

Week 3 - Case Analysis

A. Pendahuluan

Perkembangan teknologi *high-throughput* seperti *microarray* dan *Next-Generation Sequencing* (NGS) telah memungkinkan identifikasi ribuan gen secara simultan. Namun, teknologi ini menghasilkan daftar data yang sangat besar, sehingga diperlukan metode bioinformatika untuk menerjemahkan data numerik tersebut ke dalam pemahaman biologis yang bermakna. Analisis ekspresi gen sering kali menghasilkan daftar *Differentially Expressed Genes* (DEGs), yaitu gen-gen yang mengalami perubahan ekspresi signifikan secara statistik antara dua kondisi biologis yang berbeda.

Untuk memahami mekanisme di balik DEGs tersebut, digunakan pendekatan analisis fungsional melalui *Gene Ontology* (GO) dan KEGG *Pathway Analysis*.

1. **Gene Ontology (GO)**: menyediakan kosakata terstruktur untuk mendeskripsikan peran gen dalam tiga domain utama:
 - a. **Biological Process** (BP): Fenomena biologis yang dicapai melalui satu atau lebih rangkaian aktivitas molekuler (misal: mitosis, metabolisme purin, atau respon imun).
 - b. **Molecular Function** (MF): Aktivitas fisik atau kimia dari produk gen pada tingkat molekuler (misal: aktivitas katalitik, pengikatan reseptor/binding, atau aktivitas transporter).
 - c. **Cellular Component** (CC): Lokasi spesifik di dalam struktur sel tempat produk gen aktif (misal: mitokondria, ribosom, atau nukleus).
2. **KEGG Pathway**: *Kyoto Encyclopedia of Genes and Genomes* merupakan database referensi yang memetakan gen ke dalam jalur metabolisme, jalur persinyalan, dan proses seluler. KEGG merepresentasikan interaksi molekuler kompleks di dalam sel yang menggambarkan bagaimana gen-gen bekerja sama dalam satu sistem sistemik.

Aplikasi Klinis dan Personalized Drug Therapy: Integrasi dari analisis GO dan KEGG bertujuan untuk mendukung pengembangan *Personalized Drug Therapy* (Kedokteran Presisi). Dengan mengetahui jalur biologis spesifik yang terganggu pada individu tertentu, klinisi dapat menentukan target pengobatan yang lebih akurat dan personal berdasarkan profil genetik unik pasien, bukan lagi menggunakan pendekatan "satu obat untuk semua orang" (*one-size-fits-all*).

B. Metode Analisis

Metode dalam praktikum ini dibagi menjadi dua tahap utama: identifikasi fungsional gen melalui analisis *enrichment* dan visualisasi jalur biologis secara interaktif.

1. Persiapan Data dan Lingkungan Kerja (R & Bioconductor)

Sebelum melakukan analisis jalur, dilakukan persiapan data menggunakan bahasa pemrograman R untuk menghasilkan daftar gen yang terekspresi secara diferensial (*Differentially Expressed Genes* atau DEGs).

- a. **Instalasi Perangkat Lunak:** Menyiapkan **R** sebagai mesin pemrosesan data dan **RStudio** sebagai antarmuka kerja (*interface*).
- b. **Manajemen Paket (Bioconductor):** Menginstal BiocManager untuk mengakses pustaka khusus bioinformatika yang tidak tersedia di repositori CRAN standar.
- c. **Pemuatan Library Utama:**
 - 1) limma: Digunakan sebagai mesin statistik untuk membangun model linear, menghitung *P-value*, dan menentukan DEG.
 - 2) GEOquery: Digunakan untuk mengunduh dataset ekspresi gen secara langsung dari database NCBI Gene Expression Omnibus (GEO).
 - 3) ggplot2 & pheatmap: Digunakan untuk memvisualisasikan distribusi data melalui *volcano plot* dan *heatmap*.
- d. **Pra-pemrosesan Data:** Data mentah dikonversi ke skala log2 untuk menstabilkan varians dan memastikan data memenuhi asumsi model statistik.
- e. **Gene ID Conversion:** Melakukan pemetaan (*mapping*) ID probe (misalnya format Illumina atau Affymetrix) ke dalam *Gene Symbol* atau *Entrez ID* agar dapat dikenali oleh sistem anotasi fungsional.

2. Analisis Fungsional dengan g:Profiler

Analisis ini bertujuan untuk menentukan apakah suatu istilah *Gene Ontology* (GO) atau jalur KEGG muncul lebih sering dari yang diharapkan secara kebetulan (*Enrichment Test*).

- a. **Input Query:** Mengakses laman **g:Profiler**, kemudian memasukkan daftar ID gen (misal: ID Illumina) yang telah diurutkan berdasarkan nilai *Log Fold Change* (LogFC) ke dalam kolom *Query*.
- b. **Konfigurasi Parameter:** Organisme pilih *Homo sapiens*.
- f. **Eksekusi:** Klik **Run Query** untuk menjalankan algoritma statistik.
- g. **Evaluasi Hasil:** Hasil dievaluasi melalui **Manhattan Plot**. Jalur dengan nilai padj terkecil (paling signifikan) diidentifikasi,

3. Visualisasi Jalur dengan KEGG Mapper – Color Tool

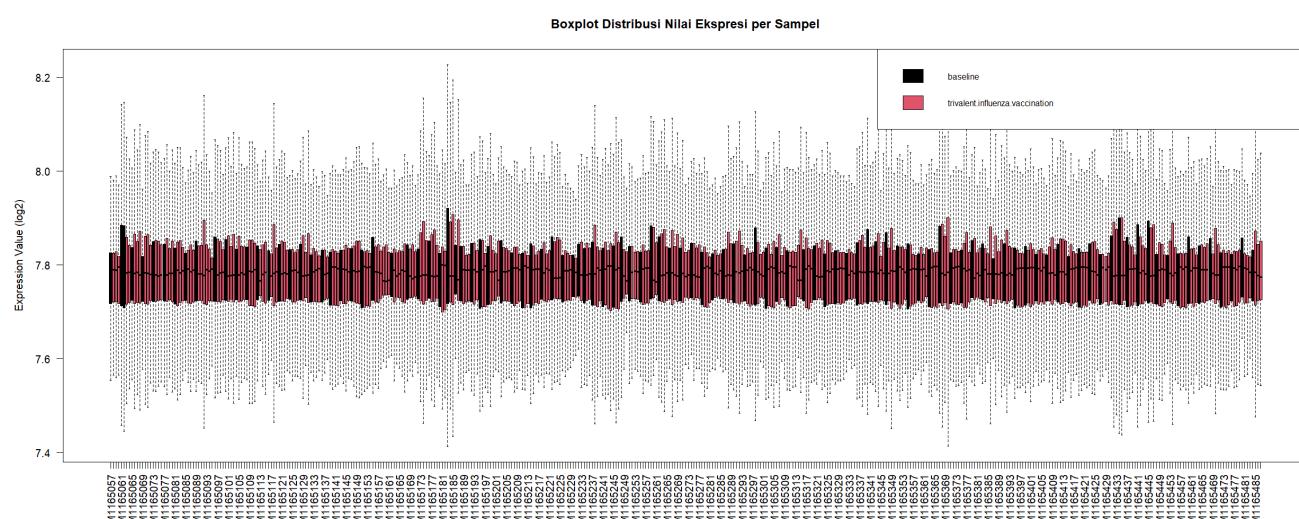
Tahap ini bertujuan untuk memetakan gen-gen terpilih ke dalam diagram jalur koordinasi seluler guna melihat interaksi antar gen secara visual.

