

## **Analisis Ekspresi Gen pada Kanker Paru**

Dataset	: GSE10072 ( <i>Lung Adenocarcinoma</i> vs Normal)
Platform	: Microarray (Affymetrix GPL96)
Tujuan	: Mengidentifikasi <i>Differentially Expressed Genes</i> (DEG)

### **1. PENDAHULUAN**

Kanker paru merupakan salah satu penyebab utama kematian akibat kanker di seluruh dunia, dengan *lung adenocarcinoma* sebagai subtipe yang paling sering ditemukan. Perkembangan penyakit ini terjadi melalui akumulasi berbagai perubahan genetik dan epigenetik yang mengganggu mekanisme normal pengaturan pertumbuhan sel, diferensiasi, serta proses apoptosis. Dalam beberapa dekade terakhir, teknologi mikroarray telah banyak dimanfaatkan untuk memprofilkan ekspresi gen secara global sehingga memungkinkan peneliti memperoleh gambaran komprehensif mengenai perubahan molekuler yang berperan dalam proses karsinogenesis (Siegel et al., 2023).

Identifikasi *Differentially Expressed Genes* (DEG) merupakan tahap krusial dalam studi genomik kanker. Analisis ini memungkinkan peneliti mendeteksi gen yang mengalami peningkatan ekspresi (*upregulated*) maupun penurunan ekspresi (*downregulated*) ketika membandingkan jaringan kanker dengan jaringan normal. Informasi mengenai DEG tidak hanya memberikan pemahaman yang lebih mendalam mengenai mekanisme patogenesis penyakit, tetapi juga berpotensi mengungkap kandidat biomarker diagnostik serta target terapi yang lebih spesifik dan efektif (Ritchie et al., 2015).

Perbandingan ekspresi gen antara jaringan kanker dan jaringan normal dianalisis menggunakan metode statistik limma untuk mengidentifikasi gen yang berubah secara signifikan. Hasil analisis divisualisasikan melalui *volcano plot* dan *heatmap*, kemudian dianalisis lebih lanjut menggunakan Gene Ontology (GO) dan KEGG pathway untuk mengetahui proses biologis yang terlibat. Penelitian ini bertujuan mengidentifikasi gen yang mengalami perubahan ekspresi signifikan pada kanker paru serta memahami peran biologisnya.

## 2. METODE

### a. Sumber Data dan Lingkungan Analisis

Data ekspresi gen diperoleh dari basis data publik Gene Expression Omnibus yang dikelola oleh National Center for Biotechnology Information dengan nomor akses GSE10072. Dataset ini berisi profil ekspresi gen dari jaringan *lung adenocarcinoma* dan jaringan paru normal yang dianalisis menggunakan platform mikroarray Affymetrix Human Genome U133A Array (GPL96). Seluruh analisis bioinformatika dilakukan menggunakan bahasa pemrograman R melalui antarmuka RStudio dengan bantuan beberapa paket, yaitu GEOquery, limma, ggplot2, pheatmap, dplyr, AnnotationDbi, dan hgu133a.db.

### b. Pre-processing Data

Dataset diunduh menggunakan paket GEOquery dalam format *ExpressionSet*. Matriks ekspresi gen kemudian diekstraksi menggunakan fungsi `exprs()`. Untuk menstabilkan varians data dan menyesuaikan dengan asumsi model linear, dilakukan transformasi `log2` terhadap nilai ekspresi apabila distribusi data menunjukkan rentang nilai yang tinggi. Informasi metadata sampel diperoleh menggunakan fungsi `pData()` dan digunakan untuk menentukan kelompok biologis antara jaringan kanker paru dan jaringan paru normal.

### c. Analisis Differentially Expressed Genes

Identifikasi gen yang mengalami perubahan ekspresi signifikan dilakukan menggunakan paket limma dengan pendekatan model linear. Matriks desain dibuat menggunakan fungsi `model.matrix()` untuk merepresentasikan kelompok sampel tanpa intercept. Selanjutnya dilakukan perbandingan antara kelompok kanker dan normal menggunakan fungsi `makeContrasts()` dan `contrasts.fit()`. Estimasi varians kemudian dimoderasi menggunakan metode *empirical Bayes* melalui fungsi `eBayes()`. Gen dengan perubahan ekspresi signifikan dipilih berdasarkan nilai *adjusted p-value* menggunakan metode *false discovery rate*.

