

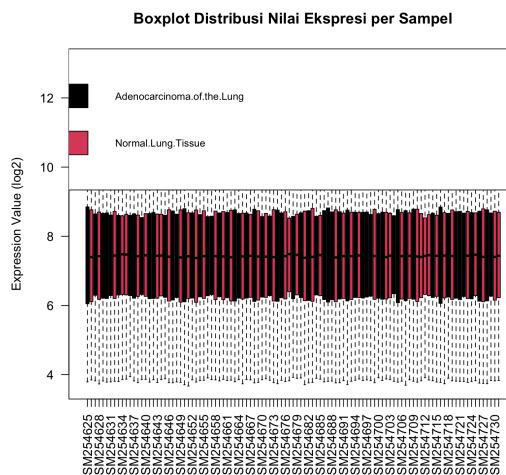
Dasar Analisis DEG menggunakan Bahasa R: Analisis Kanker Paru (GSE10072)

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	<i>Homo sapiens</i>	
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	

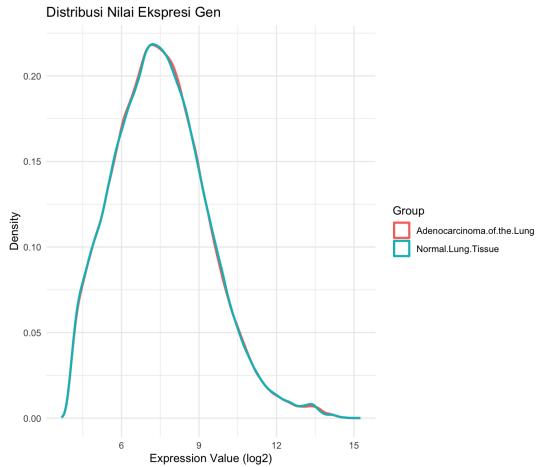
Kanker paru-paru merupakan salah satu penyakit neoplastik dengan angka mortalitas yang sangat tinggi, yang secara etiologis menunjukkan hubungan kuat dengan paparan asap rokok dan dilaporkan berkontribusi pada lebih dari 90% kasus. Walaupun bukti epidemiologis yang menghubungkan karsinogen tembakau dengan kanker paru sudah sangat kuat, mekanisme pada level molekuler yang memicu perkembangan tumor dan menentukan tingkat kelangsungan hidup pasien tetap menjadi area yang belum sepenuhnya tereksplorasi. Penelitian ini bertujuan untuk melakukan pemetaan profil transkriptomik menggunakan platform microarray Affymetrix HG-U133A pada 135 sampel jaringan tumor dan non-tumor guna mengidentifikasi gene expression signature yang dipicu oleh aktivitas merokok (Landi et al., 2008). Fokus utama analisis diarahkan pada disregulasi gen pengendali siklus sel, khususnya komponen spindel mitotik seperti NEK2, TTK, dan PRC1, yang terbukti tetap persisten pasca-henti rokok dan berperan krusial dalam progresivitas tumor serta risiko mortalitas. Identifikasi kandidat biomarker ini menjadi esensial dalam pengembangan strategi kemoprevensi dan optimasi terapi presisi bagi pasien kanker paru, terutama pada kelompok perokok.

Analisis data transkriptomik dilakukan menggunakan bahasa pemrograman R (v4.x2026.01.1+403) dengan memanfaatkan repositori Bioconductor. Data ekspresi gen diakuisisi dari database *Gene Expression Omnibus* (GEO) dengan kode akses GSE10072 melalui GEOquery. Matriks ekspresi yang diperoleh kemudian melalui tahap preprocessing berupa transformasi log2 untuk menstabilkan varians dan memastikan distribusi data memenuhi asumsi model linier. Anotasi probe Affymetrix dikonversi menjadi Gene Symbol yang valid menggunakan database spesifik platform hgu133a.db. Identifikasi *Differentially Expressed Genes* (DEGs) dilakukan dengan

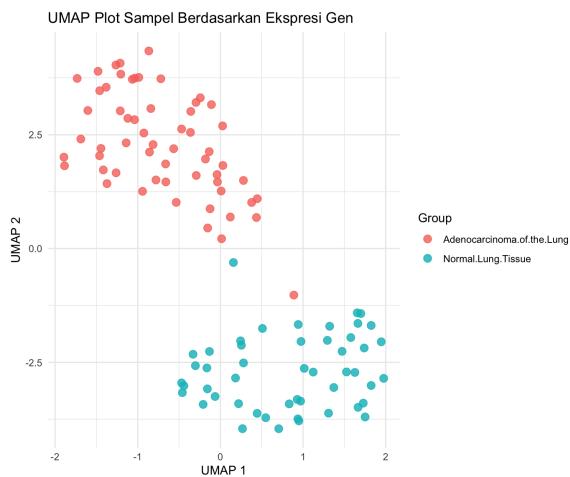
pendekatan model linier melalui paket LIMMA (*Linear Models for Microarray Data*). Estimasi varians distabilkan menggunakan metode Empirical Bayes (EB). Untuk meminimalisir temuan positif palsu akibat pengujian hipotesis simultan, nilai signifikansi dikoreksi menggunakan prosedur *False Discovery Rate* (FDR) dengan ambang batas *adjusted p-value* < 0,01. Visualisasi distribusi global dan pemisahan sampel dilakukan menggunakan analisis dimensi rendah UMAP (*Uniform Manifold Approximation and Projection*). Selanjutnya, profil ekspresi gen dipetakan secara spasial melalui *Volcano Plot* (dengan kriteria $\log_2 \text{Fold Change} > 1$) dan analisis klaster hierarki ditampilkan dalam bentuk *Heatmap* terhadap 50 DEGs teratas menggunakan metrik jarak Euclidean.



