

# Analisis Differentially Expressed Genes pada Kanker Paru Menggunakan Dataset GSE10072

Nama : Shofia Ramadhani

## 1. Pendahuluan

Kanker paru merupakan salah satu penyebab utama kematian akibat kanker di dunia dan memiliki karakteristik molekuler yang kompleks. Perubahan pola ekspresi gen berperan penting dalam proses transformasi sel normal menjadi sel kanker melalui mekanisme proliferasi tidak terkendali, gangguan siklus sel, serta perubahan interaksi sel dengan lingkungan sekitarnya. Analisis ekspresi gen memungkinkan identifikasi gen yang mengalami perubahan signifikan antara kondisi normal dan penyakit sehingga dapat memberikan gambaran mekanisme biologis yang mendasari perkembangan kanker.

Pendekatan bioinformatika berbasis *microarray* menyediakan peluang untuk menganalisis ribuan gen secara simultan. Dataset publik dari Gene Expression Omnibus (GEO) menjadi sumber data yang banyak digunakan untuk eksplorasi biomarker penyakit. Dataset GSE10072 dipilih karena berisi data ekspresi gen jaringan *adenokarsinoma* paru dan jaringan paru normal yang dianalisis menggunakan platform Affymetrix GPL96. Tujuan analisis ini adalah mengidentifikasi *Differentially Expressed Genes* (DEGs) antara jaringan kanker paru dan jaringan normal, serta menginterpretasikan fungsi biologis gen-gen tersebut melalui analisis *enrichment Gene Ontology* dan KEGG *pathway*. Hasil analisis diharapkan memberikan pemahaman mengenai proses molekuler yang terlibat dalam patogenesis kanker paru.

## 2. Metode

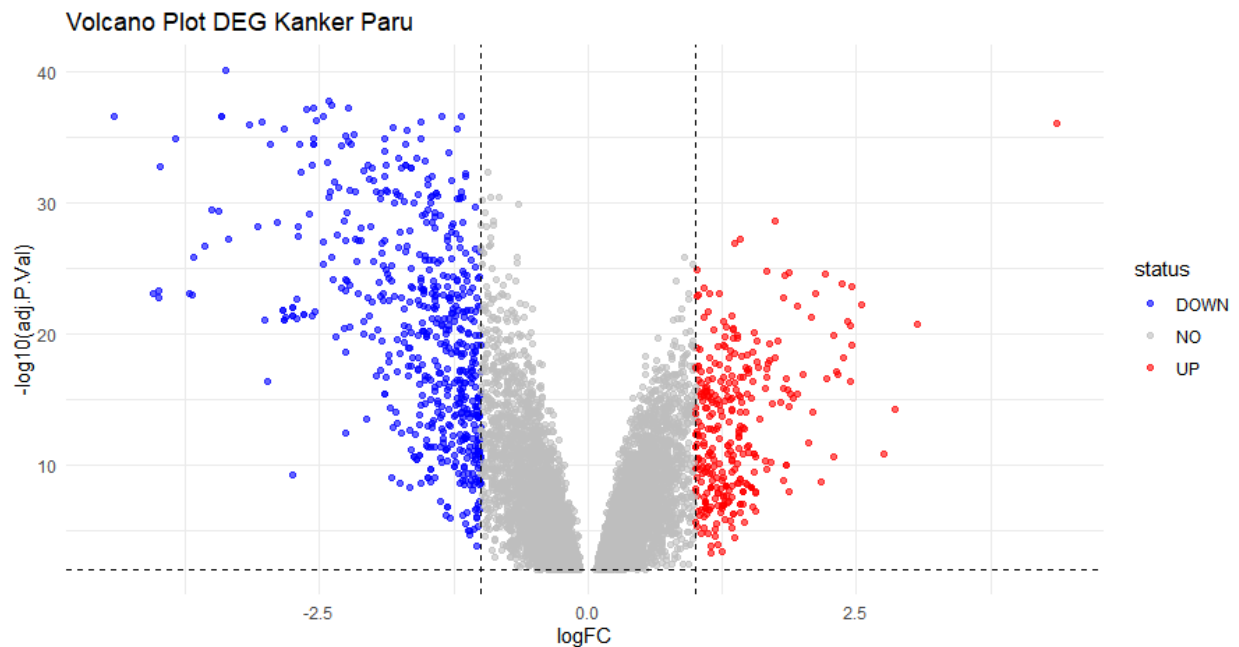
Analisis diawali dengan mengunduh dataset GSE10072 menggunakan paket GEOquery dalam format ExpressionSet yang memuat matriks ekspresi gen, metadata sampel, dan anotasi *probe*. Data ekspresi gen terlebih dahulu melalui tahap *preprocessing* untuk memastikan kualitas analisis statistik. Transformasi log2 diterapkan apabila distribusi nilai ekspresi memiliki rentang yang besar, bertujuan menstabilkan varians dan memenuhi asumsi model linear. Sampel kemudian dikelompokkan menjadi dua kondisi biologis, yaitu *adenocarcinoma of the lung* dan *normal lung tissue*. Analisis DEG dilakukan menggunakan limma yang menerapkan model linear pada setiap gen. Matriks desain dibangun tanpa *intercept* untuk membandingkan langsung kedua kelompok. Estimasi varians diperbaiki menggunakan pendekatan empirical Bayes sehingga meningkatkan stabilitas hasil pada data berukuran besar. Gen dianggap signifikan berdasarkan adjusted p-value hasil koreksi false discovery rate.

