

RESEARCH

Open Access



Integrative analysis of genetic and epigenetic profiling of lung squamous cell carcinoma (LSCC) patients to identify smoking level relevant biomarkers

Bidong Ma¹, Zhiyou Huang¹, Qian Wang², Jizhou Zhang¹, Bin Zhou¹ and Jiaohong Wu^{3*}

* Correspondence: mazheng0817@163.com

³Department of Gynecology and Oncology, Wen Zhou Medical University affiliated People's Hospital, Wen Zhou, Zhe Jiang province, People's Republic of China Full list of author information is available at the end of the article

Abstract

Background: Incidence and mortality of lung cancer have dramatically decreased during the last decades, yet still approximately 160,000 deaths per year occurred in United States. Smoking intensity, duration, starting age, as well as environmental cofactors including air-pollution, showed strong association with major types of lung cancer. Lung squamous cell carcinoma is a subtype of non-small cell lung cancer, which represents 25% of the cases. Thus, exploring the molecular pathogenic mechanisms of lung squamous cell carcinoma plays crucial roles in lung cancer clinical diagnosis and therapy.

Results: In this study, we performed integrative analyses on 299 comparative datasets of RNA-seq and methylation data, collected from 513 lung squamous cell carcinoma cases in The Cancer Genome Atlas. The data were divided into high and low smoking groups based on smoking intensity (Numbers of packs per year). We identified 1002 significantly up-regulated genes and 534 significantly down-regulated genes, and explored their cellular functions and signaling pathways by bioconductor packages GOseq and KEGG. Global methylation status was analyzed and visualized in circular plot by CIRCOS. RNA-and methylation data were correlatively analyzed, and 24 unique genes were identified, for further investigation of regional CpG sites' interactive patterns by bioconductor package coMET. AIRE, PENK, and SLC6A3 were the top 3 genes in the high and low smoking groups with significant differences.

Conclusions: Gene functions and DNA methylation patterns of these 24 genes are important and useful in disclosing the differences of gene expression and methylation profiling caused by different smoking levels.

Keywords: Lung squamous cell carcinoma, Data mining, RNA-seq, Methylation, The Cancer genome atlas, Smoking intensity

Background

Lung cancer ranked the second place in estimated new cases (approximately 234,000 new cases) and the first place in estimated deaths (approximately 160,000 new cases) in United States, 2018 [1]. Recent studies reveal several major risk factors influence the lung cancer [2, 3]. For example, environmental pollutions and tobacco consumption [4–6]. Chronic exposure to carcinogens released by smoking metabolically induced



damages on pulmonary cells and caused DNA mutation or DNA adducts, eventually leading to lung cancer [7, 8]. Seeking the connection between smoking habits and changes of genetic and epigenetic levels in all types of lung cancer is quiet promising in finding unique biomarkers for diagnosis and treatment, especially after people found out smoking showed various impacts of different types of lung cancers [9].

There are two main types of lung cancers, small-cell lung carcinoma (SCLC) and non-small-cell lung carcinoma (NSCLC). Lung squamous cell cancer (LSCC) is the second common subtype of the latter, which represented almost 25% of overall lung cancer [10]. Numerous studies had focused on genomic characteristics of NSCLC (including LSCC), especially mutated genes found in lung cancer patients. Glutathione S-Transferase Mu 1 (GSTM1) is a glutathione S-transferase, which functions as detoxifier in tobacco smoking induced carcinogenesis [11]. GSTM1 deficiency would significantly increase the risk of LSCC [12]. Cytochrome P450 Family 1 Subfamily A Member 1 (CYP1A1) was also found high polymorphic and associated with elevated risk of lung cancer [13, 14]. Epidemiology studies showed that there was an increasing risk of developing lung cancer in cigarette amount dependent manner on both GSTM1 and CYP1A1 gene mutate patients [15]. Other genes with gain or loss of function mutations, such as FGFR1, TP53 and NOTCH1, might also play important roles in developing LSCC [16–19]. The TCGA project conducted a more comprehensive study on the genomes and regulatory pathways of LSCC, and reported more mutated genes, such as CDKN2A, PTEN, PIK3CA, NFE2L2, KEAP1 and RB1 [20]. However, compared to lung adenocarcinoma, the understanding of mutated genes and pathways in LSCC is still limited, especially in terms of actionable mutations.. Next-generation sequencing and gene chip technology produce high-throughput RNA, DNA and CpG methylation data from multiple LSCC patients. With appropriated bioinformatics analysis tools, these data would be revisited/reinvestigated in a more comprehensive perspective of gene differentiation and regulatory pathways, and for the purpose to identify significant biomarkers for diagnosis and treatment.

In this study, we aimed to identify biomarker genes related to the smoking intensity in LSCC patients, and performed the integrative analysis using dataset downloaded from The Cancer Genome Atlas (TCGA). In particular, gene ontology, cellular enrichment functions and pathway analysis were performed on RNA-seq data. Global DNA methylation profiles were drawn from on methylation data. Significant biomarker genes were selected and regional CpG methylations of them were evaluated. All the bioinformatics analyses were performed using Bioconductor packages and computer software. This study is for the first time to systematically explore the genetic and epigenetic profiles of LSCC patients and look for potential genes significantly different between high- and low-smoking groups. Such information is highly valuable for establishing an effective non-invasive screening method and treatment against human LSCC.

Materials and methods

Data source, RNA-seq and DNA methylation analysis

We used publicly available data for the TCGA-LSCC cohort that includes 513 cases [21, 22]. From this cohort, we obtained 299 LSCC cases with both gene expression and epigenetic datasets. We divided these data into two equal groups according to smoking

intensity, 150 high smoking cases (average number of cigarettes consumed per year 73.6 packs with a median of 63 packs) and 149 low smoking cases (average number of cigarettes consumed per year 30.9 packs with a median of 30 packs).

The gene expression data was obtained as raw count values from TCGA public level 3 transcription profiles. We applied bioconductor package (edgeR) [23] to the transcription profiles and tested for differential expression between high-smoking and low-smoking groups. *P*-values were corrected for multiple testing by computing *q*-values (false discovery rates). Then the significant Differentially Expressed Genes (DEGs, $P < 0.01$ and fold change value > 2 or < -2) were selected for further analysis. Principle component analysis was performed to classify between high- and low-smoking groups. The Gene Ontology (GO) functional annotation of DEGs was accomplished by Biomart Database and Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway annotation of DEGs was accomplished by using BLASTP to align to KEGG database with a cutoff *e*-value of 10^{-5} [24–26]. GO enrichment analysis provides all GO terms that significantly enriched in DEGs comparing to the genome background, and filtered the DEGs that correspond to biological functions. This method firstly mapped all DEGs to GO terms in the database [27], calculating gene numbers for every term, then using hyper geometric test to find significantly enriched GO terms in DEGs comparing to the genome background. The formula is:

$$P = 1 - \sum_{i=0}^{m-1} \frac{\binom{M}{i} \binom{N-M}{n-i}}{\binom{N}{n}}$$

Where *N* is the number of all genes with GO annotation; *n* is the number of DEGs in *N*; *M* is the number of all genes that are annotated to the certain GO terms; *m* is the number of DEGs in *M*. The calculated *P*-value goes through Bonferroni Correction, taking corrected *P*-value ≤ 0.05 as a threshold. GO terms fulfilling this condition are defined as significantly enriched GO terms in DEGs. Pathway enrichment analysis identified significantly enriched metabolic pathways or signal transduction pathways in DEGs comparing with the whole genome background. And the calculating formula was the same as that in GO enrichment analysis.

The DNA methylation data was obtained as beta values from TCGA public level 3 methylation profiles. We applied bioconductor packages *CHAMP* to the methylation profiles and tested for differential methylated CpG site and region between smoking high and low group samples [28]. Individual samples and CpG sites with a high missing rate ($> 5\%$) were excluded. The overall DNA methylation status was analyzed and visualized by CIRCOS, a multi-layer highly informative circular infographics suitable for different sets of DNA methylation status [29].

Integrative analysis of RNA-seq and DNA methylation data

Integrative analysis of RNA-seq and DNA methylation data was performed and detected *cis*-related correlations of CpG methylation and RNA expression. The core set of samples was used since all samples in this set had data available across the two platforms. For analysis involving the RNA-seq datasets, a \log_2 -transformation was used in order to deal with skewness in the data. Filtered DEGs from both datasets were

picked up and showed in a Column/Dot plot with RNA expression and DNA methylation status. Given the scientific evidence that reverse regulation was shown between DNA methylation and transcription levels, all genes that following the relationship were selected for further detailed pattern analysis.

