

## EKSPLORASI DATA & ANALISIS PRAKTIS DENGAN GEO2R

### PENDAHULUAN

$\beta$ -thalassemia merupakan kelainan genetik pada sintesis rantai  $\beta$ -globin yang menyebabkan gangguan pembentukan sel darah merah dan anemia kronis. Perubahan proses eritropoiesis pada penyakit ini diduga berkaitan dengan perubahan ekspresi gen pada sel prekursor eritroid. Pada dataset (GSE56088) ini, perbedaan ekspresi gen dilaporkan muncul setelah 14 hari kultur sel eritroid, sementara pada hari ke-7 belum terdeteksi perubahan signifikan. Tujuan analisis ini adalah untuk mengidentifikasi dan menginterpretasikan gen-gen yang mengalami perubahan ekspresi (DEG) pada hari ke-14 kultur sel prekursor eritroid antara pasien  $\beta$ -thalassaemia dan kontrol sehat menggunakan metode GEO2R.

### METODE

#### Dataset

Link: <https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE56088>

Analisis menggunakan dataset ekspresi gen **sel prekursor eritroid manusia** dari:

- **6 pasien  $\beta$ -thalassemia tergantung transfusi**
- **6 kontrol sehat**
- Metode pengukuran: **microarray (expression profiling by array)**
- Waktu kultur: **14 hari** (fokus analisis)

#### Pembagian Grup

Analisis GEO2R dilakukan dengan membagi sampel menjadi dua kelompok:

- **Grup kontrol:** sel eritroid dari individu sehat (hari ke-14)
- **Grup penderita:** sel eritroid dari pasien  $\beta$ -thalassaemia (hari ke-14)

#### Parameter utama GEO2R

Kriteria penentuan DEG signifikan:

- Adjusted p-value (FDR) < 0,05
- $|\log_2 \text{Fold Change}| \geq 1$

Parameter ini digunakan untuk memastikan bahwa:

- Perubahan ekspresi **signifikan secara statistik**
- Perubahan **bermakna secara biologis**

#### Skema Replikasi

Setiap kelompok terdiri dari **6 sampel biologis independen**, sehingga analisis didasarkan pada replikasi biologis yang memadai untuk membandingkan perbedaan ekspresi gen antar kelompok.

HASIL DAN INTERPRETASI

Hasil

Analisis menggunakan GEO2R pada data hari ke-14 kultur menunjukkan adanya 277 gen yang mengalami diferensial ekspresi signifikan antara pasien  $\beta$ -thalassaemia dan kontrol sehat yang ditunjukan gambar 1.

ID	adj.P.Val	P.Value	t	B	logFC	Gene.symbol	Gene.title
ILMN_1671693	0.0000371	7.86e-10	17.31	11.67	1.599	TPGS2	tubulin polyglutamylase compl...
ILMN_3309473	0.0000531	3.30e-09	15.28	10.83	2.292	MIR101-1	microRNA 101-1
ILMN_1794976	0.0000531	3.95e-09	15.04	10.7	2.464	NUOT4P1	nudix hydrolase 4 pseudogene 1
ILMN_1656501	0.0000531	4.50e-09	14.87	10.6	2.615	DUSP5	dual specificity phosphatase 5
ILMN_3235517	0.0000743	7.86e-09	14.16	10.17	2.137	ZNF777	zinc finger protein 777
ILMN_2390859	0.000212	2.69e-08	12.69	9.18	1.651	NFKB2	nuclear factor kappa B subunit 2
ILMN_1870111	0.0002126	3.40e-08	-12.43	8.98	-0.553	FLJ41455	uncharacterized LOC441441
ILMN_1726547	0.0002126	3.60e-08	12.36	8.93	1.832	MAP3K5	mitogen-activated protein kina...
ILMN_1811258	0.0002127	4.05e-08	12.23	8.84	1.423	RELB	RELB proto-oncogene, NF- $\kappa$ B ...
ILMN_1804448	0.000269	5.70e-08	11.86	8.55	0.987	MSI2	musashi RNA binding protein 2
ILMN_3236637	0.0003403	7.93e-08	11.52	8.27	1.586	PIP4K2A	phosphatidylinositol-5-phosph...
ILMN_1729175	0.000431	1.10e-07	-11.18	7.99	-1.327	FBXO3	F-box protein 3
ILMN_1799062	0.0004858	1.34e-07	10.98	7.82	1.668	NFKB2	nuclear factor kappa B subunit 2
ILMN_1703015	0.0004873	1.44e-07	-10.9	7.75	-1.266	ZRANB2	zinc finger RANBP2-type cont...

