJ (2010) Mutational and expressional analysis of RFC3, a clamp loader in DNA replication, in gastric and colorectal cancers. Hum Pathol 41(10):1431–1437
20. Balakrishnan L, Bambara RA (2013) Flap Endonuclease 1. Annu Rev Biochem (0)
21. Yang M, Guo H, Wu C, He Y, Yu D, Zhou L, Wang F, Xu J, Tan W, Wang G (2009) Functional FEN1 polymorphisms are associated with DNA damage levels and lung cancer risk. Hum Mutat 30(9):1320–1328
22. Allera-Moreau C, Rouquette I, Lepage B, Oumouhou N, Walschaerts M, Leconte E, Schilling V, Gordien K, Brouchet L, Delisle M (2012) DNA replication stress response involving PLK1, CDC6, POLQ, RAD51 and CLASPIN upregulation prognoses the outcome of early/mid-stage non-small cell lung cancer patients. Oncogenesis 1(10):e30
23. Li H, Yin C, Zhang B, Sun Y, Shi L, Liu N, Liang S, Lu S, Liu Y, Zhang J (2013) PTTG1 promotes migration and invasion of human non-small cell lung cancer cells and is modulated by miR-186. Carcinogenesis
24. Andriani F, Roz E, Caserini R, Conte D, Pastorino U, Sozzi G, Roz L (2012) Inactivation of both FHIT and p53 cooperate in deregulating proliferation-related pathways in lung cancer. J Thorac Oncol 7(4): 631
25. Daraselia N, Wang Y, Budoff A, Lituev A, Potapova O, Monforte J, Ossovskaya V (2011) Pathway analysis of primary human non-small cell lung cancer (NSCLC). J Clin Oncol (Meeting Abstracts), p 10573
26. Yang I-P, Tsai H-L, Hou M-F, Chen K-C, Tsai P-C, Huang S-W, Chou W-W, Wang J-Y, Juo S-HH (2012) MicroRNA-93 inhibits tumor growth and early relapse of human colorectal cancer by affecting genes involved in the cell cycle. Carcinogenesis 33(8):1522–1530
27. Zhao Y, Li X, Sui X, Tang X, Qin H, Ren H (2010) Expression and significance of PCNA and Caspase-3 in the tissue of lung cancer. Chin J Cell Mol Immunol 26(2):154
28. Groeger AM, Caputi M, Esposito V, Baldi A, Rossiello R, Santini D, Mancini A, Kaiser HE, Baldi F (2000) Expression of p21 in non small cell lung cancer relationship with PCNA. Anticancer Res 20(5A):3301
29. Li C, Johnson DE (2013) Liberation of functional p53 by proteasome inhibition in human papilloma virus-positive head and neck squamous cell carcinoma cells promotes apoptosis and cell cycle arrest. Cell Cycle 12(6)