Visualisasi data dilakukan untuk mengevaluasi kualitas *preprocessing* dan interpretasi biologis hasil analisis. Volcano plot dibuat untuk menampilkan gen yang mengalami *upregulation* dan *downregulation* secara simultan berdasarkan nilai log fold change dan signifikansi statistik. Heatmap digunakan untuk menampilkan pola ekspresi dari 50 gen paling signifikan. Analisis

*enrichment* dilakukan menggunakan *Gene Ontology* yang mencakup *Biological Process*, *Cellular Component*, dan *Molecular Function*, serta analisis jalur biologis menggunakan *database KEGG*.

### 3. Hasil dan Interpretasi

#### 3.1 Identifikasi Differentially Expressed Genes

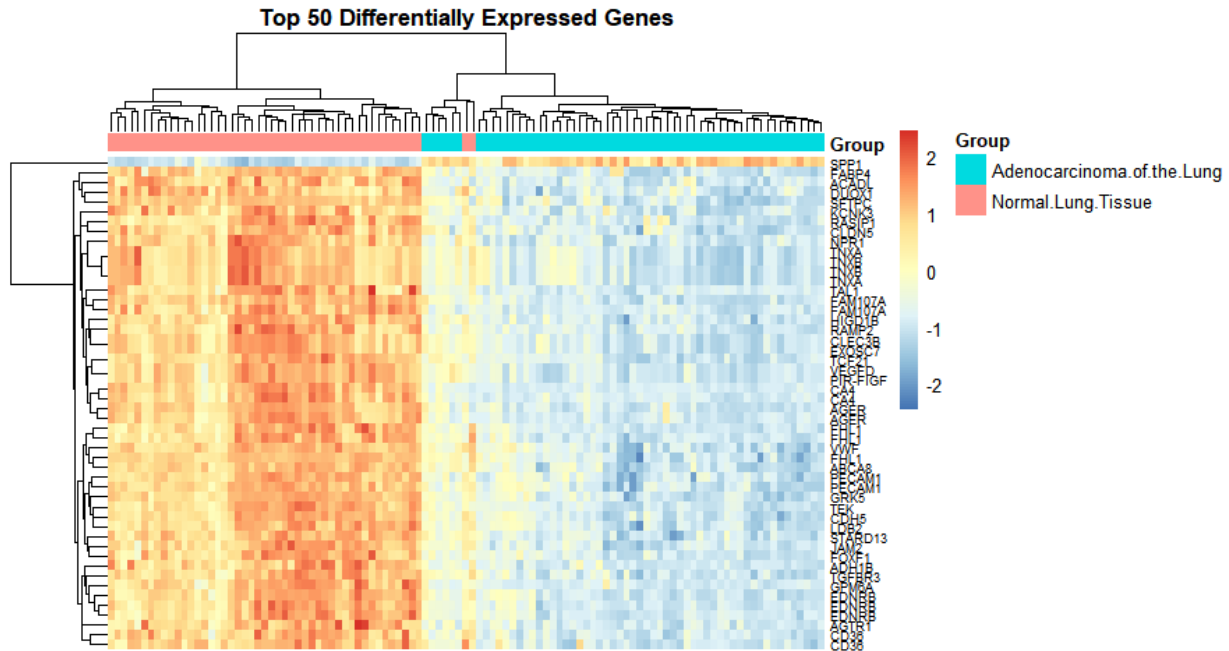


**Gambar 1.** Hasil volcano plot

Volcano plot menampilkan distribusi gen berdasarkan kriteria signifikansi statistik  $\text{adjusted } p\text{-value} < 0,01$  serta perubahan ekspresi biologis  $|\log_{FC}| > 1$ . Gen yang berada di sisi kanan plot dengan warna merah menunjukkan gen yang mengalami *upregulation* pada kanker paru, sedangkan gen di sisi kiri berwarna biru menunjukkan *downregulation*. **Gambar 1** menunjukkan bahwa dari total 9552 gen yang dianalisis, sebanyak 336 gen mengalami *upregulation* dan 584 gen mengalami *downregulation*. Sedangkan sisanya tidak signifikan secara diferensial.