### d. Anotasi Gen

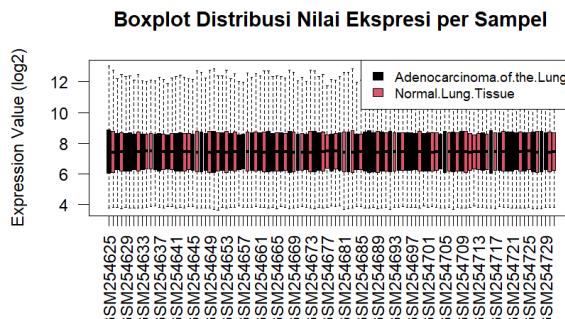
Karena unit analisis pada data mikroarray Affymetrix berupa *probe ID*, maka dilakukan proses anotasi untuk mengonversi probe menjadi simbol gen dan nama gen menggunakan paket hgu133a.db melalui fungsi `select()` dari paket AnnotationDbi.

#### e. Visualisasi Data

Distribusi nilai ekspresi gen antar sampel dievaluasi menggunakan boxplot dan density plot untuk menilai kualitas data. Selain itu, reduksi dimensi dilakukan menggunakan metode Uniform Manifold Approximation and Projection (UMAP) untuk melihat pola pengelompokan sampel berdasarkan profil ekspresi gen. Gen yang mengalami perubahan ekspresi signifikan divisualisasikan menggunakan *volcano plot*, sedangkan pola ekspresi gen paling signifikan ditampilkan menggunakan *heatmap* berdasarkan 50 gen dengan nilai signifikansi tertinggi.

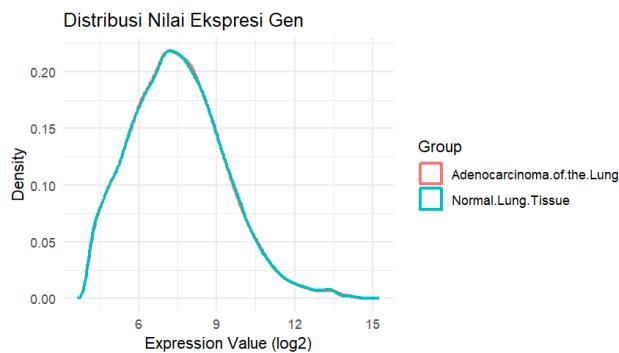
### 3. HASIL DAN INTERPRETASI

#### 1. Boxplot Distribusi Nilai Ekspresi per Sampel



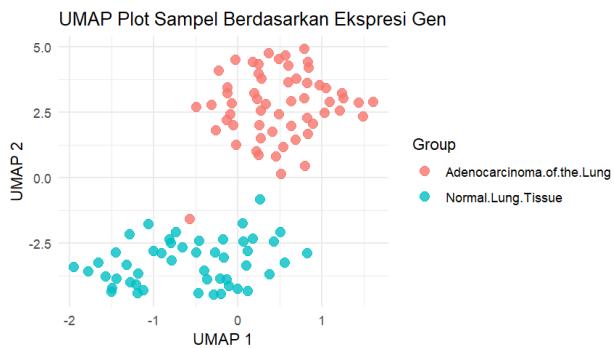
Boxplot menunjukkan distribusi nilai ekspresi gen pada setiap sampel yang digunakan dalam analisis. Sebagian besar sampel memiliki median dan rentang interkuartil yang relatif seragam, menandakan bahwa data telah terdistribusi dengan baik setelah proses normalisasi. Keseragaman distribusi ini menunjukkan bahwa tidak terdapat bias teknis yang signifikan antar sampel, sehingga data layak digunakan untuk analisis ekspresi diferensial lebih lanjut antara kelompok kanker paru dan jaringan normal.