Visualisasi boxplot distribusi nilai ekspresi pada dataset GSE10072 menunjukkan profil transkriptomik yang telah ternormalisasi secara adekuat, ditandai dengan keselarasan garis median yang konsisten di seluruh sampel adenocarcinoma dan jaringan paru normal pada skala log2. Homogenitas sebaran data serta rentang interkuartil yang seragam mengindikasikan bahwa variansi teknis antarsampel telah diminimalkan melalui prosedur transformasi data, sehingga memastikan stabilitas varians dan komparabilitas statistik yang diperlukan untuk analisis ekspresi gen diferensial (*Differentially Expressed Genes* atau DEGs) tahap lanjut. Kondisi distribusi yang stabil dan terstandar ini memberikan basis data yang reliabel untuk mengidentifikasi signifikansi biologis dari disregulasi molekuler pada sel kanker tanpa terdistorsi oleh bias sistematis atau artefak eksperimental.

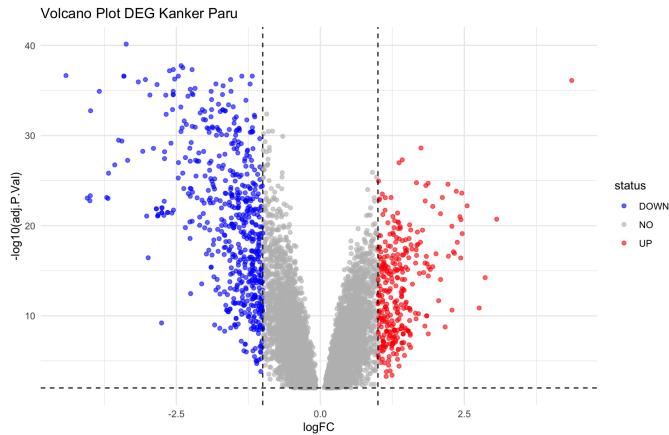


Berdasarkan visualisasi density plot yang dihasilkan, dataset GSE10072 menunjukkan profil transkriptomik yang telah ternormalisasi secara adekuat melalui logaritma , ditandai dengan keselarasan nilai median serta homogenitas rentang interkuartil yang konsisten di seluruh kohort adenokarsinoma dan jaringan paru normal. Kurva distribusi pada density plot menunjukkan tumpang tindih yang hampir sempurna antarkelompok biologis, yang mengonfirmasi bahwa stabilitas varians telah tercapai dan eliminasi bias teknis sistematis telah dilakukan secara efektif. Kondisi distribusi yang seragam dan terstandar ini menjamin komparabilitas statistik yang tinggi, sehingga menyediakan fondasi data yang reliabel untuk analisis ekspresi gen diferensial (DEGs) tahap lanjut guna memastikan bahwa signifikansi molekuler yang ditemukan nantinya murni merepresentasikan patofisiologi kanker tanpa interferensi artefak eksperimental.

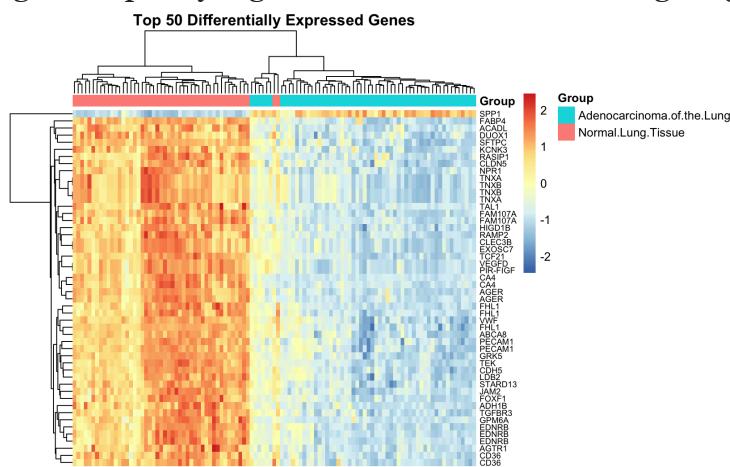


Berdasarkan visualisasi kontrol kualitas dan reduksi dimensi pada dataset GSE10072, hasil transformasi log2 menunjukkan normalisasi yang adekuat, ditandai dengan keselarasan median pada boxplot serta distribusi nilai ekspresi yang tumpang tindih secara simetris pada density plot. Stabilitas varians ini mengonfirmasi eliminasi bias teknis sistematis, sehingga data layak untuk komparasi statistik tingkat lanjut. Lebih lanjut, analisis UMAP mengungkapkan separasi spasial yang distingatif antara