Identification of regional DNA methylation pattern on selected genes

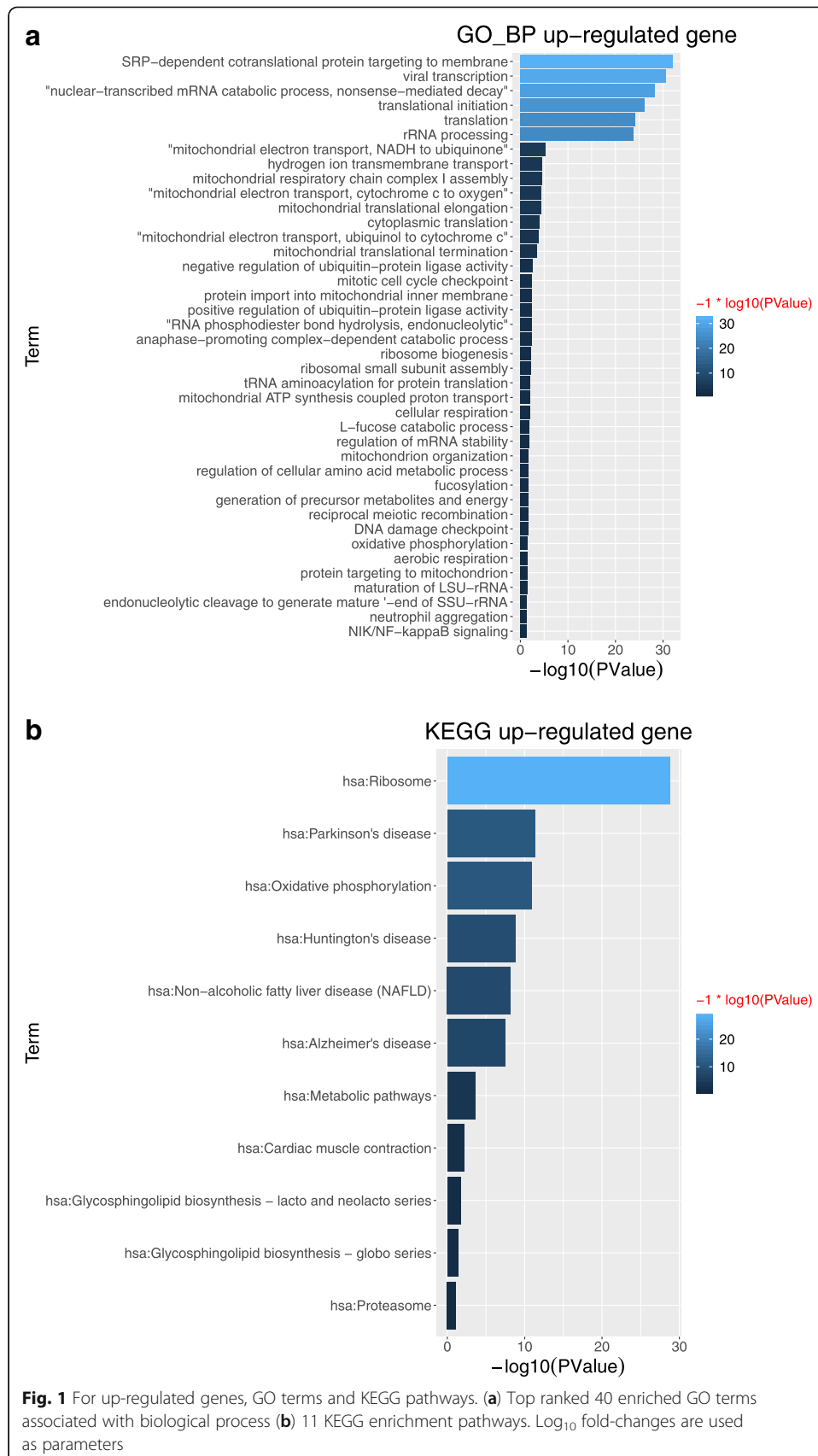
For identifying the regional DNA methylation pattern from selected 24 genes, bioconductor package *coMET* was used for analysis and data visualization [30]. The *coMET* package provides information including chromosome position of genes, regional CpG site methylation changes, genomic annotation tracks and correlation of DNA methylation in selected CpG sites. The *coMET* is valuable for interpreting paternal results in order to fully elucidate the relationship between gene epigenome and transcriptional expression.

Results

Differentially expressed genes between LSCC patients in high and low groups

In order to explore the genetic profiles and identify significant genes differentially expressed in the comparison of high and low groups, we performed principle component analysis (PCA), DEGs discovery and statistical analysis (EdgeR), Gene Ontology (GO) and KEGG analysis on all 299 RNA-seq datasets. First of all, PCA plot shows less significant variance between high and low groups and results are summarized in Additional file 1: Fig. S1. Even through the overall transcriptome signatures are not significantly different, it is very intriguing to see if individual genes are able to be identified to distinguish two groups by bioinformatical analysis. Using EdgeR, we identified a total number of 1002 up-regulated and 534 down-regulated DEGs in high smoking intensive group when compared to low group. Then we input up-regulated and down-regulated genes separately into *Goseq* package for GO and KEGG enrichment analysis to determine gene functions and pathways.

For up-regulated genes, top ranked 40 enriched GO terms and 11 KEGG pathways are summarized in Fig. 1a and b, respectively. In Fig. 1a, highly enriched GO terms with high gene numbers annotated can be roughly divided into two biological functions: regulating the mRNA formation and translation in ribosome which includes translational initiation, translation, rRNA processing, nuclear-transcribed mRNA catabolic process, nonsense-mediated decay, and viral infection to host cells (viral transcription). Several biological process GO terms such as mitochondrial electron transport, cytochrome c to oxygen, mitochondrial electron transport, NADH to ubiquinone are highly associated with mitochondrial membrane potential, which suggested that high intensive smoking might cause significant dysfunction of mitochondria. Mitochondrial membrane potential has been determined as a unique biomarker for oxidative environmental stress, especially for smoking cigarettes [31]. Intensive researches have been conducted and found out cigarette smoke not only induces significantly mitochondrial membrane potential alteration, but also disturbs the expression and distribution of membrane protein, mitochondrial depolarization [32, 33]. Reactive oxidative stress (or inflammation) associated biological process GO terms, NIK/NF-kappa B signaling, DNA damage checkpoint are also found. Nicotine has been reported to induce

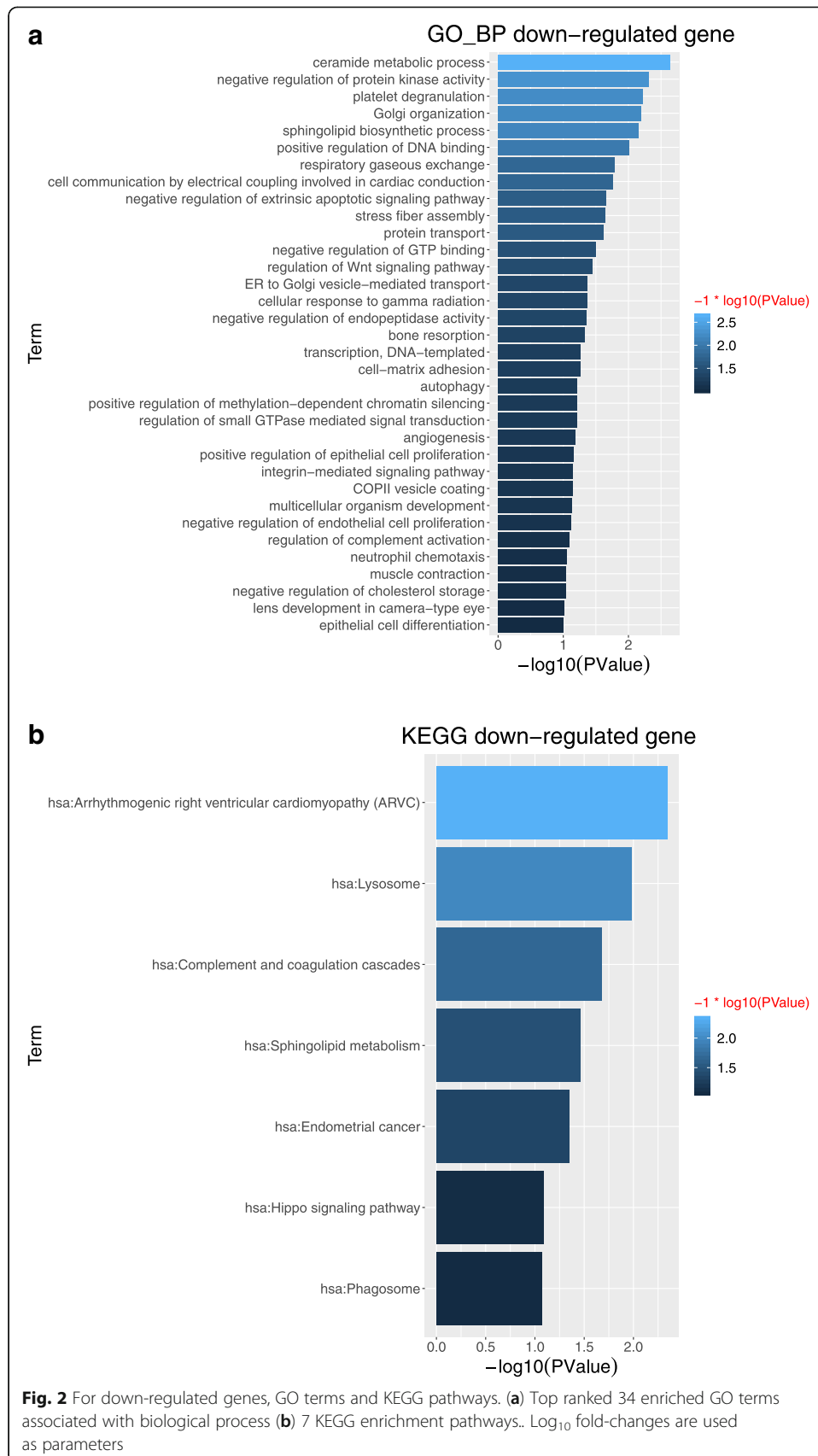


inflammation and oxidative stress through regulating the nuclear factor (NF)-kappa B signaling pathway [34]. DNA damage is also one of major carcinogenic mechanisms induced by smoking [35]. In Fig. 1b, 11 KEGG pathways are listed and ribosome pathway (and RNA transport pathway) are highly associated with biological process GO terms identified previously. Proteasome pathway involves in many cellular functions including regulating cell cycle [36], inflammation response [37] and induction of apoptosis [38] reported in cigarette smoking induced carcinogenesis. The summary of GO terms and KEGG pathways for up-regulated genes are summarized in Additional file 5: Table S1A and Additional file 6: Table S1B.