- a. **Akses Tool:** Membuka menu KEGG Mapper – Color Tool.
- b. **Pewarnaan Gen (Color Specification):** Memasukkan daftar ID gen ke kotak input dengan menambahkan keterangan warna secara manual:
 - 1) ID_GEN red: Untuk gen yang mengalami peningkatan ekspresi (*upregulated*).
 - 2) ID_GEN yellow: Untuk gen yang mengalami penurunan ekspresi (*downregulated*).
- c. **Konfigurasi Mapping:** Mencentang opsi **Include aliases** untuk memastikan ID Illumina dapat dipetakan secara akurat ke dalam ID internal database KEGG.
- d. **Visualisasi:** Klik **Exec** untuk menghasilkan peta jalur (pathway) yang berwarna, yang menunjukkan posisi gen dalam sistem biologis yang relevan.

C. Hasil dan Interpretasi

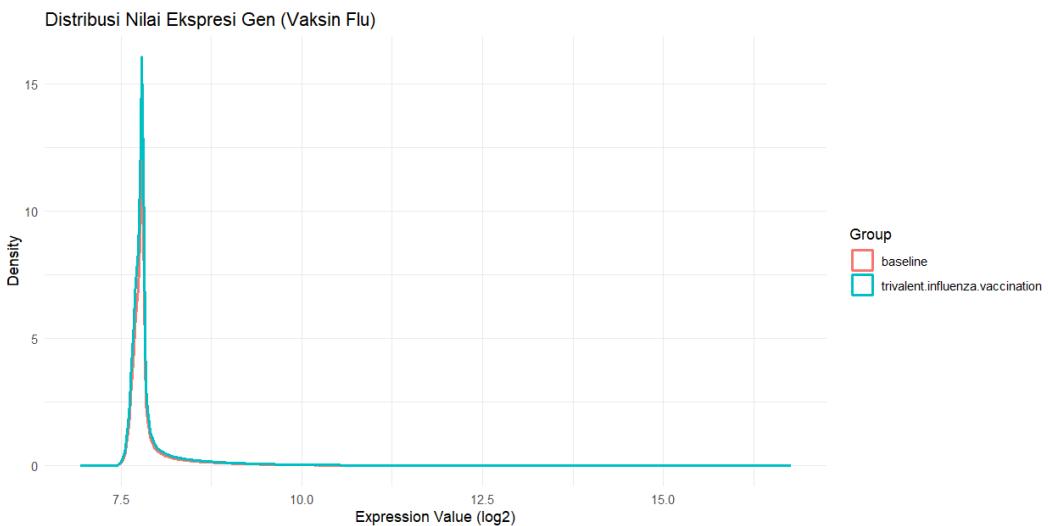
Modul	: Analisis Ekspresi Gen Respon Vaksin Flu
Dataset	: GSE48018 (Vaccine vs Baseline)
Platform	: Microarray (Illumina HumanHT-12 V4.0 - GPL10558)
Tujuan	: Mengidentifikasi Differentially Expressed Genes (DEG)

1. Boxplot



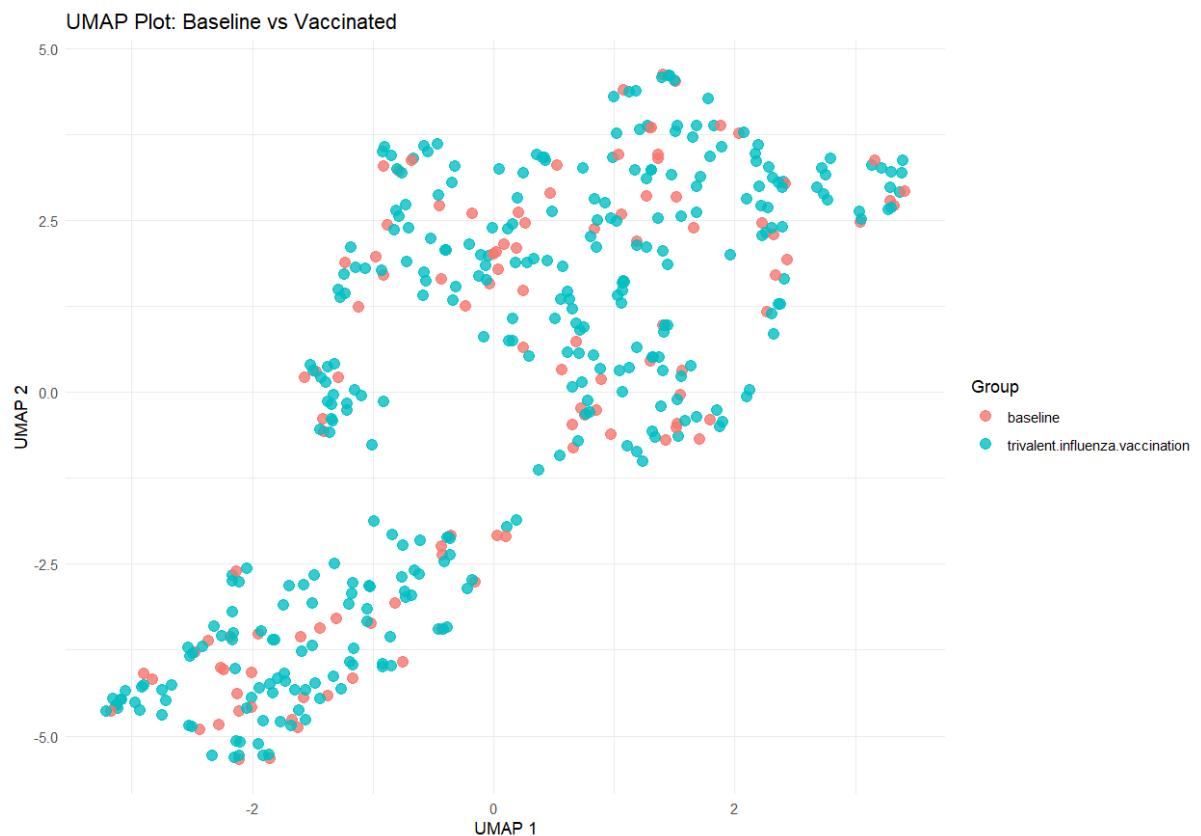
Interpretasi: Tinggi kotak dan garis median antar sampel terlihat sejajar (rata), artinya **proses normalisasi berhasil**. Hal ini memastikan bahwa perbedaan ekspresi yang muncul nanti benar-benar karena faktor biologis, bukan karena kesalahan teknis alat.

