

## ANALISIS DIFFERENTIALLY EXPRESSED GENES PADA INFEKSI TUBERKULOSIS MENGGUNAKAN DATASET GSE62525 DAN GEO2R

### PENDAHULUAN

Tuberkulosis (TB) merupakan penyakit infeksi yang disebabkan oleh *Mycobacterium tuberculosis* dan masih menjadi salah satu penyebab utama kematian akibat penyakit menular di dunia. TB memiliki dua kondisi klinis utama, yaitu tuberkulosis aktif dan infeksi laten. Pada kondisi aktif, pasien menunjukkan gejala klinis dan dapat menularkan penyakit, sedangkan pada infeksi laten bakteri berada dalam tubuh tanpa menimbulkan gejala klinis yang jelas. Tantangan utama dalam pengendalian TB adalah membedakan secara akurat antara individu sehat, infeksi laten, dan TB aktif, terutama melalui pendekatan diagnostik berbasis molekuler. Perkembangan teknologi transcriptomics memungkinkan analisis ekspresi gen secara menyeluruh untuk mengidentifikasi gen-gen yang berperan dalam respons imun terhadap infeksi TB.

Dataset publik seperti Gene Expression Omnibus (GEO) menyediakan data ekspresi gen yang dapat dianalisis ulang untuk menemukan biomarker potensial. Salah satu dataset yang tersedia adalah GSE62525, yang memuat profil ekspresi gen peripheral blood mononuclear cells (PBMC) dari individu dengan TB aktif, infeksi laten, dan individu sehat. Analisis Differentially Expressed Genes (DEG) menggunakan GEO2R memungkinkan identifikasi gen yang mengalami perubahan ekspresi signifikan antara kondisi terinfeksi dan sehat. Identifikasi gen-gen tersebut penting untuk memahami mekanisme molekuler infeksi TB serta mendukung pengembangan biomarker diagnostik. Tujuan dari analisis ini adalah untuk mengidentifikasi gen-gen yang terekspresi berbeda secara signifikan antara individu terinfeksi TB (aktif dan laten) dibandingkan dengan individu sehat menggunakan pendekatan Differential Expression Analysis berbasis GEO2R.

### METODE

#### A. Dataset

Dataset yang digunakan dalam analisis ini adalah GSE62525 yang diperoleh dari database publik Gene Expression Omnibus (GEO), National Center for Biotechnology Information (NCBI). Dataset ini merupakan data transcriptomics berbasis microarray (Platform: GPL16951, Phalanx Human OneArray Ver. 6) yang menganalisis ekspresi gen pada peripheral blood mononuclear cells (PBMC) dari individu dengan tuberkulosis aktif, infeksi laten, dan individu sehat. Total sampel dalam dataset ini berjumlah 42 sampel, terdiri atas 7 individu dengan TB aktif, 7 individu dengan infeksi laten, dan 14 individu sehat. Dalam analisis ini, sampel TB aktif dan infeksi laten digabung sebagai kelompok Infected ( $n = 28$ ), sedangkan individu sehat digunakan sebagai kelompok Healthy ( $n = 14$ ).

#### B. Analisis Differential Expression (DEG)

Analisis diferensial ekspresi gen dilakukan menggunakan tool berbasis web GEO2R yang tersedia pada platform GEO. GEO2R menggunakan metode statistik limma (Linear Models for Microarray Data) untuk membandingkan ekspresi gen antar kelompok. Parameter analisis yang digunakan adalah group 1 yaitu Infected (Active TB + Latent TB), group 2 yaitu healthy, metode koreksi multiple testing yaitu Benjamini–Hochberg (False Discovery Rate), dan kriteria signifikan yaitu adjusted p-value  $< 0.05$ . Gen dengan nilai adjusted p-value kurang dari 0.05 dianggap sebagai gen yang terekspresi berbeda secara signifikan (Differentially Expressed Genes/DEGs).

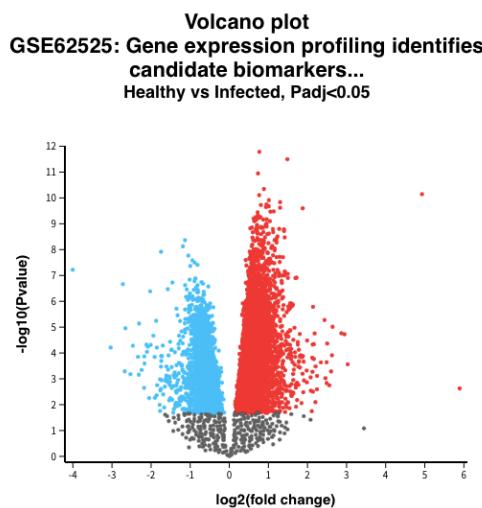
#### C. Replikasi Analisis

Dalam memastikan konsistensi hasil, analisis dilakukan sebanyak tiga kali dengan parameter yang sama dan alur yang konsisten pada GEO2R. Hasil antar replikasi menunjukkan pola yang serupa dalam jumlah gen signifikan serta arah perubahan ekspresi gen, sehingga analisis dinilai stabil dan reproducible.

