Berita Biologi
LIPI
Jurnal Ilmu-ilmu Hayati

10.14203/beritabiologi.v20i1.3991
P-ISSN 0126-1754
E-ISSN 2337-8751

ANALISIS GEN KOMPARATIF KARSINOMA SEL SKUAMOSA PARU-PARU ANTARA INDIVIDU MEROKOK DAN TIDAK MEROKOK

[*Comparative Gene Analysis of Squamous Cell Lung Carcinoma Between Smoking and Non-smoking Individuals*]

Josia Shemuel [1], Priscilla Angelique [2], Evelina Josephine [2], Daniel Ryan Fugaha [1], Vania Gabriela [3], Shaheer Alhussain [2], Arli Aditya Parikesit [1*]✉

[1]Department of Bioinformatics, Indonesia International Institute for Life Sciences (i3L), Jakarta, Indonesia
[2]Department of Biotechnology, Indonesia International Institute for Life Sciences (i3L), Jakarta, Indonesia
[3]Department of Biomedicine, Indonesia International Institute for Life Sciences (i3L), Jakarta, Indonesia
*Email: arli.parikesit@i3l.ac.id

ABSTRACT

Squamous cell lung carcinoma (SCC) is a form of non-small cell lung cancer that commonly arises in the primary airway. The development of SCC is closely linked to changes in squamous cells that line the airways, primarily caused by exposure to tobacco smoke. To gain insights into SCC, bioinformatics techniques have been employed to detect biomarkers and analyze gene expression patterns, utilizing data from the Cancer Genome Atlas (TCGA) database, which was preprocessed for analysis. By employing DESeq2, a differential gene expression analysis method, identified genes showed significant variations in expression between smoking and non-smoking groups among the 11,530 genes examined. Notably, five genes, namely CT45A1, GCGR, TPTE, ABCC2 and PI16, were found to play a significant role in tumor development and were susceptible to under- or over-expression due to smoking. The majority of these genes were found to be underexpressed rather than overexpressed. These identified genes hold potential as biomarkers for tumor development and exhibit a strong correlation between smoking history and the development of SCC. However, a limitation encountered during this analysis was the unavailability of data from normal non-tumor patients, which could have facilitated a more comprehensive analysis of differential gene expression. Furthermore this research gives a deeper implementation regarding the molecular mechanisms and genomics underlying SCC development, identifies differentially expressed genes associated with SCC and smoking history and highlights potential biomarkers that warrant further investigation.

Keywords: Squamous cell lung carcinoma, Smoking history, Gene expression analysis, Database, Biomarkers

ABSTRAK

Karsinoma sel skuamosa paru (SCC) adalah salah satu jenis kanker paru-paru sel non-kecil yang umumnya muncul di saluran udara utama. Perkembangan SCC erat kaitannya dengan perubahan pada sel skuamosa yang melapisi saluran udara, yang terutama disebabkan oleh paparan asap tembakau. Untuk memperoleh wawasan mengenai SCC, teknik bioinformatika telah digunakan untuk mendeteksi biomarker dan menganalisis pola ekspresi gen, dengan memanfaatkan data dari basis data Atlas Genom Kanker (TCGA) yang telah diproses sebelumnya untuk analisis. Dengan menggunakan DESeq2, metode analisis ekspresi gen diferensial, gen-gen yang diidentifikasi menunjukkan variasi signifikan dalam ekspresi antara kelompok perokok dan bukan perokok di antara 11.530 gen yang diperiksa. Secara khusus, lima gen, yaitu CT45A1, GCGR, TPTE, ABCC2 dan PI16, ditemukan berperan penting dalam perkembangan tumor dan rentan terhadap pengungkapan yang kurang atau lebih akibat merokok. Sebagian besar gen-gen ini ditemukan berkurang ekspresinya daripada meningkat. Gen-gen yang diidentifikasi ini memiliki potensi sebagai biomarker untuk perkembangan tumor dan menunjukkan korelasi yang kuat antara riwayat merokok dan perkembangan SCC. Namun, batasan yang dihadapi selama analisis ini adalah tidak tersedianya data dari pasien non-tumor normal, yang dapat memfasilitasi analisis ekspresi gen diferensial yang lebih komprehensif. Selain itu, penelitian ini memberikan implementasi yang lebih dalam mengenai mekanisme molekuler dan genomik yang mendasari perkembangan SCC, mengidentifikasi gen-gen yang diekspresikan secara diferensial yang terkait dengan SCC dan riwayat merokok, serta menyoroti biomarker potensial yang perlu diselidiki lebih lanjut.

Kata Kunci: Karsinoma paru sel skuamosa, Riwayat merokok, Analisis ekspresi gen, Database, Biomarker

INTRODUCTION

Squamous Cell Lung Carcinoma

Squamous cells are the thin, flat cells that line many human organs within the respiratory and digestive tracts, such as the lungs. Squamous cell carcinoma, or SCC of the lung, is sometimes referred to as squamous cell lung cancer. Squamous cell lung cancers frequently develop in the main airway, such as the left or right bronchus, or in the middle of the lung. This type of cell line is a non-small cell lung cancer, or NSCLC. According to Sabbula and Anjum (2021), lung squamous cell carcinoma is the second most prevalent type of NSCLC, most notably in females. It affects approximately 400,000 people worldwide, with a majority being identified as current or heavy smokers (Cardona *et al.*, 2020).

The oncogenesis of SCC is known to be due to the squamous cells lining the airways undergoing change. Tobacco smoke, which contains more than 300 hazardous substances and 40 possible carcinogens, is the main cause of cellular change (American Cancer Society, 2020). Keratinization and/or intercellular bridges are characteristics of transformed squamous cells, which frequently show a high level of mutation frequency. Keratinization is often followed by apoptosis: A process which ultimately leads to the progression of tumor in

*Kontributor Utama
*Diterima: 7 Juli 2023 - Diperbaiki: 11 Agustus 2023 - Disetujui: 2 Oktober 2023

patients with SCC (Park *et al.*, 2017).

Molecular Mechanism and Genomics of Squamous Cell Lung Carcinoma Development

Lung squamous cell carcinoma frequently develops from the bronchial epithelium of larger and more central airways (basal cells), with most cases occurring in the center of the lung. It can be distinguished by the formation of squamous pearls, individual cell keratinization and intercellular bridges (Yeh *et al.*, 2019). According to Zhu *et al.* (2020), once a malignant cell mass has become established, it must be able to cope with various environmental stresses in order to grow and spread to other parts of the body (metastasis). These environmental stresses include hypoxic stroma, immunological responses and hostile local cell types. Lung cancer cells have the ability to manipulate their environment, therefore allowing it to build resistance towards the environmental stresses. They can even use certain harmful impacts into signals that will help them grow, turning foes into allies. The metastasis of lung cancer is carried out by very complicated molecular mechanisms and pathways that are intimately linked to physiological processes of growth and recovery (Perlikos *et al.*, 2013).

According to The Cancer Genome Atlas Research Network (2011), a high overall mutation rate of 8.1 mutations per megabase and noticeable genomic complexity are characteristics of lung squamous cell carcinoma (SQCCs). The somatic TP53 mutation is present in virtually all lung SQCCs, much like it is in high-grade serous ovarian cancer. The CDKN2A/RB1, NFE2L2/ KEAP1/ CUL3, PI3K/AKT, SOX2/TP63/NOTCH1 and PI3K/AKT pathways were also often altered, indicating a common malfunction in cell cycle regulation, response to oxidative stress, apoptotic signaling, and/or squamous cell differentiation. In several instances, pathway changes grouped by expression-subtype, indicating that those subtypes had a biological basis (The Cancer Genome Atlas Research Network, 2012).

