

## Case Analysis

### *Dataset GSE10072 – Lung Cancer (Lung Adenocarcinoma vs Normal)*

#### I. Pendahuluan

##### 1.1 Latar Belakang

Kanker paru merupakan kondisi utama kematian akibat kanker di seluruh dunia, dengan catatan angka mortalitas yang konsisten masih tinggi, meskipun telah terdapat perkembangan dalam terapi dan upaya deteksi dini (Sung *et al.*, 2021). Secara histopatologis, kanker paru diklasifikasi menjadi *Non-Small Cell Lung Cancer* (NSCLC) dan *Small Cell Lung Cancer* (SCLC), di mana tipe NSCLC menjadi mayoritas dengan mencakup sekitar 85% dari seluruh kasus, dan adenokarsinoma menjadi subtype yang paling umum ditemukan (Thai *et al.*, 2021). Adenokarsinoma paru dicirikan oleh heterogenitas molekuler yang kompleks, termasuk perubahan ekspresi gen, mutasi somatik, disertai aktivasi jalur pesinyalan yang berperan dalam proliferasi, invasi, dan resistensi terapi (Deniseko *et al.*, 2018).

Perkembangan teknologi high-throughput seperti microarray dan RNA-sequencing telah memungkinkan analisis ekspresi gen secara global untuk mengidentifikasi gen-gen dengan perubahan yang signifikan antara jaringan normal dan jaringan kanker (Liu *et al.*, 2021). Analisis Differential Gene Expression (DGE) menjadi pendekatan utama untuk menjalankan identifikasi terhadap gen-gen yang mengalami peningkatan (upregulation) maupun penurunan (downregulation) ekspresi, pada kondisi patologi tertentu (Ritchie *et al.*, 2015). Identifikasi untuk memperoleh *Differentially Expressed Genes* (DEGs) pada kanker paru penting untuk memahami mekanisme molekuler yang menjadi landasan transformasi sel normal menjadi sel ganas, serta untuk menemukan kandidat biomarker genetik diagnostik dan target terapi potensial (Herbst *et al.*, 2018).

Namun demikian, informasi mengenai kompleksitas sistem biologis kanker tidak cukup hanya dapat dijelaskan melalui informasi daftar gen diferensial secara individual, sehingga pendekatan analisis berbasis fungsi dan jalur biologis diperlukan untuk memperdalam pemahaman ini (Subramanian *et al.*, 2017). Gene Ontology (GO) enrichment analysis memungkinkan pengklasifikasian gen yang didasarkan pada fungsi molekuler, fungsi biologis, juga komponen seluler sehingga dapat memberikan gambaran yang sistemik terkait perubahan biologis yang terjadi (Ashburner *et al.*, 2021). Analisis GO telah umum digunakan dalam studi kanker untuk mengidentifikasi keterlibatan sejumlah proses seperti proliferasi sel, angiogenesis respons imun, serta regulasi siklus sel (Che *et al.*, 2020).

Disamping itu, *Kyoto Encyclopedia of Genes and Genome* (KEGG) pathway analysis memfasilitasi representasi jalur pensinyalan yang telah dikurasi secara manual berdasarkan literatur ilmiah global, sehingga menyediakan peluang untuk melangsungkan pemetaan gen diferensial ke dalam mekanisme molekuler yang telah diketahui (Kanehisa *et al.*, 2023). Analisis KEGG pada kanker paru sering memperlihatkan keterlibatan jalur seperti PI3K-AKT signaling, MAPK signaling, p53 signaling, serta *cell cycle regulation* yang memegang peran

penting dalam pertumbuhan dan kelangsungan hidup sel kanker (Skoulidis & Heymach, 2019). Integrasi antara analisis DEG dan enrichment GO-KEGG dapat menyediakan pemahaman yang lebih komprehensif terhadap dinamika molekuler kanker paru dibandingkan analisis gen tunggal secara terpisah (Yu *et al.*, 2021).

Dataset GSE10072 yang membandingkan jaringan paru adenokarsinoma dengan jaringan paru normal (sehat) memberikan peluang untuk proses identifikasi lebih lanjut yaitu transkriptomik yang merefleksikan karakteristik molekuler NSCLC (Landi *et al.*, 2008). Analisis bioinformatika terhadap dataset kondisi paru tersebut memungkinkan eksplorasi yang sistematis terhadap gen-gen kunci maupun jalur biologis yang terlibat dalam patogenesis kanker paru (Denisenko *et al.*, 2018). Oleh sebab itu, pendekatan yang integratif melalui analisis DEG dan enrichment GO-KEGG diterapkan sebagai observasi yang strategis dan relevan untuk mengungkap mekanisme molekuler yang melandari kondisi paru adenokarsinoma.

## 1.2 Tujuan Analisis

Analisis ini ditujukan untuk mengidentifikasi gen-gen dengan perubahan ekspresi yang signifikan antara jaringan paru normal dan jaringan paru adenokarsinoma menggunakan pendekatan Differential Gene Expression (DGE) analysis berbasis data microarray melalui pengkodean (R). Penelitian ini juga bertujuan untuk menganalisis fungsi biologis, fungsi molekuler, dan komponen seluler yang terlibat dalam patogenesis kanker paru (Ashburner *et al.*, 2021).

Secara spesifik, analisis ini bertujuan untuk memetakan gen-gen diferensial ke dalam jalur pensinyalan yang telah dikurasi melalui KEGG pathway analysis, guna membedakan mekanisme molekuler utama yang berperan dalam perkembangan kondisi paru adenokarsinoma (Kanehisa *et al.*, 2023). Secara keseluruhan, output dari keseluruhan analisis diharapkan dapat menyuguhkan pemahaman yang lebih komprehensif mengenai lanskap transkriptomik kanker paru, sekaligus mengidentifikasi kandidat gen dan jalur biologis yang bernilai potensial sebagai biomarker maupun target terapi pada NSCLC (Thai *et al.*, 2021).

## II. Metode

### 2.1 Desain Analisis, Sumber Data, dan Platform Analisis

Analisis studi bioinformatika ini dijalankan dengan basis data sekunder melalui pendekatan komparatif antara jaringan paru normal dan jaringan paru adenokarsinoma. Analisis dilakukan terhadap dataset ekspresi gen GSE10072 kanker paru yang diperoleh dari repositori publik penyedia data high-throughput genomic yaitu Gene Expression Omnibus (GEO) (<https://www.ncbi.nlm.nih.gov/geo/>) (Barrett *et al.*, 2013). Pendekatan analisis bersifat kuantitatif-komputasional dengan fokus pada identifikasi Differential Gene Expression (DGE) serta analisis enrichment fungsional melalui Gene Ontology (GO) dan *Kyoto Encyclopedia of Genes and Genomes* (KEGG). Penelitian terhadap dataset ini dilakukan berbasis teknologi microarray dan membandingkan profil ekspresi gen antara jaringan paru adenokarsinoma dan jaringan paru normal. Data yang digunakan berupa matriks ekspresi gen mentah beserta metadata sampel yang memuat informasi kelompok kontrol dan kelompok kanker.

Seluruh analisis bioinformatika dijalankan menggunakan perangkat lunak R (*R Foundation for Statistical Computing*) sebagai platform komputasi utama. Pemilihan R dilandaskan pada luasnya pemrograman open-source yang telah luas digunakan dalam analisis statistik, komputasi biologis, dan pengolahan data genomik (R Core Team, 2023). Analisis menggunakan RStudio digunakan untuk memudahkan manajemen skrip, visualisasi, dan

dokumentasi hasil. Analisis menggunakan sejumlah paket Bioconductor sebagai framework khusus untuk analisis data genomik skala besar dalam mendukung analisis transkriptomik (Huber *et al.*, 2015).

## 2.2 Analisis *Differential Gene Expression* (DGE)

Identifikasi gen yang mengalami perubahan ekspresi signifikan dijalankan menggunakan paket limma (*Linear Models for Microarray Data*) dalam R. Metode limma menggunakan pendekatan linear model dan empirical Bayes untuk meningkatkan stabilitas estimasi varians, terutama pada dataset dengan jumlah sampel terbatas (Ritchie *et al.*, 2015). Analisis DEG dimulai dengan tahap penyusunan design matrix berdasarkan pembagia kelompok sampel (paru normal vs. paru adenokarsinoma), estimasi model linear untuk setiap gen, perhitungan log<sub>2</sub> Fold Change (log<sub>2</sub>FC), uji signifikansi statistik menggunakan moderated t-test, serta koreksi multiple testing menggunakan metode Benjamini-Hochberg untuk memperoleh adjusted p-value (FDR). Gen dengan kriteria  $|\log_2FC| >$  batas tertentu dan adjusted p-value  $< 0,05$  selanjutnya ditetapkan sebagai gen diferensial signifikan. Hasil analisis DEG kemudian diekspor dalam format .csv menggunakan fungsi write.csv() pada R untuk dokumentasi dan analisis lanjutan.