For down-regulated genes, top ranked 34 enriched GO terms and 7 KEGG pathways are summarized in Fig. 2a and b, respectively. In Fig. 2a, transcription, DNA-templated is defined as regulating the RNA synthesis using DNA template. For the first time, high intensive smoking has been identified not only to positively affect RNA expression and function in ribosome, but also negatively influence the transcription using DNA-templated. These findings are critical for fully understanding the mechanism of high cigarette smoking caused LSCC and other type of lung cancers, and provides guidance for future lung cancer research focusing on RNA function targeted therapies and diagnosis biomarkers. Regulation of Wnt signaling pathway, positive regulation of epithelial cell proliferation, negative regulation of apoptotic signaling pathway, negative regulation of endothelial cell proliferation can be grouped as cell fate signaling pathways. Wnt signaling pathway has been well studied as an important regulatory pathway response to cigarette smoking [39–42], and some of studies reports that Wnt pathway plays important roles in epithelial cells proliferation [43–45]. Smoking also impairs the endothelial cell proliferation through reactive oxidative stress, and inhibited cell apoptosis [46–48]. In Fig. 2b, phagosome is described as defendant system against infectious toxin induced tissue damage. Several studies show that inhalation of toxic particles during smoking is able to induce severe damages on cellular phagosome [49, 50]. Lysosome is main organelle responsible for external macromolecules degradation. Cigarette smoking is reported to impair the function of lysosome through dysfunction of NRF-2 antioxidant pathway [51] and immune-response apoptotic pathways [52], which pathways are well organized as critical regulators in many human cancer carcinogenesis [53, 54]. Hippo signaling pathway regulates the cell density and population by enhancing cell proliferation and inducing apoptosis, controlling organ size across many species including human [55]. Cohort studies suggest that hippo signaling pathways are involved and important in types of lung cancers [56]. Yes Associated Protein 1 (YAP1) and WW Domain Containing Transcription Regulator 1 (WWTR1/TAZ) are key regulatory factors of Hippo Signaling pathways, and are targeted to develop effective lung cancer therapy by their inhibition [57]. The summary of GO terms and KEGG pathways for down-regulated genes are summarized in Additional file 7: Table S2A and Additional file 8: Table S2B.

DNA methylation status visualization

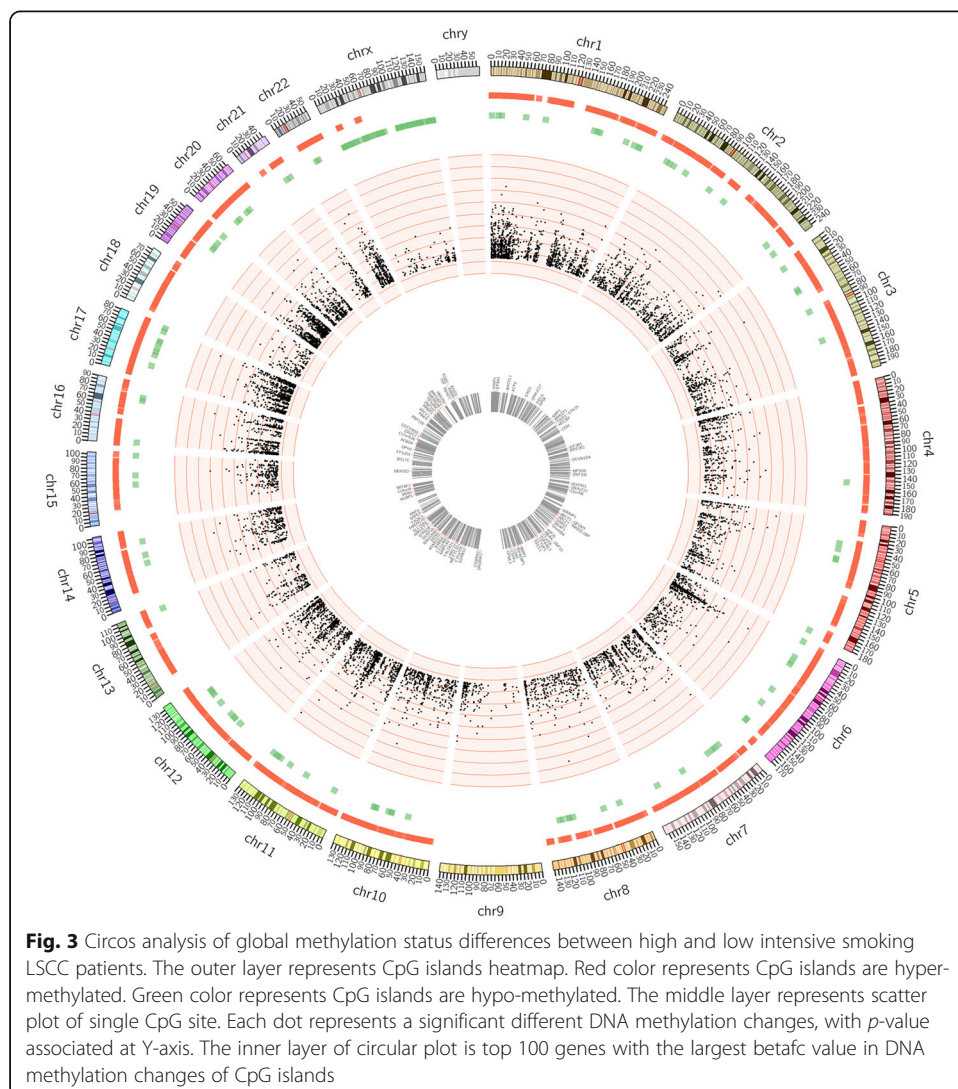
All 299 cases of methylation data were analyzed by Bioconductor package *CHAMP*. Beta value distribution and hierarchical cluster is summarized in Additional file 2: Figure S2A-B. The principle component analysis of methylation status difference between high and low



groups is summarized in Additional file 3: Fig. S3. These Figs suggest that global epigenome of LSCC patients with different smoking intensity has few significant changes in this cohort study. This observation shows consistency with our previous results from RNA-seq dataset analysis. However, significant changes of individual differentially methylated CpG site and region between high and low group are identified by *CHAMP* package. To check overall genome-wide methylation changes, CIRCOS is used to visualize data as multiple-layer circular plot in Fig. 3. The outer layer contains information includes chromosome, CpG island position and DNA methylation alteration. The middle layer is scatter plot of CpG site location. Each marker represents a CpG site which can be either hypo- or hyper-methylated and Y-axis represents the possibility. The inner layer lists the top 100 gene which show the most significant CpG island methylation changes based on beta values.

