

Analisis Ekspresi Gen Diferensial pada Kanker Payudara Dataset GSE42568

Oleh: Puspa Hening

1. Pendahuluan

Kanker payudara adalah salah satu jenis kanker yang paling sering dialami oleh perempuan di seluruh dunia dan menjadi penyebab kematian akibat kanker yang cukup tinggi. Penyakit ini terjadi karena adanya perubahan pada sel-sel payudara, terutama pada tingkat gen, yang membuat sel tumbuh secara tidak terkendali. Oleh karena itu, memahami perubahan gen yang terjadi pada kanker payudara sangat penting agar dapat membantu pengembangan metode diagnosis dan pengobatan yang lebih efektif.

Salah satu cara untuk mempelajari perubahan tersebut adalah dengan melihat pola ekspresi gen, yaitu tingkat “aktivitas” gen di dalam sel. Melalui teknologi microarray, ribuan gen dapat dianalisis sekaligus untuk mengetahui gen mana yang lebih aktif (upregulation) atau kurang aktif (downregulation) pada jaringan kanker dibandingkan jaringan normal. Dalam studi ini digunakan data publik GSE42568 dari database Gene Expression Omnibus (GEO), yang berisi 104 sampel jaringan kanker payudara dan 17 sampel jaringan payudara normal, yang dianalisis menggunakan platform Affymetrix.

Tujuan dari analisis ini adalah untuk mengetahui gen-gen apa saja yang berubah aktivitasnya pada kanker payudara, menampilkan 50 gen yang paling berbeda dalam bentuk heatmap agar perbedaannya terlihat jelas, serta melakukan analisis lebih lanjut menggunakan Gene Ontology (GO) dan KEGG untuk memahami fungsi biologis dan jalur molekuler yang terlibat. Dengan cara ini, diharapkan kita dapat memperoleh gambaran yang lebih jelas tentang bagaimana kanker payudara berkembang dan menemukan kemungkinan target baru untuk terapi di masa depan.

2. Metode

2.1 Dataset

Data yang digunakan dalam penelitian ini adalah dataset **GSE42568** yang diunduh dari database publik Gene Expression Omnibus (GEO) milik NCBI. Dataset ini berisi total 121 sampel, terdiri dari 104 sampel jaringan kanker payudara invasif dan 17 sampel jaringan payudara normal. Seluruh data dihasilkan menggunakan platform microarray Affymetrix Human Genome U133 Plus 2.0 (GPL570), yang mampu mengukur ekspresi ribuan gen secara bersamaan.

2.2 Pra-pemrosesan Data

Analisis dilakukan menggunakan software R dengan paket Bioconductor. Data diunduh dan dibaca menggunakan paket GEOquery, kemudian dikelola dalam format objek ekspresi menggunakan Biobase. Nilai ekspresi gen dinormalisasi dalam skala \log_2 agar distribusi data lebih stabil dan dapat dibandingkan antar sampel.

Untuk memastikan kualitas data, distribusi nilai ekspresi tiap sampel divisualisasikan menggunakan boxplot dan density plot. Selain itu, dilakukan analisis reduksi dimensi menggunakan metode UMAP (Uniform Manifold Approximation and Projection) untuk melihat apakah sampel kanker dan normal dapat terpisah secara jelas berdasarkan pola ekspresi gennya.

2.3 Analisis Differentially Expressed Genes (DEGs)

Identifikasi gen yang berbeda secara signifikan antara jaringan kanker dan jaringan normal dilakukan menggunakan paket limma (Linear Models for Microarray Data). Metode ini menggunakan pendekatan statistik empirical Bayes moderated t-test untuk meningkatkan akurasi estimasi varians.

Gen dikategorikan sebagai Differentially Expressed Genes (DEGs) jika memenuhi kriteria berikut:

- Adjusted p-value < 0.05 (menggunakan koreksi Benjamini–Hochberg untuk mengontrol False Discovery Rate/FDR)
- Nilai $|\log_2 \text{Fold Change}| \geq 1$ (artinya terjadi perubahan ekspresi minimal dua kali lipat)

2.4 Visualisasi

Hasil analisis divisualisasikan dalam beberapa bentuk grafik:

- Volcano plot dibuat menggunakan paket ggplot2 untuk menampilkan gen-gen signifikan berdasarkan kombinasi nilai p dan fold change.
- Heatmap dibuat untuk 50 DEGs teratas menggunakan paket pheatmap. Data dinormalisasi dalam bentuk z-score per baris untuk memperjelas perbedaan pola ekspresi antar sampel.

2.5 Analisis Enrichment

Untuk memahami makna biologis dari gen-gen yang berbeda secara signifikan, dilakukan analisis enrichment menggunakan paket clusterProfiler. Analisis ini bertujuan untuk mengetahui apakah DEGs yang ditemukan banyak terlibat dalam fungsi atau jalur biologis tertentu.

Analisis dilakukan terhadap:

- Gene Ontology (GO), khususnya kategori Biological Process (BP) dan Molecular Function (MF), menggunakan anotasi gen dari paket org.Hs.eg.db.
- KEGG Pathway, untuk mengidentifikasi jalur metabolisme dan sinyal yang berperan dalam kanker payudara.

Batas signifikansi yang digunakan adalah adjusted p-value < 0.05. Hasil analisis enrichment divisualisasikan menggunakan barplot dan dotplot dari paket enrichplot.