Differential Gene Expression in Squamous Cell Lung Carcinoma

Cancers are products of uncontrolled several genetic and epigenetic modifications, which can be induced by a variety of factors (Loiselle *et al.*, 2016). In recent years, high throughput gene expression has shown a huge potential in identifying molecular causes of lung carcinogenesis, including in the identification of biomarkers and gene expression profiles of many lung carcinoma types.

A research conducted by Shriwash *et al.* (2019), identified differentially expressed genes in small and non-small cell lung cancer and found that the differentially expressed genes (DEGs) found in non-small-cell lung cancer were mainly responsible for mitotic nuclear and cell division, enriching KEGG pathway to convert between pentose and glucuronate, as well as increasing cytochrome p450 and enriching gene ontology analysis to glucuronidate xenobiotic. Further testing also showed that 8 genes, including AGR2, ANK3, CSTA, FABP6, FGG, IL33, S100P and TRIM29 are commonly down regulated DEGs in NSCLC, whereas the genes CHL1, CXCL1 and GUSBP8 are commonly upregulated DEGs in NSCLC. Knowing the state of these genes are essential to reach an agreement of which genes affect NSCLC, which would eventually be targets of management and treatment of lung cancer.

The CHL1 gene is cell adhesion molecule L1 like, which serves as a helicase protein in cell cycle, specifically in the interphase stage (Brooker and Berkowitz, 2014). Upregulation of this gene is known to promote metastasis and invasion in many types of cancers. On the other hand, CXCL1 is a cytokine responsible for angiogenesis, tumorigenesis, inflammation and arteriogenesis (Vries *et al.*, 2014). Upregulation of CXCL1 results in a more aggressive tumor as tumorigenesis is promoted, along with enhanced angiogenesis of tumor in metastatic sites. This indicates that CXCL1 partly contributes to the aggressiveness of the cancer.

Onco Informatics Background Method of Squamous Cell Lung Carcinoma

The process of utilizing bioinformatics in oncology was done through the detection of biomarkers (Man *et al.*, 2019). A potential lead as to the biomarkers are circular RNA that has been shown to prove of some diagnostic importance in detecting squamous cell lung carcinoma (LUSC) (Wang *et al.*, 2020).

Such data can be found in the Cancer Genome Atlas (TCGA): A cutting-edge cancer genomics program, characterizing 20,000 primary cancers of 33 different types (National Cancer Institute, 2022). The program as a whole has amassed 2.5 petabytes of data, all available for public usage (The Cancer Genome Atlas (TCGA), 2017). The data available in the TCGA repository include clinical information (e.g. smoking history), molecular analyte data, as well as molecular characterization data. TCGA is included as a part of the Genome Data Commons (GDC) -a repository and cancer knowledge base comprising several cancer genome programs for research (NCI Genomic Data Commons, n.d.). The data derived from TCGA can then be subjected to analysis in R: A programming language purposed for statistical analysis.

Objective of Research

The research aims to identify the relationship between smoking history and squamous cell lung cancer development. Therefore some hypotheses were proposed for this research, which would be no significant correlation were found between smoking history and development of squamous cell lung cancer development. And, there could be significant correlation between smoking history and the development of squamous cell lung carcinoma

MATERIALS AND METHOD

The methodology utilized was based on previous papers with detection of gene expressions through the steps of retrieving the data, processing the data, and creating a differential analysis from the expressions (Bernard and Agustriawan, 2019; Agustriawan *et al.*, 2021; Ivan, Agustriawan *et al.*, 2021; Ivan, Patricia *et al.*, 2021).

Retrieval of Gene Expression and Clinical Data

The first step was to retrieve the gene expressions of lung squamous cell carcinoma patients from the TCGA database (https:// www.cancer.gov/ccg/research/genome-sequencing/ tcga) -a cancer genomics program which contains data on 20,000 primary cancers as well as its 33 different types (National Cancer Institute, 2022). The clinical data of the patients can also be extracted from the repository, providing two separate records of gene expressions and patient data. The TCGA assembler tool was utilized in this process to extract data and later conduct the data preprocessing stage.

Data Preprocessing

Preprocessing of data was required to conduct genomic analysis for a number of reasons. Firstly, the procured data needed to be compatible with the analysis tool. Secondly, upper-quartile normalization methods for gene expression were very liable to bias when a gene with high read counts was introduced, emphasizing the need of filtering. Finally, identifiers were required to calibrate the gene expression data to the clinical data (Rahman *et al.*, 2015).

Patients were referred to by their TCGA ID, wherein codes 01–09 were listed as tumorous patients while 10 and above were listed as normal patients (National Cancer Institute, n.d.). Hence, the extracted TCGA barcodes of the patients can be identified as either tumorous or normal according to a specific portion in the barcode. The barcode was also used to intersect the gene expression data with the clinical data, as they both use TCGA ID to refer to the same patients. The dataset was then split into smoking and non-smoking groups of both tumorous and normal patients. An R dataframe was constructed for each respective group.

Matrix Construction of Gene Data

The resulting data was then stored inside a matrix using the package DESeq2 available on R. Genes were internally normalized by the constructed object by dividing each sample by their mean. The differentially expressed genes were filtered as the following criteria:

1. Over 50% of the samples exhibited expression in the selected gene.
2. Original RNA sequencing data has been normalized via trimmed mean of M-values method.
3. Thresholds of Log-fold change over 1.5 and false discovery rate or adjusted p-value of under 0.01 (Yao *et al.*, 2019).

Differential Gene Expression Analysis of Smoking Group

Smoking group was subjected to differential gene expression analysis with the condition of interest being set to tumor development. The results were plotted using DESeq2's inbuilt function plotMA: A plot purposed for base means and log fold changes.

Differential Gene Expression Analysis of Tumor Development between Tumor and Normal Groups

Another differential gene expression analysis was set for tumor development with the addition of smoking as a factor that accounts for variation. Smoking status was factored in for the purpose of the design. The following design formula was used:

design = ~ smoking + tumor

Overlapping of Abnormally Expressed Genes

Genes found to be abnormally expressed in tumor development between tumor and normal groups were then overlapped with genes abnormally expressed in the tumor development of the smoking group. log 2 fold changes were set to two and three for both analyses, respectively, to select the most differentially expressed gene susceptible to smoke and implicated in tumor development. The genes were then sorted according to their log 2 fold change value.

RESULT

Distribution of Patient Data

The collected genomic and clinical data was divided into smoking and non-smoking groups. Data availability could be seen in the appendix. Table 1 displays the distribution of patients in each group.

Table 1. Distribution of patient data (*Distribusi data pasien*)

Status	Smoker	Non-smoker
Tumor	420	25
Normal	102	0

Differential Gene Expression Analysis between Tumor and Normal Patients in the Smoking Group

Differential gene expression analysis was executed exclusively on smoking patients to discover aberrantly expressed genes in relation to tumor development. To achieve the most aberrant gene expressions from 11,530 genes, the log 2 fold change threshold was set to one. 1,052 genes were identified as underexpressed in smoking tumor patients, compared to a meager 98 overexpressed genes. The adjusted p-value threshold of the analysis was set to 0.01 to reduce instances of false detections of abnormally expressed genes.