Sejumlah gen menunjukkan perubahan ekspresi yang besar disertai nilai  $\text{adjusted } p\text{-value}$  yang sangat signifikan. Dominasi titik berwarna pada kedua sisi plot menunjukkan adanya perubahan transkriptomik yang luas antara jaringan kanker dan jaringan normal. Gen *upregulated* kemungkinan berperan dalam proliferasi sel, aktivasi sinyal pertumbuhan, dan mekanisme invasi tumor. Gen *downregulated* cenderung berkaitan dengan fungsi fisiologis jaringan paru normal seperti adhesi sel dan homeostasis jaringan.

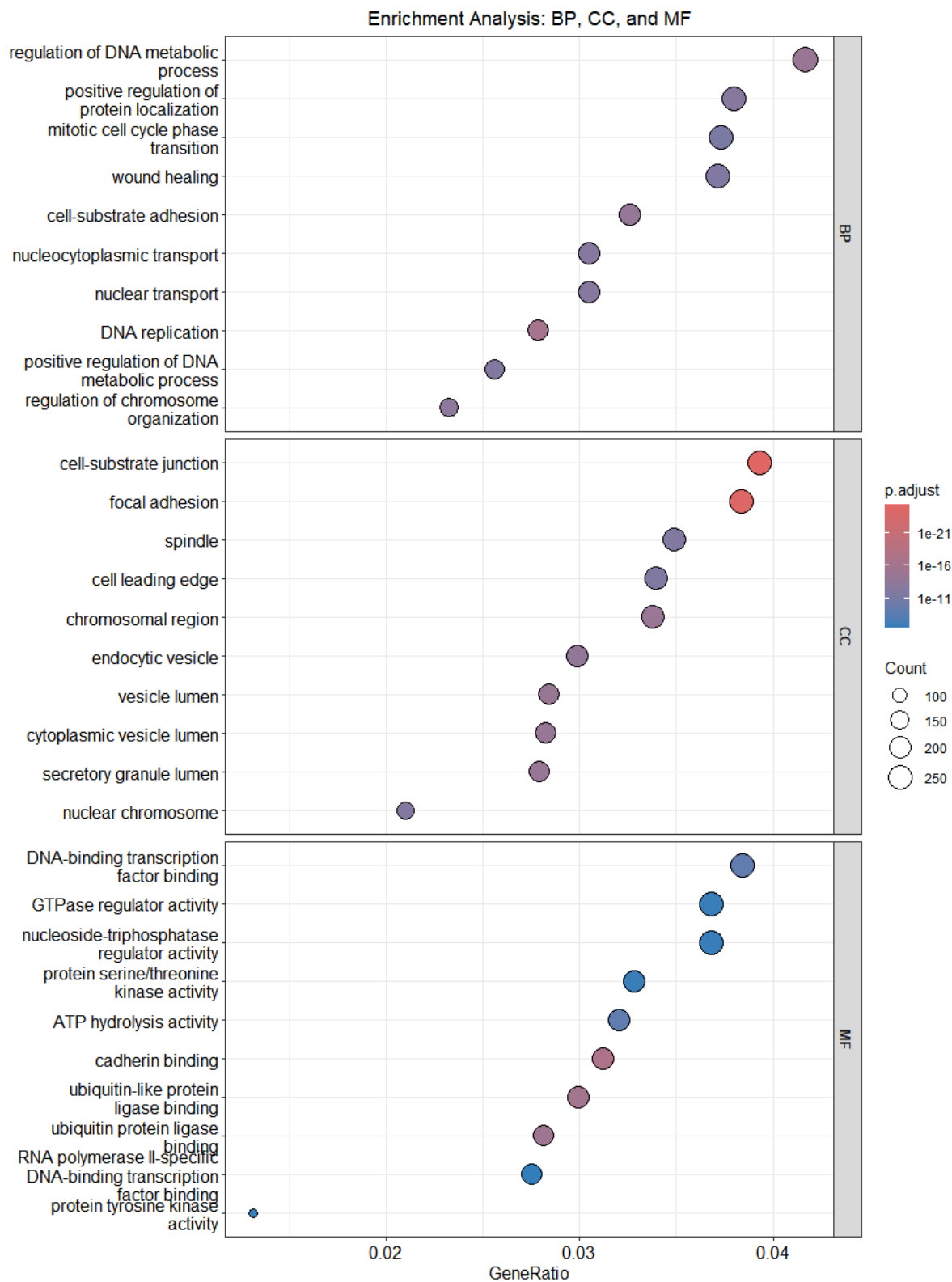
### 3.2 Analisis Top 50 Differentially Expressed Genes



