

Analisis Diferensial Ekspresi Gen pada Dataset GSE30911

Menggunakan GEO2R

Qorinatul Imaniyah

Pada era pascagenomik, *long non-coding RNA* (lncRNA) diketahui berperan sebagai regulator penting dalam perkembangan berbagai jenis kanker. Kanker payudara subtipe luminal A merupakan salah satu kanker dengan angka kejadian tertinggi pada wanita. Analisis ekspresi RNA, khususnya melalui sampel darah perifer, menjadi pendekatan yang menjanjikan dalam identifikasi biomarker yang bersifat minimal invasif. Penelitian sebelumnya pada kanker payudara *triple-negative* menunjukkan potensi keterlibatan lncRNA dalam proses patogenesis kanker. Oleh karena itu, penelitian ini bertujuan menganalisis profil ekspresi lncRNA pada sel mononuklear darah perifer (PBMC) pasien kanker payudara luminal A menggunakan metode analisis mikroarray (Minina *et al.*, 2025).

Dataset yang digunakan dalam analisis ini adalah GSE303911 yang diperoleh dari database NCBI *Gene Expression Omnibus* (GEO). Sampel dibagi menjadi dua kelompok, yaitu grup kontrol dan grup penyakit dengan total delapan sampel. Berikut merupakan tabel pembagian kelompok sampel.

Tabel 1 Pembagian kelompok sampel dataset GSE303911

No	Kode sampel	Kelompok
1	GSM9138079	Kontrol
2	GSM9138080	
3	GSM9138081	
4	GSM9138082	
5	GSM9138083	Penyakit
6	GSM9138084	
7	GSM9138085	
8	GSM9138086	

Analisis diferensial ekspresi gen dilakukan menggunakan GEO2R dengan kriteria gen signifikan berdasarkan nilai *adjusted p-value* < 0,05 dan $|\log FC| \geq 1$. Analisis GEO2R dilakukan sebanyak tiga kali pengulangan menggunakan alur kerja dan parameter yang sama untuk memastikan konsistensi hasil analisis.

Setelah dilakukan analisis dengan GEO2R, didapatkan hasil 250 gen teratas dengan perbedaan ekspresi paling signifikan berdasarkan parameter statistiknya, seperti nilai ekspresi rata-rata, \log_2 fold change, $p\text{-value}$, dan $\text{adjusted } p\text{-value}$. Berikut merupakan tabel hasil GEO2R yang mencakup 250 gen teratas.

Top differentially expressed genes 							
Download full table		Select columns					
ID	adj.P.Val	P.Value	t	B	logFC	GENE_SYMBOL	GENE_NAME
# A_23_P66732	0.026	5.29e-07	-17.31	4.00806	-2.176	GSG2	germ cell associated 2 (...)
# A_21_P0013503	0.026	1.04e-06	-15.68	3.79101	-2.709	LOC101928108	lymphocyte antigen 6A-...
# A_22_P00011597	0.0423	2.54e-06	13.75	3.45725	2.562	RAB30-AS1	RAB30 antisense RNA 1...
# A_22_P00017135	0.0898	8.41e-06	-11.51	2.91766	-2.945	ZNF385D-AS2	ZNF385D antisense RN...
# A_23_P48740	0.0898	9.92e-06	-11.23	2.83443	-2.082	DIO2	deiodinase, iodothyronin...
# A_22_P00025017	0.0898	1.08e-05	-11.09	2.79127	-2.249	LOC101929535	uncharacterized LOC10...
# A_33_P3348362	0.129	1.81e-05	10.26	2.51344	2.466	ARHGAP5-AS1	ARHGAP5 antisense R...
# A_24_P14367	0.1641	2.79e-05	9.61	2.26479	2.032	PTBP1	polyypyrimidine tract bind...
# A_21_P0007977	0.1641	3.53e-05	9.27	2.12253	2.543	LINC00412	long intergenic non-prot...
# A_23_P12503	0.1641	3.78e-05	9.17	2.08022	0.787	NUP133	nucleoporin 133kDa
# A_33_P3248903	0.1641	4.60e-05	-8.9	1.95828	-4.033	WNT7B	wingless-type MMTV int...
# A_22_P00007461	0.1641	4.62e-05	8.89	1.95512	1.961	lnc-GTF2A2-2	lnc-GTF2A2-2.1
# A_22_P00009866	0.1641	4.75e-05	8.85	1.93732	1.919		
# A_33_P3374957	0.1641	5.02e-05	-8.78	1.90251	-0.846	KLF16	Kruppel-like factor 16
# A_23_P307310	0.1641	5.32e-05	8.7	1.86539	0.739	ACAN	aggrecan
# A_33_P3709525	0.1641	5.56e-05	-8.64	1.83661	-2.515	LOC100128233	uncharacterized LOC10...
# A_21_P0003806	0.1641	5.74e-05	-8.6	1.81637	-0.834	lnc-DCTD-6	lnc-DCTD-6.1
# A_21_P0014195	0.1641	6.32e-05	-8.47	1.75395	-1.168		
# A_33_P3401782	0.1641	6.82e-05	-8.37	1.70293	-1.839	C19orf54	chromosome 19 open re...
# A_21_P0000604	0.1641	7.14e-05	8.31	1.67252	2.486	DOC2GP	double C2-like domains...
# A_22_P00008781	0.1641	7.33e-05	-8.28	1.65532	-1.394	lnc-KIF25-2	lnc-KIF25-2.2
# A_33_P3464900	0.1641	7.50e-05	-8.25	1.64054	-1.378	BFS2P2-AS1	BFS2P2 antisense RNA 1