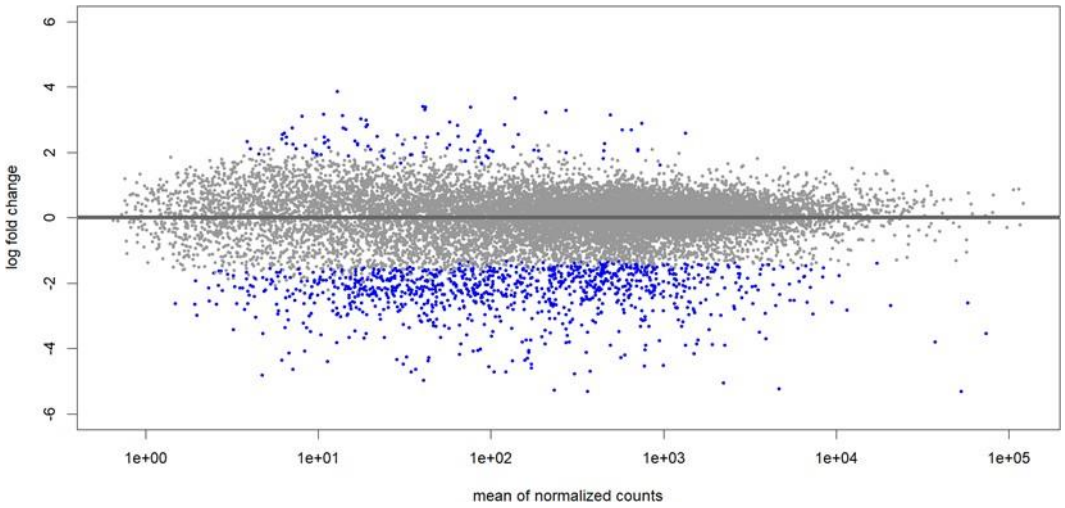

Figure 1. Differentially expressed genes including in smoking patients pertaining to tumor development. Log-fold change threshold was set to one and the p-adjusted value threshold was set to 0.01. Abnormally expressed genes are colored in blue. (*Gen yang diekspresikan secara diferensial, termasuk pada pasien perokok yang berkaitan dengan perkembangan tumor. Ambang perubahan lipat logaritmik diatur menjadi satu dan ambang nilai p yang disesuaikan menjadi 0,01. Gen yang diekspresikan secara abnormal diwarnai biru*).

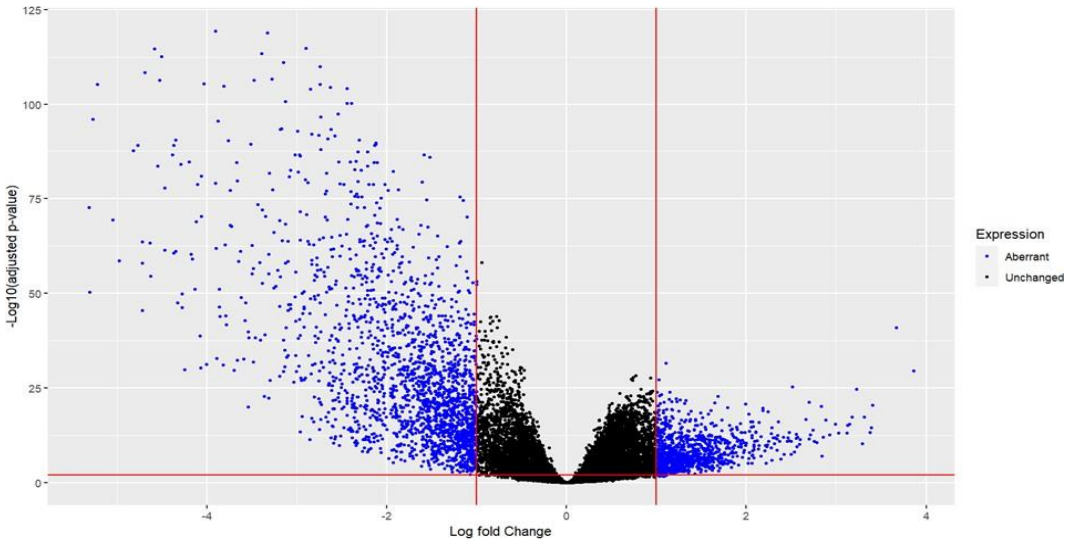

Figure 2. Volcano plot of log fold change and -log10(adjusted p-value). Absolute log fold change values above 1 are marked by the vertical red lines. Adjusted p-value above 0.01 is marked by the horizontal red line. (*Plot gunung berapi dari perubahan lipat logaritmik dan -log10(nilai p yang disesuaikan). Nilai perubahan lipat logaritmik absolut di atas 1 ditandai dengan garis vertikal merah. Nilai p yang disesuaikan di atas 0,01 ditandai dengan garis horizontal merah*).

Differential Gene Expression Analysis of Tumor Development between Smoking and Non-smoking groups

Differential gene expression analysis was executed to assess the distribution of over- and under-expressed genes accounting for the groups smoking and non-smoking. 174 genes were identified as differentially expressed, with an equal number of overexpressed and underexpressed genes (87 genes each).

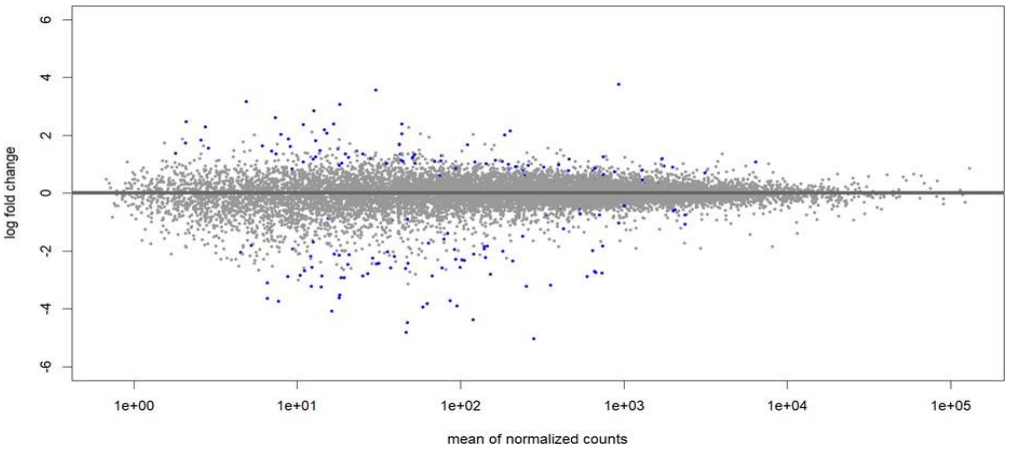

Figure 3. Differentially expressed genes including TP53, KRAS and MYC genes between smoking and non-smoking groups in tumor development. Gene expressions are plotted for the normalized mean number of overlapping reads (counts) in patient and log 2 fold change as X and Y axes respectively. Abnormally expressed genes are colored in blue. P-adjusted value is set to one. (*Gen yang diekspresikan secara diferensial, termasuk gen TP53, KRAS dan MYC, antara kelompok perokok dan bukan perokok dalam perkembangan tumor. Ekspresi gen ditampilkan untuk jumlah rata-rata ternormalisasi dari bacaan (hitungan) yang tumpang tindih pada pasien, dengan log perubahan lipat 2 sebagai sumbu X dan Y masing-masing. Gen yang diekspresikan secara abnormal diwarnai biru. Nilai p yang disesuaikan diatur menjadi satu*).