### HASIL DAN INTERPRETASI

Analisis Differentially Expressed Genes (DEG) dilakukan menggunakan GEO2R dengan membandingkan kelompok Infected (Active TB + Latent TB) terhadap kelompok Healthy, menggunakan kriteria adjusted p-value  $< 0.05$  (Benjamini–Hochberg correction). Hasil analisis

menunjukkan terdapat sejumlah gen yang mengalami perubahan ekspresi signifikan antara kedua kelompok. Berdasarkan visualisasi Volcano plot (Gambar 1), terlihat distribusi gen yang signifikan secara statistik dengan nilai  $-\log_{10}(p\text{-value})$  yang tinggi serta nilai log<sub>2</sub> fold change yang menyimpang dari nol. Gen yang ditandai dengan warna merah berada pada sisi kanan plot ( $\log_2\text{FC} > 0$ ), yang menunjukkan bahwa gen-gen tersebut mengalami peningkatan ekspresi (up-regulated) pada kelompok Infected dibandingkan Healthy. Sebaliknya, gen yang berwarna biru berada pada sisi kiri plot ( $\log_2\text{FC} < 0$ ), menunjukkan penurunan ekspresi (down-regulated) pada kelompok Infected. Distribusi titik pada volcano plot menunjukkan bahwa jumlah gen up-regulated tampak lebih dominan dibandingkan down-regulated, yang mengindikasikan adanya aktivasi transkripsi yang luas pada kondisi terinfeksi TB.



Gambar 1. Volcano plot DEG antara kelompok Infected dan Healthy (adj p-value < 0.05).

Berdasarkan tabel hasil GEO2R (Gambar 2), gen-gen dengan adjusted p-value yang sangat kecil (misalnya  $< 10^{-6}$  hingga  $< 10^{-8}$ ) menunjukkan signifikansi statistik yang sangat kuat. Nilai log<sub>2</sub> fold change yang positif dengan nilai  $> 1$  pada beberapa gen menunjukkan peningkatan ekspresi lebih dari dua kali lipat pada kelompok terinfeksi dibandingkan kelompok sehat. Sebaliknya, gen dengan log<sub>2</sub> fold change negatif menunjukkan adanya supresi atau penurunan ekspresi gen pada kondisi terinfeksi. Kombinasi antara signifikansi statistik yang tinggi (adjusted p-value kecil) dan besarnya perubahan ekspresi (log<sub>2</sub>FC tinggi secara absolut) memperkuat bahwa gen-gen tersebut merupakan kandidat DEG yang relevan secara biologis.