**Gambar 2.** Hasil heatmap

Heatmap dari 50 DEGs teratas berdasarkan adjusted p-value pada **Gambar 2** memperlihatkan pola ekspresi yang kontras antara kedua kelompok sampel. Sampel kanker menunjukkan dominasi ekspresi tinggi pada sejumlah gen tertentu yang ditampilkan dalam warna merah, sedangkan jaringan normal menunjukkan ekspresi lebih rendah pada gen yang sama. *Clustering* hierarkis memisahkan sampel berdasarkan kondisi biologis tanpa informasi label sebelumnya. Pemisahan ini menunjukkan bahwa gen-gen terpilih memiliki kemampuan diskriminatif yang kuat dalam membedakan jaringan kanker dan normal. Pola tersebut mengindikasikan adanya *signature* genetik khas pada *adenokarsinoma* paru. Mayoritas gen teratas menunjukkan pola *downregulation* pada kanker paru, terlihat dari dominasi warna biru pada kelompok *adenokarsinoma* setelah normalisasi Z-score per gen. Pola ekspresi yang konsisten antar sampel menunjukkan stabilitas biologis hasil DEG dan memperkuat validitas analisis statistik.

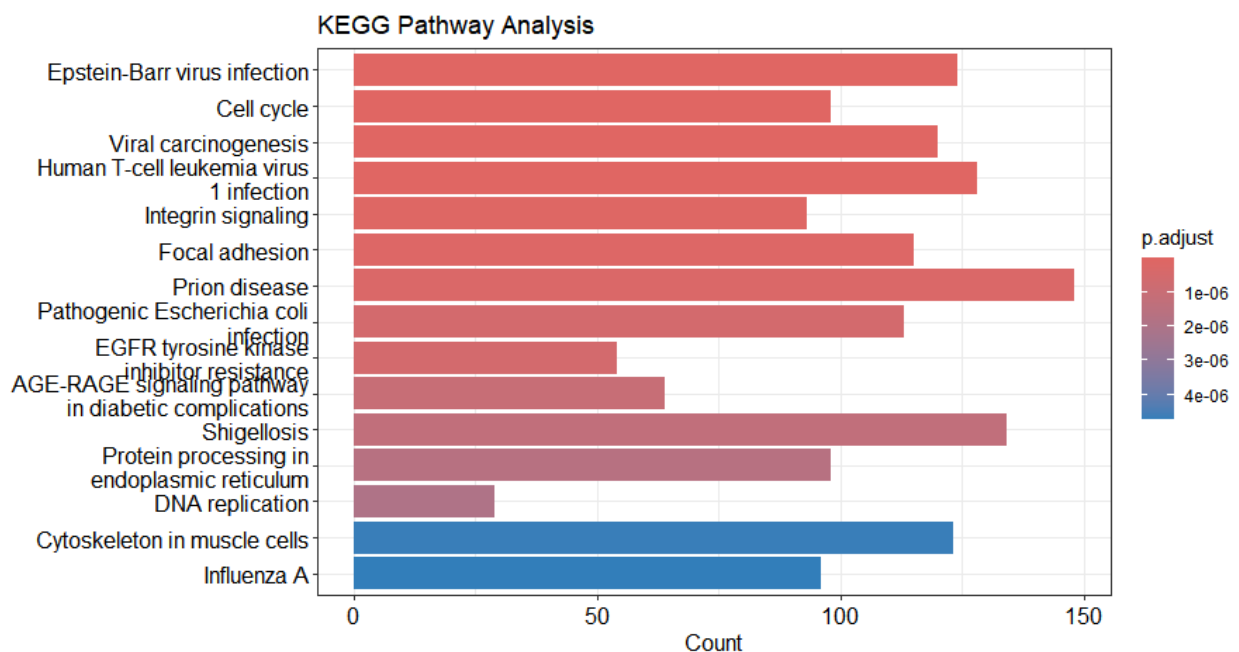
3.3 Analisis Enrichment Gene Ontology



Gambar 3. Hasil enrichment Gene Oncology

Analisis *Gene Ontology* yang ditampilkan pada **Gambar 3** memberikan gambaran fungsi biologis yang diperkaya oleh gen-gen diferensial. Kategori *Biological Process* (BP) menunjukkan pengayaan pada proses regulasi metabolisme DNA, transisi fase siklus sel mitotik, replikasi DNA, serta organisasi kromosom. Hasil ini konsisten dengan karakteristik utama sel kanker yang mengalami proliferasi cepat dan replikasi DNA berlebihan. Kategori *Cellular Component* (CC) memperlihatkan keterlibatan struktur seperti spindle, chromosomal region, focal adhesion, dan vesicle lumen. Aktivasi komponen ini berkaitan dengan pembelahan sel aktif serta perubahan interaksi sel dengan matriks ekstraseluler yang sering terjadi pada proses metastasis. Kategori *Molecular Function* (MF) menunjukkan pengayaan aktivitas pengikatan faktor transkripsi, protein kinase activity, ATP hydrolysis activity, serta ubiquitin ligase binding. Fungsi tersebut berhubungan erat dengan regulasi sinyal seluler dan kontrol degradasi protein yang sering mengalami disregulasi pada kanker. Hasil *enrichment* secara keseluruhan menunjukkan bahwa perubahan ekspresi gen pada *adenokarsinoma* paru terutama berkaitan dengan peningkatan aktivitas proliferasi dan regulasi transkripsi.

