

## ANALISIS EKSPRESI GEN MENGGUNAKAN R

### 1. Pendahuluan

Kanker paru-paru merupakan penyebab utama kematian akibat kanker di dunia, di mana 90% kasus dipicu oleh kebiasaan merokok. Meskipun demikian, perubahan molekuler spesifik yang disebabkan oleh rokok dalam perkembangan kanker dan pengaruhnya terhadap kelangsungan hidup pasien masih belum sepenuhnya dipahami.

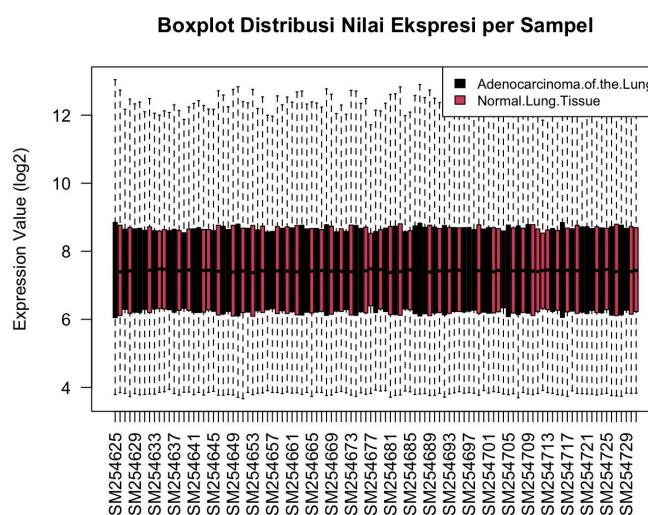
Analisis *transcriptomics* menggunakan *microarray* dilakukan pada sampel penderita adenokarsinoma untuk mengidentifikasi *Differentially Expressed Genes* (DEGs) antara kelompok perokok aktif, mantan perokok, dan bukan perokok. Identifikasi gen-gen ini bertujuan untuk memahami mekanisme molekuler karsinogenesis serta dampaknya terhadap prognosis pasien kanker paru.

Tujuan analisis ini adalah untuk mengidentifikasi gen-gen yang memiliki perbedaan ekspresi signifikan antara jaringan tumor dan jaringan normal guna mengonfirmasi peran "jejak genetik" merokok dalam perkembangan kanker paru-paru.

### 2. Metode

Metode yang digunakan adalah analisis ekspresi gen dari dataset **GSE10072** menggunakan *software R* dan RStudio.

### 3. Hasil dan Interpretasi

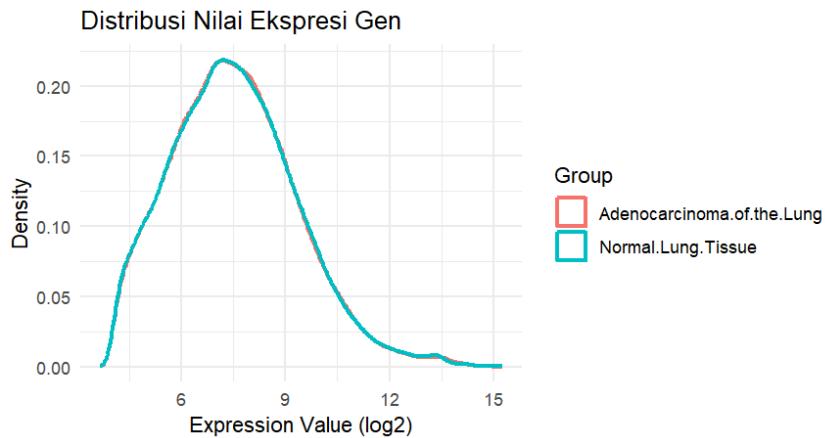


Gambar 1. Boxplot Distribusi Nilai Ekspresi Per Sampel

Visualisasi ini menunjukkan distribusi nilai intensitas ekspresi (dalam skala log2) untuk setiap sampel yang diuji.

- Normalisasi Data: Seluruh sampel memiliki nilai median dan rentang interkuartil yang sejajar dan konsisten. Hal ini mengindikasikan bahwa data telah melalui proses normalisasi yang baik, sehingga layak untuk dibandingkan antar grup.