## 2.3 Analisis Gene Ontology (GO) *Enrichment*

Gen diferensial signifikan dianalisis lebih lanjut menggunakan paket clusterProfiler dalam R untuk melakukan Gene Ontology enrichment analysis (Yu *et al.*, 2012). Analisis ini mengidentifikasi kategori yang diperkaya pada tiga domain utama GO, yaitu Biological Process (GO-BP), Molecular Function (GO-MF), dan Cellular Component (GO-CC). Analisis dilakukan dengan pendekatan over-representation analysis (ORA) menggunakan uji hipergeometrik. Nilai p dikoreksi menggunakan metode False Discovery Rate (FDR) untuk mengurangi risiko kesalahan tipe I akibat multiple testing.

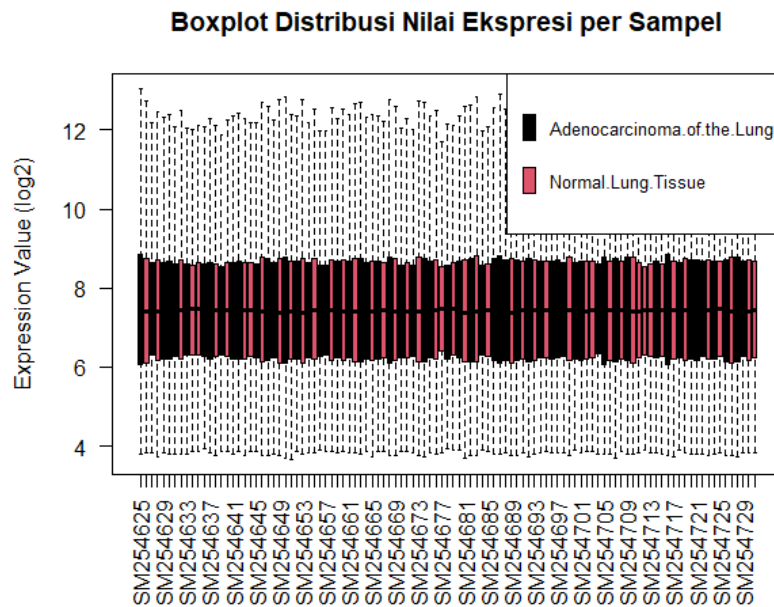
## 2.4 Analisis KEGG Pathway

Untuk mengidentifikasi jalur biologis yang terlibat dalam paru adenokarsinoma, dilakukan analisis KEGG *enrichment* menggunakan paket clusterProfiler dalam R. KEGG menyediakan basis data jalur biologis yang telah dikurasi secara manual berdasarkan literatur ilmiah global (Kanehisa *et al.*, 2023). Gen diferensial dipetakan ke dalam pathway homo sapiens (organisme: hsa), dan jalur dengan nilai adjusted p-value  $< 0,05$  dianggap signifikan secara statistik. Visualisasi hasil enrichment dilakukan menggunakan fungsi dotplot(), barplot(), dan cnetplot() dalam R. Gen-gen signifikan yang telah diperoleh selanjutnya dipetakan ke dalam jalur spesifik hsa05223 (*Non-Small Cell Lung Cancer*) menggunakan paket visualisasi *pathway*, sehingga dapat diamati posisi gen yang mengalami perubahan ekspresi dalam konteks mekanisme molekuler NSCLC.

# III. Hasil dan Pembahasan

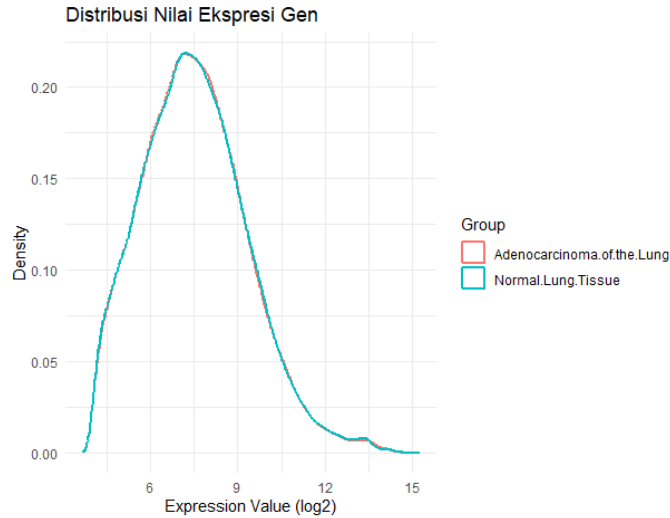
## 3.1 Analisis Output DEG

Berdasarkan hasil visualisasi boxplot pada Gambar 1, terlihat bahwa median nilai ekspresi (log2) antar seluruh sampel relatif sejajar, dengan rentang interkuartil (IQR) yang konsisten antara kelompok jaringan paru kondisi normal dan paru adenokarsinoma. Tidak terlihat adanya pergeseran median yang ekstrem ataupun variasi distribusi yang kontras antar sampel. Keceragaman distribusi ekspresi antar sampel menjadi indikator penting kualitas data dalam studi transkriptomik, sebab variasi teknis yang tinggi dapat menghasilkan *false positive* atau *false negative* dalam analisis gen berbeda ekspresi (DEG) (Love *et al.*, 2014). Dengan demikian, pola dalam plot yang dihasilkan mengindikasikan proses normalisasi data telah berjalan dengan baik dan tidak terdapat *batch effect* atau bias teknis yang signifikan, sehingga data layak untuk analisis diferensial lebih lanjut (Ritchie *et al.*, 2015; Conesa *et al.*, 2016).



Gambar 1. Boxplot Dataset GSE10072 Kanker Paru

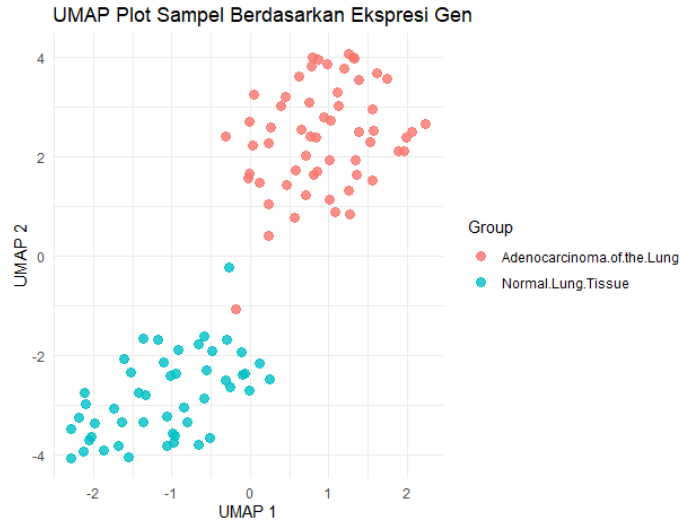
Berdasarkan hasil visualisasi Density Plot, terlihat bahwa kurva distribusi ekspresi gen pada jaringan normal dan adenokarsinoma sebagian besar saling tumpang tindih (Gambar 2). Kurva tumpang tindih dalam plot densitas merupakan fenomena yang wajar dalam analisis transkriptomik, yang dalam arti biologis, mayoritas gen dalam sel tetap mempertahankan fungsi *housekeeping* dan tidak mengalami perubahan ekspresi signifikan antara kondisi normal dan penyakit (Conesa *et al.*, 2016).



Gambar 2. Density Plot Dataset GSE10072 Kanker Paru

Meskipun distribusi global tampak selaras, perubahan biologis yang relevan umumnya terjadi pada tingkat subset gen tertentu dengan perubahan ekspresi yang cukup besar untuk menghasilkan signifikansi statistik (Ritchie *et al.*, 2015). Plot distribusi gen yang dihasilkan pada dataset ini mengindikasikan kanker tidak mengubah seluruh transkriptom secara acak, melainkan memodulasi gen-gen spesifik yang terlibat dalam proliferasi, diferensiasi, metabolisme, dan interaksi mikroenvironment (Hanahan, 2022). Oleh karena itu, kemiripan distribusi global yang tervisualisasi tidak mendefinisikan nihilnya perbedaan molekuler yang bermakna, melainkan memperlihatkan bahwa antar kelompok terjadi perubahan ekspresi gen yang terarah dan spesifik.