Integrative analysis for biomarker gene identification

The first step of integrative analysis is to select common genes showed in both RNA-seq and DNA methylation datasets, after gene filtration using core set presented

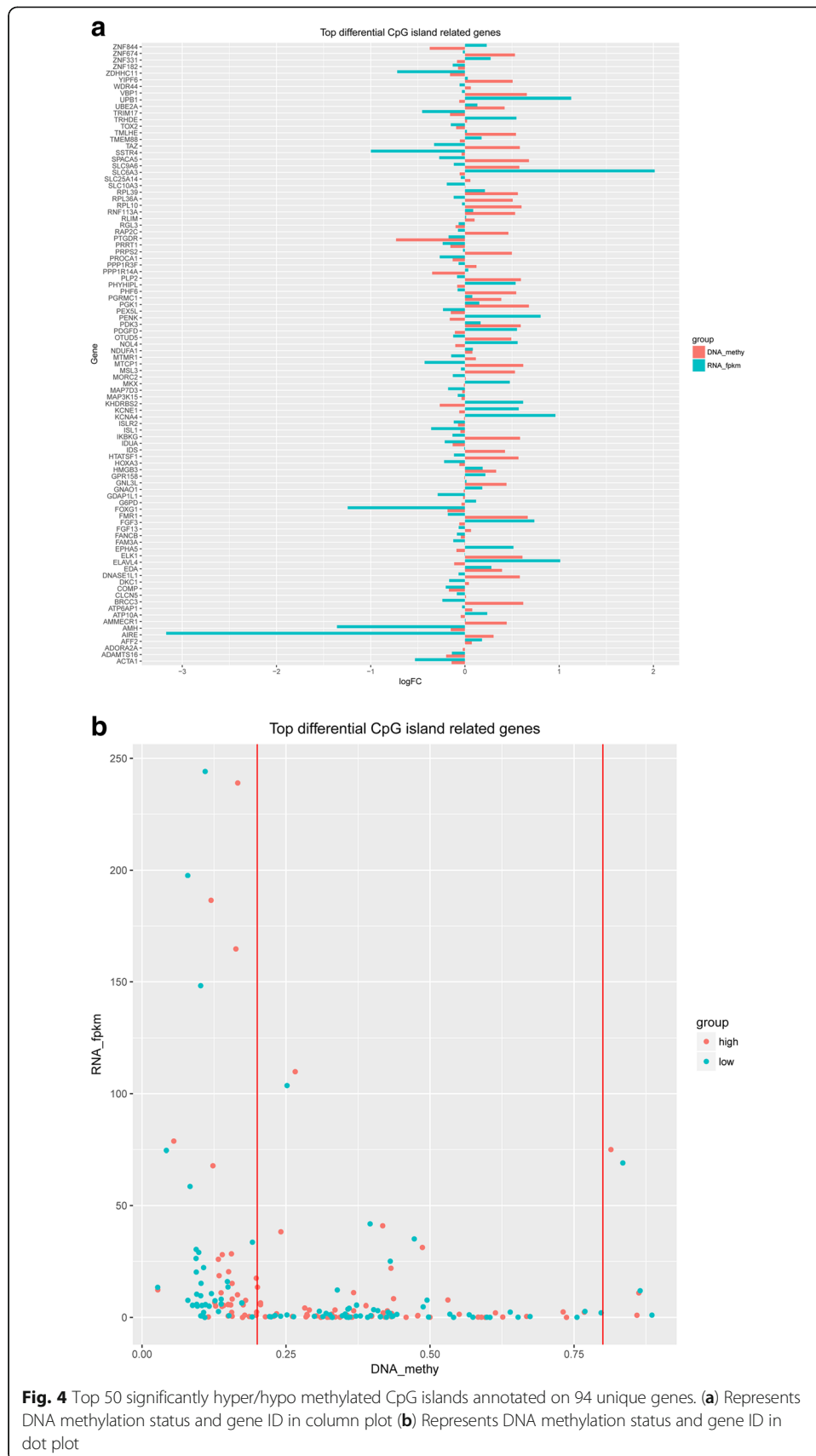


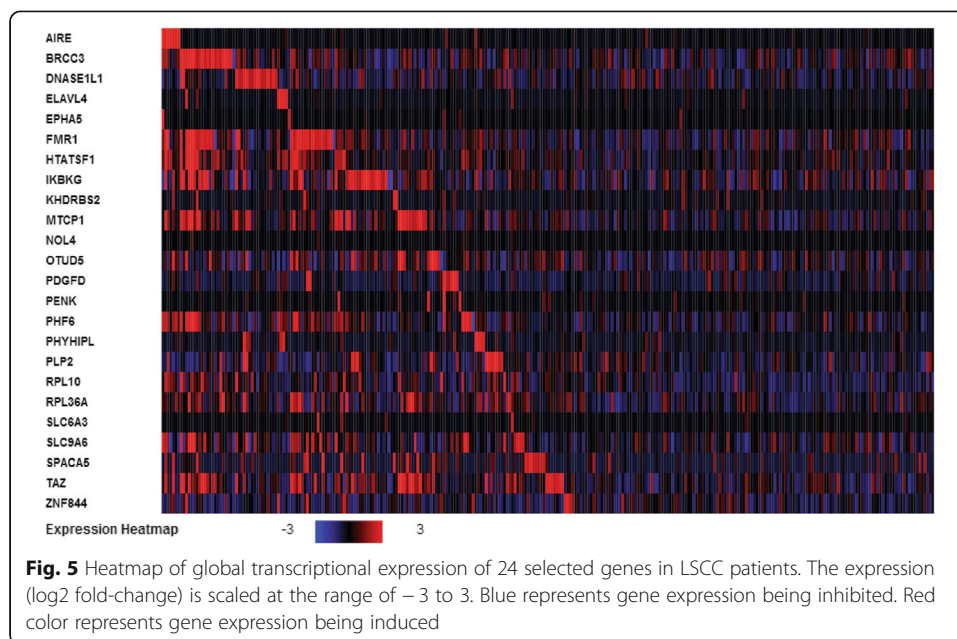
previously in this study. First of all, top 50 hyper and hypo-methylated CpG islands on 94 genes were selected from methylation data, with a p -value < 0.05 . They are summarized and visualized in Fig. 4a and b. In Fig. 4a, Column growing to right means CpG island (red column) hyper-methylated and gene expression (green column) increased. Column growing to left means CpG island hypo-methylated and gene expression decreased. Evidentially, genes with red and green columns growing to different directions are collected for further analysis. In Fig. 4b, all 94 genes expression and CpG island methylation status data from both high and low groups are extracted and presented as scatter plot. All dots close to right bottom corner are described as gene, which expression is lower and associated CpG island is hyper-methylated. These genes are considered having the most biological meaning in gene transcriptional pattern and potential to distinguish the difference between high and low groups.

In order to further explore the genomic and epigenomic difference between high and low group, the second step of integrative analysis is to select gene with high transcriptional expression and low CpG island methylation status, or low transcriptional expression and high CpG island methylation status. 24 genes have been found to match such requirements and their transcriptional expression heatmap across all LSCC patients are presented in Fig. 5. In Fig. 5, Y-axis is 24 gene names and X-axis is LSCC patient RNA expression level arranged by clustering order. The expression scale is set up at \log_2 (fold-change) from -3.0 to 3.0 . From the heatmap, it is clear that some LSCC patients show positively correlated and other LSCC patients show negative correlation on the transcriptional level. This interesting finding leads us to further studying gene promoter region of these 24 genes CpG methylation correlation pattern in order to elucidate the possible reasons. In addition, the integrative analysis of DNA methylation status and transcriptional expression of 24 genes are also presented in Additional file 4: Figure S4A-B.

DNA methylation correlation analysis on specific biomarker genes

Methylation alternation in nearby CpG sites shows strong correlation and has been identified in recent studies [58–60]. CoMet package is used to graphically display regional CpG site methylation changes and related information. Autoimmune Regulator (AIRE), Proenkephalin (PENK) and Solute Carrier Family 6 Member 3 (SLC6A3) are top 3 genes in the integrative analysis above, and may show potential for clinical diagnosis and therapies of LCSS. Their information are summarized in Fig. 6a, b, and c respectively. The entire coMet Fig contains three layer: top layer shows chromosome position of gene, and CpG site methylation changes with $-\log(p\text{-value})$ plot. Each dot represents a CpG site. Higher dot means this CpG site has more chance to be methylated. Middle layer contains information like gene ENSEMBL, CpG islands, Broad ChromHMM, SNP from USCS etc. The bottom layer shows the CpG site name and heatmap of nearby CpG site methylation correlation. The scale of heatmap is from -1.0 to 1.0 , in which -1.0 /blue means negatively correlated and 1.0 /red means positively correlated. For AIRE gene, cg09510531 to cg00495713, and cg01351072 to cg18876487 show strong positive correlation. CpG site range cg11923631 to cg27251412 shows strong negative correlation with CpG site range cg04878385 to cg21616420. For PENK gene, cg04612444 to cg06066137, and cg11060276 to cg27531336 show strong positive correlation. CpG site range cg04612444 to cg00468400 shows strong negative correlation with CpG site range





cg11060276 to cg27531336. For SLC6A3, cg04073265 to cg15999077 and cg17306747 to cg27580375 show strong positive correlation. CpG site range cg04073265 to cg16526509 shows strong negative correlation with CpG site range cg16614020 to cg27580375.