2. Density Plot



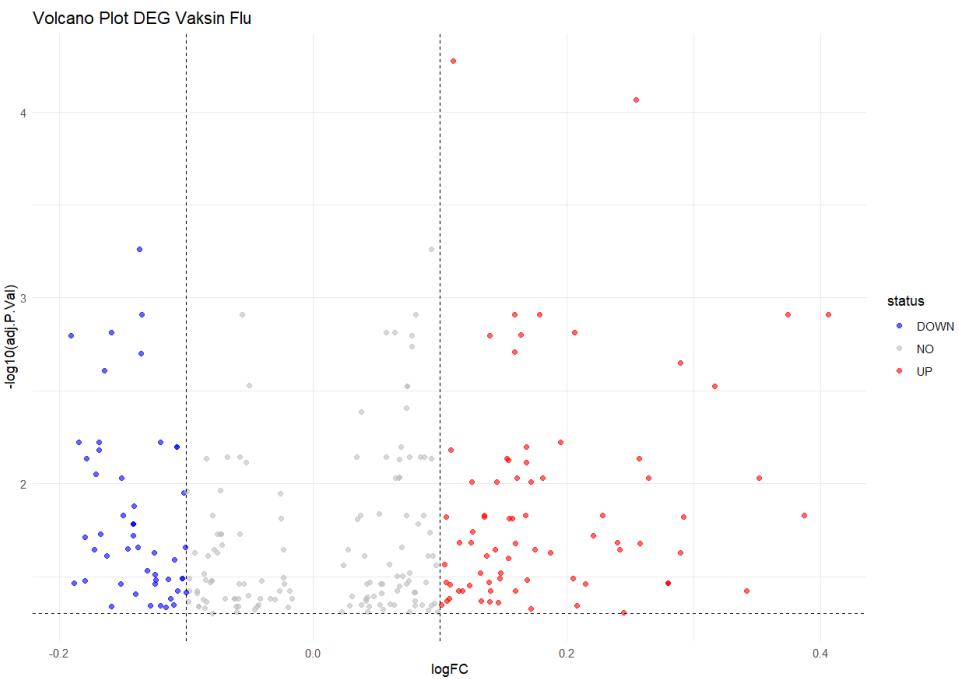
Interpretasi: Kurva menumpuk di area yang sama, berarti **distribusi data sudah seragam** dan siap untuk dianalisis lebih lanjut menggunakan model linear (*limma*).

3. UMAP Plot



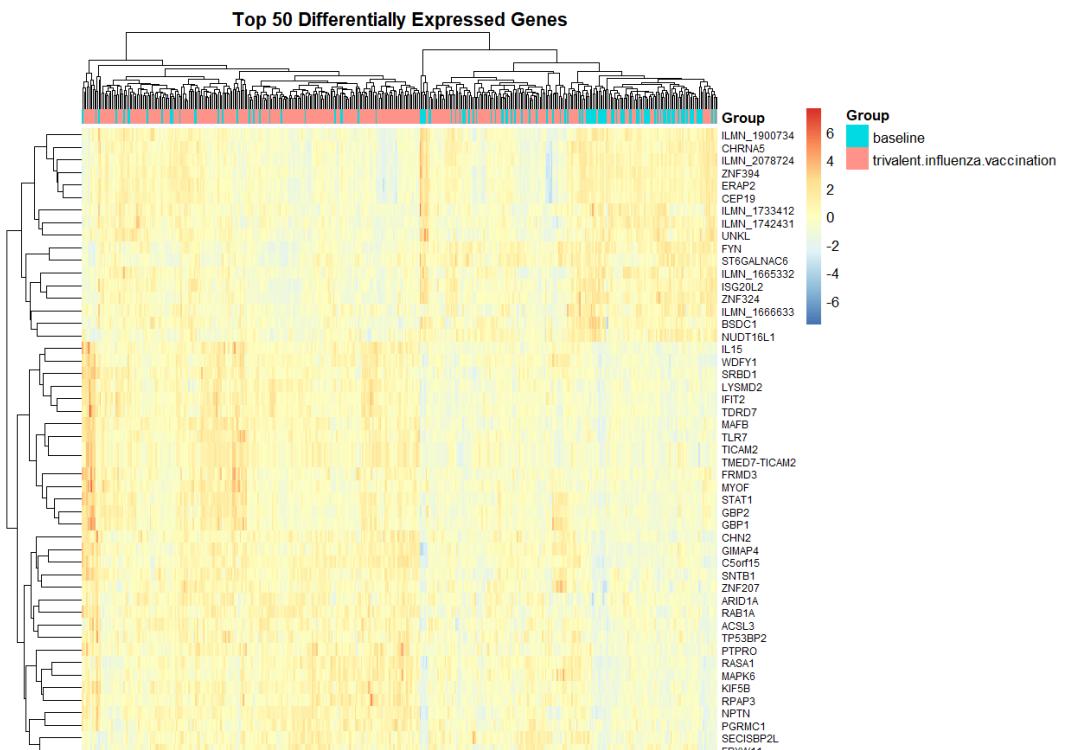
Interpretasi: Titik-titik berkelompok sesuai warnanya, berarti **ada perbedaan profil genetik yang nyata** antar kelompok tersebut. Ini membuktikan bahwa perlakuan memberikan efek yang konsisten.

4. Volcano Plot



Interpretasi: Gen yang berada di pojok kanan atas (merah) adalah gen yang naik signifikan, sedangkan di pojok kiri atas (biru) adalah gen yang turun signifikan. Titik abu-abu adalah gen yang tidak berubah secara bermakna.

5. Heatmap (urutan DEG teratas)



Interpretasi: Heatmap memungkinkan kita melihat pola genetik. Terdapat blok warna merah yang konsisten pada kelompok pasien tertentu, gen-gen tersebut bisa dijadikan **biomarker** untuk diagnosis atau target terapi terpersonalisasi.

6. Gen Up-regulated

No	ID	adj,P,Val	P,Value	t	B	logFC	Gene,symbol	Gene,title
1	ILMN_2123312	0,00159	5,89E-07	5,07	509.181	0,191	ERAP2	endoplasmic reticulum aminopeptidase 2
2	ILMN_1716105	0,03427	1,15E-04	3,89	0,06112	0,1886	NLRP12	NLR family pyrin domain containing 12
3	ILMN_2288483	0,006	3,71E-06	4,69	3.325	0,1848	CEP19	centrosomal protein 19
4	ILMN_2070052	0,03325	1,05E-04	3,92	0,14386	0,1803		
5	ILMN_1903914	0,01945	4,03E-05	4,15	104.942	0,1798		
6	ILMN_2361104	0,00730	7,41E-06	4,54	266.291	0,1788	ZMAT3	zinc finger matrin-type 3
7	ILMN_1784737	0,02262	5,45E-05	4,08	0,7643	0,1728	S1PR4	sphingosine-1-phosphate receptor 4
8	ILMN_2235745	0,00885	1,02E-05	4,47	235.916	0,1714		
9	ILMN_2249920	0,006	3,99E-06	4,67	325.539	0,1692	FYN	FYN proto-oncogene, Src family tyrosine kinase
10	ILMN_2115949	0,00661	5,20E-06	4,61	300.184	0,1692	ZNF394	zinc finger protein 394