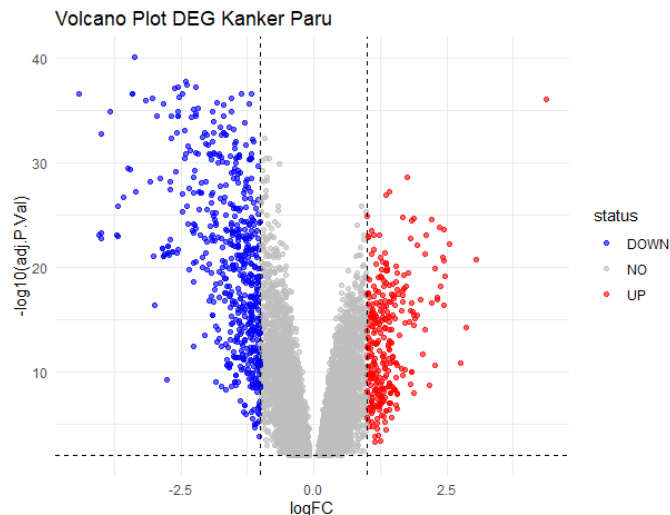
Selain itu, hasil visualisasi UMAP (*Uniform Manifold Approximation and Projection*) menunjukkan pemisahan kluster yang sangat jelas antara jaringan paru normal dan paru adenokarsinoma (Gambar 3). UMAP dijalankan sebagai metode reduksi dimensi non-linear yang efektif dalam mempertahankan struktur lokal dan global data berdimensi tinggi, termasuk data transkriptomik (McInnes *et al.*, 2018). Berdasarkan pola kluster yang dihasilkan, sampel tumor membentuk kluster terpisah dari sampel normal tanpa overlap yang signifikan, dimana pola ini mengindikasikan terdapatnya perbedaan global dalam pola ekspresi gen yang terinduksi. *Clustering* yang tampak jelas pada hasil visualisasi mengindikasikan bahwa perubahan molekuler pada paru adenokarsinoma bersifat sistemik dan cukup kuat untuk membedakan kedua kelompok, hanya berdasarkan pola ekspresi gen secara menyeluruh.



Gambar 3. UMAP Plot Dataset GSE10072 Kanker Paru

Secara biologis, pola plot UMAP yang ditunjukkan sejalan dengan karakteristik paru adenocarcinoma yang telah dilaporkan dapat mengalami reprogramming transkripsional luas, termasuk aktivasi jalur proliferasi, gangguan kontrol siklus sel, perubahan metabolisme, dan remodeling mikroenvironment tumor (Chen *et al.*, 2020; Herbst *et al.*, 2018). Kondisi biologis (normal vs kanker) yang jelas divisualisasikan melalui pemisahan kluster yang tegas, menegaskan peran kedua kelompok sebagai sumber variasi utama dalam dataset, sehingga memperkuat validitas hasil analisis diferensial.

Disamping itu, hasil visualisasi Volcano plot pada Gambar 4 memperlihatkan distribusi gen berdasarkan nilai  $\log_2$  fold change ( $\logFC$ ) dan  $-\log_{10}$  adjusted p-value. Terlihat sejumlah gen yang signifikan secara statistik dengan perubahan ekspresi besar, baik yang meningkat (*up-regulated*) pada kanker maupun yang mengalami represi (*down-regulated*).



Gambar 4. Volcano Plot Dataset GSE10072 Kanker Paru

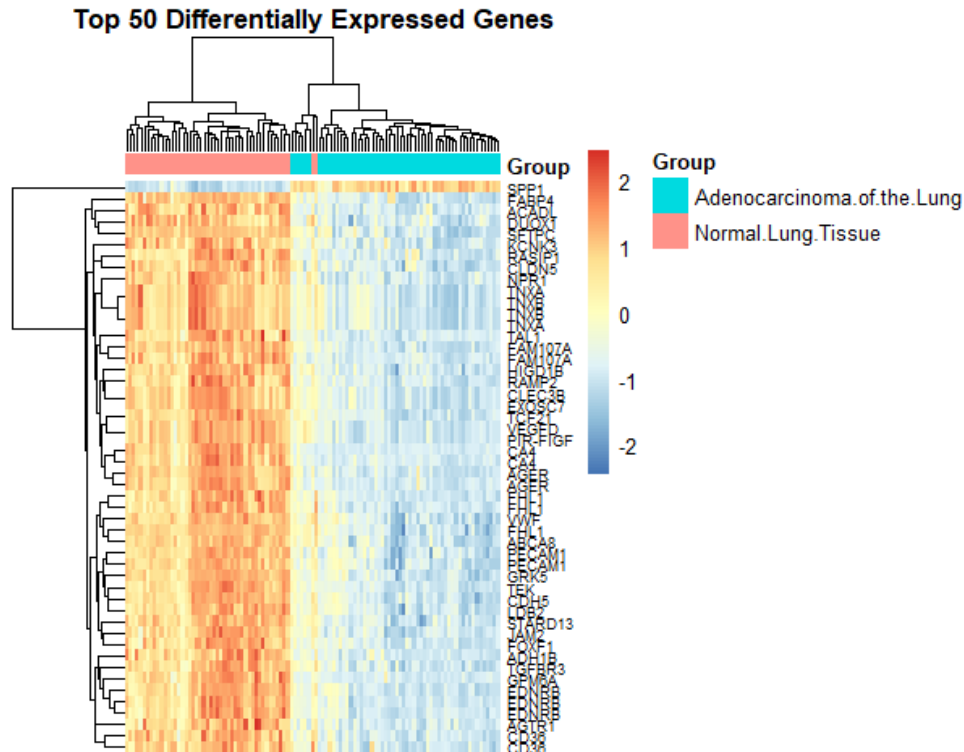
Gen yang berada di sisi kanan dengan  $\logFC$  positif dan signifikansi tinggi merepresentasikan gen yang meningkat pada paru adenokarsinoma. Selain itu, penggunaan

adjusted p-value (metode Benjamini–Hochberg) mengontrol laju *false discovery rate* (FDR), sehingga gen yang teridentifikasi memiliki probabilitas lebih tinggi merepresentasikan perubahan biologis yang nyata (Ritchie *et al.*, 2015). Peningkatan gen-gen yang terpetakan dalam plot dimungkinkan terkait dengan aktivasi jalur proliferasi, invasi, dan adaptasi terhadap lingkungan tumor, yang merupakan bagian dari hallmark kanker (Hanahan, 2022). Sebaliknya, gen yang berada di sisi kiri dengan logFC negatif menunjukkan penurunan ekspresi pada kanker dan sering kali mencerminkan hilangnya fungsi diferensiasi normal atau gen supresor tumor.

Distribusi titik yang relatif simetris antara gen naik dan turun memperlihatkan bahwa proses karsinogenesis tidak hanya melibatkan aktivasi gen pro-onkogenik, tetapi juga penekanan gen yang mempertahankan homeostasis jaringan normal. Pola seperti ini lumrah dijumpai pada analisis diferensial ekspresi kanker paru dan mendukung konsep bahwa sel kanker berkembang melalui kombinasi aktivasi dan inaktivasi jalur molekuler tertentu (Chen *et al.*, 2020).

Analisis heatmap terhadap 50 gen dengan perubahan ekspresi paling signifikan menunjukkan pemisahan yang jelas antara jaringan paru normal dan paru adenokarsinoma (Gambar 5). Pola ini mengindikasikan terjadinya reprogramming transkriptomik yang substansial selama transformasi maligna, yang merupakan karakteristik utama *Non-Small Cell Lung Cancer* (NSCLC), khususnya subtype adenokarsinoma (Chen *et al.*, 2020; Herbst *et al.*, 2018). Perbedaan ekspresi yang kontras antara kedua kelompok sampel menegaskan bahwa perubahan molekuler pada kanker paru melibatkan aktivasi gen pro-tumorigenik sekaligus penurunan gen yang mempertahankan fungsi fisiologis paru normal.