Discussion

In this study, we found 24 identical genes in LSCC patients, which exhibit significant epigenetic and genetic patterns between high and low intensive smokers. AIRE, PENK and SLC6A3 are Top 3 candidate genes, may be involved in the development of LSCC. AIRE is a transcriptional regulator gene, which guides the production of tissue specific antigens [61, 62]. Mutation of AIRE genes caused the dysfunction of T cell development and the immune system self-tolerance breakdown [63]. Recently, AIRE genes are found expressed in both human androgen-sensitive prostate cancer LNCaP cells and human androgen-insensitive prostate cancer PC-3 cells [64]. Interleukin-6 (IL-6), which acts as both pro-inflammatory and anti-inflammatory cytokine, is proposed to be regulated directly by AIRE genes in PC-3 cells. In addition, AIRE is found highly expressed in PC-3 cells and functions as prostate tumor promoter gene in vivo. AIRE deficiency finds to decrease the expression of Tyrosinase Related Protein 1 (TYRP1) in thymus, yet increases the T cell immune responses against melanoma development [65]. AIRE is also involved in the development of tumor-associated Foxp3 + regulatory T cells (Tregs), which do not respond to tumor-specific antigens [66]. Epidemiological study shows that AIRE gene expression is strongly associated with tumor prognosis in ER positive breast cancer patients [67]. There is no direct evidence to demonstrate the role of AIRE in the development of human lung cancer, however based on your observations in this study and previous research, it is highly possible that AIRE regulates the transcriptional proteins overexpressed to induce G1 cell cycle arrest and apoptosis, and may also be indirectly involved in the immune escape of tumor cells. PENK (also known as proenkephalin A) regulates the expression of a serious of opioid polypeptides which modulates the immune system [68]. The mRNA level of PENK gene is

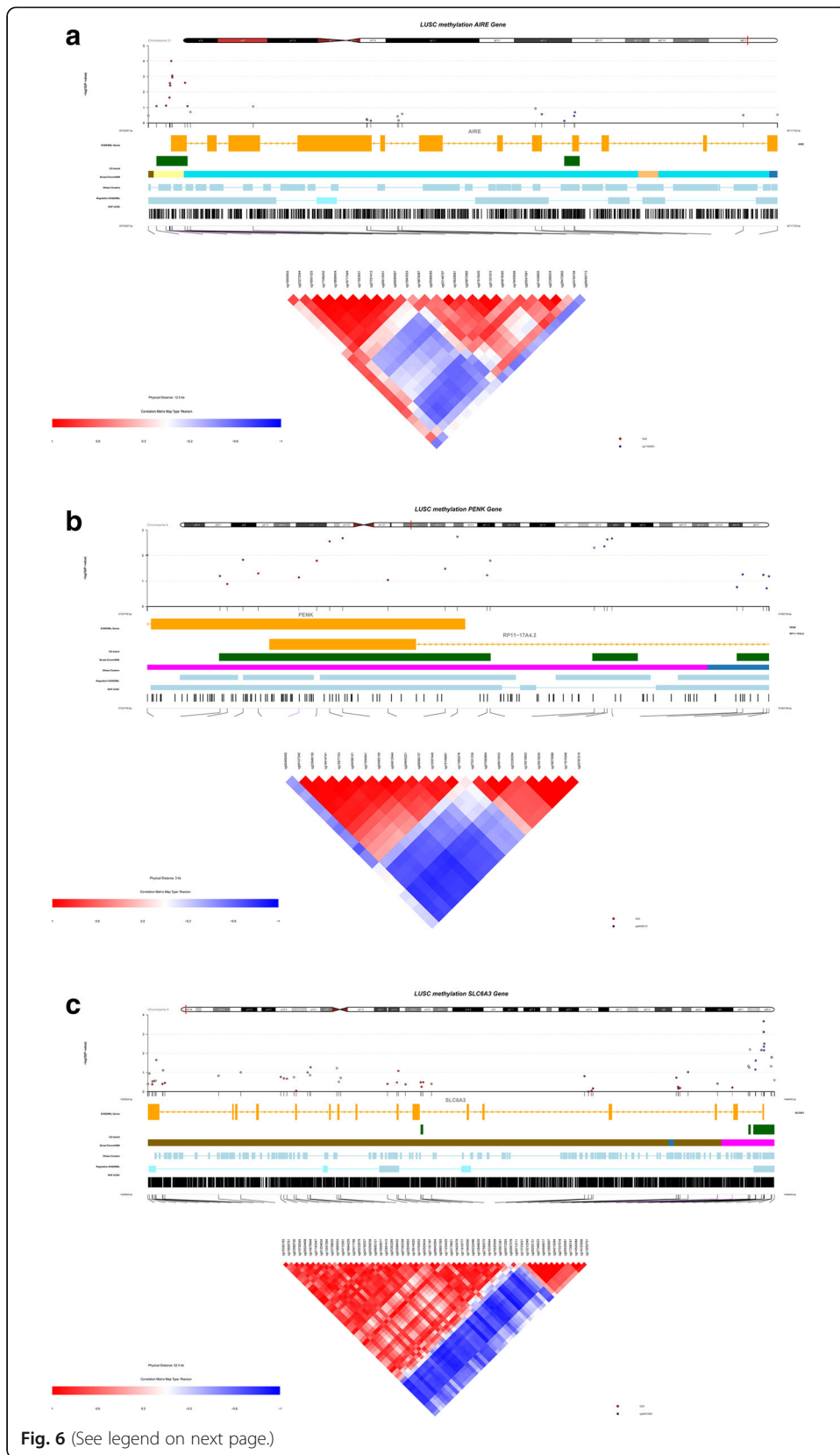


Fig. 6 (See legend on next page.)

(See figure on previous page.)

Fig. 6 CoMET plot showing the regional CpG methylation patterns on gene AIRE, PENK and SLC6A3. (a) AIRE, (b) PENK, (c) SLC6A3 are organized by p-value, gene location, CpG island (green), Gene ENSEMBL (yellow), Broad ChromHMM (pink), DNase Clusters, SPN from UCSC etc. and the correlation heatmap of Spearman correlation values among CpGs in the gene regulatory regions

significantly induced by the activation of T-Helper cell [69]. It has been identified that PENK was associated with has-miR126, which was found significantly down regulated in NSCLC [70]. The hyper-methylation of PENK was also found significantly in prostate cancer, colorectal cancer and lung adenocarcinoma [71–73]. Research suggests that PENK gene may be involved in activating the NF- κ B and p53 pathways by transcriptional repression [74]. SLC6A3 is dopamine transporter encoding gene, which plays important roles in neuronal disorder diseases [75–78]. It was found that the polymorphisms of SLC6A3 was highly associated with body mass index [79], which the later one have been strongly suggested as risk factor of many types of cancer including lung [80–82]. Recently, SLC6A3 is proposed as poor prognostic biomarker in human triple negative breast cancer tissues [83]. Methylation status in promoter region of SLC6A3 gene shows significant strong correlation with overall survival, with significant probability difference over 40%. SLC6A3 is also found highly expressed in clear cell renal cell carcinoma, which indicates as powerful therapeutic biomarkers for human renal cancer diagnosis and treatment [84]. Moreover, SLC6A3 c.-1476 T > G polymorphism may increase the risk of NSCLC, and the gain of 5p15.33 (which containing SLC6A3) is the most frequent genetic event in early stage NSCLC [85, 86]. It suggests that SLC6A3 may play an important role in the development of LSCC.

Integrative analysis of complex transcriptome, genome, methylome, proteome dataset and clinical profiles for cancer research are widely applied [87–91]. Whole-exome sequence reveals several driver mutations, including TP53, PTEN, NFE2L2 and KEAP1, whether in White or Asian LSCC patients [92]. Another cohort study determines that BRF2 RNA Polymerase III Transcription Initiation Factor Subunit (BRF2) is a novel oncogene showing strong lineage association with specificity of LSCC in an integrative analysis of genomic and gene expression microarray profiles [93]. The WWTR1/TAZ gene is also identified by genomic, epigenomic and proteomic integrative analysis in LSCC cohort study [94]. Most researches are separating LSCC patients by smokers or not, but not significantly focusing on smoking habit (intensity) or regional CpG methylation pattern analysis. Therefore, for the first time, we categorized the LSCC patients by smoking intensity and divided them as high (> 59 packs per year) and low (< 38 packs per year) group in this study. We performed computational analysis to reveal RNA-seq and DNA methylation datasets with clinical records obtained from TCGA database. Transcriptome dataset are analyzed by multiple Bioconductor packages including EdgeR, Goseq and KEGG. Epigenetic datasets are analyzed by CHAMP Bioconductor package, and regional epigenome correlation pattern are revealed and visualized by coMet Bioconductor package. Most of results are visualized by R. The whole analytic procedure can be used as guidance for handling multiple-omics datasets for future cohort genomic and epigenetic studies.

There are several confounding factors might influence including gender, ethnic and race etc. For example, gender variation has great impact on the risk of lung

cancer [95, 96]. Specific gene mutation such as CYP1A1 significantly increased the risk of female lung cancer [97]. Among all American cigarette smokers, African and Native Hawaiians are the more susceptible to lung cancer than white, Asian and Latino Americans [98]. In addition, some patients had quit smoking, or received targeted therapy in our datasets. These confounding factors should also be considered as well as smoking intensity. Important clinical information of the patients are summarized in Additional file 9: Table S3.

Dysregulation of gene CpG methylation is one of the hallmarks of human carcinogenesis, therefore making local and global epigenome serving as good sensor to response environmental impact, such as smoking patterns, duration and host immune response. In this study, 24 unique genes are identified their role as distinguisher between high and low intensive smoking habits and their CpG sites methylation shows strong correlation patterns. Based on all the results, it is possible to improve current LSCC early diagnosis and treatment by utilizing these genes.

Supplementary information

Supplementary information accompanies this paper at <https://doi.org/10.1186/s13040-019-0207-y>.

Additional file 1: Fig. S1. Principle Components Analysis (PCA) of 299 cases of RNA-seq with LSCC patients clinical data. Green dots represent cases of low groups; Red dots represent cases of low groups. There is no significant difference between high and low smoking intensity in LSCC patients.

