

## ANALISIS GEO2R

### I. Pendahuluan

Kanker paru-paru memiliki angka tertinggi di negara-negara berkembang, di mana kebiasaan merokok paling banyak terjadi, dengan variasi angka kejadian lebih dari 20 kali lipat antarwilayah. Sementara kanker prostat merupakan kanker yang paling umum di kalangan pria di 104 negara, kanker paru-paru menjadi yang paling umum di 37 negara, termasuk Rusia, Tiongkok, serta sebagian besar wilayah Eropa Timur, Timur Tengah, dan Asia Tenggara. Perkembangan kanker paru ditandai oleh perubahan ekspresi gen yang menyebabkan gangguan regulasi siklus sel, proliferasi tidak terkendali, serta perubahan jalur pensinyalan seluler (Thandara dkk. 2021).

Analisis *transcriptomics* memungkinkan identifikasi gen-gen yang mengalami perubahan ekspresi antara jaringan kanker dan jaringan normal. Identifikasi *Differentially Expressed Genes* (DEGs) dapat membantu memahami mekanisme molekuler yang mendasari perkembangan kanker paru.

Tujuan analisis ini adalah untuk mengidentifikasi gen-gen yang berbeda secara signifikan antara jaringan kanker paru dan jaringan paru normal menggunakan data publik dari GEO.

### II. Metode

#### 2.1. Dataset

Dataset yang digunakan adalah GSE10072, diperoleh dari *Gene Expression Omnibus* (NCBI GEO).

Karakteristik dataset:

- Platform: GPL96 (*Affymetrix Human Genome U133A Array*)
- Total sampel: 107 sampel (58 sampel tumor paru, 49 sampel jaringan paru normal)
- Organisme: *Homo sapiens*

Analisis dilakukan untuk membandingkan ekspresi gen antara jaringan kanker paru dan jaringan normal. Replikasi dilakukan dengan menjalankan ulang analisis GEO2R dengan fitur “*Reanalyze*”.

NCBI

**GEO**  
Gene Expression Omnibus

HOME | SEARCH | SITE MAP | GEO Publications | FAQ | MIAME | Email GEO | Not logged in | Login ?

NCBI > GEO > Accession Display ?

Scope: Self Format: HTML Amount: Quick GEO accession: GSE10072 | GO

**Series GSE10072** Query DataSets for GSE10072

Status	Public on Feb 20, 2008
Title	Gene expression signature of cigarette smoking and its role in lung adenocarcinoma development and survival
Organism	<a href="#">Homo sapiens</a>
Experiment type	Expression profiling by array
Summary	Tobacco smoking is responsible for over 90% of lung cancer cases, and yet the precise molecular alterations induced by smoking in lung that develop into cancer and impact survival have remained obscure. We performed gene expression analysis using HG-U133A Affymetrix chips on 135 fresh frozen tissue samples of adenocarcinoma and paired noninvolved lung tissue from current, former and never smokers, with biochemically validated smoking information. ANOVA analysis adjusted for potential confounders, multiple testing procedure, Gene Set Enrichment Analysis, and GO-functional classification were conducted for gene selection. Results were confirmed in independent adenocarcinoma and non-tumor tissues from two studies. We identified a gene expression signature characteristic of smoking that includes cell cycle genes, particularly those involved in the mitotic spindle formation (e.g., NEK2, TTK, PRC1). Expression of these genes strongly differentiated both smokers from non-smokers in lung tumors and early stage tumor tissue from non-tumor tissue ( $p<0.001$ and fold-change $>1.5$ , for each comparison), consistent with an important role for this pathway in lung carcinogenesis induced by smoking. These changes persisted many years after smoking cessation. NEK2 ( $p<0.001$ ) and TTK ( $p=0.002$ ) expression in the noninvolved lung tissue was also associated with a 3-fold increased risk of mortality from lung adenocarcinoma in smokers. Our work provides insight into the smoking-related mechanisms of lung neoplasia, and shows that the very mitotic genes known to be involved in cancer development are induced by smoking and affect survival. These genes are candidate targets for chemoprevention and treatment of lung cancer in smokers.
Keywords:	comparative genomics

## 2.2. Analisis DEG Menggunakan GEO2R

Analisis dilakukan menggunakan tool berbasis web GEO2R yang tersedia pada halaman dataset GEO.

Parameter utama yang digunakan:

a. Pembagian kelompok

- Group 1: Tumor (kanker paru) sebanyak 58 sampel
- Group 2: Normal (jaringan paru sehat) sebanyak 49 sampel

The screenshot shows the GEO2R interface with the following details:

- NCBI > GEO > GEO2R > GSE10072**
- GEO accession:** GSE10072
- Sample Groups:**
  - Lung Cancer (58 samples)
  - Healthy (49 samples)
- Table Headers:** Source name, Gender, Age at Diagnosis, Cigarette Smoking Status, Stage.
- Data Rows:** A sample table showing 107 rows of data, with columns for Source name, Gender, Age at Diagnosis, Cigarette Smoking Status, and Stage.