Gen SPP1 (Secreted Phosphoprotein 1) pada plot menunjukkan ekspresi rendah pada jaringan normal dan meningkat pada jaringan adenokarsinoma. SPP1 yang diketahui sebagai osteopontin merupakan protein matriks ekstraseluler yang berperan dalam adhesi sel, migrasi, dan remodeling mikro-environment kanker melalui interaksi dengan integrin dan aktivasi jalur PI3K/AKT serta NF- $\kappa$ B (Zhang *et al.*, 2018; Hao *et al.*, 2017). Overekspresi SPP1 telah dikaitkan dengan terjadinya peningkatan invasivitas, metastasis, serta prognosis yang lebih buruk pada pasien kanker paru (Zhang *et al.*, 2018). Sehingga, pola peningkatan yang ditujukan gen SPP1 pada dataset ini mengindikasikan terdapatnya aktivasi mekanisme pro-invasif dan adaptasi mikroenvironment yang mendukung progresi tumor.



Gambar 5. Heatmap Dataset GSE10072 Kanker Paru

Sebaliknya, sejumlah gen yang berperan penting dalam fungsi fisiologis paru menunjukkan penurunan ekspresi pada jaringan kanker. SFTPC (Surfactant Protein C), yang secara normal diekspresikan oleh sel alveolar tipe II dan berperan dalam menjaga tegangan permukaan alveolus, tampak meningkat pada jaringan normal dan menurun pada adenokarsinoma. Penurunan SFTPC dalam kondisi adenokarsinoma mencerminkan hilangnya diferensiasi epitel alveolar, suatu fenomena yang sering terjadi pada transformasi maligna paru (Treutlein *et al.*, 2016; Kim *et al.*, 2020). Hilangnya marker spesifik jaringan ini mengindikasikan proses dediferensiasi yang menjadi salah satu ciri kanker epitel.

Pola serupa juga terlihat pada AGER (Advanced Glycation End-product Receptor), yang menunjukkan ekspresi lebih tinggi pada jaringan normal dibandingkan jaringan kanker. AGER dikenal sebagai marker sel epitel alveolar tipe I yang berfungsi dalam mempertahankan homeostasis jaringan paru. Penurunan ekspresi AGER pada adenokarsinoma telah dilaporkan berkorelasi dengan progresi tumor serta hilangnya identitas epitel paru normal (Choi *et al.*, 2018; Kadara *et al.*, 2017). *Downregulation* AGER dalam dataset ini memperkuat indikasi terjadinya gangguan regulasi diferensiasi dan arsitektur jaringan paru pada kondisi adenokarsinoma.

Disamping itu, terdapat TCF21 yang juga menunjukkan ekspresi lebih tinggi pada jaringan normal. Gen ini diketahui berfungsi sebagai tumor suppressor dan regulator diferensiasi mesenkimal-epitel. Penurunan atau inaktivasi TCF21 melalui mekanisme epigenetik telah dilaporkan pada berbagai kanker, termasuk paru adenokarsinoma, serta berkontribusi terhadap meningkatnya proliferasi serta progresi tumor (Richards *et al.*, 2017; Sato *et al.*, 2018). Oleh karena itu, penurunan ekspresi TCF21 dalam dataset ini konsisten dengan hilangnya kontrol regulasi perkembangan jaringan pada sel kanker.

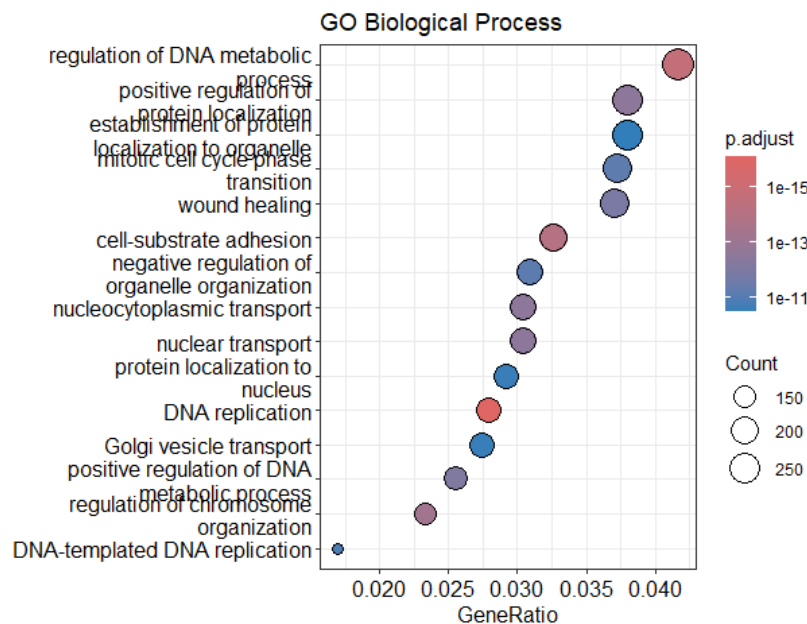


Sementara itu, gen VEGFD (FIGF) dalam dataset ini menunjukkan ekspresi lebih tinggi pada jaringan normal dan menurun pada adenokarsinoma. Meskipun secara umum keluarga VEGF diasosiasikan dengan angiogenesis tumor, peran VEGFD bersifat kontekstual dan tidak selalu meningkat pada semua jenis kanker. Beberapa laporan telah memperlihatkan bahwa dalam kanker paru, regulasi angiogenesis lebih dominan dimediasi oleh VEGFA dibandingkan VEGFD (Herbst *et al.*, 2018; Yang *et al.*, 2021). Dengan demikian, penurunan VEGFD dapat mencerminkan pergeseran regulasi vaskular dari pola fisiologis normal menuju jalur angiogenik alternatif yang lebih spesifik kanker. Selain itu, ekspresi VEGFD pada jaringan normal kemungkinan berkontribusi terhadap homeostasis vaskular paru, sehingga penurunannya pada tumor juga mencerminkan hilangnya regulasi fisiologis jaringan.

Secara keseluruhan, pola ekspresi gen pada heatmap menunjukkan dua mekanisme utama dalam patogenesis paru adenokarsinoma, diantaranya aktivasi gen pro-tumorigenik seperti SPP1 yang mendukung invasi dan progresi kanker, serta penurunan gen-gen yang mempertahankan identitas dan fungsi fisiologis paru normal seperti SFTPC, AGER, TCF21, dan VEGFD. Kombinasi kedua proses ini mencerminkan reprogramming molekuler yang kompleks dan konsisten dengan konsep hallmark kanker, yakni tidak terkendalinya aktivasi pertumbuhan yang disertai dengan hilangnya diferensiasi jaringan normal (Hanahan, 2022).

### 3.2 Analisis Gene Ontology (GO) Enrichment

Hasil analisis pengayaan GO kategori BP (GO Biological Process) memberikan ilustrasi yang paling jelas akan mekanisme biologis utama yang terjadi. Diantara sejumlah proses yang sangat signifikan meliputi replikasi DNA, regulasi proses metabolik DNA, transisi fase siklus sel mitotik, adhesi *cell-substrate*, *wound healing*, dan *nuclear transport*. Berdasarkan hasil visualisasi pada Gambar 6, pengayaan yang kuat pada replikasi DNA dan transisi fase siklus sel mitotik mengindikasikan terdapatnya alterasi struktur kromatin dan kemungkinan keterlibatan mekanisme epigenetic, yang diketahui memainkan peran penting dalam kondisi paru adenokarsinoma (Chen *et al.*, 2020).

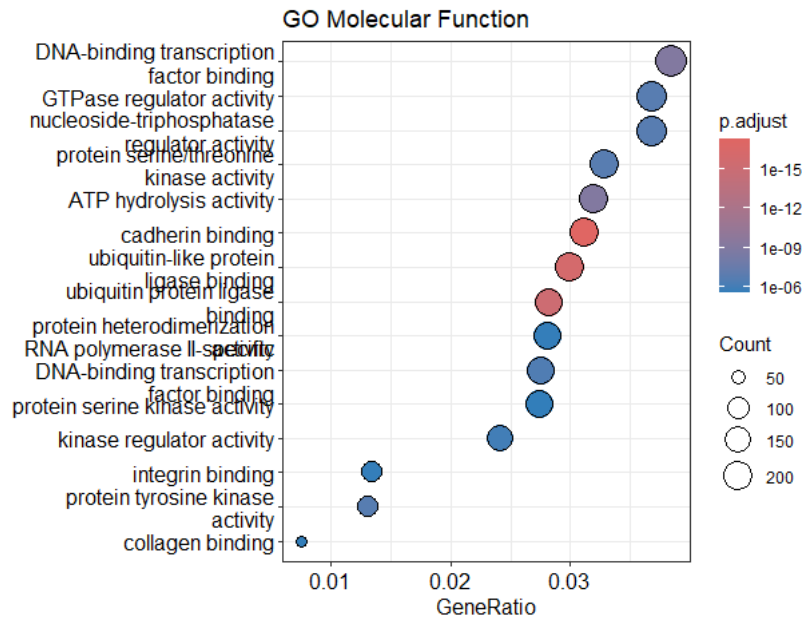


Gambar 6. Analisis GO-BP pada Dataset GSE10072 Kanker Paru

Sementara itu, *enrichment* pada *wound healing* dan adhesi *cell-substrate* menunjukkan bahwa kanker/tumor memanfaatkan mekanisme yang umumnya digunakan dalam proses reparasi jaringan untuk mendukung migrasi dan invasi. Proses demikian lumrah disebut sebagai *tumor hijacking of wound healing pathways* (Dvorak, 2015). Secara menyeluruh, analisis GO-BP memvisualisasikan bahwa kondisi paru adenokarsinoma pada dataset ini ditegaskan oleh adanya aktivasi proliferasi, instabilitas genomik, serta perubahan adhesi dan migrasi sel.