ID	adj.P.Val	P.Value	t	B	logFC	ORF	SPOT_ID
* PH_hs_0026007	4.88e-08	4.88e-08	1.65e-12	9.72	18.19	0.768	57585
* PH_hs_0042056	4.88e-08	4.88e-08	3.18e-12	9.51	17.57	1.485	57562
* PH_hs_0023535	1.16e-07	1.16e-07	1.13e-11	9.11	16.37	0.736	23225
* PH_hs_0049138	3.46e-07	3.46e-07	4.50e-11	8.68	15.06	0.887	57608
* PH_hs_0020122	4.00e-07	4.00e-07	7.17e-11	8.53	14.62	4.928	57493
* PH_hs_0061205	4.00e-07	4.00e-07	7.81e-11	8.51	14.54	0.764	7329
* PH_hs_0047026	5.36e-07	5.36e-07	1.22e-10	8.37	14.12	1.012	100131096
* PH_hs_0060355	5.51e-07	5.51e-07	1.46e-10	8.32	13.95	1.299	3298
* PH_hs_0009202	5.51e-07	5.51e-07	1.79e-10	8.25	13.75	0.949	9013
* PH_hs_0043676	5.51e-07	5.51e-07	1.88e-10	8.24	13.71	0.807	3320
* PH_hs_0031073	5.51e-07	5.51e-07	2.09e-10	8.21	13.6	1.009	10286
* PH_hs_0024679	5.51e-07	5.51e-07	2.26e-10	8.18	13.53	0.939	7088
* PH_hs_0061687	5.51e-07	5.51e-07	2.40e-10	8.16	13.47	1.298	130026
* PH_hs_0045214	5.51e-07	5.51e-07	2.51e-10	8.15	13.43	1.876	677847
* PH_hs_0030842	6.53e-07	6.53e-07	3.34e-10	8.06	13.16	1.136	10609
* PH_hs_0062304	6.53e-07	6.53e-07	3.42e-10	8.06	13.14	1.051	85108511
* PH_hs_0015893	6.53e-07	6.53e-07	3.61e-10	8.04	13.08	0.724	10212
* PH_hs_0004553	7.89e-07	7.89e-07	4.84e-10	7.95	12.81	0.74	8165
* PH_hs_0044905	7.89e-07	7.89e-07	4.88e-10	7.95	12.8	0.734	267004
* PH_hs_0033107	7.99e-07	7.99e-07	5.58e-10	7.91	12.67	0.797	5116
* PH_hs_0025776	7.99e-07	7.99e-07	5.64e-10	7.9	12.66	0.672	9020
* PH_hs_0008282	7.99e-07	7.99e-07	5.79e-10	7.9	12.64	0.77	PH_hs_0008282
* PH_hs_0039021	7.99e-07	7.99e-07	5.98e-10	7.89	12.6	0.733	220906
* PH_hs_0033538	8.00e-07	8.00e-07	6.25e-10	7.87	12.56	0.619	54093
* PH_hs_0061197	8.02e-07	8.02e-07	6.74e-10	7.85	12.49	0.952	23157
* PH_hs_0031569	8.02e-07	8.02e-07	6.88e-10	7.84	12.47	0.868	8914
* PH_hs_00044905	8.02e-07	8.02e-07	7.06e-10	7.84	12.45	0.666	4070

Gambar 2. Cuplikan Top Differentially Expressed Genes berdasarkan adjusted p-value.

Secara interpretasi biologis, infeksi *Mycobacterium tuberculosis* memicu respons imun yang kompleks, terutama melalui aktivasi sel mononuklear darah perifer (PBMC), termasuk limfosit dan monosit. Peningkatan ekspresi gen pada kelompok Infected kemungkinan merefleksikan aktivasi jalur inflamasi, produksi sitokin proinflamasi, aktivasi respons imun bawaan dan adaptif serta mekanisme pertahanan terhadap patogen intraseluler. Sementara itu, gen yang mengalami penurunan ekspresi dapat berkaitan dengan regulasi negatif jalur tertentu atau perubahan keseimbangan homeostasis seluler akibat infeksi. Penggabungan TB aktif dan laten dalam kelompok Infected menunjukkan bahwa bahkan pada kondisi laten, sistem imun tetap menunjukkan perubahan transkripsional dibandingkan individu sehat. Hal ini mendukung hipotesis bahwa infeksi laten bukan kondisi “benar-benar diam”, melainkan tetap melibatkan aktivitas imunologis yang dapat dideteksi melalui analisis ekspresi gen. Secara keseluruhan, hasil ini menunjukkan adanya perbedaan profil ekspresi gen yang jelas antara individu terinfeksi TB dan individu sehat, serta menguatkan potensi penggunaan transcriptomics sebagai pendekatan untuk identifikasi biomarker molekuler tuberkulosis.

## KESIMPULAN

Analisis Differentially Expressed Genes (DEG) menggunakan GEO2R pada dataset GSE62525 menunjukkan adanya perbedaan profil ekspresi gen yang signifikan antara kelompok terinfeksi tuberkulosis (TB aktif dan laten) dan individu sehat. Dengan menggunakan kriteria adjusted p-value < 0,05, diperoleh sejumlah gen yang mengalami peningkatan (up-regulated) maupun penurunan ekspresi (down-regulated) pada kelompok terinfeksi. Visualisasi melalui volcano plot memperlihatkan distribusi gen signifikan yang jelas pada kedua sisi log2 fold change, dengan dominasi gen yang mengalami peningkatan ekspresi pada kelompok terinfeksi. Hal ini mengindikasikan adanya aktivasi respons imun yang kuat pada peripheral blood mononuclear cells (PBMC) sebagai respons terhadap infeksi *Mycobacterium tuberculosis*. Hasil ini menunjukkan bahwa pendekatan transcriptomics berbasis microarray mampu mengidentifikasi perubahan molekuler yang relevan secara biologis antara kondisi terinfeksi dan sehat. Dengan demikian, analisis DEG berpotensi mendukung identifikasi biomarker molekuler untuk deteksi dan pemahaman mekanisme imunologis pada tuberkulosis