### 3.4 Analisis KEGG Pathway



**Gambar 4.** Hasil KEGG pathway

Analisis *KEGG pathway* pada **Gambar 4** mengidentifikasi beberapa jalur biologis yang signifikan, termasuk cell cycle, viral carcinogenesis, integrin signaling, focal adhesion, DNA replication, dan EGFR tyrosine kinase inhibitor resistance. Terlihat pada **Gambar 4** bahwa jumlah gen yang terlibat pada beberapa pathway mencapai lebih dari 100 gen, menunjukkan keterlibatan sistem biologis berskala besar dalam perkembangan kanker paru. Jalur cell cycle dan DNA replication menunjukkan bahwa mekanisme pembelahan sel merupakan proses utama yang

mengalami aktivasi pada kanker paru. Jalur focal adhesion dan integrin signaling menunjukkan perubahan komunikasi antara sel kanker dan lingkungan mikro sekitarnya yang berperan dalam migrasi dan invasi sel tumor. Kemunculan jalur viral carcinogenesis dan infeksi virus tertentu tidak selalu menunjukkan keberadaan virus, melainkan kesamaan mekanisme molekuler yang digunakan sel kanker dan proses infeksi dalam mengendalikan proliferasi sel.

#### 4. Kesimpulan

Analisis DEGs menggunakan dataset GSE10072 berhasil mengidentifikasi perubahan ekspresi gen yang signifikan antara jaringan *adenokarsinoma* paru dan jaringan paru normal. Visualisasi distribusi data menunjukkan kualitas *preprocessing* yang baik serta pemisahan biologis yang jelas antar kelompok sampel. Volcano plot menunjukkan distribusi gen yang mengalami *upregulation* dan *downregulation* secara signifikan, yang mengindikasikan adanya perubahan regulasi transkripsional yang jelas pada kondisi biologis yang dianalisis. Heatmap top 50 DEGs memperlihatkan pola klusterisasi yang mampu memisahkan sampel secara konsisten sesuai kelompoknya, menunjukkan bahwa gen-gen tersebut memiliki kontribusi penting dalam membedakan kondisi biologis. Analisis *enrichment* mengindikasikan bahwa perubahan ekspresi gen terutama berkaitan dengan regulasi siklus sel, replikasi DNA, aktivitas transkripsi, serta interaksi sel dengan matriks ekstraseluler. Analisis *enrichment* melalui *Gene Ontology* (GO) mengungkap keterlibatan gen signifikan dalam proses biologis seperti regulasi proliferasi sel, respons imun, dan aktivitas metabolik, sementara analisis KEGG *pathway* menunjukkan pengayaan pada jalur pensinyalan sel, regulasi siklus sel, serta mekanisme terkait perkembangan penyakit. Secara keseluruhan, hasil ini menunjukkan bahwa perubahan ekspresi gen yang teridentifikasi tidak terjadi secara acak, melainkan berkaitan dengan jalur molekuler spesifik yang berperan dalam mekanisme biologis kondisi yang dianalisis. Selain itu, temuan ini menunjukkan bahwa perubahan transkriptomik berperan penting dalam patogenesis kanker paru dan berpotensi menjadi dasar identifikasi biomarker maupun target terapi di masa mendatang.