#### 2. Density Plot Distribusi Nilai Ekspresi Gen



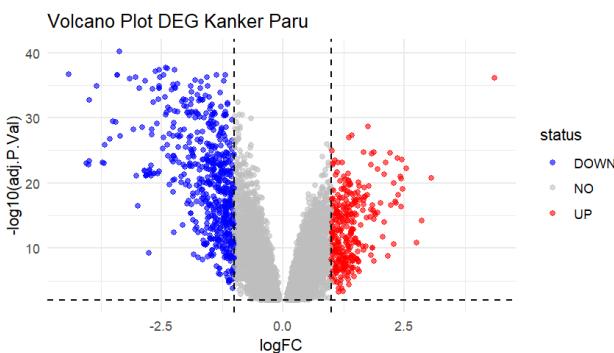
Density plot menggambarkan pola distribusi keseluruhan nilai ekspresi gen pada seluruh sampel. Kurva yang terbentuk menunjukkan distribusi yang relatif normal dan konsisten antar kelompok sampel, menandakan bahwa proses preprocessing dan normalisasi data telah berjalan dengan baik. Pola distribusi yang serupa ini penting untuk memastikan bahwa perbedaan yang ditemukan pada analisis selanjutnya benar-benar disebabkan oleh perbedaan biologis, bukan oleh variasi teknis dalam data.

### 3. UMAP Plot Sampel Berdasarkan Ekspresi Gen



UMAP plot digunakan untuk melakukan reduksi dimensi data ekspresi gen sehingga pola pengelompokan sampel dapat divisualisasikan secara jelas. Pada plot terlihat bahwa sampel kanker paru dan jaringan normal membentuk dua kelompok yang terpisah dengan cukup jelas. Pemisahan ini menunjukkan adanya perbedaan profil ekspresi gen yang signifikan antara kedua kondisi tersebut, yang mengindikasikan adanya perubahan regulasi gen yang berkaitan dengan proses perkembangan kanker paru.

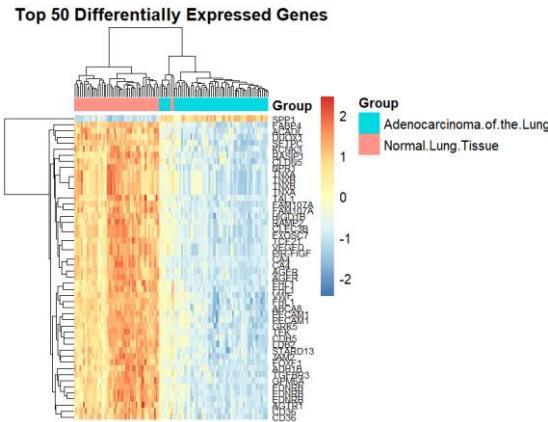
### 4. Volcano Plot Differentially Expressed Genes



Analisis ekspresi diferensial dilakukan untuk mengidentifikasi gen yang mengalami perubahan ekspresi antara jaringan kanker paru dan jaringan normal. Hasil analisis divisualisasikan menggunakan volcano plot yang menampilkan hubungan antara nilai log<sub>2</sub> fold change dan  $-\log_{10}$  p-value. Gen yang berada di sisi

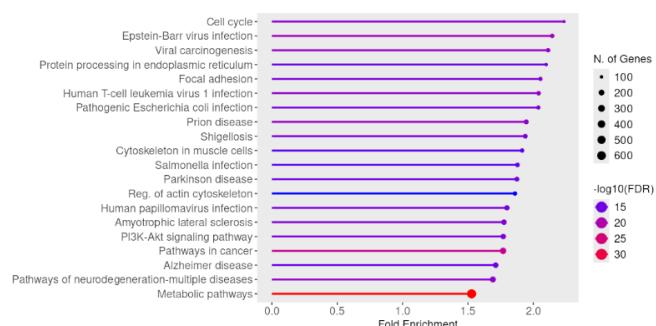
kanan plot menunjukkan upregulation (peningkatan ekspresi pada jaringan kanker), sedangkan gen di sisi kiri menunjukkan downregulation (penurunan ekspresi). Gen dengan nilai adjusted p-value < 0,05 dan  $|\log_{2}FC| > 1$  dianggap mengalami perubahan ekspresi signifikan, sehingga gen-gen yang berada pada area tersebut merupakan gen yang paling berbeda ekspresinya dan berpotensi berperan dalam mekanisme molekuler kanker paru.