Additional file 2: Fig. S2. Principle Components Analysis (PCA) of 299 cases of methylation with LSCC patients clinical data. Blue dots represent cases of low groups; Red dots represent cases of low groups. There is no significant difference between high and low smoking intensity in LSCC patients.

Additional file 3: Fig. S3. Global DNA methylation status with Hierarchical clustering (a) DNA methylation hierarchical clustering (b) Beta value distribution of global DNA methylation.

Additional file 4: Fig. S4. Hyper/hypo methylated CpG islands status and transcriptional expression of 24 unique genes. (a) Represents DNA methylation status and gene ID in column plot (b) Represents DNA methylation status and gene ID in dot plot.

Additional file 5: Table S1A. The summary of GO terms for up-regulated genes by RNA-seq.

Additional file 6: Table S1B. The summary of KEGG pathways for up-regulated genes by RNA-seq.

Additional file 7: Table S2A. The summary of GO terms for down-regulated genes by RNA-seq.

Additional file 8: Table S2B. The summary of KEGG pathways for down-regulated genes by RNA-seq.

Additional file 9: Table S3. Summary of clinical information of 299 LSCC cases involved in this study.

Abbreviations

AIRE: Autoimmune Regulator; CYP1A1: Cytochrome P450 Family 1 Subfamily A Member 1; DEGs: Differentially expressed genes; GO: Gene Ontology; GSTM1: Glutathione S-Transferase Mu 1; KEGG: Kyoto Encyclopedia of Genes and Genomes; LSCC: lung squamous cell carcinoma; NSCLC: non-small cell lung cancer; PENK: Proenkephalin; SLC6A3: Solute Carrier Family 6 Member 3; TCGA: The Cancer Genome Atlas; WWTR1/TAZ: WW Domain Containing Transcription Regulator 1

Acknowledgements

We would like to thank the Wenzhou Science and Technology Bureau (2017Y0849). Ma Bidong and want to thank the Wenzhou Hospital of Traditional Chinese Medicine for the workmate fellowship. Wu Jiaohong wished to thank the People's Hospital of Wenzhou.

Authors' contributions

Conceived and designed the ideas: BM, JW. Analyzed the data: BM, JW, JZ. Wrote the paper: BM, JW, QW, JZ, BZ, ZH. All authors read and approved the final manuscript.

Funding

Not applicable.

Availability of data and materials

The datasets analyzed during the current study are available in the TCGA public data, <https://www.cbioportal.org/> and <https://portal.gdc.cancer.gov/> [21, 22].

Ethics approval and consent to participate

Because the human data of our manuscript is all from TCGA public data, we believe that we do not need additional ethics approval.

Consent for publication

Not applicable.

Competing interests

The authors declare that they have no competing interests.

Author details

¹Department of Medical Oncology, Zhe Jiang Chinese Medicine University affiliated Chinese Medicine Hospital, Wen Zhou, Zhe Jiang province, People's Republic of China. ²Tianjia Genomes Tech CO., LTD., No. 6 Longquan Road, Anhui Chaohu economic develop zone, Hefei 238014, People's Republic of China. ³Department of Gynecology and Oncology, Wen Zhou Medical University affiliated People's Hospital, Wen Zhou, Zhe Jiang province, People's Republic of China.

Received: 6 June 2019 Accepted: 12 September 2019

Published online: 21 October 2019

References

1. Siegel RL, Miller KD, Jemal A. Cancer statistics, 2018. *CA Cancer J Clin*. 2018;68(1):7–30.
2. Simkovich SM, Goodman D, Roa C, Crocker ME, Gianella GE, Kirenga BJ, et al. The health and social implications of household air pollution and respiratory diseases. *NPJ Prim Care Respir Med*. 2019;29(1):12.
3. Murphy SE, Park SL, Balbo S, Haiman CA, Hatsukami DK, Patel Y, et al. Tobacco biomarkers and genetic/epigenetic analysis to investigate ethnic/racial differences in lung cancer risk among smokers. *Npj Precis Oncol*. 2018;2.
4. Kutikova L, Bowman L, Chang S, Long SR, Obasaju C, Crown WH. The economic burden of lung cancer and the associated costs of treatment failure in the United States. *Lung Cancer*. 2005;50(2):143–54.
5. Beane J, Vick J, Schembri F, Anderlind C, Gower A, Campbell J, et al. Characterizing the impact of smoking and lung cancer on the airway transcriptome using RNA-Seq. *Cancer Prev Res*. 2011;4(6):803–17.
6. Pesch B, Kendzia B, Gustavsson P, Jockel KH, Johnen G, Pohlbeln H, et al. Cigarette smoking and lung cancer—relative risk estimates for the major histological types from a pooled analysis of case-control studies. *Int J Cancer*. 2012;131(5):1210–9.
7. Hecht SS. Cigarette smoking and lung cancer: chemical mechanisms and approaches to prevention. *The Lancet Oncology*. 2002;3(8):461–9.
8. Pope CA 3rd, Burnett RT, Thun MJ, Calle EE, Krewski D, Ito K, et al. Lung cancer, cardiopulmonary mortality, and long-term exposure to fine particulate air pollution. *Jama*. 2002;287(9):1132–41.
9. Khuder SA. Effect of cigarette smoking on major histological types of lung cancer: a meta-analysis. *Lung Cancer*. 2001;31(2–3):139–48.
10. Kim Y, Hammerman PS, Kim J, Yoon JA, Lee Y, Sun JM, et al. Integrative and comparative genomic analysis of lung squamous cell carcinomas in east Asian patients. *Journal of clinical oncology : official journal of the American Society of Clinical Oncology*. 2014;32(2):121–8.
11. McWilliams JE, Sanderson BJS, Harris EL, Richertboe KE, Henner WD. Glutathione-S-transferase M1 (Gstm1) deficiency and lung-Cancer risk. *Cancer Epidem Biomar*. 1995;4(6):589–94.
12. Hirvonen A, Husgafvelpursiainen K, Anttila S, Vainio H. The Gstm1 null genotype as a potential risk modifier for squamous-cell carcinoma of the lung. *Carcinogenesis*. 1993;14(7):1479–81.
13. Kellermann G, Shaw CR, Luyten-Kellerman M. Aryl hydrocarbon hydroxylase inducibility and bronchogenic carcinoma. *N Engl J Med*. 1973;289(18):934–7.
14. Kouri RE, McKinney CE, Slomiany DJ, Snodgrass DR, Wray NP, Mclemore TL. Positive correlation between high aryl-hydrocarbon hydroxylase-activity and primary lung-Cancer as analyzed in cryopreserved lymphocytes. *Cancer Res*. 1982;42(12):5030–7.
15. Kihara M, Kihara M, Noda K. Risk of smoking for squamous and small cell carcinomas of the lung modulated by combinations of CYP1A1 and GSTM1 gene polymorphisms in a Japanese population. *Carcinogenesis*. 1995;16(10):2331–6.
16. Heist RS, Mino-Kenudson M, Sequist LV, Tammireddy S, Morrissey L, Christiani DC, et al. FGFR1 amplification in squamous cell carcinoma of the lung. *Journal of thoracic oncology : official publication of the International Association for the Study of Lung Cancer*. 2012;7(12):1775–80.
17. Jiang T, Gao G, Fan G, Li M, Zhou C. FGFR1 amplification in lung squamous cell carcinoma: a systematic review with meta-analysis. *Lung Cancer*. 2015;87(1):1–7.
18. Monaco SE, Rodriguez EF, Mahaffey AL, Dacic S. FGFR1 amplification in squamous cell carcinoma of the lung with correlation of primary and metastatic tumor status. *Am J Clin Pathol*. 2016;145(1):55–61.
19. Wang NJ, Sanborn Z, Arnett KL, Bayston LJ, Liao W, Proby CM, et al. Loss-of-function mutations in notch receptors in cutaneous and lung squamous cell carcinoma. *Proc Natl Acad Sci U S A*. 2011;108(43):17761–6.
20. Cancer Genome Atlas Research N. Comprehensive genomic characterization of squamous cell lung cancers. *Nature*. 2012;489(7417):519–25.
21. Cerami E, Gao J, Dogrusoz U, Gross BE, Sumer SO, Aksoy BA, et al. The cBio Cancer genomics portal: an open platform for exploring multidimensional Cancer genomics data. *Cancer Discovery*. 2012;2(5):401–4.
22. Genomic Data Commons Data Portal [Internet]. National Cancer Institute. 2018 [cited June 1]. Available from: <https://portal.gdc.cancer.gov/>.
23. Robinson MD, McCarthy DJ, Smyth GK. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics*. 2010;26(1):139–40.