## b. Metode statistik

- Metode: limma (*Linear Models for Microarray Data*)
- Koreksi *multiple testing*: Benjamini-Hochberg *False Discovery Rate* (FDR)

## c. Kriteria signifikan

Gen dianggap signifikan apabila memenuhi kriteria:

- *Adjusted p-value* (adj.P.Val)  $< 0.05$
- $|\log_2 \text{Fold Change}| (\text{logFC}) \geq 1$

The screenshot shows the GEO2R Options tab with the following settings:

- Apply adjustment to the P-values.** More...  
Selected: Benjamini & Hochberg (False discovery rate)
- Apply log transformation to the data.** More...  
Selected: Auto-detect
- Category of Platform annotation to display on results.**  
Selected: NCBI generated
- Plot displays.** More...  
Significance level cut-off: 0.05  
Log 2 fold change threshold: 0
- Force normalization.** More...  
Selected: No
- Volcano and Mean-difference plot contrasts (select up to 5).**  
0 selected (clear)  
Lung Cancer vs Healthy

## III. Hasil dan Pembahasan

Berdasarkan kriteria:  $\text{adj.P.Val} < 0.05$  dan  $|\log_2 \text{FC}| \geq 1$ , telah diperoleh sejumlah gen yang berbeda secara signifikan antara jaringan kanker paru dan jaringan normal.

Secara umum ditemukan gen yang mengalami peningkatan ekspresi (*up-regulated*) pada kanker paru dan gen yang mengalami penurunan ekspresi (*down-regulated*) pada kanker paru.

Down regulated		
LogFC	Gene symbol	Gene title
-4.4174695	AGER	advanced glycosylation end-product specific receptor
-4.0471792	SFTPC	surfactant protein C
-3.8385413	FABP4	fatty acid binding protein 4
-3.7097964	CYP4B1	cytochrome P450 family 4 subfamily B member 1
-3.6867095	WIF1	WNT inhibitory factor 1

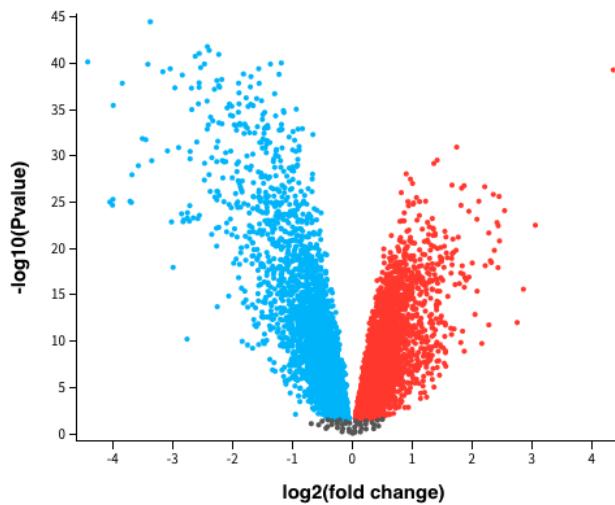
  

Up regulated		
LogFC	Gene symbol	Gene title
4.3644151	SPP1	secreted phosphoprotein 1
3.061522	COL11A1	collagen type XI alpha 1 chain
2.8620356	MMP1	matrix metallopeptidase 1
2.7583995	SPINK1	serine peptidase inhibitor, Kazal type 1
2.5483627	GREM1	gremlin 1, DAN family BMP antagonist

Gen-gen dengan nilai logFC positif tinggi menunjukkan peningkatan aktivitas gen yang berkaitan dengan proses tumorigenesis, sedangkan gen dengan logFC negatif menunjukkan penurunan fungsi gen normal jaringan paru.

Gen-gen yang mengalami peningkatan ekspresi (*up-regulated*) pada kanker paru umumnya berhubungan dengan proliferasi sel, regulasi siklus sel, replikasi DNA, dan jalur pensinyalan pertumbuhan sel. Hal ini konsisten dengan karakteristik biologis kanker, yaitu proliferasi sel yang tidak terkontrol dan gangguan regulasi *checkpoint* siklus sel. Sebaliknya, gen-gen yang mengalami penurunan ekspresi (*down-regulated*) sering kali berkaitan dengan diferensiasi sel normal, fungsi fisiologis jaringan paru, dan homeostasis jaringan. Penurunan ekspresi gen-gen tersebut menunjukkan hilangnya fungsi normal jaringan akibat transformasi sel menjadi sel kanker.

**Volcano plot**  
**GSE10072: Gene expression signature of**  
**cigarette smoking and its role...**  
**Lung Cancer vs Healthy, Padj<0.05**



#### IV. Kesimpulan

Secara keseluruhan, hasil analisis DEG mendukung pemahaman bahwa kanker paru melibatkan aktivasi jalur proliferatif dan gangguan regulasi gen yang mempertahankan fungsi normal jaringan.

#### References

Thandra KC, Barsouk A, Saginala K, Aluru JS, Barsouk A. Epidemiology of lung cancer. Contemp Oncol (Pozn). 2021;25(1):45-52. doi: 10.5114/wo.2021.103829. Epub 2021 Feb 23. PMID: 33911981; PMCID: PMC8063897.