## 5. Heatmap Top 50 Differentially Expressed Genes



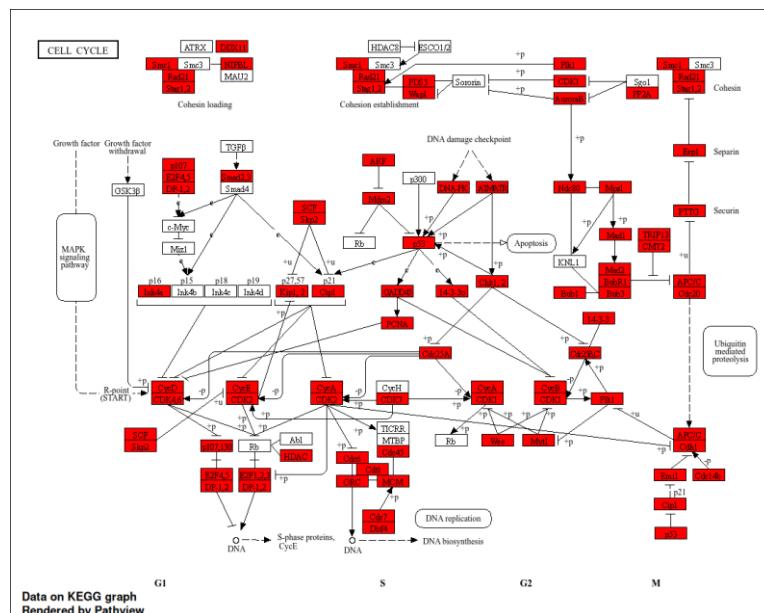
Untuk melihat pola ekspresi gen secara lebih rinci, dilakukan visualisasi terhadap 50 Differentially Expressed Genes (DEGs) teratas menggunakan heatmap. Heatmap menampilkan tingkat ekspresi gen pada setiap sampel dengan representasi warna yang menunjukkan tingkat ekspresi relatif. Pola pengelompokan pada heatmap menunjukkan bahwa sampel kanker paru dan jaringan normal memiliki profil ekspresi gen yang berbeda secara jelas. Perbedaan pola ekspresi ini mengindikasikan adanya gen-gen tertentu yang mengalami perubahan regulasi secara konsisten pada kondisi kanker, sehingga berpotensi berperan dalam proses tumorigenesis dan perkembangan kanker paru.

## 6. Analisis Enrichment: Gene Ontology (GO)



Analisis Gene Ontology (GO) dilakukan untuk mengidentifikasi fungsi biologis yang terkait dengan gen-gen yang mengalami perubahan ekspresi signifikan. Hasil analisis GO menunjukkan bahwa gen-gen tersebut terlibat dalam berbagai proses biologis penting seperti regulasi proliferasi sel, respons terhadap stimulus biologis, serta proses metabolisme seluler. Visualisasi hasil analisis GO ditampilkan dalam bentuk plot enrichment yang menunjukkan kategori fungsi biologis dengan tingkat signifikansi tertinggi. Hasil ini menunjukkan bahwa perubahan ekspresi gen pada kanker paru berkaitan dengan gangguan regulasi proses biologis yang berperan dalam perkembangan kanker.

## 7. Analisis Enrichment: KEGG Pathway



#### **4. KESIMPULAN**

Analisis menunjukkan adanya gen yang mengalami upregulation dan downregulation antara jaringan kanker paru dan normal. Visualisasi volcano plot dan heatmap 50 DEGs menunjukkan perbedaan pola ekspresi gen yang jelas, sedangkan analisis GO dan KEGG mengindikasikan keterlibatan gen-gen tersebut dalam proses biologis dan jalur molekuler yang berkaitan dengan perkembangan kanker paru.

#### **DAFTAR PUSTAKA**

- Ritchie, M. E., Phipson, B., Wu, D., Hu, Y., Law, C. W., Shi, W., & Smyth, G. K. (2015). limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic Acids Research*, 43(7), e47. <https://doi.org/10.1093/nar/gkv007>
- Siegel, R. L., Miller, K. D., Wagle, N. S., & Jemal, A. (2023). Cancer statistics, 2023. *CA: A Cancer Journal for Clinicians*, 73(1), 17–48. <https://doi.org/10.3322/caac.21763>
- Yu, G., Wang, L. G., Han, Y., & He, Q. Y. (2012). clusterProfiler: an R package for comparing biological themes among gene clusters. *OMICS*, 16(5), 284–287. <https://doi.org/10.1089/omi.2011.0118>