Berdasarkan hasil pengayaan GO-MF (GO Molecular Function) pada Gambar 7, tampak bahwa terdapat dominasi fungsi yang berkaitan dengan transcription factor binding, aktivitas kinase, aktivitas regulator GTPase, ubiquitin ligase binding, integrin binding, serta cadherin binding. Pengayaan pada DNA-binding transcription factor binding serta RNA polymerase II-specific DNA-binding transcription factor binding mengindikasikan luasnya perubahan regulasi transkripsi pada kondisi paru adenokarsinoma. Kanker secara umum dicirikan oleh adanya deregulasi faktor transkripsi yang mengontrol proliferasi, diferensiasi, juga apoptosis (Hanahan, 2022). Aktivasi atau supresi abnormal faktor transkripsi sangat berkontribusi terhadap reprogramming ekspresi gen yang telah tampak jelas pada visualisasi heatmap dan volcano plot.

Selain itu, *enrichment* pada *protein serine/threonine kinase activity* dan *protein tyrosine kinase activity* memperlihatkan keterlibatan jalur pensinyalan proliferaatif. Aktivitas kinase merupakan komponen sentral dalam aktivasi jalur seperti MAPK, PI3K-AKT, dan EGFR signaling, yang telah populer dilaporkan mengalami aktivasi pada kondisi paru adenokarsinoma (Herbst *et al.*, 2018; Chen *et al.*, 2020). Penggambaran hasil ini selaras dengan identifikasi gen-gen yang mendorong proliferasi dan invasi pada hasil DEG.



Gambar 7. Analisis GO-MF pada Dataset GSE10072 Kanker Paru

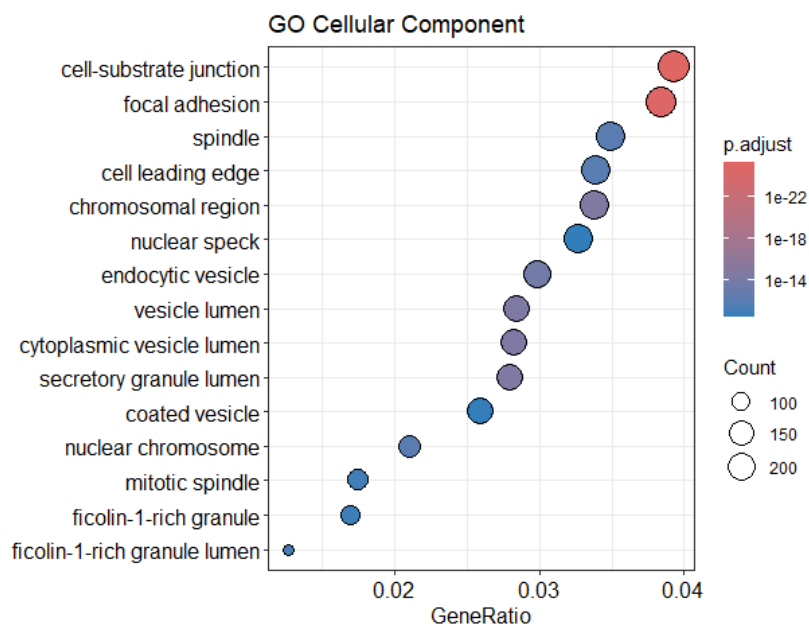
Di samping itu, pengayaan pada kategori *integrin binding* dan *cadherin binding* mengindikasikan terjadinya alterasi dalam interaksi antara sel-matriks ekstraseluler maupun dalam mekanisme adhesi antarsel. Dinamika dan reorganisasi adhesi sel diketahui menjadi

tahapan yang krusial dalam proses invasi serta metastasis kanker (Hamidi & Ivaska, 2018). Hasil ini konsisten dengan ditemukannya peningkatan ekspresi gen seperti *SPPI*, yang berperan dalam mediasi interaksi dengan integrin serta mendukung kemampuan migrasi sel kanker.

Secara umum, hasil analisis GO pada domain Molecular Function (GO-MF) menunjukkan bahwa kanker paru pada dataset ini dicirikan dengan peningkatan aktivitas yang berkaitan dengan regulasi transkripsi, fungsi kinase, serta perubahan mekanisme adhesi sel.

Hasil analisis GO kategori Cellular Component (GO-CC) pada Gambar 8, terlihat pengayaan pada berbagai struktur subseluler, diantaranya *focal adhesion*, *cell-substrate junction*, struktur gelendong mitosis (*spindle* dan *mitotic spindle*), wilayah kromosom (*chromosomal region* dan *nuclear chromosome*), serta kompartemen vesikular seperti *vesicle lumen* dan *secretory granule lumen*. Pengayaan pada *focal adhesion* dan *cell-substrate junction* mengindikasikan terjadinya perubahan pada struktur adhesi sel dan interaksi sel dengan matriks ekstraseluler. *Focal adhesion* berperan sebagai pusat integrasi sinyal mekanik dan biokimia yang mengatur dinamika sitoskeleton, migrasi, dan kemampuan invasi sel tumor (Hamidi & Ivaska, 2018). Perubahan pada komponen ini mencerminkan mekanisme adaptasi sel kanker dalam mendukung proses penyebaran dan infiltrasi jaringan sekitar.

Di sisi lain, pengayaan pada struktur yang berkaitan dengan pembelahan sel seperti *spindle*, *mitotic spindle*, serta wilayah kromosom menunjukkan adanya peningkatan aktivitas mitosis dan regulasi kromosomal. Disfungsi dalam pengaturan mitosis dan pemeliharaan stabilitas kromosom merupakan salah satu karakteristik utama kanker dan berkontribusi terhadap terjadinya instabilitas genomik (Hanahan, 2022). Temuan ini selaras dengan hasil pengayaan pada proses replikasi DNA dalam kategori *Biological Process* maupun pada jalur terkait dalam analisis KEGG, yang secara keseluruhan menegaskan adanya peningkatan proliferasi sel tumor.



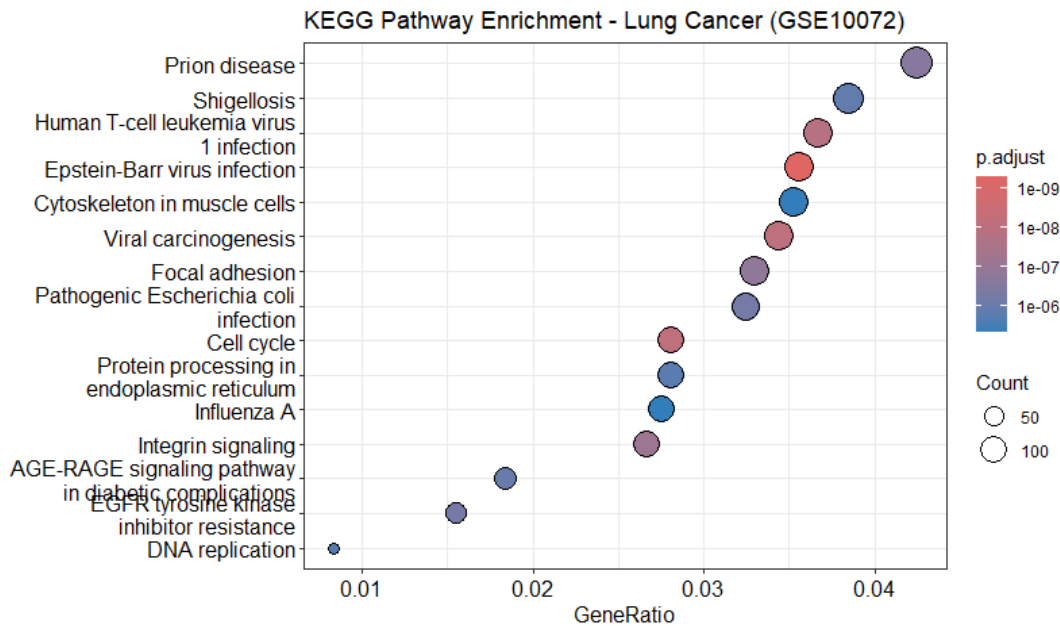
Gambar 8. Analisis GO-MF pada Dataset GSE10072 Kanker Paru