Seluruh proses analisis dilakukan menggunakan bahasa pemrograman R. Script telah diunggah pada link berikut:

<https://github.com/puspahening/BRSP-Omicslite/blob/main/Script%20R%20-%20Breast%20Cancer%203%20and%20Enrichment%20R.R>

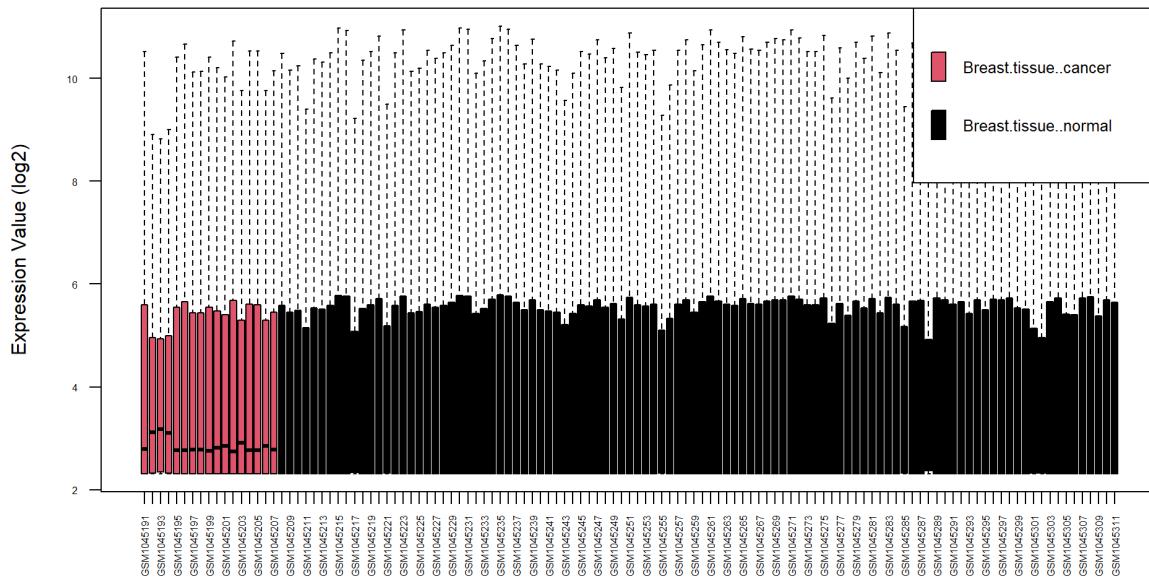
3. Hasil dan Interpretasi

3.1 Eksplorasi Kualitas Data

Boxplot Distribusi Nilai Ekspresi per Sampel

Boxplot menunjukkan distribusi nilai ekspresi gen (skala log₂) pada seluruh 121 sampel. Sampel kanker payudara (merah/pink) memiliki median ekspresi yang sedikit lebih rendah dibandingkan jaringan normal (hitam), namun rentang nilai keseluruhan relatif serupa antar sampel. Hal ini mengindikasikan bahwa normalisasi data telah berjalan dengan baik dan tidak terdapat batch effect yang mencolok antar sampel.

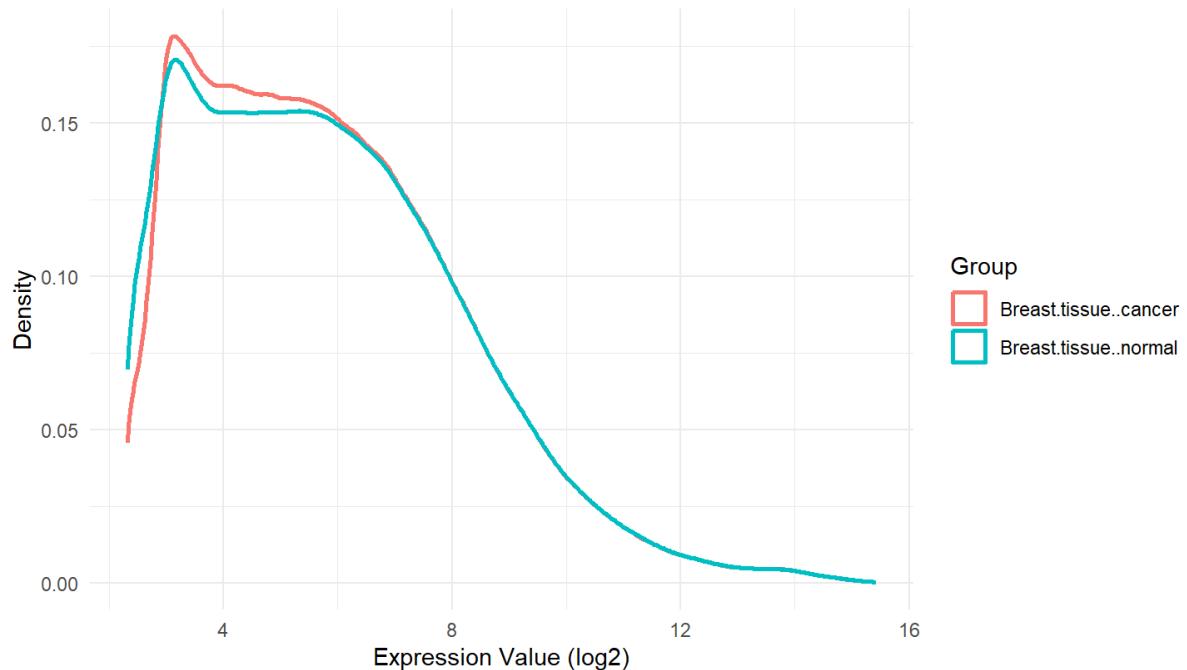
Boxplot Distribusi Nilai Ekspresi per Sampel