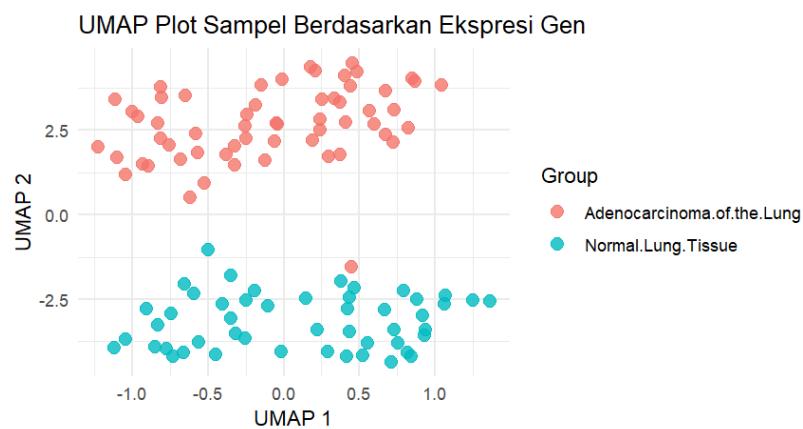
- Konsistensi: Tidak ditemukan adanya sampel *outlier* yang ekstrem, baik pada kelompok adenocarcinoma (hitam) maupun jaringan normal (merah muda).



Gambar 2. Distribusi Nilai Ekspresi Gen

Grafik densitas ini menggambarkan sebaran frekuensi nilai ekspresi gen secara keseluruhan.

- Kesamaan Distribusi: Garis merah (adenocarcinoma) dan biru (normal) hampir berhimpit secara sempurna. Ini menunjukkan bahwa secara global, profil distribusi ekspresi gen pada kedua kelompok memiliki karakteristik statistik yang serupa.
- Puncak Ekspresi: Konsentrasi ekspresi gen terbanyak berada pada nilai log2 sekitar 7 hingga 8.



Gambar 3. UMAP Plot Sampel Ekspresi Gen

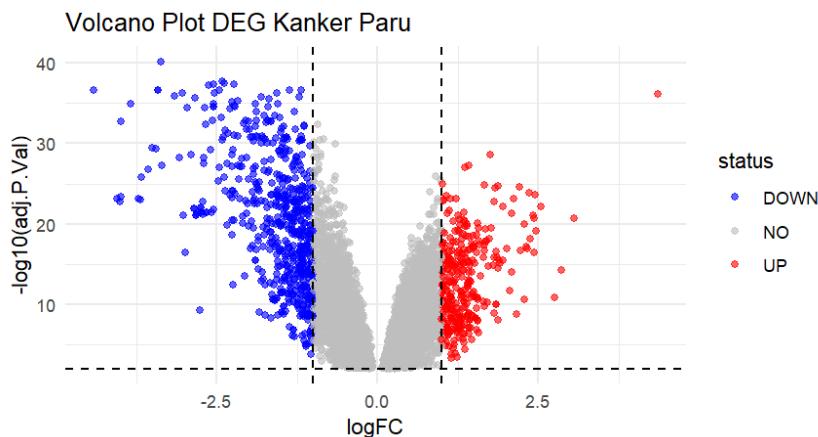
UMAP (*Uniform Manifold Approximation and Projection*) digunakan untuk mereduksi dimensi data guna melihat kemiripan antar sampel.

- Pemisahan Kelompok: Terlihat pemisahan visual yang sangat jelas antara kelompok penderita *adenocarcinoma* (merah) dan jaringan normal (biru).

- Interpretasi Biologis: Pemisahan yang tegas ini membuktikan bahwa terdapat perbedaan profil transkriptomik (ekspresi gen) yang signifikan dan konsisten antara jaringan sehat dan jaringan kanker.

### Jawaban dari pertanyaan :

1. Gen apa saja yang mengalami upregulation dan downregulation, ditampilkan dalam bentuk volcano plot.



Gambar 4. Volcano Plot DEG Kanker Paru

Plot ini memetakan signifikansi statistik ( $-\log_{10} \text{adj.P.Val}$ ) terhadap besarnya perubahan ekspresi (log FC).

- Differentially Expressed Genes (DEGs):
  - Titik Merah (UP): Gen yang mengalami peningkatan ekspresi secara signifikan pada jaringan kanker.
  - Titik Biru (DOWN): Gen yang mengalami penurunan ekspresi secara signifikan pada jaringan kanker.
- Karakteristik: Banyaknya titik yang berada di atas garis ambang batas menunjukkan terdapat ratusan gen yang ekspresinya terganggu akibat kanker paru.