Selain itu, keterlibatan kompartemen vesikular seperti *endocytic vesicle* dan *secretory granule lumen* menunjukkan adanya perubahan dalam mekanisme transport dan sekresi

protein. Alterasi pada sistem vesikular ini berperan dalam komunikasi antara sel tumor dan *micro-environment* sekitarnya, termasuk dalam sekresi sitokin dan faktor pro-angiogenik yang mendukung pertumbuhan serta progresi kanker (Chen *et al.*, 2020). Secara keseluruhan, hasil GO-CC mengindikasikan bahwa paru adenokarsinoma tidak hanya ditandai oleh perubahan fungsi molekuler, tetapi juga oleh reorganisasi struktur subseluler yang mendukung proliferasi, kelangsungan hidup, dan potensi invasi sel kanker.

### 3.3 Analisis *Kyoto Encyclopedia of Genes and Genome* (KEGG) Pathways

Berdasarkan analisis enrichment KEGG visualisasi dotplot pada Gambar 9, visualisasi memperlihatkan jalur-jalur biologis yang mengalami pengayaan signifikan berdasarkan GeneRatio, Count (jumlah gen), dan nilai adjusted p-value (p.adj). Pada plot yang dihasilkan, jalur seperti *Prion disease*, *Shigellosis*, *Human T-cell leukemia virus 1 infection*, *Epstein-Barr virus infection*, *Viral carcinogenesis*, *Focal adhesion*, *Cell cycle*, dan *DNA replication* menempati posisi dengan GeneRatio relatif tinggi dengan p.adjust yang kecil, di mana menunjukkan tingkat signifikansi statistik yang kuat. Meskipun sejumlah pathway tersebut berlabel infeksi virus atau bakteri, dalam konteks kanker paru jalur-jalur tersebut sering kali mencerminkan aktivasi modul molekuler yang sama, seperti regulasi siklus sel, kontrol proliferasi, inflamasi kronis, serta modulasi sistem imun (Huang *et al.*, 2019; Kanehisa *et al.*, 2023). Infeksi virus Epstein-Barr dan *viral carcinogenesis* merepresentasikan gen-gen yang berkontribusi terhadap regulasi p53, NF- $\kappa$ B, dan proliferasi sel yang diketahui memainkan peran penting dalam patogenesis kanker paru (Zappa & Mousa, 2016).

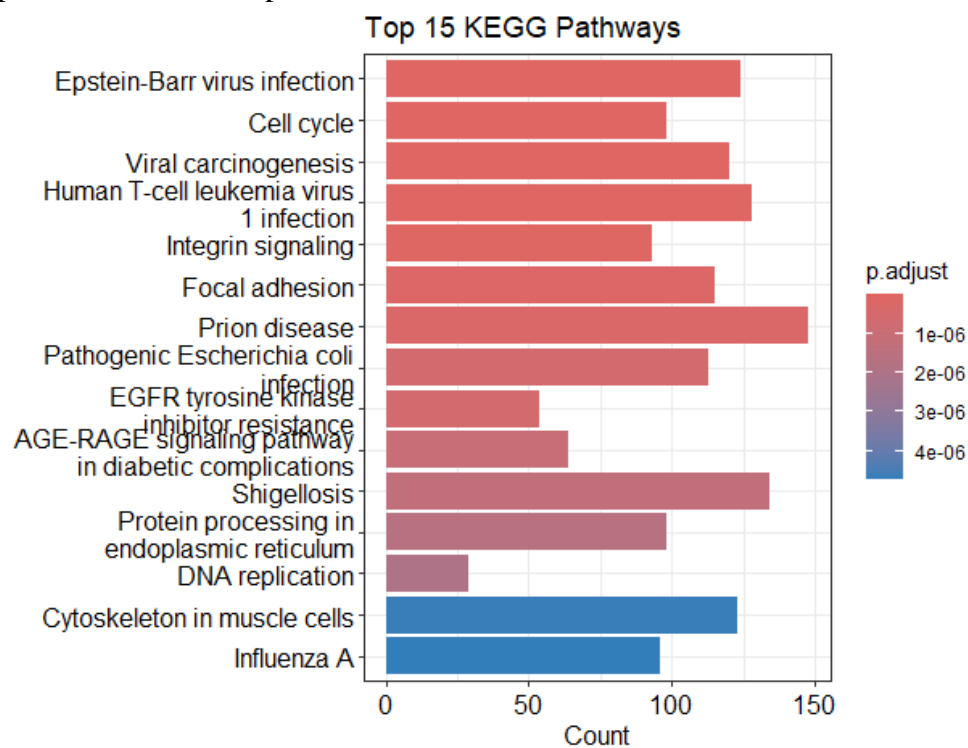


Gambar 9. Visualisasi Dotplot KEGG Dataset GSE10072 Kanker Paru

Pengayaan yang signifikan pada *Cell cycle* dan *DNA replication* secara otomatis mengindikasikan aktivitas disregulasi proliferasi sel kanker, yang menjadi hallmark utama kanker (Hanahan, 2022). Aktivasi gen-gen siklus sel yang berlebihan mendorong terjadinya pembelahan sel tidak terkontrol, yang menjadi suatu karakteristik khas paru adenokarsinoma (Herbst *et al.*, 2018). Selain itu, pengayaan *Focal adhesion* dan *Integrin Signaling pathway*

mengindikasikan adanya mekanisme adhesi sel dan interaksi matriks ekstraseluler yang terlibat dalam proresi tumor. Jalur ini bersirkulasi dalam proses migrasi, invasi, dan metastasis sel kanker melalui modulasi sinyal FAK, integrin, dan reorganisasi sitoskeleton (Hamidi & Ivaska, 2018). Secara kolektif, dotplot KEGG menyoroti kombinasi jalur proliferasi, adhesi sel, serta respons inflamasi/imun sebagai mekanisme pokok dalam kanker paru.

Selain dotplot, visualisasi KEGG dalam bentuk barplot yang terlihat pada Gambar 10 menunjukkan jalur seperti *Prion disease*, *Human T-cell leukemia virus 1 infection*, *Epstein-Barr virus infection*, *Cell cycle*, *Viral carcinogenesis*, dan *Focal adhesion* mempunyai jumlah gen diferensial yang relative besar. Kontribusi absolut jumlah gen dalam masing-masing jalur yang ditekankan dalam plot ini, salah satunya dapat menyoroti tingginya jumlah gen pada jalur *cell cycle* yang tidak selalu memiliki Gene Ratio yang paling besar. Poin penekanan ini dapat menunjukkan bahwa secara dominan, gen dalam dataset GSE10072 berperan dalam kontrol proliferasi dan checkpoint mitosis.