Gambar 1. Hasil GEO2R

Hasil ini divisualisasikan menggunakan volcano plot visualisasi lainnya, yang menunjukkan distribusi gen berdasarkan nilai log2 fold change dan signifikansi statistic.

- Gen dengan log2FC positif dikategorikan sebagai upregulated, menandakan peningkatan ekspresi pada pasien ditunjukan gambar 2.

ID	adj.P.Val	P.Value	t	B	logFC	GI	Gene.symbol	Gene.title
ILMN_1671693	0.0000371	7.86E-10	17,3141126	11,86895	1,59888856	431822383	TPGS2	tubulin polyglutamylase complex subunit 2
ILMN_1656501	0.0000531	4.50E-09	14,868719	10,59609	2,6148653	62865889	DUSP5	dual specificity phosphatase 5
ILMN_1794976	0.0000531	3.95E-09	15,0399976	10,69531	2,46440728	526253109	NUDT4P1	nudix hydrolase 4 pseudogene 1
ILMN_3309473	0.0000531	3.30E-09	15,2796679	10,83139	2,29239392	262205766	MIR101-1	microRNA 101-1
ILMN_3235517	0.0000743	7.86E-09	14,1562524	10,165	2,13652833	118918410	ZNF777	zinc finger protein 777
ILMN_2390859	0.000212	2.69E-08	12,6897859	9,17506	1,65077869	570359567	NFKB2	nuclear factor kappa B subunit 2
ILMN_1726547	0.0002126	3.60E-08	12,3642355	8,93455	1,83151642	21536459	MAP3K5	mitogen-activated protein kinase kinase 5
ILMN_1811258	0.0002127	4.05E-08	12,2336532	8,83578	1,42311049	317575693	RELB	RELB proto-oncogene, NF- $\kappa$ B subunit
ILMN_3236637	0.0003403	7.93E-08	11,5156538	8,26814	1,5863061	156416001	PIP4K2A	phosphatidylinositol-5-phosphate 4-kinase type 2 alpha
ILMN_1799062	0.0004858	1.34E-07	10,9807732	7,81657	1,66813278	570359558	NFKB2	nuclear factor kappa B subunit 2
ILMN_1673682	0.0005047	1.77E-07	10,700467	7,56948	1,14820941	164519145	GATAD2A	GATA zinc finger domain containing 2A
ILMN_2094061	0.0005047	1.82E-07	10,6770677	7,54852	1,13811429	319996748	IMPA2	inositol monophosphatase 2

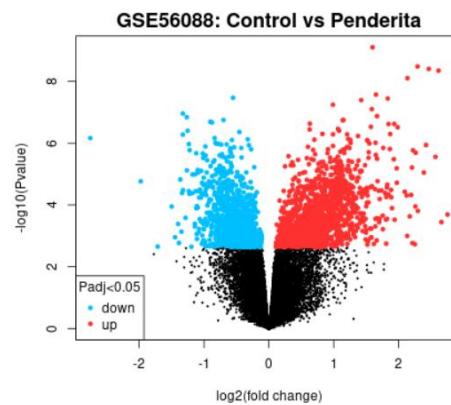
Gambar 2. Upregulated

- Gen dengan log2FC negatif dikategorikan sebagai downregulated, menandakan penurunan ekspresi pada pasien ditunjukan gambar 3.