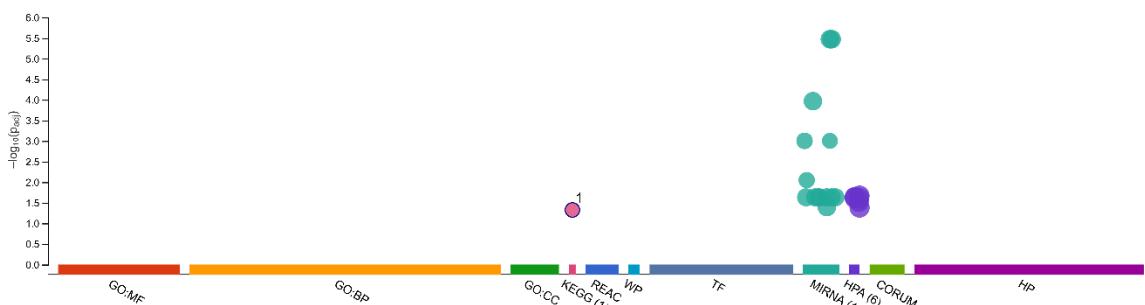
7. Gen Down-Regulated

No	ID	adj,P,Val	P,Value	t	B	logFC	Gene,symbol	Gene,title
1	ILMN_1739428	0,00123	2,60E-07	-5,2	588.061	-0,406	IFIT2	interferon induced protein with tetratricopeptide repeats 2
2	ILMN_2388547	0,01477	2,44E-05	-4,3	152.606	-0,387	EPSTI1	epithelial stromal interaction 1 (breast)
3	ILMN_1777325	0,00123	2,08E-07	-5,3	609.853	-0,374	STAT1	signal transducer and activator of transcription 1
4	ILMN_2114568	0,00933	1,23E-05	-4,4	218.193	-0,351	GBP5	guanylate binding protein 5
5	ILMN_1729749	0,03767	1,39E-04	-3,8	-0,12266	-0,341	HERC5	HECT and RLD domain containing E3 ubiquitin protein ligase 5
6	ILMN_1701114	0,00297	1,65E-06	-4,9	410.236	-0,316	GBP1	guanylate binding protein 1
7	ILMN_2148785	0,01517	2,63E-05	-4,3	145.414	-0,291	GBP1	guanylate binding protein 1

8	ILMN_1769520	0,02354	5,83E-05	-4,1	0,69971	-0,289	UBE2L6	ubiquitin conjugating enzyme E2 L6
9	ILMN_1774077	0,00224	1,06E-06	-5	452.613	-0,289	GBP2	guanylate binding protein 2
10	ILMN_2261600	0,03427	1,14E-04	-3,9	0,06527	-0,279	FCGR1B	Fc fragment of IgG receptor Ib

8. Gene Ontology (GO)

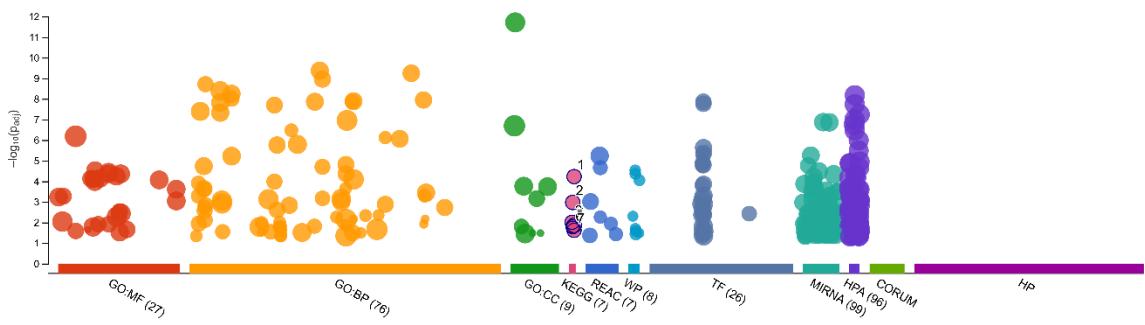
a. Manhattan Plot dari gen Down-Regulated



version
date
organism

g:Profiler

b. Manhattan Plot dari gen Down-Regulated



version
date
organism

g:Profiler

Manhattan plot ini memvisualisasikan hasil pengayaan fungsional dari daftar gen (DEGs) yang dianalisis.

- a. **Sumbu X:** Menunjukkan sumber database fungsional yang digunakan, seperti GO (Biological Process, Molecular Function, Cellular Component) dan KEGG.
- b. **Sumbu Y:** Menunjukkan tingkat signifikansi statistik dalam skala $-\log_{10}(\text{padj})$. Semakin tinggi titik pada grafik, semakin kecil nilai p-value-nya, yang berarti jalur tersebut semakin signifikan secara biologis.
- c. **Interpretasi:** Titik tertinggi pada area berwarna pink (KEGG) menunjukkan jalur **Influenza-A** dan **NOD-like receptor signaling pathway** sebagai jalur yang paling dominan dipengaruhi oleh daftar gen yang dimasukkan.