24. Young MD, Wakefield MJ, Smyth GK, Oshlack A. Gene ontology analysis for RNA-seq: accounting for selection bias. *Genome Biol.* 2010;11(2):R14.
25. ensembl [Internet]. European Molecular Biology Laboratory's European Bioinformatics Institute. 2018 [cited June 1]. Available from: <https://asia.ensembl.org/biomart/martview>.
26. KEGG [Internet]. Kanehisa Labs. 2018 [cited June 1]. Available from: <https://www.kegg.jp/>.
27. The Gene Ontology Resource [Internet]. Gene Ontology Consortium. 2018 [cited June 1]. Available from: <http://geneontology.org/>.
28. Morris TJ, Butcher LM, Feber A, Teschendorff AE, Chakravarthy AR, Wojdacz TK, et al. ChAMP: 450k Chip analysis methylation pipeline. *Bioinformatics.* 2014;30(3):428–30.
29. Krzywinski M, Schein J, Birol I, Connors J, Gascoyne R, Horsman D, et al. Circos: an information aesthetic for comparative genomics. *Genome Res.* 2009;19(9):1639–45.
30. Martin TC, Yet I, Tsai PC, Bell JT coMET: visualisation of regional epigenome-wide association scan results and DNA co-methylation patterns *Bmc Bioinformatics.* 2015;16:131.
31. Vayssier-Taussat M, Kreps SE, Adrie C, Dall'Ava J, Christiani D, Polla BS. Mitochondrial membrane potential: a novel biomarker of oxidative environmental stress. *Environ Health Perspect.* 2002;110(3):301–5.
32. Slebos DJ, Ryter SW, van der Toorn M, Liu F, Guo F, Baty CJ, et al. Mitochondrial localization and function of heme oxygenase-1 in cigarette smoke-induced cell death. *Am J Respir Cell Mol Biol.* 2007;36(4):409–17.
33. Banzet N, Francois D, Polla BS. Tobacco smoke induces mitochondrial depolarization along with cell death: effects of antioxidants. *Redox Rep.* 1999;4(5):229–36.
34. Sugano N, Shimada K, Ito K, Murai S. Nicotine inhibits the production of inflammatory mediators in U937 cells through modulation of nuclear factor-kappa B activation. *Biochem Bioph Res Co.* 1998;252(1):25–8.
35. Pfeifer GP, Denissenko MF, Olivier M, Tretyakova N, Hecht SS, Hainaut P. Tobacco smoke carcinogens, DNA damage and p53 mutations in smoking-associated cancers. *Oncogene.* 2002;21(48):7435–51.
36. Kim SY, Lee JH, Huh JW, Ro JY, Oh YM, Lee SD, et al. Cigarette smoke induces Akt protein degradation by the ubiquitin-proteasome system. *J Biol Chem.* 2011;286(37):31932–43.
37. Adenuga D, Yao H, March TH, Seagrave J, Rahman I. Histone deacetylase 2 is phosphorylated, Ubiquitinated, and degraded by cigarette smoke. *Am J Respir Cell Mol Biol.* 2009;40(4):464–73.
38. Rangasamy T, Misra V, Zhen LJ, Tankersley CG, Tudor RM, Biswal S. Cigarette smoke-induced emphysema in a/J mice is associated with pulmonary oxidative stress, apoptosis of lung cells, and global alterations in gene expression. *Am J Phys Lung Cell Mol Phys.* 2009;296(6):L888–900.
39. Heijink IH, de Bruin HG, van den Berge M, Bennink LJ, Brandenburg SM, Gosens R, et al. Role of aberrant WNT signalling in the airway epithelial response to cigarette smoke in chronic obstructive pulmonary disease. *Thorax.* 2013;68(8):709–16.
40. Guo L, Wang T, Wu Y, Yuan Z, Dong J, Li X, et al. Wnt/beta-catenin signaling regulates cigarette smoke-induced airway inflammation via the PPARDelta/p38 pathway. *Lab Invest.* 2016;96(2):218–29.
41. Wang Z, Zhang JN, Hu XF, Chen XL, Wang XR, Zhao TT, et al. Effects of pentoxifylline on Wnt/beta-catenin signaling in mice chronically exposed to cigarette smoke. *Chin Med J.* 2010;123(19):2688–94.
42. Stewart DJ. Wnt signaling pathway in non-small cell lung cancer. *J Natl Cancer Inst.* 2014;106(1):djt356.
43. Heijink IH, de Bruin HG, Dennebos R, Jonker MR, Noordhoek JA, Brandsma CA, et al. Cigarette smoke-induced epithelial expression of WNT-5B: implications for COPD. *Eur Respir J.* 2016;48(2):504–15.
44. Liu F, Killian JK, Yang M, Walker RL, Hong JA, Zhang M, et al. Epigenomic alterations and gene expression profiles in respiratory epithelia exposed to cigarette smoke condensate. *Oncogene.* 2010;29(25):3650–64.
45. Hao YQ, Su ZZ, Lv XJ, Li P, Gao P, Wang C, et al. RNA-binding motif protein 5 negatively regulates the activity of Wnt/beta-catenin signaling in cigarette smoke-induced alveolar epithelial injury. *Oncol Rep.* 2015;33(5):2438–44.
46. Michaud SE, Dussault S, Groleau J, Haddad P, Rivard A. Cigarette smoke exposure impairs VEGF-induced endothelial cell migration: role of NO and reactive oxygen species. *J Mol Cell Cardiol.* 2006;41(2):275–84.
47. Nana-Sinkam SP, Lee LD, Sotto-Santiago S, Stearman RS, Keith RL, Choudhury Q, et al. Prostacyclin prevents pulmonary endothelial cell apoptosis induced by cigarette smoke. *Am J Resp Crit Care.* 2007;175(7):676–85.
48. Vayssier-Taussat M, Camilli T, Aron Y, Meplan C, Hainaut P, Polla BS, et al. Effects of tobacco smoke and benzo [a] pyrene on human endothelial cell and monocyte stress responses. *Am J Phys Heart Circ Phys.* 2001;280(3):H1293–H300.
49. Advani J, Subbannayya Y, Patel K, Khan AA, Patil AH, Jain AP, et al. Long-term cigarette smoke exposure and changes in miRNA expression and proteome in non-small-cell lung cancer. *OmicS.* 2017;21(7):390–403.
50. Moller W, Barth W, Pohlit W, Rust M, Siekmeier R, Stahlhofen W, et al. Smoking impairs alveolar macrophage activation after inert dust exposure. *Toxicol Lett.* 1996;88(1–3):131–7.
51. Tan WSD, Liao W, Peh HY, Vila M, Dong J, Shen HM, et al. Andrographolide simultaneously augments Nrf2 antioxidant defense and facilitates autophagic flux blockade in cigarette smoke-exposed human bronchial epithelial cells. *Toxicol Appl Pharmacol.* 2018;360:120–30.
52. Park EJ, Lee HS, Lee SJ, Park YJ, Park SI, Chang J, et al. Cigarette smoke condensate may disturb immune function with apoptotic cell death by impairing function of organelles in alveolar macrophages. *Toxicol in Vitro.* 2018;52:351–64.
53. Guo Y, Wu RY, Gaspar JM, Sargsyan D, Su ZY, Zhang CY, et al. DNA methylome and transcriptome alterations and cancer prevention by curcumin in colitis-accelerated colon cancer in mice. *Carcinogenesis.* 2018;39(5):669–80.
54. Ramirez CN, Li WJ, Zhang CY, Wu RY, Su S, Wang C, et al. In Vitro-In Vivo Dose Response of Ursolic Acid, Sulforaphane, PEITC, and Curcumin in Cancer Prevention (vol 20, 2017). *Aaps Journal.* 2018;20(2).
55. Pan DJ. The hippo signaling pathway in development and Cancer. *Dev Cell.* 2010;19(4):491–505.
56. Dhanasekaran SM, Balbin OA, Chen GA, Nadal E, Kalyana-Sundaram S, Pan JC, et al. Transcriptome meta-analysis of lung cancer reveals recurrent aberrations in NRG1 and hippo pathway genes. *Nat Commun.* 2014;5.
57. Lo Sardo F, Strano S, Blandino G. YAP and TAZ in Lung Cancer: Oncogenic Role and Clinical Targeting. *Cancers.* 2018;10(5).
58. Feng W, Shen L, Wen S, Rosen DG, Jelinek J, Hu X, et al. Correlation between CpG methylation profiles and hormone receptor status in breast cancers. *Breast Cancer Res.* 2007;9(4).
59. Zou B, Chim CS, Zeng H, Leung SY, Yang Y, Tu SP, et al. Correlation between the single-site CpG methylation and expression silencing of the XAF1 gene in human gastric and colon cancers. *Gastroenterology.* 2006;131(6):1835–43.