klaster Adenocarcinoma dan jaringan paru normal, yang merefleksikan perbedaan profil transkriptomik global yang signifikan secara biologis. Pemisahan kelompok yang tegas ini memvalidasi integritas data dan memperkuat potensi identifikasi gen diferensial (DEGs) sebagai kandidat biomarker yang reliabel bagi kanker paru.



Hasil analisis transkriptomik pada dataset GSE10072 menunjukkan profil data yang terstandardisasi secara optimal, tercermin dari distribusi nilai ekspresi yang homogen pada boxplot dan density plot pasca-normalisasi log2. Reduksi dimensi menggunakan UMAP mengonfirmasi separasi spasial yang distingtif antara klaster Adenocarcinoma dan jaringan paru normal, yang merefleksikan divergensi biologis global yang signifikan antara kedua kelompok sampel. Lebih lanjut, Volcano Plot secara statistik mengidentifikasi populasi gen fungsional yang mengalami upregulasi (merah) dan downregulasi (biru) secara signifikan ($P_{adj} < 0,01$; $|\log_2 FC| > 1$), memberikan basis data yang reliabel untuk identifikasi kandidat biomarker serta pemahaman mekanisme molekuler karsinogenesis paru yang diinduksi oleh faktor etiologi lingkungan.



Analisis terhadap dataset GSE10072 menunjukkan profil transkriptomik yang telah ternormalisasi secara optimal, tercermin dari distribusi nilai ekspresi yang homogen pada boxplot dan density plot pasca-transformasi log2. Teknik reduksi dimensi menggunakan UMAP mendemonstrasikan separasi spasial yang tegas antara klaster Adenocarcinoma dan jaringan paru normal, yang mengonfirmasi adanya

divergensi biologis global yang signifikan antara kedua kelompok sampel. Analisis ekspresi gen diferensial (DEGs) melalui Volcano Plot dan Heatmap berhasil mengidentifikasi disregulasi masif pada gen-gen krusial, seperti upregulasi SPP1 dan downregulasi AGER serta FABP4, yang secara konsisten membedakan jaringan tumor dari kontrol dengan signifikansi statistik yang tinggi ($P_{adj} < 0,01$; $|\log_2 FC| > 1$). Keseluruhan data ini memvalidasi integritas molekuler dataset tersebut sebagai basis yang reliabel untuk studi karsinogenesis dan identifikasi biomarker spesifik pada adenokarsinoma paru.

Berdasarkan analisis transkriptomik pada dataset GSE10072, dapat disimpulkan bahwa penggunaan platform microarray Affymetrix HG-U133A yang dikombinasikan dengan pendekatan bioinformatika menggunakan paket limma berhasil mengidentifikasi profil molekuler yang membedakan adenokarsinoma paru dari jaringan normal secara signifikan. Hasil normalisasi log₂ dan reduksi dimensi via UMAP memvalidasi integritas data dengan menunjukkan separasi klaster yang tegas serta distribusi ekspresi yang homogen. Identifikasi gene expression signature melalui Volcano Plot dan Heatmap mengungkapkan disregulasi masif pada gen-gen krusial, termasuk upregulasi signifikan pada onkogen potensial seperti SPP1 serta induksi gen pengendali siklus sel dan spindel mitotik (NEK2, TTK, PRC1) yang tetap persisten pasca-henti rokok. Penemuan kandidat biomarker ini, yang juga mencakup downregulasi gen AGER dan FABP4, memberikan fondasi molekuler yang reliabel bagi pengembangan strategi kemoprevensi dan optimasi terapi presisi untuk menurunkan risiko mortalitas pada pasien kanker paru, khususnya kelompok perokok.

Referensi

- Landi, M. T., Dracheva, T., Rotunno, M., Figueroa, J. D., Liu, H., Dasgupta, A., Mann, F. E., Fukuoka, J., Hames, M., Bergen, A. W., Murphy, S. E., Yang, P., Pesatori, A. C., Consonni, D., Bertazzi, P. A., Wacholder, S., Shih, J. H., Caporaso, N. E., & Jen, J. (2008). Gene Expression Signature of Cigarette Smoking and Its Role in Lung Adenocarcinoma Development and Survival. *PLoS ONE*, 3(2), e1651. <https://doi.org/10.1371/journal.pone.0001651>