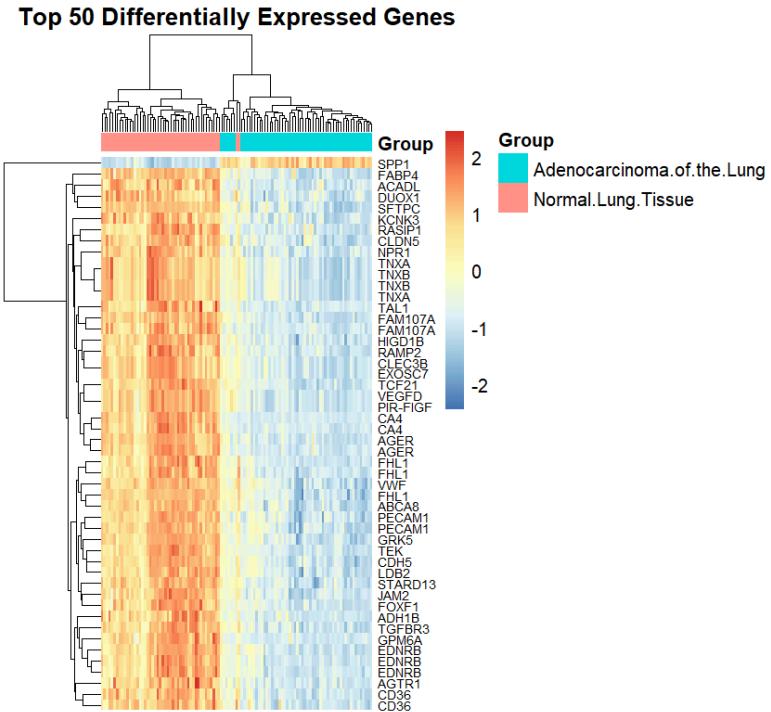
Gen yang mengalami upregulation, misalnya :

- *GREM1*
- *CEP55*
- *TNFRSF21*

Gen yang mengalami downregulation :

- *GAS6*
- *SPTBN1*
- *DPYSL2*

2. 50 Differentially Expressed Genes (DEGs) teratas, ditampilkan dalam bentuk heatmap.

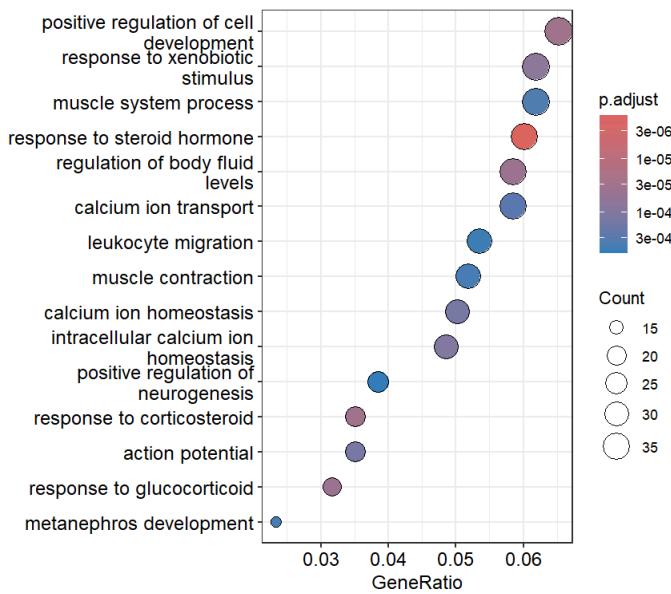


Gambar 5. Heatmap Top 50 Differentially Expressed Genes  
Representasi visual dari 50 gen dengan perbedaan ekspresi paling kontras.

- Pola Ekspresi: Terjadi polaritas yang jelas; gen yang aktif (merah/oranye) di jaringan normal cenderung tidak aktif (biru) di jaringan kanker, dan sebaliknya.
- Contoh Gen: Beberapa gen spesifik seperti *SPP1* menunjukkan intensitas tinggi pada kanker, sementara gen seperti *AGER* atau *CA4* menunjukkan penurunan ekspresi yang tajam dibandingkan jaringan normal.
- Klasterisasi: *Dendrogram* di bagian atas menunjukkan bahwa algoritma berhasil mengelompokkan sampel ke dalam dua grup besar secara otomatis berdasarkan kemiripan ekspresi gennya.

### 3. Analisis enrichment, yang mencakup:

- Gene Ontology (GO)

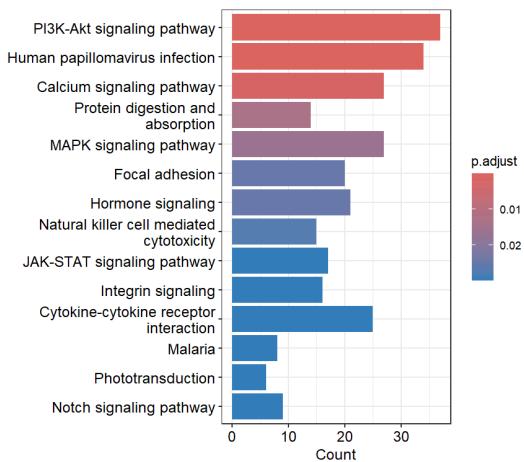


Gambar 6. Analisis *Gene Ontology* (GO)

Gambar 6 memetakan fungsi biologis gen-gen tersebut pada level proses seluler.