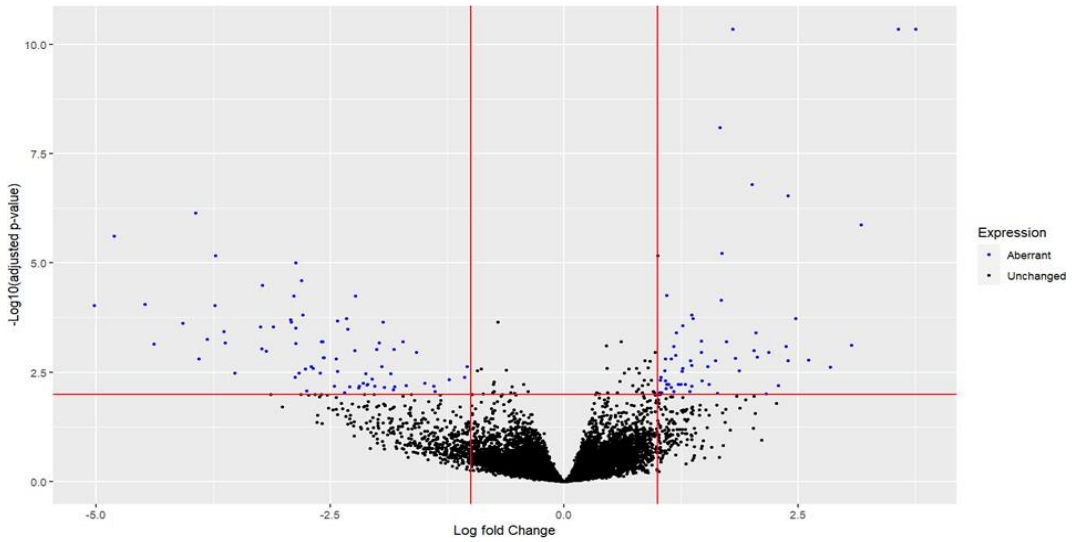

Figure 4. Volcano plot of log fold change and -log10(adjusted p-value). Log fold change values above 1 and under 1 are marked by the vertical red lines. Adjusted p-value above 0.01 is marked by the horizontal red line. (*Plot Gunung Berapi dari perubahan lipat logaritmik dan -log10(nilai p yang disesuaikan). Nilai perubahan lipat logaritmik di atas 1 dan di bawah 1 ditandai dengan garis merah vertikal. Nilai p yang disesuaikan di atas 0,01 ditandai dengan garis merah horizontal*).

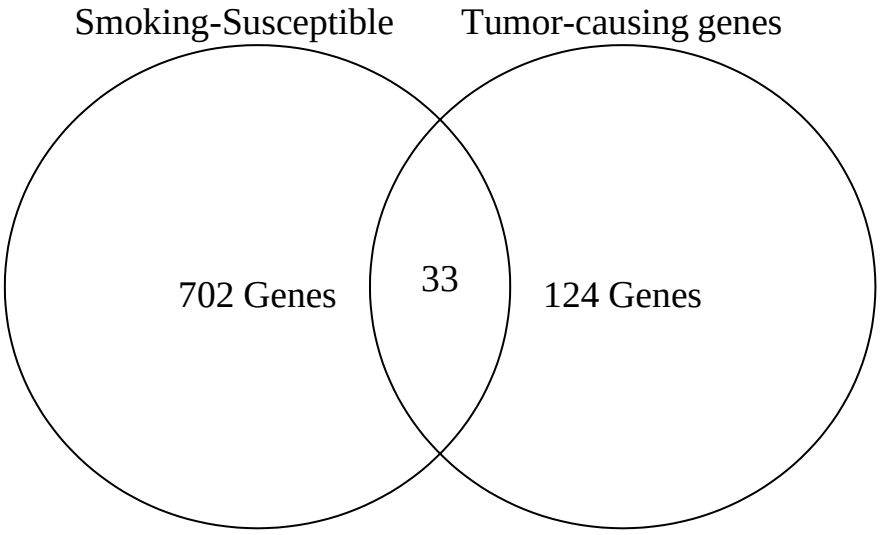

Figure 5. Venn diagram for the number of smoking-susceptible and tumor-causing genes along with their overlap. (*Diagram Venn untuk jumlah gen yang rentan terhadap merokok dan gen penyebab tumor beserta tumpang tindihnya*).

Overlapping of Genes Susceptible to Smoking History and Involved in Tumor Development

Genes identified as abnormally expressed in the previous two analyses were then overlapped to implicate tumor-inducing smoke-susceptible genes. A total of 33 genes were initially identified for their involvement in tumor development and susceptibility to smoking. However, for the purpose of comprehensive analyses of the most differentially expressed genes, the results were filtered down to five by introducing an absolute log fold change threshold value of 2.

Table 2. Differentially expressed genes involved in tumor development and susceptible to smoking. Absolute log 2 fold change minimum value was set to 2 for tumor involvement analysis (Figure 1) and smoking susceptibility analysis (Figure 3). (*Gen yang diekspresikan secara berbeda yang terlibat dalam perkembangan tumor dan rentan terhadap merokok. Nilai mutlak minimum perubahan 2 kali lipat log ditetapkan menjadi 3 untuk analisis keterlibatan tumor (Gambar 1) dan 3 untuk analisis kerentanan merokok (Gambar 3)*).

Gene	Base Mean	Log 2 Fold Change (LFC)	Standard Error of LFC	Wald Test Value	Adjusted P-value
CT45A1	279.92	-5.02	0.95	-5.31	9.27e-5
GCGR	14.09	-2.87	0.60	-4.75	6.87e-4
TPTE	10.41	-2.84	0.67	-4.27	3.31e-3
ABCC2	152.12	-2.81	0.50	-5.61	2.56e-5
PI16	15.18	2.06	0.45	4.55	1.41e-3

Table 2 shows the found genes along with their LFC values, with CT45A1, GCGR, TPTE and ABCC2 being all underexpressed and PI16 being overexpressed.

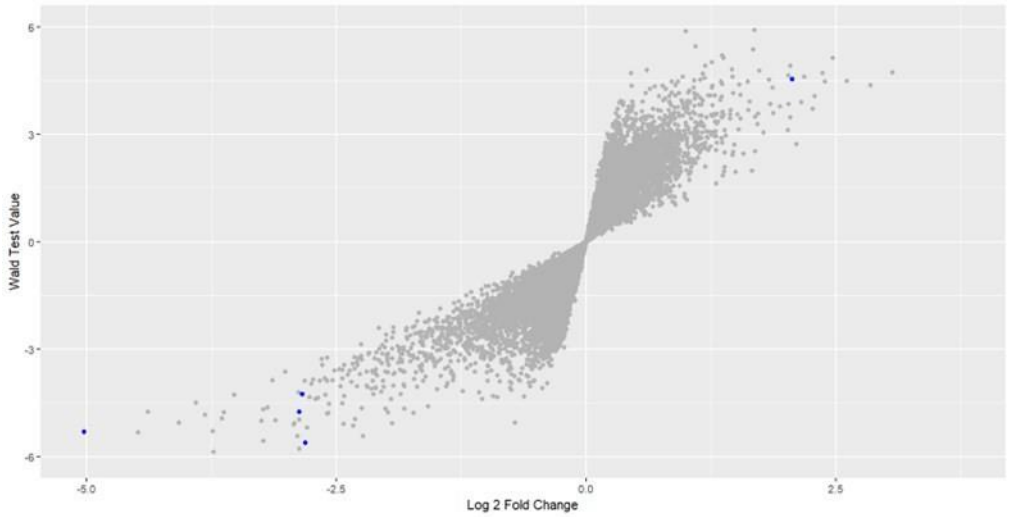

Figure 6. Scatter plot of log 2 fold change vs. Wald test value of genes. The five implicated genes are colored in blue. (*Plot sebar perubahan log 2 kali lipat vs nilai uji gen Wald. Lima gen yang terlibat diwarnai dengan warna biru*).

Figure 6 plots gene expressions based on their Wald test value and log 2 fold change. Wald test value is utilized to represent statistical evidence for the abnormal expression of genes. From the figure, four are noticeably underexpressed, while one is overexpressed.

DISCUSSION

Identification of Genes Susceptible to Smoking

Figure 1 displays the genes abnormally expressed in smoking vs nonsmoking patients. It was found that a large majority of genes were underexpressed and this could be explained by the nitrosamine metabolism pathway being affected by the individual's smoking habit which inhibits the breakdown of harmful compounds in the cigarette (Tu *et al.*, 2018). Furthermore, a small amount of genes were also overexpressed because the DNA repair pathways are constantly trying to repair DNA that may have been altered by the deleterious effects of the cigarette compounds (Laporte *et al.*, 2018). However, the trend is mostly in the underexpression which could cause mechanisms meant to repress tumors or cancer to also be underexpressed, hence causing the increase in likelihood of developing cancer (Yukimatsu *et al.*, 2019). Moreover, the overexpression of the DNA repair pathway could mean the failed attempts at repairing DNA damage, as such promoting further deleterious mutations (Majidinia and Yousefi, 2017).