3. KEGG Pathway

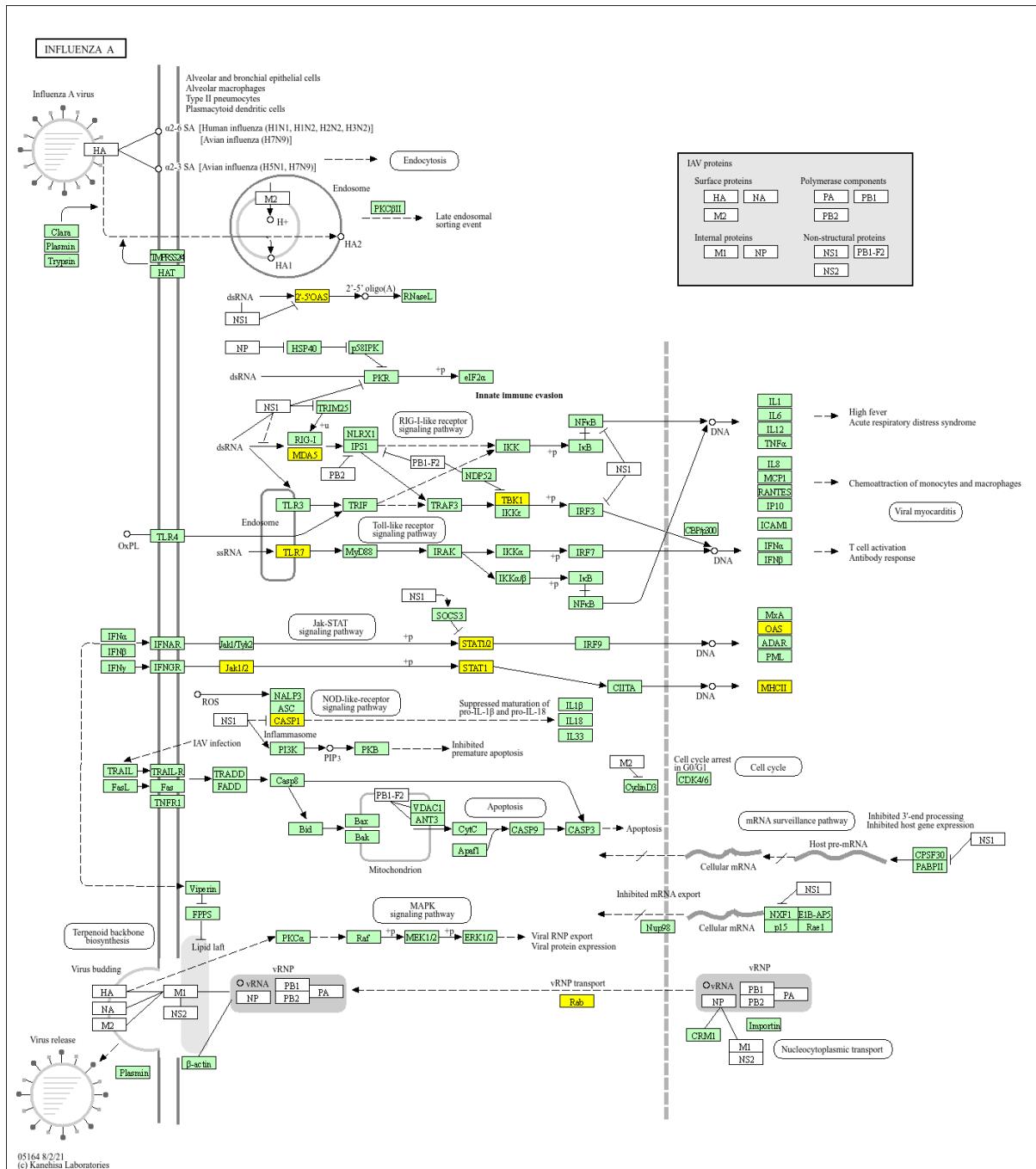
Gen up-regulated berwarna merah

Gen down-regulated berwarna kuning

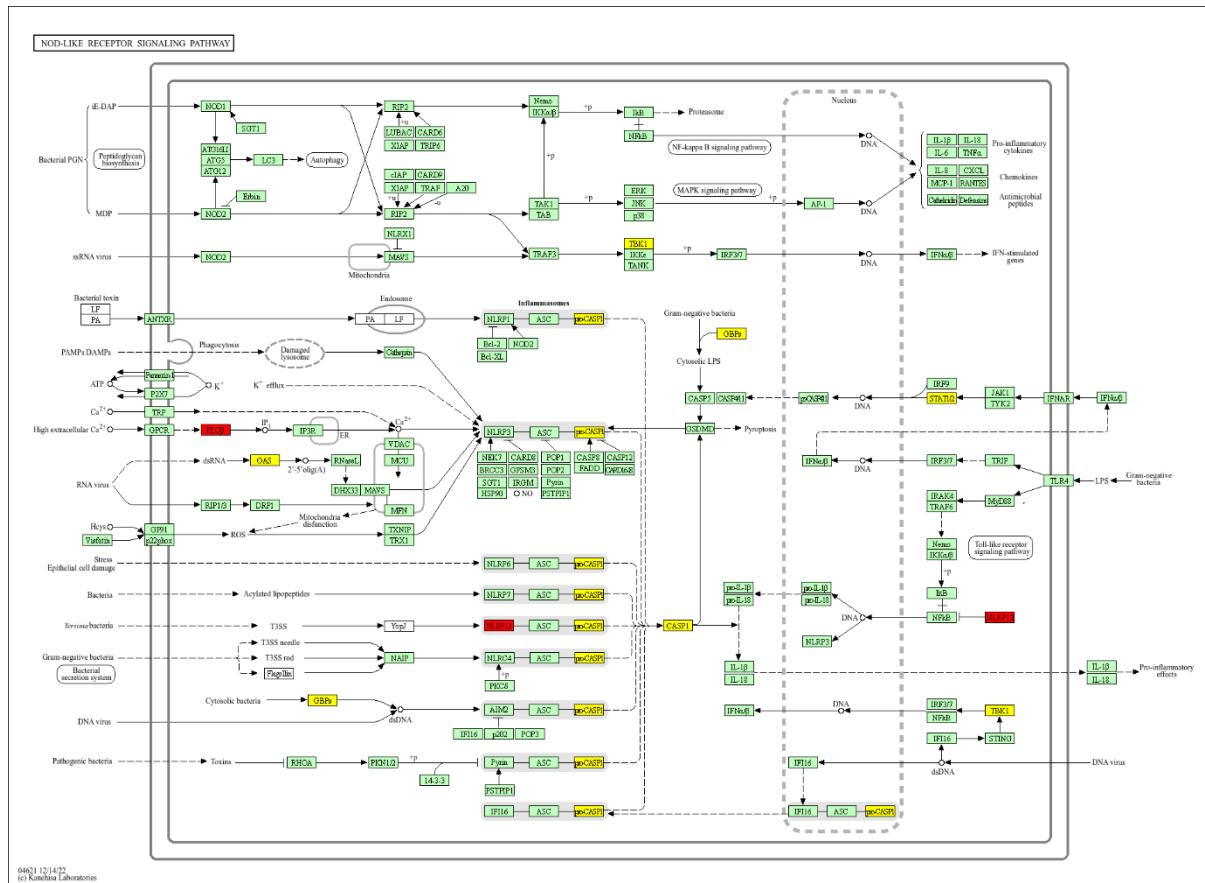
Visualisasi ini menggunakan *KEGG Color Tool* untuk memetakan DEGs ke dalam jalur metabolisme spesifik.

- a. **Warna Merah:** Merepresentasikan gen yang mengalami peningkatan ekspresi (*upregulated*) atau memiliki nilai LogFC positif.
- b. **Warna Biru:** Merepresentasikan gen yang mengalami penurunan ekspresi (*downregulated*) atau memiliki nilai LogFC negatif.
- c. **Kotak Kuning/Tanpa Warna:** Menunjukkan komponen dalam jalur tersebut yang tidak ditemukan dalam daftar gen input atau tidak mengalami perubahan ekspresi yang signifikan.
- d. **Interpretasi:** Sebaran warna pada peta jalur memberikan gambaran visual apakah suatu mekanisme seluler sedang diaktifkan atau ditekan secara keseluruhan.

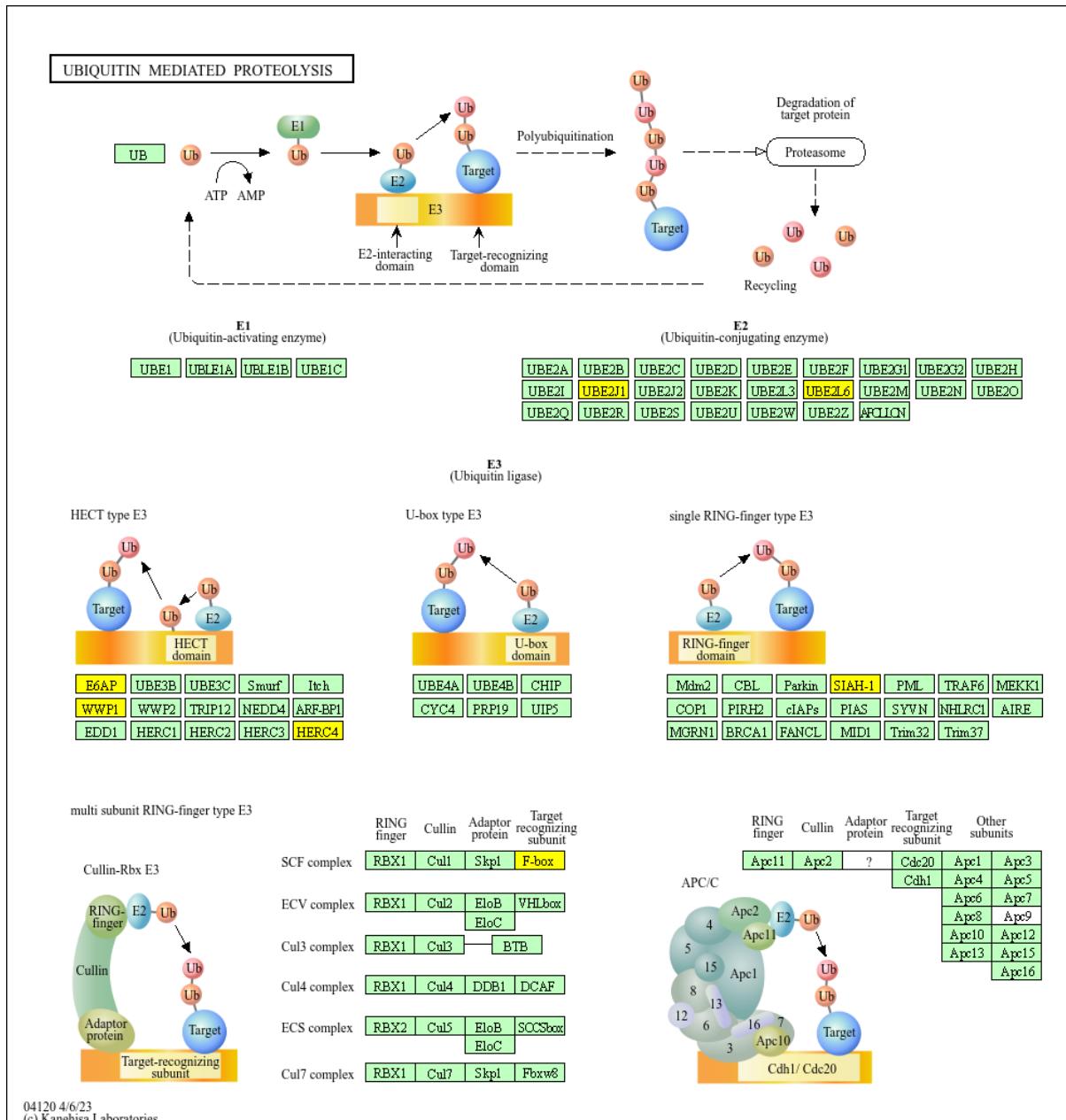
a. hsa05164 Influenza A (10)