Gambar 1. Tabel 250 gen teratas dari GSE303911

(Dokumentasi pribadi, 2026)

Dengan menerapkan kriteria signifikan berupa $\text{adjusted } p\text{-value} < 0,05$ dan $|\log FC| \geq 1$, didapatkan gen-gen yang memiliki perubahan ekspresi signifikan secara statistik. Berikut merupakan tabel hasil DEGs signifikan.

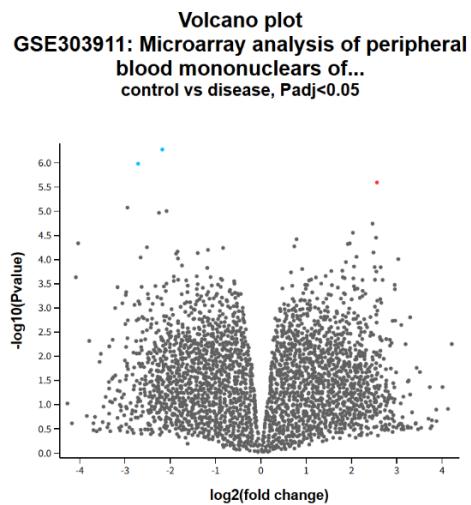
Tabel 2. Hasil DEGs signifikan dari GSE303911

ID	adj.P.Val	P.Value	logFC	GENE_SYMBOL	GENE NAME	DIRECTION
A_23_P66732	0.026	5.29E-07	-2.1763908	GSG2	germ cell associated 2 (haspin)	down regulated
A_21_P0013503	0.026	1.04E-06	-2.7085109	LOC101928108	lymphocyte antigen 6A-2/6E-1-like	down regulated
A_22_P00011597	0.0423	2.54E-06	2.56240214	RAB30-AS1	RAB30 antisense RNA 1 (head to head)	up regulated

Dari tabel di atas, dapat diketahui bahwa terdapat tiga gen yang mengalami perubahan ekspresi signifikan antara sampel kontrol dan penyakit kanker payudara luminal A. Dari jumlah tersebut, satu gen mengalami *up-regulation* yang menandakan peningkatan aktivitas ekspresi gen pada kondisi kanker payudara luminal A, sedangkan dua gen lainnya menunjukkan pola *down-regulation* yang menandakan tingkat ekspresi lebih rendah dibandingkan kontrol. Hal ini sesuai dengan literatur yang menyatakan bahwa *up-regulated* menunjukkan gen dengan tingkat ekspresi lebih tinggi pada sampel perlakuan dibandingkan kontrol, sedangkan *down-regulated*

menunjukkan gen dengan ekspresi lebih rendah pada sampel perlakuan dibandingkan kontrol (Rahmawati & Shabrina, 2024). Dominasi gen yang mengalami *down-regulation* pada sampel kanker menunjukkan adanya penurunan aktivitas beberapa fungsi sel normal. Hal ini sejalan dengan karakteristik biologis kanker payudara luminal A yang memiliki tingkat proliferasi relatif rendah dan regulasi pertumbuhan yang dipengaruhi oleh jalur hormon estrogen (Dai *et al.*, 2015).

Selain itu, visualisasi menggunakan *volcano plot* dilakukan untuk menggambarkan distribusi perubahan ekspresi gen antara sampel kontrol dan sampel penyakit. Plot ini menampilkan nilai logFC pada sumbu X serta $-\log_{10} p\text{-value}$ pada sumbu Y. Berikut merupakan visualisasi *volcano plot* dari GSE303911.