Density Plot Distribusi Nilai Ekspresi Gen

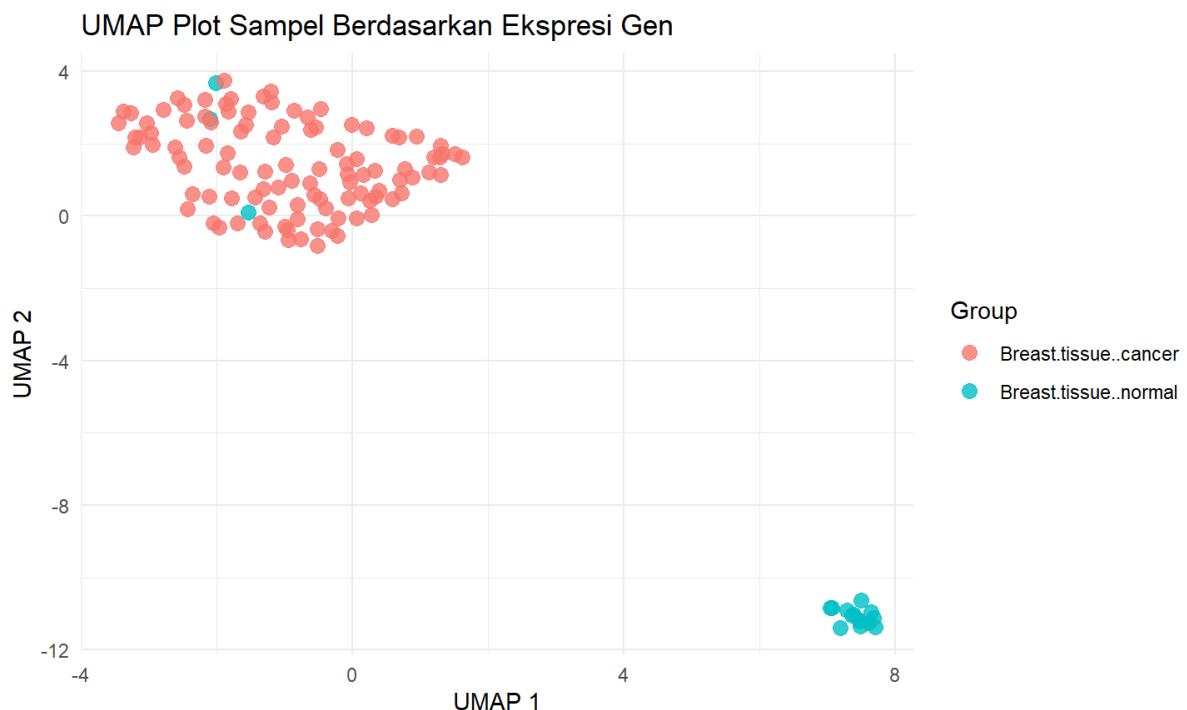
Density plot memperlihatkan distribusi bimodal yang khas untuk data microarray: puncak pertama mencerminkan gen-gen yang tidak atau rendah terekspresi, sedangkan ekor distribusi yang lebih landai mencerminkan gen-gen aktif. Kedua kelompok (kanker vs. normal) menunjukkan pola distribusi yang sebanding, meskipun terdapat sedikit pergeseran pada jaringan kanker di region ekspresi rendah - mengindikasikan bahwa sejumlah besar gen mengalami penekanan ekspresi pada kondisi kanker.

Distribusi Nilai Ekspresi Gen (Filtered)



UMAP Plot Sampel Berdasarkan Ekspresi Gen

Visualisasi UMAP secara jelas memisahkan sampel kanker payudara dari jaringan normal dalam dua kluster yang terpisah jauh di ruang dimensi rendah. Pemisahan yang tegas ini mengkonfirmasi bahwa terdapat perbedaan profil transkriptomik yang besar antara kedua kondisi. Beberapa sampel normal yang tampak berada dekat kluster kanker mungkin mencerminkan heterogenitas biologis atau variasi kualitas sampel individual.



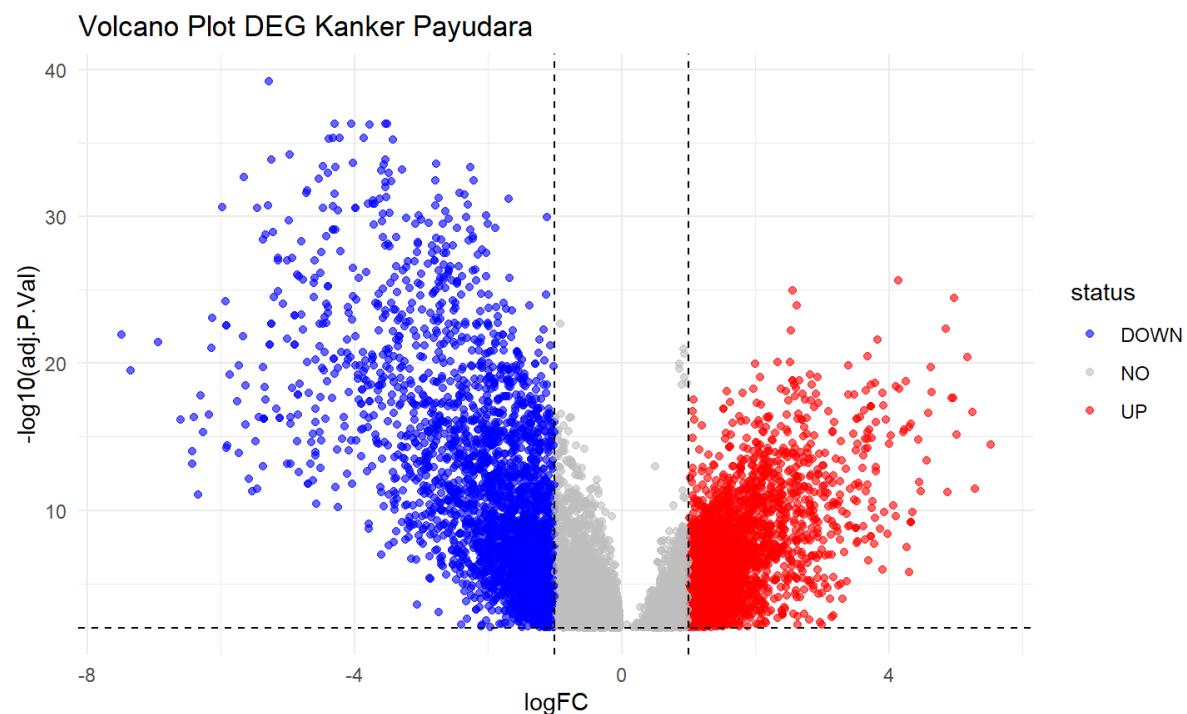
3.2 Volcano Plot – Upregulation dan Downregulation