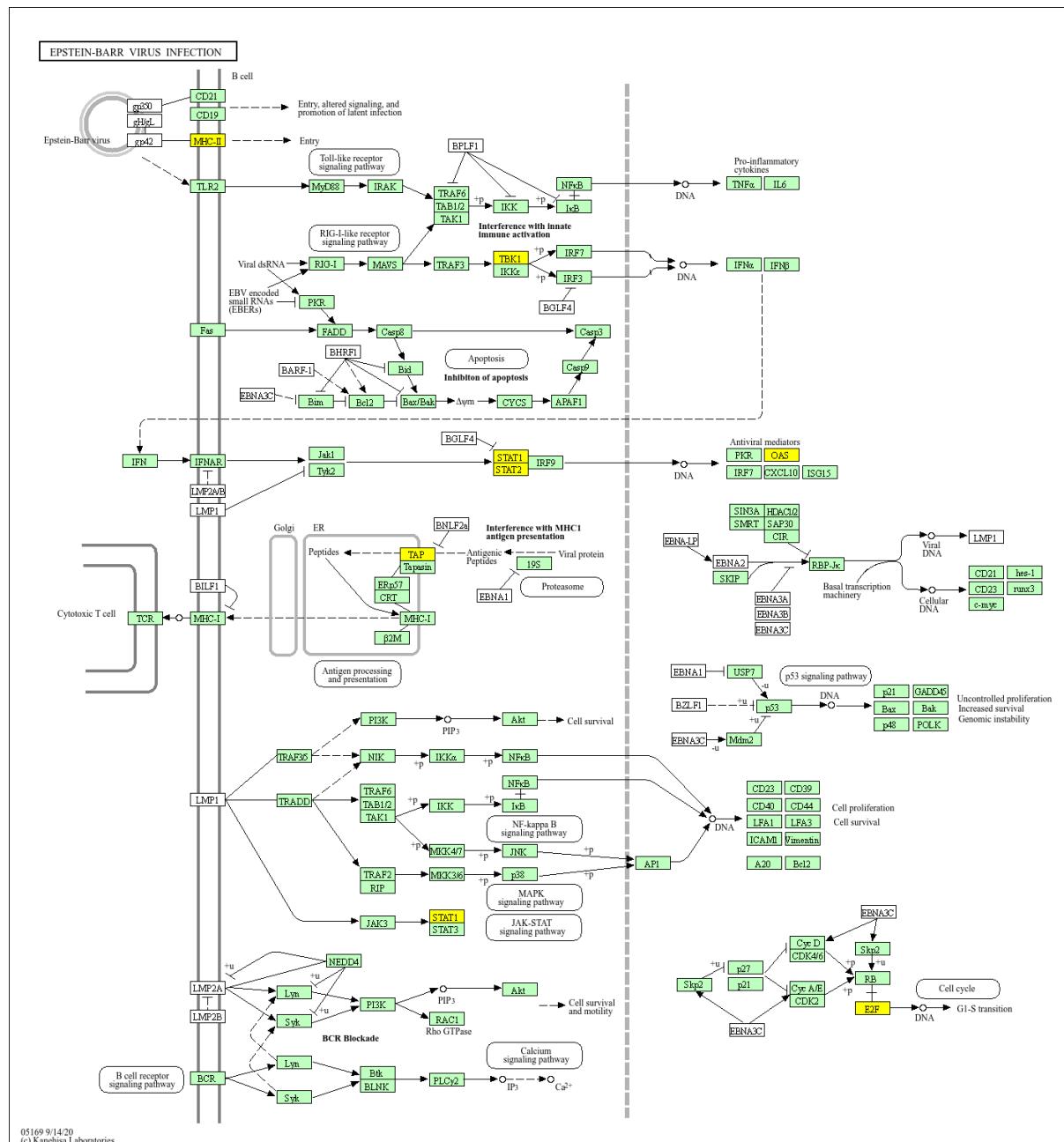
b. hsa04621 NOD-like receptor signaling pathway (11)



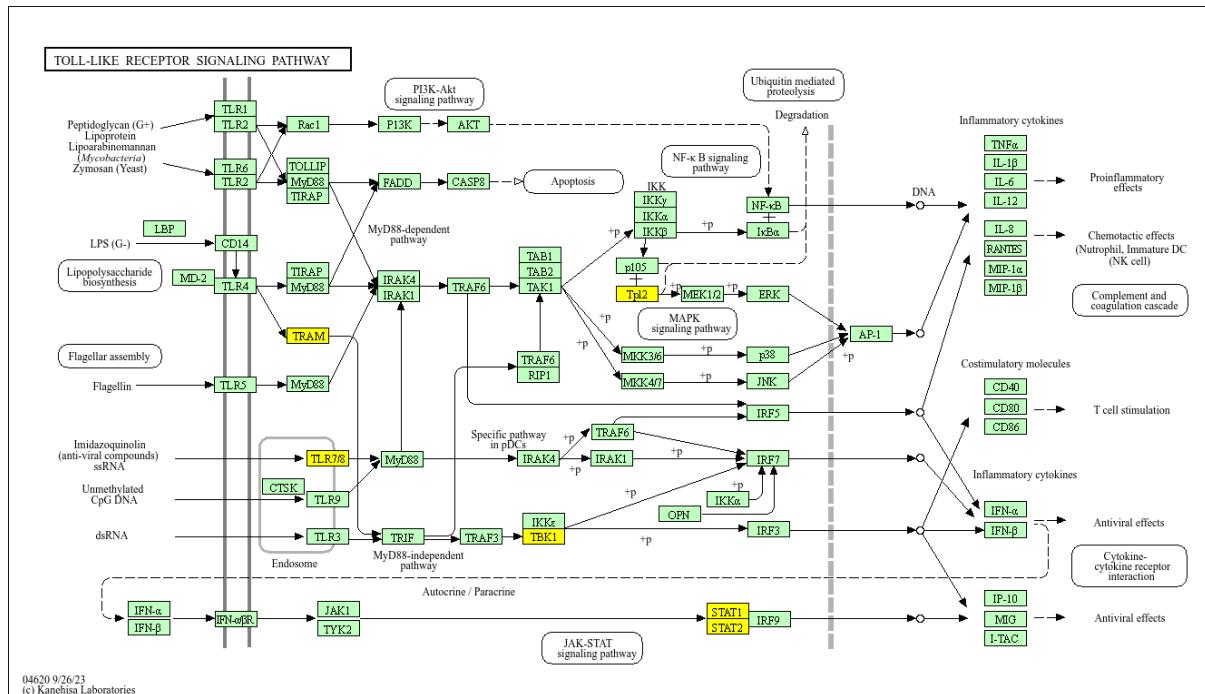
c. hsa04120 Ubiquitin mediated proteolysis (7)