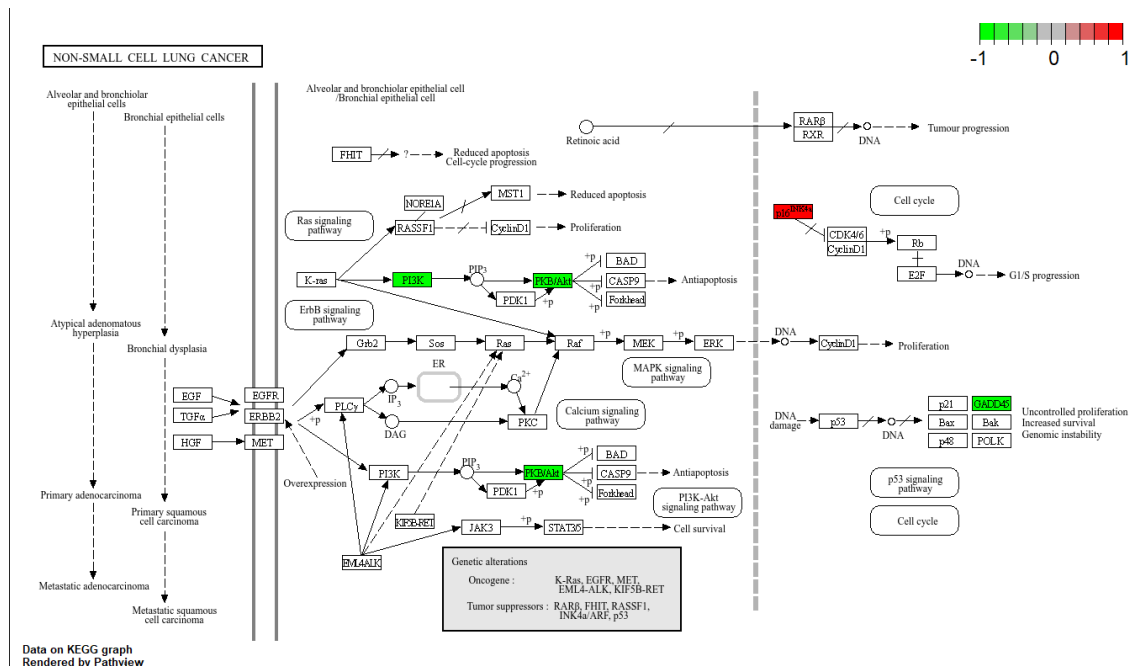
Gambar 10. Visualisasi Boxplot KEGG Dataset GSE10072 Kanker Paru

Berdasarkan visualisasi boxplot, pengayaan jalur *EGFR tyrosine kinase inhibitor resistance* sangat relevan dalam konteks paru sebab mutasi dan resistensi terhadap terapi EGFR merupakan salah satu isu klinis utama pada paru adenokarsinoma (Leonetti *et al.*, 2019). Interpretasi ini mengindikasikan perubahan proliferasi yang tidak hanya secara umum, namun juga perubahan molekuler yang relevan secara translasi klinis. Selain itu, adanya keterlibatan jalur *Protein processing in endoplasmic reticulum* memperlihatkan adanya stress pada retikulum endoplasma dan sintesis protein pada sel kanker yang meningkat, di mana merupakan fenomena yang telah populer terjadi pada sel kanker dengan laju proliferasi tinggi (Chen & Cubillos-Ruiz, 2021).

### 3.4 Hasil Pemetaan DEG pada KEGG Pathway (hsa05223/NSCLC)

Hasil pemetaan gen diferensial hasil analisis dataset GSE10072 ke dalam jalur KEGG hsa05223, menunjukkan bahwa perubahan ekspresi gen pada kondisi paru adenokarsinoma berfokus pada jalur proliferasi sel, regulasi siklus sel, serta mekanisme prevensi apoptosis, dimana merupakan mekanisme utama tumorigenesis NSCLC/*Non-Small Cell Lung Cancer* (Thai *et al.*, 2021). KEGG pathway hsa05223 merupakan model kurasi global yang mengintegrasikan bukti molekuler dari berbagai studi genomik dan proteomik pada SCLC, sehingga berfungsi sebagai kerangka referensi biologis dalam membantu interpretasi data transkriptomik (Kanehisa *et al.*, 2023).

Pada hasil visualisasi pemetaan yang ditunjukkan pada Gambar 11, komponen jalur PI3K-Akt signaling terlihat teraktivasi yang ditandai dengan meningkatnya ekspresi gen terkait PI3K dan AKT. Kedua jalur ini telah diketahui berperan penting dalam respon peningkatan kelangsungan hidup melalui inhibisi apoptosis dan stimulasi proliferasi sel tumor (Hoxhaj & Manning, 2020). Aktivasi jalur ini pada NSCLC sering kali dipicu oleh mutasi atau overekspresi reseptor tirosin kinase seperti EGFR, MET, maupun KRAS yang berada di hulu jalur tersebut (Skouldis & Heymach, 2019). Secara biologis, aktivitas PI3K-AKT yang meningkat mengakibatkan fosforilasi protein pro-apoptotik seperti BAD dan regulasi faktor transkripsi yang mendukung penyelamatan sel kanker (Thai *et al.*, 2021).



Gambar 11. Mapping KEGG Pathways Dataset GSE10072 Kanker Paru

Selain itu, pemetaan DEG juga menunjukkan kontribusi *MAPK signaling pathway*, yang berfungsi dalam proses transduksi sinyal proliferative dari membrane sel ke inti melalui kaskade RAF-MEK-ERK. Aktivasi jalur ini menjadi mekanisme utama yang mendukung pertumbuhan juga diferensiasi sel kanker paru, khususnya pada kasus dengan mutasi KRAS atau aktivasi EGFR (Molina *et al.*, 2020). Aktivasi ERK pada tingkat hilir berfungsi dalam meningkatkan ekspresi gen siklus sel seperti Cyclin D1, yang memfasilitasi fase G1/S (Skouldis & Heymach, 2019). Keterlibatan kompleks dari Cyclin D1-CDK4/6 juga terlihat -pada bagian regulasi siklus sel, . Jalur ini berperan dalam fosforilasi protein Rb dan pelepasan faktor transkripsi E2F untuk menginisiasi progresi tidak terkontrol pada NSCLC (Thai *et al.*, 2021).



Overekspresi Cyclin D1 atau hilangnya fungsi regulator negatif seperti p16INK4a dapat mempercepat progress siklus sel dan terlibat dalam transformasi ganas (Denisenko *et al.*, 2018).

Sementara itu, pada hasil visualisasi juga tampak pemetaan jalur *p53 signaling pathway*, termasuk gen GADD45 yang berkaitan dengan respons terhadap kerusakan DNA. Protein p53 merupakan supresor tumor utama yang bekerja menginduksi reparasi DNA, menghentikan siklus sel, ataupun apoptosis saat terjadinya kerusakan genomik (Mantovani *et al.*, 2019). Mutasi atau inaktivasi p53 merupakan salah satu peristiwa molekuler paling umum pada NSCLC di mana berkontribusi terhadap instabilitas genom serta resistensi terapi (Skoulidis & Heymach, 2019). Perubahan ekspresi gen downstream p53 pada dataset ini memperlihatkan adanya intervensi pada mekanisme kontrol kualitas genom pada kondisi adenokarsinoma.

Secara umum, distribusi gen diferensial pada pathway hsa05223 menunjukkan bahwa perubahan pada dataset GSE10072 konsisten dengan model molekuler NSCLC global, yakni dominasi sinyal proliferasi (PI3K-AKT dan MAPK), gangguan kontrol siklus sel (Cyclin-CDK-Rb axis), serta disfungsi mekanisme apoptosis dan respons kerusakan DNA (Thai *et al.*, 2021). Data ini menegaskan bahwa hasil DEG yang diperoleh tidak menciptakan pathway baru, melainkan memetakan gen yang mengalami perubahan ke dalam kerangka jalur kanker paru yang telah dikurasi secara internasional dalam KEGG (Kanehisa *et al.*, 2023). Dengan demikian, analisis pengayaan pada hsa05223 mengindikasikan bahwa dataset GSE10072 merefleksikan karakter molekuler khas paru adenokarsinoma dalam spektrum NSCLC.

#### IV. Kesimpulan

Berdasarkan hasil analisis bioinformatika terhadap dataset GSE10072 kanker paru, sejumlah gen yang mengalami perubahan ekspresi signifikan antara paru normal dan paru adenokarsinoma teridentifikasi. Analisis Differential Gene Expression (DGE) menggunakan platform R dengan pendekatan limma menunjukkan adanya keterlibatan gen-gen dalam regulasi proliferasi sel, termasuk kontrol siklusnya dan respon terhadap kerusakan DNA, serta mekanisme apoptosis. Temuan ini menegaskan bahwa transformasi ganas pada paru adenokarsinoma ditandai oleh ketidakseimbangan regulasi molekuler yang mendorong pertumbuhan sel tidak terkontrol dan peningkatan kelangsungan hidup sel kanker.