Volcano plot menampilkan semua gen yang diuji dalam koordinat logFC (sumbu x) versus $-\log_{10}(\text{adjusted p-value})$ (sumbu y), dengan pembagian:

- Merah (UP): $\log FC > 1$ dan $\text{adj.p} < 0.05 \rightarrow$ gen upregulated pada kanker
- Biru (DOWN): $\log FC < -1$ dan $\text{adj.p} < 0.05 \rightarrow$ gen downregulated pada kanker
- Abu-abu (NO): tidak memenuhi kriteria signifikansi

Distribusi DEGs pada volcano plot bersifat asimetrik, dengan jumlah gen downregulated yang secara visual lebih banyak dan memiliki nilai perubahan ekspresi yang lebih ekstrem ($\log FC$ hingga -7 hingga -8) dibandingkan gen upregulated ($\log FC$ maksimal $\sim +5$ hingga $+6$). Hal ini mengindikasikan bahwa penekanan ekspresi gen merupakan fenomena dominan pada jaringan kanker payudara dibandingkan jaringan normal.

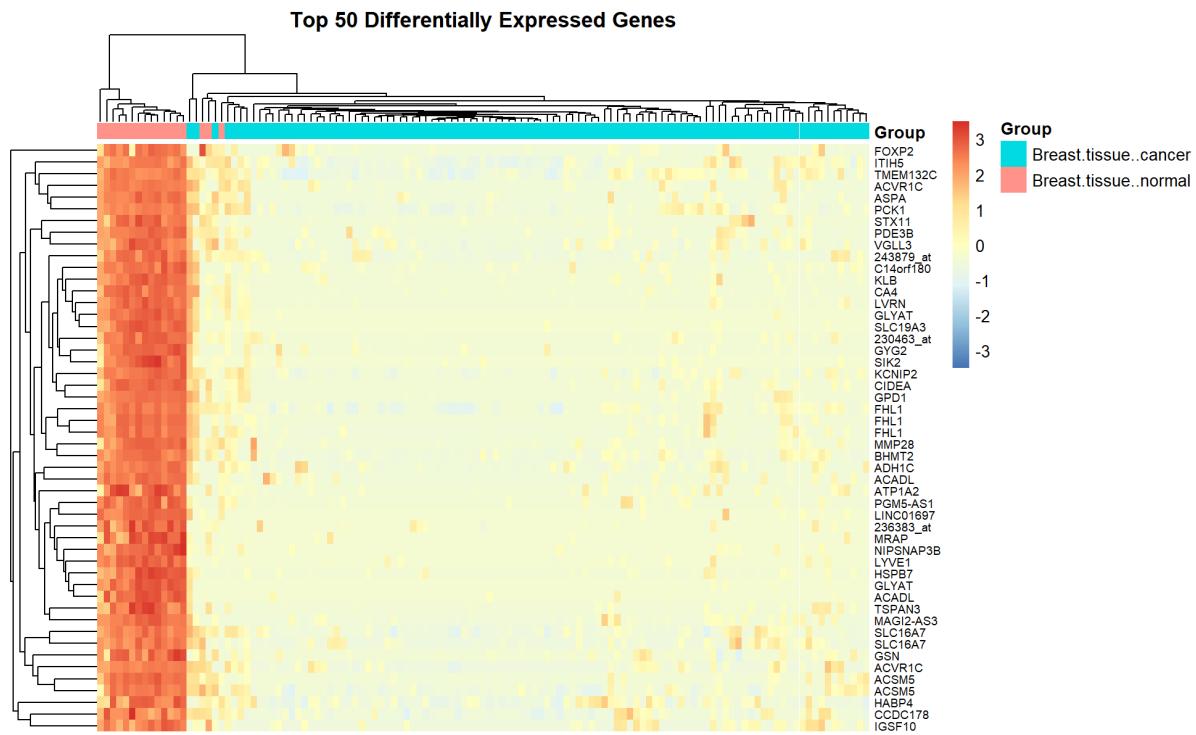
Gen-gen dengan nilai $-\log_{10}(p)$ tertinggi (mendekati 40) adalah kandidat biomarker yang paling robust secara statistik. Gen downregulated dengan $\log FC$ sangat negatif dan p-value sangat kecil kemungkinan besar merepresentasikan gen-gen supresor yang hilang fungsinya selama transformasi keganasan, seperti gen-gen yang mengontrol diferensiasi epitel dan homeostasis jaringan payudara normal.



3.3 Heatmap dengan 50 DEGs Teratas

Heatmap menampilkan pola ekspresi 50 DEGs teratas (berdasarkan adjusted p-value terkecil) dengan klusterisasi hierarki pada baris (gen) dan kolom (sampel). Pola yang terlihat sangat konsisten: sebagian besar dari 50 DEGs ini menunjukkan ekspresi tinggi pada jaringan normal (merah pekat, kluster pink kecil di kiri) dan ekspresi rendah pada jaringan kanker (kuning-

putih, kluster biru besar di kanan). Ini menegaskan kembali bahwa DEGs teratas didominasi oleh gen yang mengalami penekanan ekspresi pada kondisi kanker.