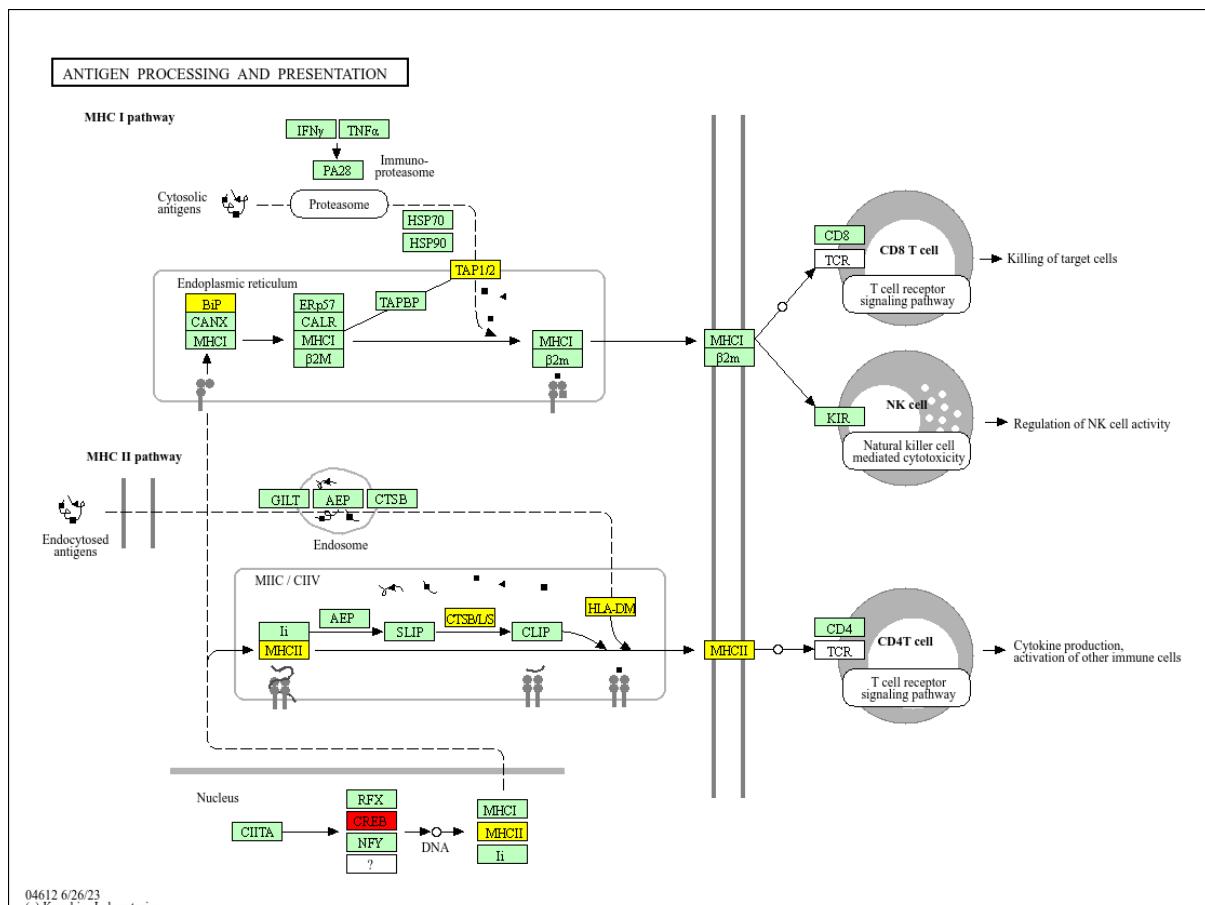
d. hsa05169 Epstein-Barr virus infection (8)



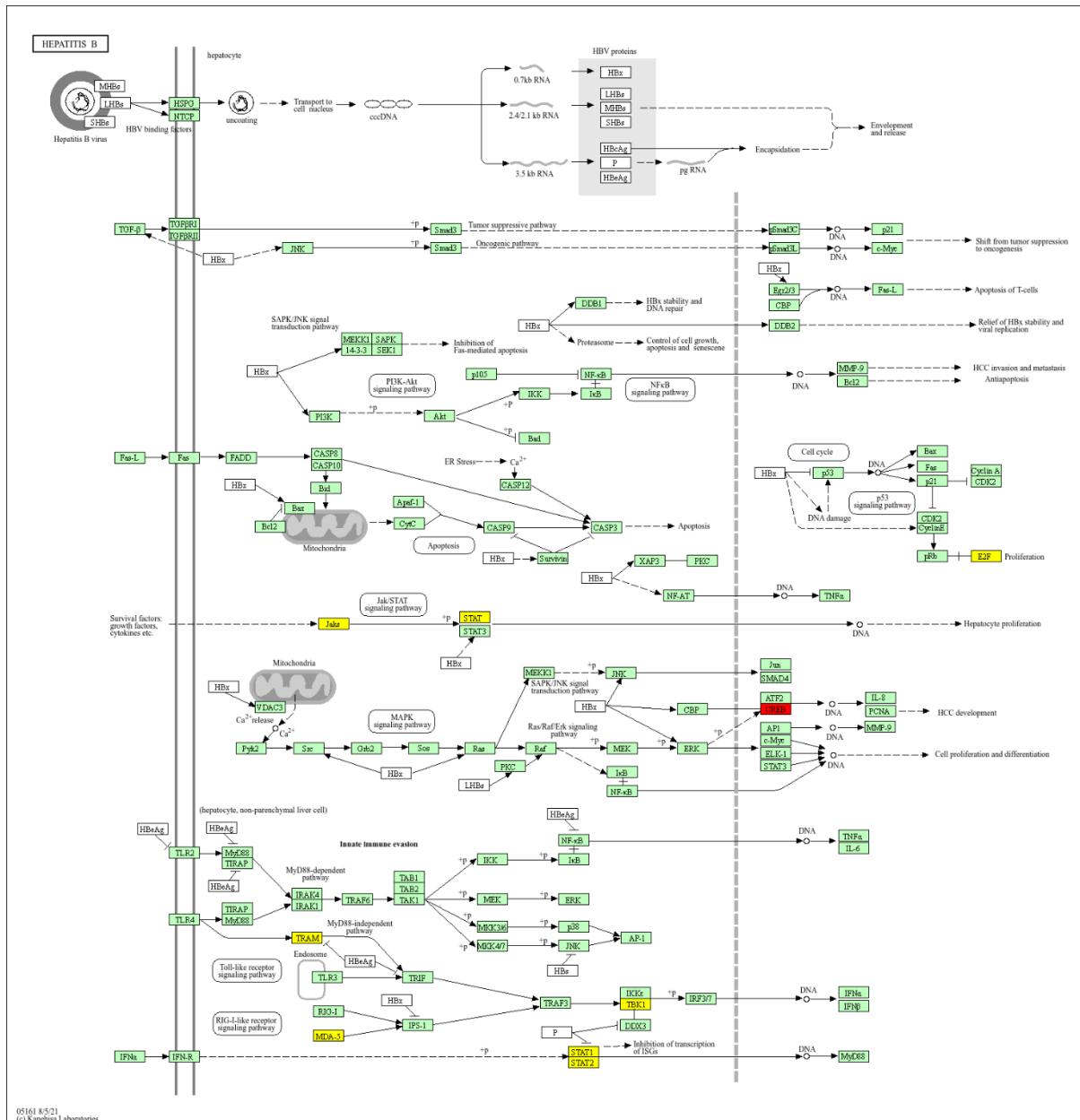
e. hsa04620 Toll-like receptor signaling pathway (6)