Identification of the Abnormally Expressed Genes

Numerous genes have been found to express improperly in smokers, and smoking is known to have major effects on gene expression. A notable abnormally expressed gene when an individual smokes is the CYP1A1 gene, which is involved in the metabolism of toxins, especially toxic materials that can be found in cigarettes (Zajda *et al.*, 2017). Another aberrant gene is NRF2, a transcription factor that plays a role in cellular defense against oxidative stress due to smoking. On the other hand, numerous genes may exhibit aberrant patterns of expression when a person develops a tumor. One such gene that is improperly expressed when someone develops a tumor is the TP53 gene. According to Greathouse *et al.* (2018), this gene is a tumor suppressor gene that is crucial in preventing tumor development. Many different malignancies, including lung cancer, exhibit TP53 mutations or aberrant expression. Another example of a gene that is commonly associated with abnormal expression in tumor development is KRAS and MYC, which are oncogenes that regulate cell growth and cell division by maintaining their proliferation (Bouillez *et al.*, 2016). Therefore, the mutation of one or both genes can lead to overexpression that will result in angiogenesis and resistance to cell death. Our analysis shows that the CT45A1, GCGR, TPTE, ABCC2 and PI16 genes are susceptible to abnormal expression when exposed to smoking and are among the most abnormally expressed in the development of squamous cell lung carcinoma.

CT45A1

The CT45A1 gene, also known as cancer/testis antigen family 45 member A1 is present in the testis and various types of cancer. CT45A1 is absent in normal lung tissue and is overexpressed in cell lung cancer, making the gene a marker for lung cancer (Zhou, 2019). This is proven by multiple studies; one is according to the research by Tang *et al.* (2017), where the CT45A1 was positive in multiple lung cancer cells and negative in normal cells. CT45A1 gene overexpression showed an increase in oncogenic and metastatic genes, which promote cell invasion, metastasis, tumorigenesis and stemness (Shang *et al.*, 2014). On the other hand, silencing the gene showed a reduced migration and invasion of the cancer cells. This is due to the fact that CT45A1 diminishes the ERK/ CREB signaling pathway, causing the suppression of the lung cancer invasion, metastasis and proliferation. In addition, the result showed that CT45A1 has the strongest correlation between smoking and tumor growth and the degree of CT45A1 gene expression among other genes that were analyzed.

GCGR

The GCGR gene is a protein-coding gene that is involved in blood glucose regulation. The glucagon receptor is mostly found in the kidneys and liver. Mutations of the GCGR gene cause diseases like Hypoglycemia, Type 2 diabetes, Mahvash diseases (NIH, 2023). Not a lot of studies talk about how smoking impacts the GCGR gene. However, one study discovered the GCGR gene is expressed in epicardial adipose tissue, that is linked to genes that promote FFA (Free Fatty Acid) transport and is lower in former or non-smokers (Alexis *et al.*, 2023). Another study found that a genetic mutation affecting the glucagon gene (GCGR) from smoking can increase the chance of developing type 2 diabetic mellitus (T2DM) (Li *et al.*, 2014). Furthermore, the results showed that the GCGR gene is eight times under-expressed when there is a tumor and four times under-expressed for smoking. This further proves that the expression level of the GCGR gene is heavily correlated with both tumor development and smoking.

TPTE

Lung cancer is only one of the malignancies that exhibit abnormal expression of the transmembrane phosphatase with tensin homology (TPTE) gene. A protein restricted to germ cells called TPTE is excessively expressed in human malignancies of the liver, prostate, and lung (Cao *et al.*, 2020). Only the copy on chromosome 21

appears to be expressed out of the copies of the TPTE gene that are present on chromosomes 13, 15, 21, 22 and Y. The PTEN gene-related transmembrane tyrosine phosphatase known as the TPTE gene may function in signal transduction pathways (Forés-Martos *et al.,* 2015). However, PTEN differs from TPTE in that it has an N-terminal extension made up of three transmembrane domains. PTEN has phosphatase activity, while TPTE does not and it is unknown what biological purpose it serves (Kuemmel *et al.,* 2015). Male individuals with lung adenocarcinoma who do not smoke had noticeably more mutations than female patients including in the TPTE gene (Levy *et al.,* 2023). However, the findings revealed that smoking causes a four times increase in TPTE gene expression, whereas tumors cause an eight times increase. This demonstrates the strong correlation between smoking and tumor growth and the degree of TPTE gene expression.

ABCC2

On chromosome 10q24, the adenosine triphosphate (ATP)-binding cassette subfamily C member 2 (ABCC2) gene, commonly known as the multidrug resistance-associated protein 2 (MRP2), encodes the human canalicular multispecific organic anion transporter (Wu *et al.,* 2018). It was discovered that ABCC2 was overexpressed in a number of human malignancies. Additionally, cisplatin (DDP)-resistant A549 cells (A549/DDP) showed increased ABCC2 expression. Increased ABCC2 expression was substantially related to progression-free survival in lung cancer from 982 samples, according to research by Chen *et al.* (2021). These findings imply that lung cancer patients with greater levels of ABCC2 have a worse prognosis. Changes in ABCC2 in asthmatics who smoke may be both advantageous for controlling oxidative stress and harmful for decreasing the intracellular bioavailability of pharmacological medicines (Aguiar *et al.,* 2019). In addition, according to the findings, smoking increases ABCC2 gene expression by four times, whereas tumors increase it by eight times. The degree of ABCC2 gene expression and tumor development are strongly correlated, as shown by the result.

PI16

PI16 belongs to the superfamily of cysteine-rich secretory proteins, antigen 5 and pathogenesis-related 1 proteins (CAP), also commonly known as peptidase inhibitor 16, PSPBP, or CRISP-1 (Hazell *et al.,* 2016). In a research by Paxson *et al.* (2013), it was shown that lung-derived mesenchymal stromal cells possess extremely high quantities more than 200-fold of a rare gene that is identified as PI16. For pulmonary matrix fibroblasts, PI16 is one of the most highly discriminatory extracellular expressing genes, but its transcript levels are modest (Xie *et al.,* 2018). Therefore according to research by Pradhan *et al.* (2021), fibroblasts exhibit high levels of many genes, including PI16, SFRP1 and IGF2, which are similar to the adventitial fibroblasts observed in mouse lungs. Therefore, the results show that tumors raise PI16 gene expression by eight times, whilst smoking increases it by four times. The results demonstrate a substantial correlation between tumor development and PI16 gene expression levels.

CONCLUSION

Lung squamous cell carcinoma has been associated with substances contained within cigarettes. A differential gene expression analysis was conducted on smoking and non-smoking patients to investigate smoking-induced changes in gene expression and the genes involved in tumor development. 702 genes were found to be susceptible to abnormal expression when an individual smokes, whereas 124 genes were identified to be correlated to tumor development when taking into account an individual's smoking history. 33 of these identified genes overlapped, with CT45A1, GCGR, TPTE, ABCC2 and PI16 being both susceptible to smoking and abnormally expressed in tumor development. The identified genes can serve as biomarkers for tumor development and exhibit a significant correlation between smoking history and tumor development. The limitation encountered in this analysis was the nonavailability of normal non-tumor patients, as such a more extensive differential gene expression analysis could be possible if the data is available. Future research could use a more extensive database and conduct differential gene expression analysis using other more conservative approaches.