Beberapa gen menarik dalam heatmap beserta relevansi biologisnya:

Gen	Fungsi	Relevansi Kanker
FOXP2	Faktor transkripsi perkembangan	Regulasi diferensiasi sel; hilangnya ekspresi dikaitkan dengan dediferensiasi
CA4	Carbonic Anhydrase 4	Regulasi pH dan homeostasis; perubahan ekspresi terkait adaptasi metabolismik kanker
ACADM	Acyl-CoA Dehydrogenase	Oksidasi asam lemak mitokondria; penurunan ekspresi konsisten dengan efek Warburg
CIDEA	Cell death-inducing DFFA-like effector A	Apoptosis dan metabolisme lipid; downregulation dikaitkan dengan resistensi apoptosis
GSN	Gelsolin	Protein pengikat aktin; hilangnya ekspresi dikaitkan dengan invasivitas tumor
SLC16A7	Monocarboxylate transporter 2	Transpor laktat dan metabolisme energi sel
ACSM5	Acyl-CoA Synthetase Medium Chain	Aktivasi asam lemak; perubahan metabolisme lipid pada kanker

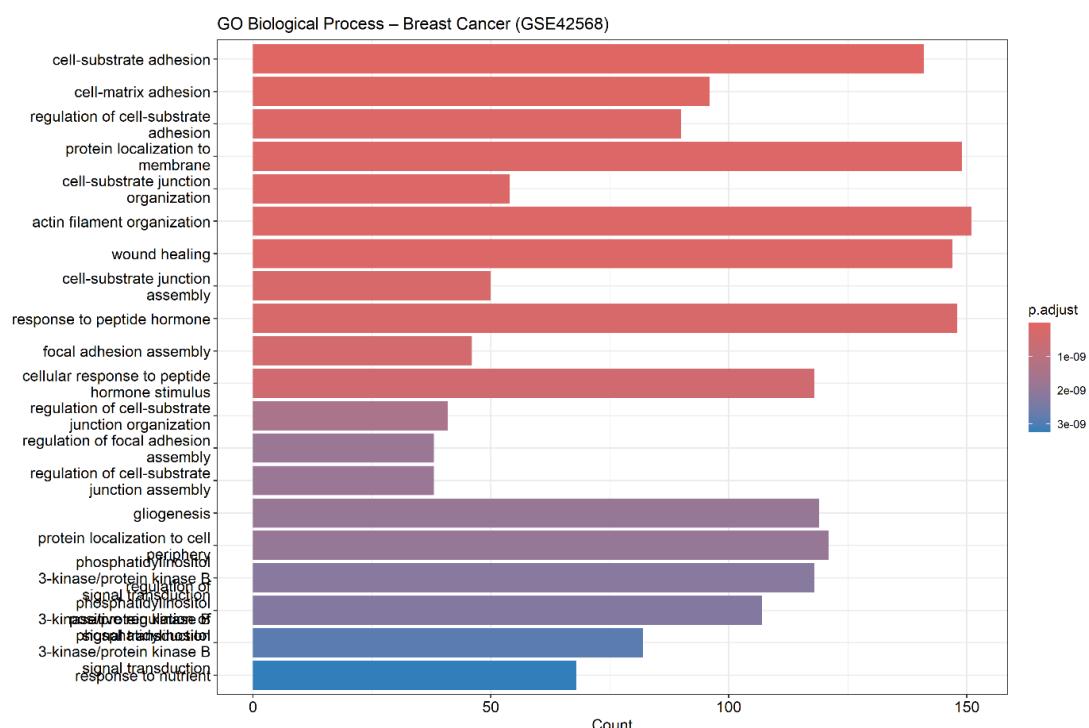
Gen	Fungsi	Relevansi Kanker
ADH1C	Alcohol dehydrogenase 1C	Metabolisme alkohol dan retinol; penurunan ekspresi sering ditemukan pada kanker payudara

Pola downregulation masif ini mencerminkan hilangnya program transkripsi normal sel epitel payudara selama proses transformasi keganasan.

3.4 Analisis Enrichment

3.4.1 Gene Ontology – Biological Process (GO-BP)

Barplot menampilkan jumlah gen (Count) yang terlibat dalam setiap term GO, diwarnai berdasarkan adjusted p-value. Dotplot menambahkan dimensi GeneRatio (proporsi gen DEG yang terlibat terhadap total gen dalam term tersebut).