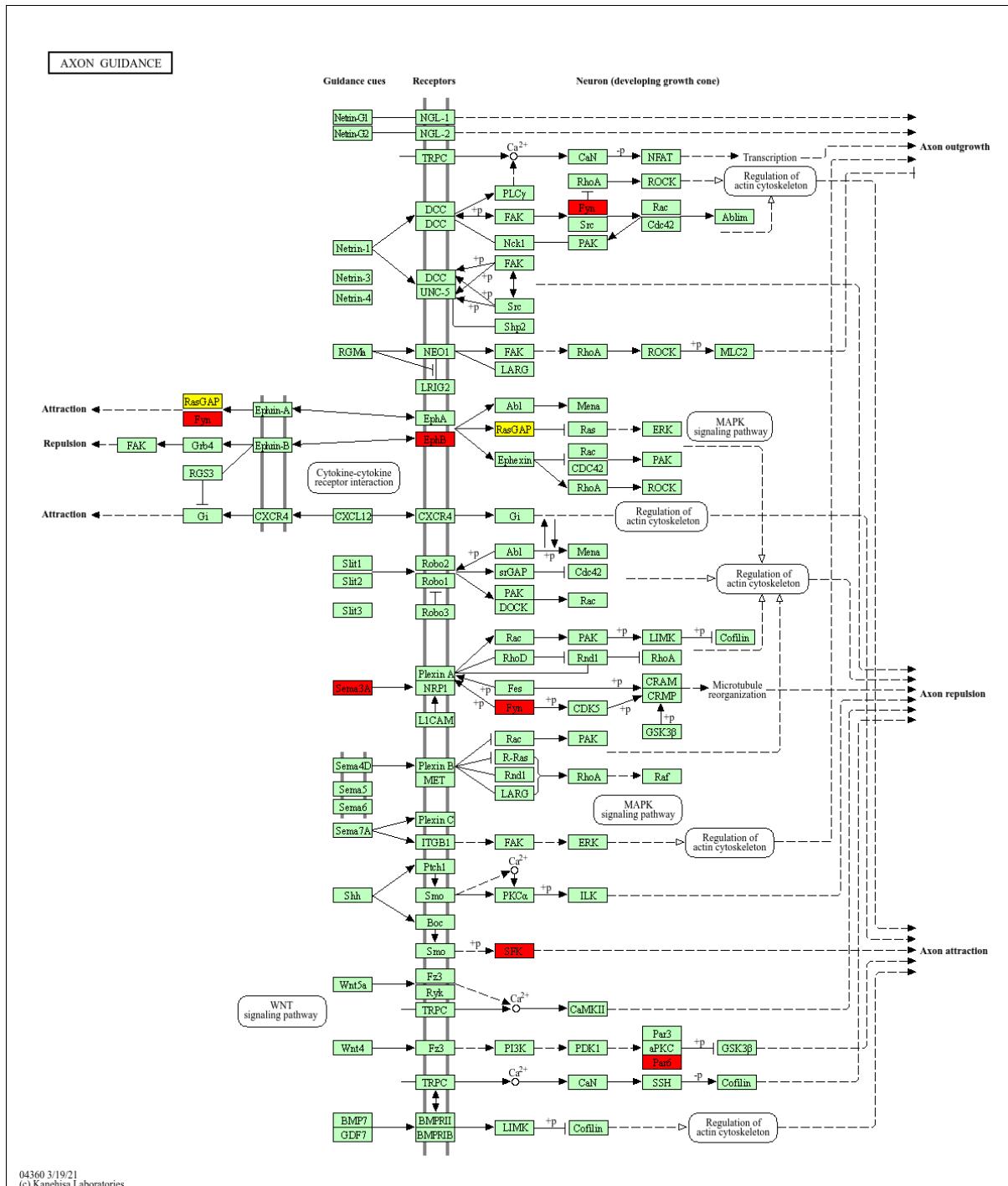
f. hsa04612 Antigen processing and presentation (6)



g. hsa05161 Hepatitis B (8)



h. hsa04360 Axon guidance (6)



Berdasarkan hasil analisis fungsional yang dilakukan menggunakan g:Profiler, jalur **NOD-like receptor (NLR) signaling pathway** teridentifikasi sebagai jalur yang paling signifikan dengan nilai **padj=1,916×10⁻⁵**. Secara biologis, NLR merupakan reseptor intraseluler yang berfungsi sebagai sensor terhadap pola molekuler terkait patogen (PAMPs) dan sinyal bahaya seluler (DAMPs).

Dominansi jalur ini menunjukkan bahwa sampel mengalami respon imun bawaan (*innate immunity*) yang sangat kuat. Aktivasi jalur NLR biasanya memicu pembentukan kompleks *inflamasome*, yang bertanggung jawab atas pelepasan sitokin pro-inflamasi seperti IL-1 β dan IL-18. Temuan ini sangat relevan dalam konteks pengembangan vaksin atau terapi, karena memberikan petunjuk bahwa modulasi pada jalur peradangan ini dapat menjadi target utama.

Selain itu, munculnya jalur lain seperti *Influenza A*, *Measles*, dan *Epstein-Barr virus infection* menguatkan indikasi bahwa profil ekspresi gen sampel didominasi oleh mekanisme pertahanan antiviral. Hal ini menunjukkan adanya koordinasi sistem imun seluler dalam upaya mengenali dan mengeliminasi ancaman asing. Dalam kerangka **Personalized Drug Therapy**, informasi ini sangat berharga untuk menentukan apakah seorang pasien akan merespon lebih baik terhadap agen anti-inflamasi spesifik atau imunomodulator tertentu berdasarkan aktivitas jalur NLR mereka.

D. Kesimpulan

1. **Analisis Enrichment:** Jalur **NOD-like receptor signaling pathway** merupakan jalur biologis yang paling signifikan dipengaruhi oleh daftar gen yang dianalisis, mengindikasikan aktivitas respon imun inflamasi yang dominan.
2. **Visualisasi Jalur:** Penggunaan *KEGG Color Tool* berhasil memetakan distribusi gen *upregulated* (merah) dan *downregulated* (biru) di dalam jalur koordinasi seluler, memberikan gambaran mekanisme molekuler yang terjadi secara visual.
3. **Aplikasi Klinis:** Hasil ini dapat dijadikan dasar dalam strategi **Personalized Drug Therapy**. Identifikasi jalur NLR sebagai target utama memungkinkan pendekatan terapi yang lebih presisi, di mana pemberian obat dapat disesuaikan untuk menargetkan molekul spesifik dalam jalur tersebut guna meminimalkan efek samping dan meningkatkan efektivitas pengobatan pada pasien.