ID	adj.P.Val	P.Value	t	B	logFC	GI	Gene.symbol	Gene.title
ILMN_1729175	0.000431	1.10E-07	-11,1823926	7,98979	-1,32744678	48928040	FBXO3	F-box protein 3
ILMN_1703015	0.0004873	1.44E-07	-10,9036226	7,74929	-1,26570527	300797236	ZRANB2	zinc finger RANBP2-type containing 2
ILMN_1723414	0.0006704	3.97E-07	-9,9324832	6,85321	-1,2540102	548923676	HACL1	2-hydroxyacyl-CoA lyase 1
ILMN_1740171	0.0007345	5.29E-07	-9,6715728	6,59601	-1,32658005	205277446	DUSP11	dual specificity phosphatase 11
ILMN_1726153	0.0008761	6.88E-07	-9,4358667	6,35728	-2,75418202	NA		
ILMN_1776337	0.0009917	1.07E-06	-9,0472364	5,94997	-1,23392614	221316565	CHORDC1	cysteine and histidine rich domain containing 1
ILMN_1730824	0.001115	1.68E-06	-8,6673418	5,53463	-1,22196889	404501449	ZNF512	zinc finger protein 512
ILMN_1794505	0.0012534	2.12E-06	-8,4763975	5,31921	-1,12025166	5453639	SHFM1	split hand/foot malformation (ectrodactyly) type 1
ILMN_1749641	0.0012581	2.16E-06	-8,4632314	5,30419	-1,01886805	48928041	FBXO3	F-box protein 3
ILMN_1812070	0.0014655	3.02E-06	-8,1917977	4,98955	-1,01287246	318037598	ABCB1	ATP binding cassette subfamily B member 1
ILMN_1704557	0.0021903	7.63E-06	-7,4762074	4,1133	-1,26431402	440546393	RPS6KB1	ribosomal protein S6 kinase B1
ILMN_1750256	0.0021903	7.40E-06	-7,4995964	4,14305	-1,09751352	209364519	ALS2	ALS2, alsin Rho guanine nucleotide exchange factor
ILMN_1853116	0.0021903	7.90E-06	-7,4500186	4,07989	-1,04336739	22658379		

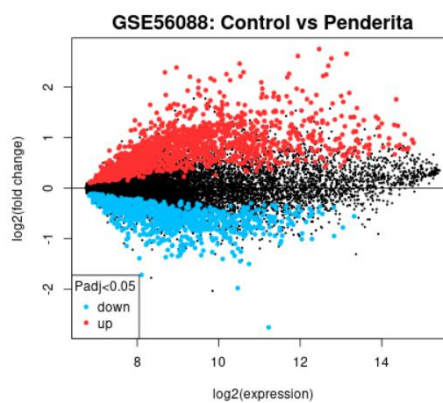
Gambar 3. Downregulated

- Gen yang berada di bagian atas kiri dan kanan volcano plot merupakan gen yang paling signifikan dan memiliki perubahan ekspresi terbesar ditunjukan gambar 4.



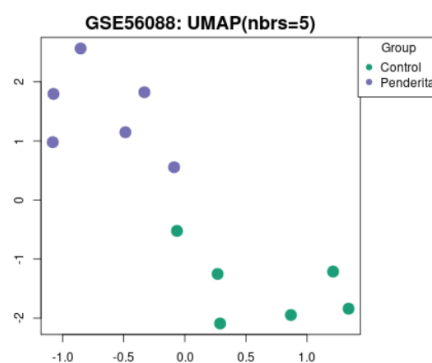
Gambar 4. Volcano Plot

- Gen yang berada di dekat garis tengahnya tidak berubah, jauh di atas/bawah artinya DEG kuat dan sebaran simetris artinya normalisasi data baik