ACKNOWLEDGEMENT

The authors would like to thank the Department of Research and Community Service of Indonesia International Institute for Life Sciences (i3L) for supporting this project.

REFERENCE

Aguiar, J.A., Tamminga, A., Lobb, B., Huff, R.D., Nguyen, J.P., Kim, Y., Dvorkin-Gheva, A., Stampfli, M.R., Doxey, A.C and Hirota, J.A., 2019. The impact of cigarette smoke exposure, COPD, or asthma status on ABC transporter gene expression in human airway epithelial cells. *Scientific Reports*, *9*(1). https://doi.org/10.1038/s41598-018-36248-9

Agustriawan, D., Parikesit, A.A., Nurdiansyah, R., Ivan, J and Ramanto, K.N., 2021. Correlation

and transcriptomic analysis revealing potential microRNA-gene interactions associated with breast cancer formation. *Research Journal of Biotechnology*, *16*(2), 16–23.

American Cancer Society., 2020, October 28. *Harmful chemicals in tobacco products*. Www.cancer.org. https://www.cancer.org/cancer/risk-prevention/tobacco/carcinogens-found-in-tobacco-products.html

Bernard, S and Agustriawan, D., 2019. Identification of microRNA targeting cancer gene of colorectal carcinoma in caucasian population. *2019 International Conference on Information and Communications Technology (ICOIACT)*. https://doi.org/10.1109/icoiact46704.2019.8938488

Bouillez, A., Rajabi, H., Pitroda, S., Jin, C., Alam, M., Kharbanda, A., Tagde, A., Wong, K.K and Kufe, D., 2016. Inhibition of MUC1-C suppresses MYC expression and attenuates malignant growth in KRAS mutant lung adenocarcinomas. *Cancer Research*, *76*(6), 1538–1548. https://doi.org/10.1158/0008-5472.CAN-15-1804

Brooker, A.S and Berkowitz, K.M., 2014. The roles of cohesins in mitosis, meiosis, and human health and disease. *Methods in Molecular Biology*, 229–266. https://doi.org/10.1007/978-1-4939-0888-2_11

Cao, F., Wang, Z., Feng, Y., Zhu, H., Yang, M., Zhang, S and Wang, X., 2020. lncRNA TPTEP1 competitively sponges miR-328-5p to inhibit the proliferation of non-small cell lung cancer cells. *Oncology Reports*, *43*(5), 1606–1618.

Cardona, A.F., Ruiz-Patiño, A., Arrieta, O., Ricaurte, L., Zatarain-Barrón, Z.L., Rodriguez, J., Avila, J., Rojas, L., Recondo, G., Barron, F., Archila, P., Sotelo, C., Bravo, M., Zamudio, N., Corrales, L., Martín, C., Rolfo, C., Viola, L., Carranza, H and Vargas, C., 2020. Genotyping squamous cell lung carcinoma in Colombia (geno1.1-clicap). *Frontiers in Oncology*, *10*. https://doi.org/10.3389/fonc.2020.588932

Chen, Y., Zhou, H., Yang, S and Su, D., 2021. Increased ABCC2 expression predicts cisplatin resistance in non-small cell lung cancer. *Cell Biochemistry and Function*, *39*(2), 277-286.

Forés-Martos, J., Cervera-Vidal, R., Chirivella, E., Ramos-Jarero, A and Climent, J., 2015. A genomic approach to study down syndrome and cancer inverse comorbidity: untangling the chromosome 21. *Frontiers in physiology*, *6*, 10.

Gao M, Kong W, Huang Z and Xie Z., 2020. Identification of key genes related to lung squamous cell carcinoma using bioinformatics analysis. Int J Mol Sci. Apr 23;21(8):2994. doi: 10.3390/ijms21082994. PMID: 32340320; PMCID: PMC7215920.

Greathouse, K.L., White, J.R., Vargas, A.J., Bliskovsky, V.V., Beck, J.A., von Muhlinen, N., Polley, E.C., Bowman, E.D., Khan, M.A., Robles, A.I., Cooks, T., Ryan, B.M., Padgett, N., Dzutsev, A.H., Trinchieri, G., Pineda, M.A., Bilke, S., Meltzer, P.S., Hokenstad, A.N and Stickrod, T.M., 2018. Interaction between the microbiome and TP53 in human lung cancer. *Genome Biology*, *19*(1). https://doi.org/10.1186/s13059-018-1501-6

Guttapadu, R., Katte, T., Sayeeram, D., Bhatia, S., Abraham, A.R., Rajeev, K., Amara, A.R.R., Siri, S., Bommana, K., Rasalkar, A.A., Malempati, R., Mustak, M.S., Narayanan, P and Reddy, S.D.N., 2023. Identification of novel biomarkers for lung squamous cell carcinoma. *3 Biotech*, *13*(2). https://doi.org/10.1007/s13205-023-03489-z

Hazell, G.G., Peachey, A.M., Teasdale, J.E., Sala-Newby, G.B., Angelini, G.D., Newby, A.C and White, S.J., 2016. PI16 is a shear stress and inflammation-regulated inhibitor of MMP2. *Scientific reports*, *6*(1), 39553.

Honkala, A.T., Tailor, D and Malhotra, S.V., 2020. Guanylate-Binding Protein 1: An emerging target in Inflammation and Cancer. *Frontiers in Immunology*, *10*. https://doi.org/10.3389/fimmu.2019.03139

Ivan, J., Agustriawan, D., Parikesit, A.A and Nurdiansyah, R., 2021. MiRNA-Regulated HspB8 as potent biomarkers in low-grade gliomas. *Research Journal of Biotechnology*, *16*(1), 17–25.

Ivan, J., Patricia, G and Agustriawan, D., 2021. In silico study of cancer stage-specific DNA methylation pattern in White breast cancer patients based on TCGA dataset. *Computational Biology and Chemistry*, *92*, 107498. https://doi.org/10.1016/j.compbiolchem.2021.107498

Kuemmel, A., Simon, P., Breitkreuz, A., Röhlig, J., Luxemburger, U., Elsäßer, A and Buhl, R., 2015. Humoral immune responses of lung cancer patients against the Transmembrane Phosphatase with TEnsin homology (TPTE). *Lung Cancer*, *90*(2), 334-341.

Lau, S.C.M., Pan, Y., Velcheti, V and Wong, K.K., 2022. Squamous cell lung cancer: Current landscape and future therapeutic options. *Cancer Cell*, *40*(11), 1279–1293. https://doi.org/10.1016/j.ccell.2022.09.018

Laporte, G.A., Leguisamo, N.M., Kalil, A.N and Saffi, J., 2018. Clinical importance of DNA repair in sporadic colorectal cancer. *Critical Reviews in Oncology/Hematology*, *126*, 168–185. https://doi.org/10.1016/

j.critrevonc.2018.03.017

Levy, C., Elkoshi, N., Parikh, S., Mahameed, S., Meidan, A and Rubin, E., 2023. Binary classification machine-learning unveils sex-dependent mutated gene signatures in melanoma. *bioRxiv*, 2023–04.