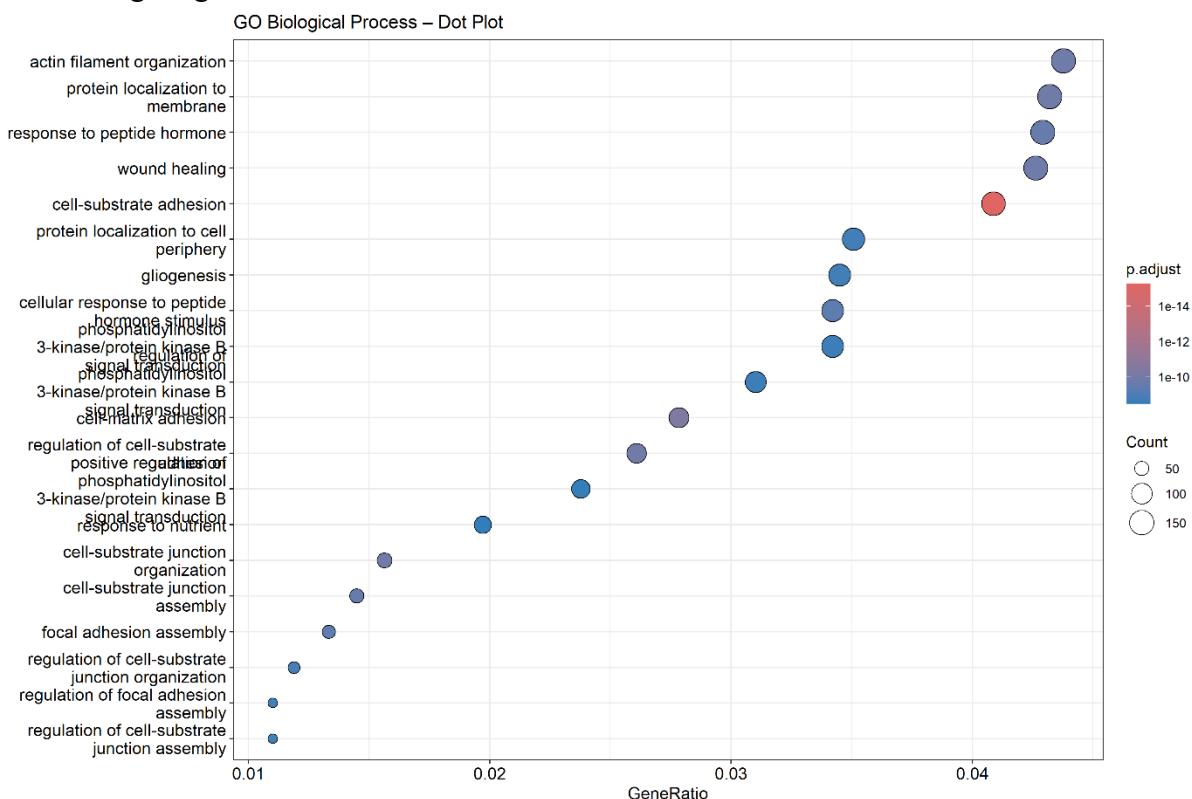


Proses biologis yang paling signifikan diperkaya:

- Actin filament organization (Count ~155, GeneRatio tertinggi ~0.042) merupakan term paling menonjol. Reorganisasi filamen aktin adalah ciri khas sel yang mengalami Epithelial-Mesenchymal Transition (EMT) - proses di mana sel epitel kehilangan sifat epitelialnya dan memperoleh fenotip mesenkimal yang lebih invasif dan motil.
- Cell-substrate adhesion dan cell-matrix adhesion mencerminkan perubahan luas pada interaksi sel kanker dengan matriks ekstraselular (ECM). Perubahan ini sangat penting dalam kemampuan sel tumor untuk melepaskan diri dari tumor primer, bermigrasi, dan membentuk metastasis di organ yang jauh.
- Wound healing (Count ~150) berbagi banyak jalur sinyal dengan invasi tumor - konsep ini dikenal sebagai "tumor sebagai luka yang tidak sembuh", di mana sel kanker

mengeksplorasi mekanisme penyembuhan luka untuk mendorong proliferasi dan angiogenesis.

- Phosphatidylinositol 3-kinase/protein kinase B signal transduction (jalur PI3K/AKT) teridentifikasi pada beberapa term yang saling terkait. Jalur PI3K/AKT merupakan salah satu jalur yang paling sering teraktivasi pada kanker payudara, mengontrol pertumbuhan sel, kelangsungan hidup, dan metabolisme. Aktivasinya sering terjadi melalui mutasi PIK3CA atau hilangnya PTEN.
- Protein localization to membrane dan gliogenesis mencerminkan reorganisasi arsitektur seluler dan plastisitas fenotipik sel kanker dalam beradaptasi terhadap mikrolingkungan tumor.



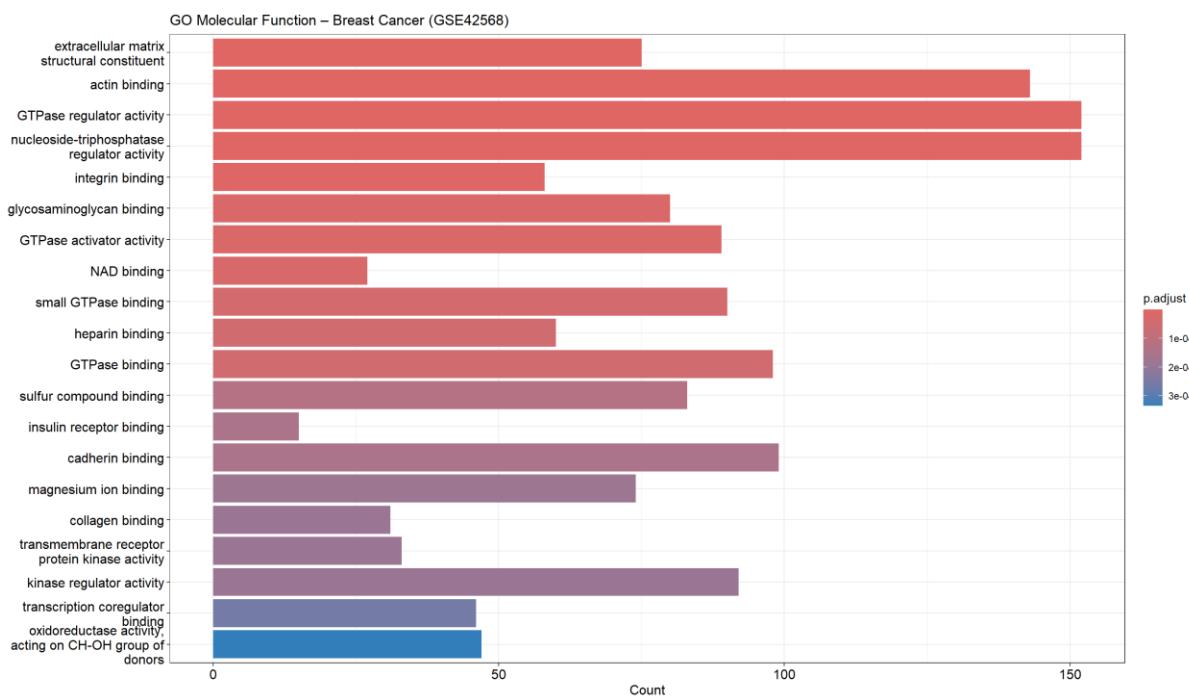
3.4.2 Gene Ontology – Molecular Function (GO-MF)

Fungsi molekuler yang paling signifikan diperkaya:

- Actin binding (Count ~145) berada di posisi teratas, konsisten dengan temuan GO-BP tentang reorganisasi sitoskeleton aktin. Gen-gen dengan fungsi pengikatan aktin mencakup berbagai protein sitoskeletal yang berperan dalam motilitas sel, pembentukan lamellipodia, dan invasi tumor.
- GTPase regulator activity dan GTPase activator activity (Count >150) mengindikasikan keterlibatan luas protein-protein GTPase kecil dari famili Rho dan Ras. Protein-protein ini merupakan regulator kunci motilitas sel, adhesi, dan proliferasi - dan sering mengalami disregulasi pada kanker melalui mutasi onkogen (misalnya KRAS, HRAS) atau perubahan ekspresi regulator GTPase.
- Extracellular matrix structural constituent mencerminkan perubahan masif pada komposisi ECM, yang berkontribusi pada pembentukan desmoplastic stroma khas

kanker payudara - suatu lingkungan yang secara aktif mendukung pertumbuhan dan invasi tumor.