Gambar 5. Mean Difference (MD) Plot

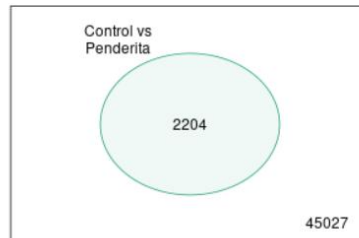
- Titik kontrol dan penderita tampak terpisah cukup jelas menunjukkan perubahan global ekspresi gen antara control dan penderita



Gambar 6. UMAP Plot

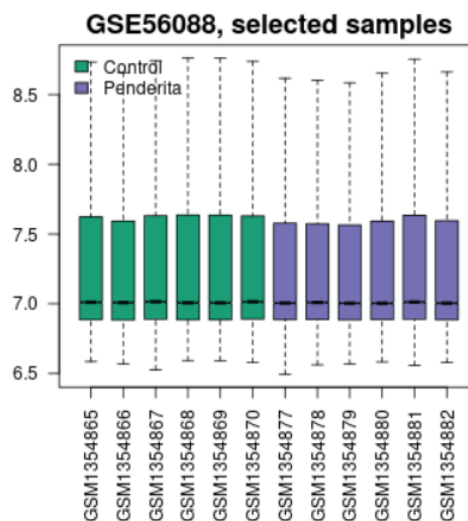
- Tertera sekitar 2204 gen signifikan ( $\text{Padj} < 0,05$ ). Jumlah total probe jauh lebih besar ( $\sim 45.000$ ). Hanya sebagian kecil gen yang benar-benar berubah, yang kemudian disaring lagi menjadi DEG utama (277 gen) dengan kriteria  $\log\text{FC}$ .

**GSE56088: limma,  $\text{Padj} < 0.05$**



*Gambar 7. Venn Diagram*

- Median semua sampel hampir sejajar dan variasi antar sampel seragam artinya Data berkualitas baik dan layak untuk analisis DEG.



*Gambar 8. Box Plot*

Distribusi ini menandakan adanya perubahan transkriptomik nyata pada tahap lanjut diferensiasi eritroid.

### Interpretasi Biologis

Munculnya DEG pada hari ke-14 tetapi tidak pada hari ke-7 menunjukkan bahwa perubahan molekuler pada  $\beta$ -thalassaemia berkaitan dengan tahap pematangan eritroid lanjut, bukan tahap awal diferensiasi. Gen-gen yang mengalami upregulasi kemungkinan terlibat dalam respons stres seluler, gangguan metabolisme eritrosit, atau mekanisme kompensasi akibat ketidakseimbangan rantai globin.

Sebaliknya, gen-gen yang mengalami downregulasi dapat berkaitan dengan terganggunya fungsi normal eritropoiesis dan pematangan sel darah merah.

Dengan demikian, DEG yang teridentifikasi berpotensi mencerminkan mekanisme patologis utama pada sel eritroid pasien  $\beta$ -thalassaemia. Adapun gen-gen DEG tersebut berpotensi menjadi biomarker molekuler  $\beta$ -thalassemia, target terapi baru dan indikator mekanisme patogenesis eritropoiesis tidak efektif

## **KESIMPULAN**

Analisis DEG menggunakan GEO2R pada data mikroarray sel prekursor eritroid hari ke-14 menunjukkan adanya perbedaan ekspresi gen yang signifikan antara pasien  $\beta$ -thalassaemia dan individu sehat. Sebanyak 277 gen teridentifikasi sebagai DEG, yang mengindikasikan adanya perubahan regulasi gen pada tahap lanjut diferensiasi eritroid. Hasil ini mendukung hipotesis bahwa gangguan eritropoiesis pada  $\beta$ -thalassaemia lebih nyata pada fase pematangan sel, dan analisis transkriptomik dapat digunakan untuk mengungkap mekanisme molekuler yang mendasari penyakit ini. Temuan DEG ini memberikan dasar penting untuk memahami mekanisme penyakit, mengembangkan biomarker diagnostik, serta eksplorasi target terapi molekuler di masa depan.