60. Kang JH, Kim SJ, Noh DY, Park IA, Choe KJ, Yoo OJ, et al. Methylation in the p53 promoter is a supplementary route to breast carcinogenesis: correlation between CpG methylation in the p53 promoter and the mutation of the p53 gene in the progression from ductal carcinoma in situ to invasive ductal carcinoma. *Lab Invest*. 2001;81(4):573–9.
61. Anderson MS, Su MA. AIRE expands: new roles in immune tolerance and beyond. *Nat Rev Immunol*. 2016;16(4):247–58.
62. Goodnow CC, Sprent J, Fazekas de St Groth B, Vinuesa CG. Cellular and genetic mechanisms of self tolerance and autoimmunity. *Nature*. 2005;435(7042):590–7.
63. Peterson P, Org T, Rebane A. Transcriptional regulation by AIRE: molecular mechanisms of central tolerance. *Nat Rev Immunol*. 2008;8(12):948–57.
64. Kalra R, Bhagyaraj E, Tiwari D, Nanduri R, Chacko AP, Jain M, et al. AIRE promotes androgen-independent prostate cancer by directly regulating IL-6 and modulating tumor microenvironment. *Oncogenesis*. 2018;7(5):43.
65. Zhu ML, Nagavalli A, Su MA. Aire deficiency promotes TRP-1-specific immune rejection of melanoma. *Cancer Res*. 2013;73(7):2104–16.
66. Malchow S, Leventhal DS, Nishi S, Fischer BI, Shen L, Paner GP, et al. Aire-dependent thymic development of tumor-associated regulatory T cells. *Science*. 2013;339(6124):1219–24.
67. Bianchi F, Sommariva M, De Cecco L, Triulzi T, Romero-Cordoba S, Tagliabue E, et al. Expression and prognostic significance of the autoimmune regulator gene in breast cancer cells. *Cell Cycle*. 2016;15(23):3220–9.
68. Vallejo R, de Leon-Casasola O, Benyamin R. Opioid therapy and immunosuppression: a review. *Am J Ther*. 2004;11(5):354–65.
69. Zurawski G, Benedik M, Kamb BJ, Abrams JS, Zurawski SM, Lee FD. Activation of mouse T-helper cells induces abundant preproenkephalin mRNA synthesis. *Science*. 1986;232(4751):772–5.
70. Gao W, Yu Y, Cao HL, Shen H, Li XD, Pan SY, et al. Deregulated expression of miR-21, miR-143 and miR-181a in non small cell lung cancer is related to clinicopathologic characteristics or patient prognosis. *Biomed Pharmacother*. 2010;64(6):399–408.
71. Roperch JP, Incitti R, Forbin S, Bard F, Mansour H, Mesli F, et al. Aberrant methylation of NPY, PENK, and WIF1 as a promising marker for blood-based diagnosis of colorectal cancer. *BMC Cancer*. 2013;13:566.
72. Ashour N, Angulo JC, Andres G, Alelu R, Gonzalez-Corpas A, Toledo MV, et al. A DNA hypermethylation profile reveals new potential biomarkers for prostate cancer diagnosis and prognosis. *Prostate*. 2014;74(12):1171–82.
73. Chung JH, Lee HJ, Kim BH, Cho NY, Kang GH. DNA methylation profile during multistage progression of pulmonary adenocarcinomas. *Virchows Archiv : an international journal of pathology*. 2011;459(2):201–11.
74. McTavish N, Copeland LA, Saville MK, Perkins ND, Spruce BA. Proenkephalin assists stress-activated apoptosis through transcriptional repression of NF-kappaB- and p53-regulated gene targets. *Cell Death Differ*. 2007;14(9):1700–10.
75. Vandenberg DJ, Persico AM, Uhl GR. A human dopamine transporter cDNA predicts reduced glycosylation, displays a novel repetitive element and provides racially-dimorphic TaqI RFLPs. *Brain Res Mol Brain Res*. 1992; 15(1–2):161–6.
76. Kurian MA, Zhen J, Cheng SY, Li Y, Mordekar SR, Jardine P, et al. Homozygous loss-of-function mutations in the gene encoding the dopamine transporter are associated with infantile parkinsonism-dystonia. *J Clin Invest*. 2009;119(6):1595–603.
77. Navaroli DM, Stevens ZH, Uzelac Z, Gabriel L, King MJ, Lifshitz LM, et al. The plasma membrane-associated GTPase Rin interacts with the dopamine transporter and is required for protein kinase C-regulated dopamine transporter trafficking. *J Neurosci*. 2011;31(39):13758–70.
78. Lin M, Pedrosa E, Shah A, Hrabovsky A, Maqbool S, Zheng D, et al. RNA-Seq of human neurons derived from iPSCs reveals candidate long non-coding RNAs involved in neurogenesis and neuropsychiatric disorders. *PLoS One*. 2011;6(9):e23356.
79. Azzato EM, Morton LM, Bergen AW, Wang SS, Chatterjee N, Kvale P, et al. SLC6A3 and body mass index in the prostate, lung, Colorectal and Ovarian Cancer Screening Trial. *BMC Med Genet*. 2009;10:9.
80. Renehan AG, Tyson M, Egger M, Heller RF, Zwahlen M. Body-mass index and incidence of cancer: a systematic review and meta-analysis of prospective observational studies. *Lancet*. 2008;371(9612):569–78.
81. Reeves GK, Pirie K, Beral V, Green J, Spencer E, Bull D, et al. Cancer incidence and mortality in relation to body mass index in the million women study: cohort study. *BMJ*. 2007;335(7630):1134.
82. Kabat GC, Wynder EL. Body mass index and lung cancer risk. *Am J Epidemiol*. 1992;135(7):769–74.
83. Stirzaker C, Zotenko E, Song JZ, Qu W, Nair SS, Locke WJ, et al. Methylome sequencing in triple-negative breast cancer reveals distinct methylation clusters with prognostic value. *Nat Commun*. 2015;6:5899.
84. Hansson J, Lindgren D, Nilsson H, Johansson E, Johansson M, Gustavsson L, et al. Overexpression of functional SLC6A3 in clear cell renal cell carcinoma. *Clinical cancer research : an official journal of the American Association for Cancer Research*. 2017;23(8):2105–15.
85. Kang JU, Koo SH, Kwon KC, Park JW, Kim JM. Gain at chromosomal region 5p15.33, containing TERT, is the most frequent genetic event in early stages of non-small cell lung cancer. *Cancer Genet Cytogenet*. 2008;182(1):1–11.
86. Campa D, Zienolddiny S, Lind H, Ryberg D, Skaug V, Canzian F, et al. Polymorphisms of dopamine receptor/transporter genes and risk of non-small cell lung cancer. *Lung Cancer*. 2007;56(1):17–23.
87. Huang da W, Sherman BT, Lempicki RA. Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nat Protoc*. 2009;4(1):44–57.
88. Han ZD, Zhang YQ, He HC, Dai QS, Qin GQ, Chen JH, et al. Identification of novel serological tumor markers for human prostate cancer using integrative transcriptome and proteome analysis. *Med Oncol*. 2012;29(4):2877–88.
89. Horie M, Miyashita N, Mattsson JSM, Mikami Y, Sandelin M, Brunnstrom H, et al. An integrative transcriptome analysis reveals a functional role for thyroid transcription factor-1 in small cell lung cancer. *J Pathol*. 2018.
90. Zhang Q, Burdette JE, Wang JP. Integrative network analysis of TCGA data for ovarian cancer. *BMC Syst Biol*. 2014;8:1338.
91. Gao JJ, Aksoy BA, Dogrusoz U, Dresdner G, Gross B, Sumer SO, et al. Integrative Analysis of Complex Cancer Genomics and Clinical Profiles Using the cBioPortal. *Sci Signal*. 2013;6(269).
92. Kim Y, Hammerman PS, Kim J, Yoon JA, Lee Y, Sun JM, et al. Integrative and Comparative Genomic Analysis of Lung Squamous Cell Carcinomas in East Asian Patients. *Journal of Clinical Oncology*. 2014;32(2):121–+.
93. Lockwood WW, Chari R, Coe BP, Thu KL, Garnis C, Malloff CA, et al. Integrative Genomic Analyses Identify BRF2 as a Novel Lineage-Specific Oncogene in Lung Squamous Cell Carcinoma. *Plos Med*. 2010;7(7).

94. Noguchi S, Saito A, Horie M, Mikami Y, Suzuki HI, Morishita Y, et al. An integrative analysis of the tumorigenic role of TAZ in human non-small cell lung Cancer. *Clin Cancer Res.* 2014;20(17):4660–72.
95. Brownson RC, Chang JC, Davis JR. Gender and histologic type variations in smoking-related risk of lung cancer. *Epidemiology.* 1992;3(1):61–4.
96. Visbal AL, Williams BA, Nichols FC, Marks RS, Jett JR, Aubry MC, et al. Gender differences in non-small-cell lung cancer survival: an analysis of 4,618 patients diagnosed between 1997 and 2002. *Ann Thorac Surg.* 2004;78(1):209–15.
97. Dresler CM, Fratelli C, Babb J, Everley L, Evans AA, Clapper ML. Gender differences in genetic susceptibility for lung cancer. *Lung Cancer.* 2000;30(3):153–60.
98. Haiman CA, Stram DO, Wilkens LR, Pike MC, Kolonel LN, Henderson BE, et al. Ethnic and racial differences in the smoking-related risk of lung cancer. *New Engl J Med.* 2006;354(4):333–42.

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Ready to submit your research? Choose BMC and benefit from:

- fast, convenient online submission
- thorough peer review by experienced researchers in your field
- rapid publication on acceptance
- support for research data, including large and complex data types
- gold Open Access which fosters wider collaboration and increased citations
- maximum visibility for your research: over 100M website views per year

At BMC, research is always in progress.

Learn more biomedcentral.com/submissions