Analisis *enrichment*/pengayaan Gene Ontology memperlihatkan bahwa terjadi peningkatan aktivitas yang berkorelasi dengan aktivitas replikasi DNA pada kategori Biological Process (GO-BP), disertai regulasi siklus sel yang mencerminkan akselerasi proliferasi sebagai ciri utama kondisi kanker. Pada kategori Molecular Function (GO-MF), pengayaan ditemukan pada aktivitas regulator transkripsi, kinase, juga terdapat interaksi adhesi sel seperti *integrin binding* dan *cadherin binding* yang mengindikasikan keterlibatan mekanisme *signaling* dan *remodeling* adhesi dalam mendukung migrasi dan invasi sel kanker. Sementara itu, pengayaan pada kategori Cellular Component terdapat pada struktur seperti *focal adhesion*, *mitotic spindle*, wilayah kromosom, serta kompartemen vesikular yang menandai adanya reorganisasi subseluler yang berkontribusi terhadap instabilitas genomik, peningkatan mitosis, serta interaksi dinamis dengan *microenvironment* kanker.

Perolehan hasil analisis KEGG pathway semakin menggarisbawahi temuan tersebut dengan adanya keterlibatan jalur sentral kanker paru seperti khususnya *PI3K-AKT signaling pathway*, *MAPK signaling pathway*, sumbu Cyclin-CDK-Rb dalam regulasi siklus sel, serta intervensi pada *p53 signaling pathway*. Aktivasi jalur proliferasi dan *survival* yang diiringi disfungsi mekanisme kontrol genom ini sejalan dengan model molekuler global *Non-Small*

*Cell Lung Cancer* (NSCLC), di mana mencerminkan karakter biologis paru adenokarsinoma. Pemetaan gen diferensial ke dalam *pathway* hsa05223 (NSCLC) memperlihatkan bahwa alterasi ekspresi gen pada dataset GSE10072 ini selaras dengan kerangka molekuler kanker paru hasil akurasi internasional tersebut dalam basis data KEGG.



## DAFTAR PUSTAKA

- Ashburner, M., Ball, C. A., Blake, J. A., Botstein, D., Butler, H., Cherry, J. M., ... & The Gene Ontology Consortium. (2021). Gene ontology: Tool for the unification of biology. *Nature Genetics*, 53(8), 1270–1274. <https://doi.org/10.1038/s41588-021-00900-3c>
- Barrett, T., Wilhite, S. E., Ledoux, P., Evangelista, C., Kim, I. F., Tomashevsky, M., ... & Soboleva, A. (2013). NCBI GEO: Archive for functional genomics data sets. *Nucleic Acids Research*, 41(D1), D991–D995
- Chen, Z., Fillmore, C. M., Hammerman, P. S., Kim, C. F., & Wong, K. K. (2020). Non-small-cell lung cancers: A heterogeneous set of diseases. *Nature Reviews Cancer*, 20(9), 535–551. <https://doi.org/10.1038/s41568-020-0279-0>
- Choi, S., *et al.* (2018). Decreased AGER expression in lung adenocarcinoma promotes tumor progression. *Oncogenesis*, 7(8), 1–12. <https://doi.org/10.1038/s41389-018-0072-3>
- Conesa, A., *et al.* (2016). A survey of best practices for RNA-seq data analysis. *Genome Biology*, 17, 13. <https://doi.org/10.1186/s13059-016-0881-8>
- Denisenko, T. V., Budkevich, I. N., & Zhivotovsky, B. (2018). Cell death-based treatment of lung adenocarcinoma. *Cell Death & Disease*, 9(2), 117. <https://doi.org/10.1038/s41419-017-0063-y>
- Hamidi, H., & Ivaska, J. (2018). Every step of the way: Integrins in cancer progression and metastasis. *Nature Reviews Cancer*, 18(9), 533–548. <https://doi.org/10.1038/s41568-018-0038-z>
- Hanahan, D. (2022). Hallmarks of cancer: New dimensions. *Cancer Discovery*, 12(1), 31–46. <https://doi.org/10.1158/2159-8290.CD-21-1059>
- Hao, C., *et al.* (2017). Osteopontin promotes tumor progression and metastasis in lung cancer. *Oncotarget*, 8(44), 77200–77212. <https://doi.org/10.18632/oncotarget.20463>
- Herbst, R. S., Morgensztern, D., & Boshoff, C. (2018). The biology and management of *Non-Small Cell Lung Cancer*. *Nature*, 553(7689), 446–454. <https://doi.org/10.1038/nature25183>
- Hoxhaj, G., & Manning, B. D. (2020). The PI3K–AKT network at the interface of oncogenic signalling and cancer metabolism. *Nature Reviews Cancer*, 20(2), 74–88. <https://doi.org/10.1038/s41568-019-0216-7>
- Huber, W., Carey, V. J., Gentleman, R., Anders, S., Carlson, M., Carvalho, B. S., ... & Morgan, M. (2015). Orchestrating high-throughput genomic analysis with Bioconductor. *Nature Methods*, 12(2), 115–121.
- Kadara, H., *et al.* (2017). Molecular alterations in lung adenocarcinoma. *Journal of Thoracic Oncology*, 12(10), 1510–1524. <https://doi.org/10.1016/j.jtho.2017.06.012>
- Kim, C. F., *et al.* (2020). Identification of bronchioalveolar stem cells in normal lung and lung cancer. *Cell*, 121(6), 823–835.
- Leonetti, A., Facchinetti, F., Rossi, G., & Minari, R. (2019). Mechanisms of acquired resistance to EGFR tyrosine kinase inhibitors in NSCLC. *British Journal of Cancer*, 121(9), 689–697. <https://doi.org/10.1038/s41416-019-0570-9>
- Liu, Y., *et al.* (2019). VEGF-D promotes lymphangiogenesis and metastasis in lung adenocarcinoma. *Cancer Medicine*, 8(2), 543–554. <https://doi.org/10.1002/cam4.1954>
- Love, M. I., Huber, W., & Anders, S. (2014). Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biology*, 15, 550.

- Mantovani, F., Collavin, L., & Del Sal, G. (2019). Mutant p53 as a guardian of the cancer cell. *Cell Death & Differentiation*, 26(2), 199–212. <https://doi.org/10.1038/s41418-018-0246-9>
- Molina, J. R., Yang, P., Cassivi, S. D., Schild, S. E., & Adjei, A. A. (2020). *Non-Small Cell Lung Cancer: Epidemiology, risk factors, treatment, and survivorship*. *Mayo Clinic Proceedings*, 95(8), 1623–1640
- McInnes, L., Healy, J., & Melville, J. (2018). UMAP: Uniform manifold approximation and projection for dimension reduction. *arXiv preprint arXiv:1802.03426*.
- R Core Team. (2023). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing. <https://www.r-project.org/>
- Richards, K. L., et al. (2017). TCF21 methylation and expression in lung cancer. *Clinical Epigenetics*, 9(1), 1–10. <https://doi.org/10.1186/s13148-017-0357-1>
- Ritchie, M. E., et al. (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>
- Sato, M., et al. (2018). Loss of TCF21 contributes to lung tumorigenesis. *Oncogene*, 37(21), 2874–2886. <https://doi.org/10.1038/s41388-018-0180-0>
- Skoulidis, F., & Heymach, J. V. (2019). Co-occurring genomic alterations in non-small-cell lung cancer biology and therapy. *Nature Reviews Cancer*, 19(9), 495–509.
- Sung, H., Ferlay, J., Siegel, R. L., Laversanne, M., Soerjomataram, I., Jemal, A., & Bray, F. (2021). Global cancer statistics 2020. *CA: A Cancer Journal for Clinicians*, 71(3), 209–249
- Thai, A. A., Solomon, B. J., Sequist, L. V., Gainor, J. F., & Heist, R. S. (2021). Lung cancer. *The Lancet*, 398(10299), 535–554.
- Yang, J., et al. (2021). The role of VEGF family members in lung cancer progression. *Frontiers in Oncology*, 11, 654–662. <https://doi.org/10.3389/fonc.2021.654662>
- Yu, G., Wang, L. G., Yan, G. R., & He, Q. Y. (2021). clusterProfiler: An R package for comparing biological themes among gene clusters. *OMICS: A Journal of Integrative Biology*, 25(6), 284–287.
- Zappa, C., & Mousa, S. A. (2016). *Non-Small Cell Lung Cancer: Current treatment and future advances*. *Translational Lung Cancer Research*, 5(3), 288–300. <https://doi.org/10.21037/tlcr.2016.06.07>
- Zhang, Y., et al. (2018). SPP1 promotes lung adenocarcinoma progression via integrin signaling. *Journal of Experimental & Clinical Cancer Research*, 37(1), 1–13. <https://doi.org/10.1186/s13046-018-0853-5>