- Integrin binding dan glycosaminoglycan binding terkait dengan perubahan interaksi sel-ECM melalui reseptor integrin. Perubahan ekspresi integrin mengatur kemampuan adhesi, migrasi, dan invasi sel kanker, serta mengaktifkan jalur sinyal intraseluler pro-survival.
- Insulin receptor binding dan cadherin binding mengindikasikan keterlibatan sinyal insulin/IGF-1 (yang berinteraksi dengan jalur PI3K/AKT) serta kehilangan E-cadherin - penanda klasik EMT yang menghancurkan sambungan adhesi antar-sel epitel dan membebaskan sel untuk bermigrasi.
- Oxidoreductase activity (CH-OH group donors) mencerminkan perubahan aktivitas enzim metabolismik, konsisten dengan reprogramming metabolismik sel kanker.



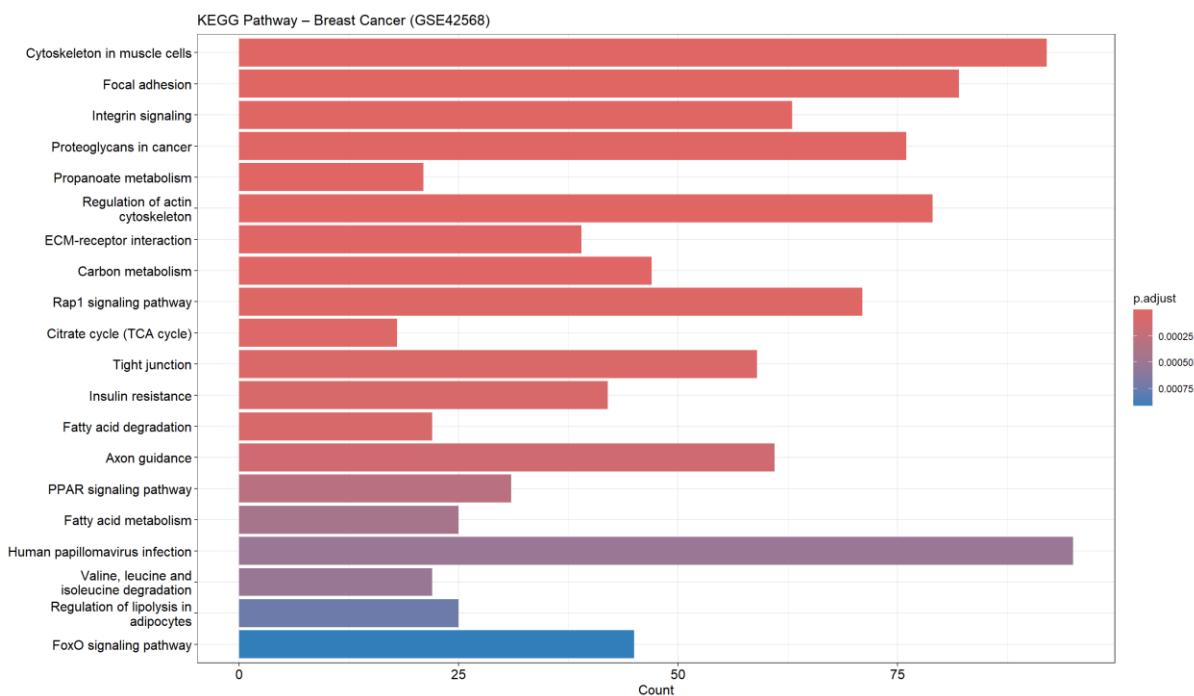
3.4.3 KEGG Pathway

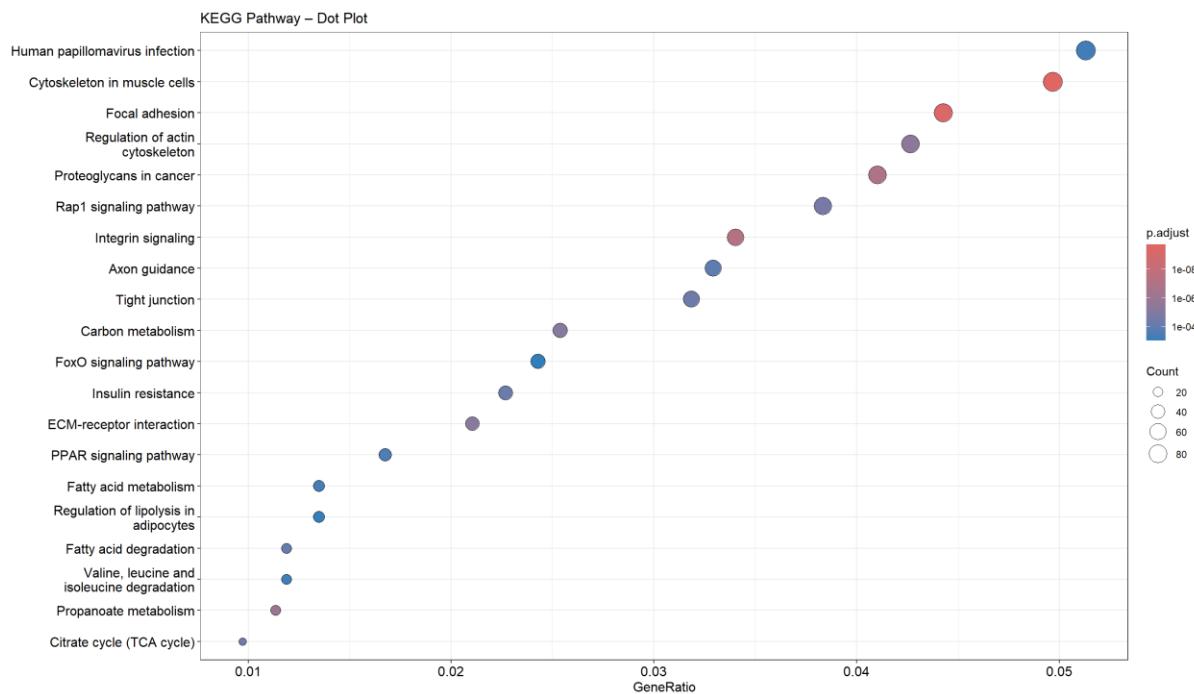
Jalur KEGG yang paling signifikan diperkaya:

- Cytoskeleton in muscle cells (Count ~90, GeneRatio ~0.051) dan Regulation of actin cytoskeleton (Count ~80) berada di posisi teratas, memperkuat temuan GO tentang pentingnya reorganisasi sitoskeleton dalam patobiologi kanker payudara ini.
- Focal adhesion (Count ~85, GeneRatio ~0.044) merupakan jalur yang sangat relevan dengan kanker. Focal adhesion adalah kompleks protein yang menghubungkan ECM (melalui integrin) dengan sitoskeleton aktin intraseluler, dan sekaligus berfungsi sebagai platform sinyal untuk jalur FAK (Focal Adhesion Kinase), Src, dan PI3K. Aktivasi berlebihan jalur ini mendukung kelangsungan hidup sel kanker dan invasi.
- Proteoglycans in cancer (Count ~75) secara langsung relevan dengan konteks kanker payudara. Proteoglikan berperan dalam remodeling ECM, modulasi sinyal

pertumbuhan (termasuk FGF, Wnt, Hedgehog), dan pembentukan niche metastasis tulang yang menjadi target metastasis umum kanker payudara.

- Integrin signaling dan Rap1 signaling pathway (Count ~70) saling terkait: Rap1 adalah GTPase kecil yang mengatur aktivasi integrin dan adhesi sel, serta berinteraksi dengan cadherin dalam mengontrol sambungan antar-sel.
- ECM-receptor interaction menggambarkan keseluruhan disregulasi interaksi antara sel kanker dan ECM-nya - komponen fundamental dari invasi dan metastasis kanker.
- FoxO signaling pathway (Count ~45) menarik karena faktor transkripsi FoxO berperan sebagai tumor suppressor yang mengontrol apoptosis, penghentian siklus sel, dan respons terhadap stres oksidatif. Penekanan jalur ini (sering terjadi akibat aktivasi PI3K/AKT) berkontribusi pada resistensi apoptosis sel kanker.
- Insulin resistance dan PPAR signaling pathway mencerminkan perubahan regulasi metabolisme glukosa dan lipid. Bersama dengan Fatty acid metabolism, Fatty acid degradation, dan Citrate cycle (TCA cycle) yang juga signifikan, temuan ini secara kolektif mengindikasikan adanya reprogramming metabolik yang luas pada sel kanker payudara - termasuk penurunan oksidasi asam lemak mitokondria dan pergeseran ke metabolisme glikolitik aerobik (efek Warburg).
- Human papillomavirus infection (Count ~90, GeneRatio tertinggi ~0.051) muncul dengan jumlah gen yang besar. Pengayaan jalur ini perlu diinterpretasikan dengan hati-hati; kemungkinan besar mencerminkan tumpang-tindih dengan gen-gen yang terlibat dalam regulasi siklus sel dan respons terhadap stres genotoksik, yang merupakan proses umum dalam karsinogenesis, bukan berarti adanya infeksi HPV aktif pada sampel-sampel ini.





4. Kesimpulan

Analisis ekspresi gen diferensial pada dataset GSE42568 menghasilkan beberapa kesimpulan utama:

- Pemisahan transkriptomik yang tegas antara sampel kanker dan normal (dikonfirmasi UMAP), dengan kualitas data yang baik berdasarkan distribusi nilai ekspresi.
- Downregulation mendominasi lanskap DEGs pada kanker payudara - lebih banyak gen yang mengalami penekanan ekspresi dibandingkan peningkatan, mencerminkan hilangnya program transkripsi normal sel epitel payudara.
- 50 DEGs teratas didominasi oleh gen-gen yang berhubungan dengan adhesi sel, metabolisme lipid, dan dinamika sitoskeleton, yang semuanya mengalami penurunan ekspresi kuat pada jaringan kanker.
- Analisis GO mengidentifikasi proses biologis kunci: reorganisasi filamen aktin, adhesi sel-matriks, penyembuhan luka, dan sinyal PI3K/AKT; serta fungsi molekuler utama: pengikatan aktin, regulasi GTPase, dan komponen struktural ECM.
- Analisis KEGG memperkuat temuan GO dengan jalur focal adhesion, regulasi sitoskeleton aktin, proteoglikan dalam kanker, sinyal Rap1, FoxO signaling, dan reprogramming metabolisme lipid sebagai jalur yang terganggu secara konsisten.

Secara keseluruhan, temuan ini konsisten dengan hallmarks kanker payudara yang dikenal dalam literatur, yaitu disregulasi adhesi sel, peningkatan motilitas dan invasivitas (melalui EMT), aktivasi sinyal pro-survival (PI3K/AKT), serta reprogramming metabolismik. Gen-gen dan jalur yang teridentifikasi merupakan kandidat menarik untuk penelitian lanjutan sebagai biomarker diagnostik maupun target terapeutik potensial.