Li, L., Gao, K., Zhao, J., Feng, T., Yin, L., Jinjin, W., Wang, C., Li, C., Wang, Y., Wang, Q., Zhai, Y., You, H., Hu, D., Wang, B and Hu, D., 2014. Glucagon gene polymorphism modifies the effects of smoking and physical activity on risk of type 2 diabetes mellitus in Han Chinese. *Gene*, *534*(2), 352–355. https://doi.org/10.1016/j.gene.2013.09.121

Liao, Y., Xiao, H., Cheng, M and Fan, X. 2020. Bioinformatics analysis reveals biomarkers with cancer stem cell characteristics in lung squamous cell carcinoma. *Frontiers in Genetics*, *11*. https://doi.org/10.3389/fgene.2020.00427

Lim, S.B., Tan, S.J., Lim, W.T and Lim, C.T., 2018. A merged lung cancer transcriptome dataset for clinical predictive modeling. *Scientific Data*, *5*, 180136. https://doi.org/10.1038/sdata.2018.136

Loiselle, J.J., Roy, J.G and Sutherland, L.C., 2016. RBM5 reduces small cell lung cancer growth, increases cisplatin sensitivity and regulates key transformation-associated pathways. *Heliyon*, *2*(11), e00204. https://doi.org/10.1016/j.heliyon.2016.e00204

Majidinia, M and Yousefi, B., 2017. DNA repair and damage pathways in breast cancer development and therapy. *DNA Repair*, *54*, 22–29. https://doi.org/10.1016/j.dnarep.2017.03.009

Malavazos, A.E., Iacobellis, G., Dozio, E., Basilico, S., Di Vincenzo, A., Dubini, C., Menicanti, L., Corsi, M., Meregalli, C., Ruocco, C., Ragni, M., Secchi, F., Spagnolo, P., Castelvecchio, S., Morricone, L., Buscemi, S., Giordano, A., Goldberger, J.J., Carruba, M.O and Cinti, S., 2023. Human epicardial adipose tissue expresses glucose-dependent insulinotropic polypeptide, glucagon and glucagon-like peptide-1 receptors as potential targets of pleiotropic therapies. *European Journal of Preventive Cardiology*, *30*(8), 680–693. https://doi.org/10.1093/eurjpc/zwad050

Paxson, J.A., Gruntman, A.M., Davis, A.M., Parkin, C.M., Ingenito, E.P and Hoffman, A.M., 2013. Age dependence of lung mesenchymal stromal cell dynamics following pneumonectomy. *Stem cells and development*, *22*(24), 3214–3225.

Perlikos, F., Harrington, K.J and Syrigos, K.N., 2013. Key molecular mechanisms in lung cancer invasion and metastasis: A comprehensive review. *Critical Reviews in Oncology/Hematology*, *87*(1), 1–11. doi:10.1016/j.critrevonc.2012.12

Pradhan, R.N., Krishnamurty, A.T., Fletcher, A.L., Turley, S.J and Müller, S., 2021. A bird's eye view of fibroblast heterogeneity: A pan-disease, pan-cancer perspective. *Immunological Reviews*, *302*(1), 299–320.

Man, J., Zhang, X., Dong, H., Li, S., Yu, X., Meng, L., Gu, X., Yan, H., Cui, J and Lai, Y., 2019. Screening and identification of key biomarkers in lung squamous cell carcinoma by bioinformatics analysis. *Oncology Letters*, *18*, 5185–5196. https://doi.org/10.3892/ol.2019.10873

National Cancer Institute., 2022, May 13. *The cancer genome atlas program (TCGA) - NCI*. Www.cancer.gov. https://www.cancer.gov/ccg/research/genome-sequencing/tcga#:~:text=The%20Cancer%20Genome%20Atlas%20(TCGA

National Cancer Institute. (n.d.). *Sample type codes*. Gdc.cancer.gov. Retrieved May 21, 2023, from https://gdc.cancer.gov/resources-tcga-users/tcga-code-tables/sample-type-codes

NCI genomic data commons. (n.d.). Gdc.cancer.gov. Retrieved June 1, 2022, from https://gdc.cancer.gov/

Park, H.J., Cha, Y.J., Kim, S.H., Kim, A., Kim, E.Y and Chang, Y.S., 2017. Keratinization of lung squamous cell carcinoma is associated with poor clinical outcome. *Tuberculosis and Respiratory Diseases*, *80*(2), 179. https://doi.org/10.4046/trd.2017.80.2.179

Rahman, M., Jackson, L.K., Johnson, W.E., Li, D.Y., Bild, A.H and Piccolo, S.R., 2015. Alternative preprocessing of RNA-sequencing data in the cancer genome atlas leads to improved analysis results. *Bioinformatics*, *31*(22), 3666–3672. https://doi.org/10.1093/bioinformatics/btv377

Sabbula, B.R and Anjum, F., 2021. *Squamous Cell Lung Cancer*. PubMed; StatPearls Publishing. https://www.ncbi.nlm.nih.gov/books/NBK564510/

Shang, B., Gao, A., Pan, Y., Zhang, G., Tu, J., Zhou, Y., Yang, P., Cao, Z., Wei, Q., Ding, Y., Zhang, J., Zhao, Y and Zhou, Q., 2014. CT45A1 acts as a new proto-oncogene to trigger tumorigenesis and cancer metastasis. *Cell Death and Disease*, *5*(6), e1285–e1285. https://doi.org/10.1038/cddis.2014.244

Shriwash, N., Singh, P., Arora, S., Ali, S.M., Ali, S and Dohare, R., 2019. Identification of differentially expressed genes in small and non-small cell lung cancer based on meta-analysis of mRNA. *Heliyon*, *5*(6),

e01707. https://doi.org/10.1016/j.heliyon.2019.e01707

Tang, F., Tang, S., Guo, X., Yang, C and Bi, T., 2017. *C*T45A1 siRNA silencing suppresses the proliferation, metastasis and invasion of lung cancer cells by downregulating the ERK/CREB signaling pathway. *16*(5), 6708–6714. https://doi.org/10.3892/mmr.2017.7466

The Cancer Genome Atlas Research Network., 2011. Integrated genomic analyses of ovarian carcinoma. Nature *474*, 609–615.

The Cancer Genome Atlas Research Network., 2012. Comprehensive genomic characterization of squamous cell lung cancers. Nature, *489*(7417), 519–525. doi:10.1038/nature11404

The cancer genome atlas (TCGA)., 2017. Genome.gov. https://www.genome.gov/Funded-Programs-Projects/Cancer-Genome-Atlas

Tu, C., Xiong, W., Qi, P., Li, X., Guo, C., Xiong, F., Xiang, B., Zhou, M., Liao, Q., Yu, J., Li, Y., Li, X., Li, G and Xiong, W., 2018. Identification of genomic alterations in nasopharyngeal carcinoma and nasopharyngeal carcinoma-derived Epstein–Barr virus by whole-genome sequencing. *Carcinogenesis*, *39*(12), 1517–1528. https://doi.org/10.1093/carcin/bgy108

U.S. National Library of Medicine. (n.d.). *GCGR glucagon receptor [homo sapiens (human)] - gene - NCBI*. National Center for Biotechnology Information. https://www.ncbi.nlm.nih.gov/gene/2642

Vries, M.H.M., Wagenaar, A., Verbruggen, S.E.L., Molin, D.G.M and Post, M.J., 2014. CXCL1 promotes arteriogenesis through enhanced monocyte recruitment into the peri-collateral space. *Angiogenesis*, *18*(2), 163–171. https://doi.org/10.1007/s10456-014-9454-1

Wang, Y., Zhang, H., Wang, J., Li, B and Wang, X., 2020. Circular RNA expression profile of lung squamous cell carcinoma: Identification of potential biomarkers and therapeutic targets. *Bioscience Reports*, *40*(4). https://doi.org/10.1042/bsr20194512

Wu, L., Zhang, W., Jia, S., Zhao, X., Zhou, D., Xu, A., Duan, W., Wu, Z., Li, H., Zheng, S., Nan, Y., Jia, J., Huang, J and Ou, X., 2018. Mutation analysis of the ABCC2 gene in Chinese patients with Dubin-Johnson syndrome. *Experimental and Therapeutic Medicine*, *16*(5), 4201–4206.