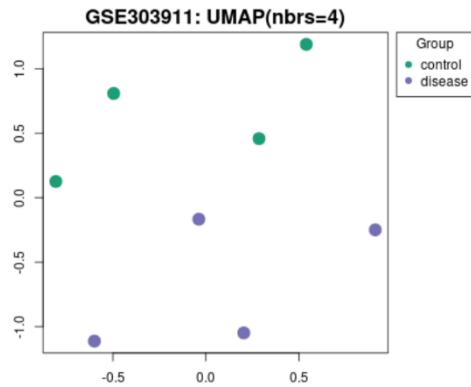


Gambar 2. *Volcano plot* dari GSE303911

(Dokumentasi pribadi, 2026)

Dari gambar *volcano plot* tersebut, posisi titik memudahkan identifikasi antara gen yang meningkat dan menurun ekspresinya. Gen yang mengalami *up-regulation* ditunjukkan oleh titik yang berada di sisi kanan garis nol (logFC positif), sedangkan gen yang mengalami *down-regulation* berada di sisi kiri garis nol (logFC negatif). Pada hasil analisis ini, terlihat satu gen signifikan berada di sisi kanan dan dua gen signifikan berada di sisi kiri plot. Titik berwarna merah menandakan gen signifikan yang bersifat *up-regulated*, sedangkan titik berwarna biru menandakan gen signifikan yang bersifat *downregulated*. Sementara itu, titik berwarna hitam merepresentasikan gen yang tidak memenuhi kriteria signifikansi berdasarkan nilai *adjusted p-value* dan *log fold change* (Rahmawati & Shabrina, 2024).

Selanjutnya adalah visualisasi dengan *Uniform Manifold Approximation and Projection (UMAP) plot* yang menunjukkan adanya pemisahan kelompok sampel berdasarkan kemiripan biologis dalam ruang dua dimensi. Sampel penyakit dan kontrol membentuk klaster terpisah yang menandakan adanya perbedaan pola ekspresi gen di antara kedua kelompok sampel (Armstrong *et al.*, 2021). Berikut merupakan visualisasi UMAP plot dari GSE303911.



Gambar 3. UMAP plot dari GSE303911

(Dokumentasi pribadi, 2026)

Dari visualisasi UMAP plot di atas, dapat diketahui bahwa sampel kontrol dan penyakit membentuk dua klaster yang relatif terpisah. Sampel kontrol mengelompok pada bagian atas plot, sedangkan sampel penyakit berada pada bagian bawah. Pemisahan klaster ini mengindikasikan adanya perbedaan pola ekspresi gen antara kedua kelompok sampel (Armstrong *et al.*, 2021).

Berdasarkan analisis dataset GSE303911 menggunakan metode mikroarray, ditemukan adanya perbedaan pola ekspresi lncRNA antara sampel kontrol dan kanker payudara luminal A. Analisis diferensial ekspresi gen menunjukkan tiga gen yang mengalami perubahan ekspresi signifikan, terdiri atas satu gen *up-regulated* dan dua gen *down-regulated*. Visualisasi *volcano plot* memberikan gambaran distribusi gen berdasarkan arah perubahan ekspresi, sedangkan UMAP plot menunjukkan pemisahan klaster yang mengindikasikan perbedaan karakteristik molekuler antara kedua kelompok sampel. Hasil ini menunjukkan bahwa perubahan ekspresi lncRNA pada PBMC berpotensi menjadi indikator biologis yang berkaitan dengan kondisi kanker payudara luminal A serta mendukung pemanfaatannya sebagai kandidat biomarker minimal invasif.

Daftar Pustaka

- Armstrong, G., Martino, C., Rahman, G., Gonzalez, A., Vázquez-Baeza, Y., Mishne, G., & Knight, R. (2021). Uniform manifold approximation and projection (UMAP) reveals composite patterns and resolves visualization artifacts in microbiome data. *mSystems*, 6(1), e00691-21. <https://doi.org/10.1128/msystems.00691-21>.
- Dai, X., Li, T., Bai, Z., Yang, Y., Liu, X., Zhan, J., & Shi, B. (2015). Breast cancer intrinsic subtype classification, clinical use and future trends. *American Journal of Cancer Research*, 5(10), 2929–2943.
- Minina, V. I., Titov, R. A., Buslaev, V. Y., Savchenko, R. R., Sleptcov, A. A., Gavrineva, N. A., Bakanova, M. L., Zakharova, Y. A., & Glushkov, A. N. (2025). Microarray analysis of differential expression of long non-coding RNAs in peripheral blood mononuclear cells in luminal A breast cancer. *Gene Expression*, 24(4), e00021. <https://doi.org/10.14218/GE.2025.00021>.
- Rahmawati, D., & Shabrina, H. (2024). *Analisis molekuler dan bioinformatika*. Bogor: SEAMEO BIOTROP.