- Proses Dominan: Proses *Positive regulation of cell development* menempati urutan teratas dengan *GeneRatio* tertinggi. Ini menunjukkan bahwa mayoritas gen tersebut terlibat aktif dalam mendorong perkembangan dan diferensiasi sel yang abnormal pada adenokarsinoma.
- Respon Terhadap Stimulus: Terdapat keterlibatan signifikan gen dalam *Response to xenobiotic stimulus* dan *Response to steroid hormone/corticosteroid*. Hal ini relevan dengan penderita kanker paru perokok, di mana paru-paru terus-menerus terpapar zat asing (xenobiotik) dari asap rokok.
- Homeostasis Ion: Banyak gen yang terlibat dalam *Calcium ion transport* dan Homeostasis. Gangguan pada regulasi kalsium sering dikaitkan dengan kemampuan sel kanker untuk bermigrasi dan menyerang jaringan lain (metastasis).
- Ukuran Dot & Warna: Ukuran lingkaran yang besar menunjukkan lebih banyak gen yang terlibat dalam proses tersebut (misalnya pada proses pengembangan sel), sementara warna merah yang pekat menunjukkan tingkat signifikansi statistik yang sangat tinggi.

- KEGG Pathway: Hasil analisis enrichment wajib disertai visualisasi/plot yang relevan.



Gambar 7. Analisis KEGG

Gambar 7 menunjukkan jalur biokimia atau sistem pensinyalan yang paling banyak dipengaruhi oleh 920 gen signifikan.

- Jalur Utama (*Top Pathways*): Jalur PI3K-Akt *signaling pathway* dan *Human papillomavirus infection* memiliki jumlah gen (*Count*) terbanyak (lebih dari 30 gen). Jalur PI3K-Akt sangat krusial dalam kanker karena mengatur proliferasi sel dan penghambatan apoptosis (kematian sel).
- Pensinyalan Seluler: Terlihat pengayaan pada jalur *Calcium signaling*, *MAPK signaling*, dan *JAK-STAT signaling*. Ini mengindikasikan adanya gangguan komunikasi antar sel yang mendorong pertumbuhan tumor yang tidak terkendali.
- Signifikansi Statistik: Warna merah pada batang (seperti pada PI3K-Akt) menunjukkan nilai *p.adjust* yang paling signifikan (paling kecil/kuat), sedangkan warna biru menunjukkan signifikansi yang lebih rendah namun tetap di bawah ambang batas 0.05.

#### 4. Kesimpulan

Berdasarkan analisis ekspresi gen pada dataset GSE10072, dapat disimpulkan bahwa terdapat 920 gen yang memiliki perbedaan ekspresi signifikan (DEGs) antara jaringan adenokarsinoma paru dan jaringan normal, dengan pemisahan profil transkriptomik yang tegas sebagaimana terlihat pada visualisasi UMAP dan *heatmap*. Identifikasi gen-gen tersebut, seperti *SPP1* yang mengalami *upregulation* serta *AGER* dan *CA4* yang mengalami *downregulation*, mengonfirmasi adanya perubahan molekuler spesifik yang memicu karsinogenesis dan perkembangan tumor. Analisis pengayaan fungsional lebih lanjut menunjukkan bahwa gen-gen ini mendominasi proses regulasi pengembangan sel serta respon terhadap zat asing (xenobiotik) dan hormon, yang terintegrasi dalam jalur pensinyalan kritis seperti PI3K-Akt dan MAPK, sehingga memberikan gambaran jelas mengenai mekanisme "jejak genetik" akibat paparan rokok terhadap prognosis pasien kanker paru.

References :

Landi MT, Dracheva T, Rotunno M, Figueroa JD, Liu H, Dasgupta A, Mann FE, Fukuoka J, Hames M, Bergen AW, Murphy SE, Yang P, Pesatori AC, Consonni D, Bertazzi PA, Wacholder S, Shih JH, Caporaso NE, Jen J. Gene expression signature of cigarette smoking and its role in lung adenocarcinoma development and survival. *PLoS One*. 2008 Feb 20;3(2):e1651. doi: 10.1371/journal.pone.0001651. PMID: 18297132; PMCID: PMC2249927.