Xie, T., Wang, Y., Deng, N., Huang, G., Taghavifar, F., Geng, Y., Liu, N., Kulur, V., Yao, C., Chen, P., Liu, Z., Stripp, B., Tang, J., Liang, J., Noble, P.W and Jiang, D., 2018. Single-Cell deconvolution of fibroblast heterogeneity in mouse pulmonary fibrosis. *Cell Reports*, *22*(13), 3625–3640. https://doi.org/10.1016/j.celrep.2018.03.010

Yao, Y., Zhang, T., Qi, L., Liu, R., Liu, G., Wang, X., Li, J., Li, J and Sun, C., 2019. Competitive endogenous RNA network construction and comparison of lung squamous cell carcinoma in smokers and nonsmokers. *Disease Markers*, *2019*, 1–14. https://doi.org/10.1155/2019/5292787

Yeh, S.J., Chang, C.A., Li, C.W., Wang, L.H.C and Chen, B.S., 2019. Comparing progression molecular mechanisms between lung adenocarcinoma and lung squamous cell carcinoma based on genetic and epigenetic networks: big data mining and genome-wide systems identification. *Oncotarget*, *10(38)*, *3760–3806*. doi:10.18632/oncotarget.26940

Yukimatsu, N., Gi, M., Okuno, T., Fujioka, M., Suzuki, S., Kakehashi, A., Yanagiba, Y., Suda, M., Koda, S., Nakatani, T and Wanibuchi, H., 2019. Promotion effects of acetoaceto-o-toluidide on N-butyl-N-(4-hydroxybutyl) nitrosamine-induced bladder carcinogenesis in rats. *Archives of Toxicology*, *93*(12), 3617–3631. https://doi.org/10.1007/s00204-019-02605-4

Zajda, K., Ptak, A., Rak, A., Fiedor, E., Grochowalski, A., Milewicz, T and Gregoraszczuk, E.L., 2017. Effects of human blood levels of two PAH mixtures on the AHR signalling activation pathway and CYP1A1 and COMT target genes in granulosa non-tumor and granulosa tumor cell lines. *Toxicology*, *389*(1), 1–12. https://doi.org/10.1016/j.tox.2017.07.003

Zhou, D., Sun, Y., Jia, Y., Liu, D., Wang, J., Chen, X., Zhang, Y and Ma, X., 2019. Bioinformatics and functional analyses of key genes in smoking-associated lung adenocarcinoma. *Oncology Letters*, *18*(4). https://doi.org/10.3892/ol.2019.10733

Zhu, T., Bao, X., Chen, M., Lin, R., Zhuyan, J., Zhen, T., Xing, K., Zhou, W and Zhu, S., 2020. Mechanisms and future of non-small cell lung cancer metastasis. *Frontiers in Oncology*, *10*. https://doi.org/10.3389/fonc.2020.585284

Appendix A
Data Availability

The data for this study was acquired from clinical and sequencing reads dataset found in the Genomic Data Commons portal at https://portal.gdc.cancer.gov/projects/TCGA-LUSC.

Appendix B
Genes Susceptible to Smoking History and Involved in Tumor Development

33 genes were marked as statistically significant when overlapping genes susceptible to smoking history and genes involved in tumor development. Absolute log fold change threshold was set to 2 and p-adjusted value threshold was set to 0.1.

Table 2. Differentially expressed genes involved in tumor development and susceptible to smoking. (*Gen yang diekspresikan secara diferensial yang terlibat dalam perkembangan tumor dan rentan terhadap merokok*).

Gene Name| Entrez ID	baseMean	log2FoldChange	lfcSE	stat	pvalue	padj
CT45A1|541466	279.9182953	-5.024736212	0.9471185875	-5.305287298	1.12E-07	9.27E-05
CTAG1B|1485	118.8325932	-4.383071499	0.9236473099	-4.745395187	2.08E-06	0.000703536 8907
FAM133A|286499	58.78405481	-3.938665031	0.6244618787	-6.307294593	2.84E-10	7.13E-07
MPPED1|758	86.10200162	-3.726934562	0.635944196	-5.860474214	4.62E-09	6.76E-06
ZMAT4|79698	14.05620375	-3.248539127	0.6495375026	-5.001311108	5.69E-07	0.000286010 234
KRT31|3881	253.2512138	-3.228075879	0.5809059312	-5.556968359	2.75E-08	3.22E-05
GABRA5|2558	47.49470512	-3.13720967	0.8098817485	-3.873663873	0.00010721 12576	0.010187967 08
ASCL1|429	19.38218755	-2.924934225	0.5723308791	-5.110565115	3.21E-07	0.000201665 672
NOS2|4843	594.5922914	-2.889522678	0.533004102	-5.421201576	5.92E-08	5.78E-05
GCGR|2642	14.08737515	-2.868083179	0.603293199	-4.75404527	1.99E-06	0.000687296 5184
TPTE|7179	10.4066142	-2.83771475	0.6653051454	-4.265283036	2.00E-05	0.0033111589
ABCC2|1244	152.1177768	-2.810677597	0.5011235272	-5.608752023	2.04E-08	2.56E-05
ADH1C|126	730.5137121	-2.766174535	0.6377459319	-4.337424038	1.44E-05	0.002667761 918
FDCSP|260436	670.8994225	-2.750932519	0.6930008875	-3.969594511	7.20E-05	0.008326795 478
SLC38A11|151258	48.94023374	-2.729373761	0.7074602254	-3.85798899	0.00011432 37966	0.010522577 72
ESRG|790952	80.92901464	-2.638862984	0.7670632822	-3.440215489	0.00058125 11987	0.029878351 09
PNMA5|114824	18.64823611	-2.593201805	0.6678803014	-3.882734375	0.00010328 83798	0.010153479 51
NPSR1|387129	12.28481686	-2.573802015	0.5680888984	-4.530632481	5.88E-06	0.001482478 118
OLFM4|10562	99.27157862	-2.569994003	0.5673480391	-4.52983676	5.90E-06	0.001482478 118

Table 2. Differentially expressed genes involved in tumor development and susceptible to smoking. (*Gen yang diekspresikan secara diferensial yang terlibat dalam perkembangan tumor dan rentan terhadap merokok*).

Gene Name| Entrez ID	baseMean	log2FoldChange	lfcSE	stat	pvalue	padj
IL19|29949	7.000651239	-2.52020797	0.6998667048	-3.600982805	0.0003170165454	0.02090359021
OTX2|5015	17.46249179	-2.471578423	0.6939874485	-3.561416605	0.0003688592713	0.02251578469
SLC16A12|387700	16.67611386	2.390528327	0.3695955282	6.467957929	9.93E-11	2.91E-07
IRS4|8471	39.82876316	-2.368153367	0.6592296254	-3.592304222	0.0003277668848	0.02132570147
FABP7|2173	72.18377576	-2.316084045	0.6538976247	-3.541967363	0.0003971546101	0.02374822464
PMP2|5375	6.206885801	-2.287300108	0.6647387604	-3.440900763	0.0005797811315	0.02987835109
FAM71F1|84691	12.0209807	-2.264981688	0.6136848926	-3.690789386	0.0002235591898	0.01688805472
NELL1|4745	190.4193433	-2.243476585	0.682379142	-3.287727375	0.001009995932	0.04261167577
SLC6A10P|386757	141.697064	-2.231599714	0.4112647852	-5.426187202	5.76E-08	5.78E-05
PIWIL2|55124	18.04562397	-2.149526162	0.522816928	-4.111431835	3.93E-05	0.005666134443
CPB2|1361	10.32776629	-2.093714093	0.5600134559	-3.738685331	0.0001849850959	0.01491760544
PI16|221476	15.18184549	2.064601875	0.454082149	4.54675851	5.45E-06	0.001408428047
KRT83|3889	5.615102587	-2.045815355	0.6033732598	-3.390629799	0.0006973222503	0.03405256989
UGT1A8|54576	57.72882731	-2.013447296	0.6824030739	-2.950524951	0.003172344206	0.07899406678