

ویژگی‌هایی مولکولی کم‌خونی آپلاستیک در کودکان

فرزان رحمانی
مریم صادق آبادی
محمد رضا مزروعی
بهمن 1403

فهرست

- معرفی بیماری کم خونی آپلاستیک
- معرفی مجموعه داده ها
- پیاده سازی مقاله
- ادغام دیتاست ها

تعريف مسئله

- کم خونی آپلاستیک یک بیماری نادر با کمبود سلولهای خونی و مغز استخوان که علائمی مانند خستگی و خونریزی دارد و نیاز به مراقبت حمایتی سریع دارد.
- عوامل محیطی مانند تماس با مواد شیمیایی و عفونتها، به علاوه عوامل ژنتیکی، ممکن است خطر ابتلا به بیماری را افزایش دهند. کودکان مبتلا به AA، به دلیل وجود نقایص ارثی متعدد و حساسیت به درمانها، نیاز به تشخیص و رویکردهای درمانی متفاوت دارند.

هدف مقاله و کار ما

- نقش تلومرها و مطالعات ژنتیکی: تلومرها در AA نقش حفاظتی دارند و طول آنها شاخصی برای تشخیص و درمان است. مطالعات ژنومی تفاوت‌های قابل توجهی در بیان ژن‌های متابولیسم سلولی و ایمنی بین بیماران و افراد سالم نشان داده‌اند.
- سلول‌های MSCs مشتق از مغز استخوان در هماتوپوییز و ایمنی نقش دارند، اما در بیماران AA دچار اختلال تکثیر و تغییرات ژنی مرتبط با چرخه سلولی و ایمنی می‌شوند، که از ویژگی‌های کلیدی بیماری است.
- مطالعه حاضر به بررسی تاثیر مسیرهای نگهداری تلومر و وضعیت بیان ژن در MSCs می‌پردازد تا ارتباط آنها با نتایج درمانی در کودکان مبتلا به AA تعیین شود.

مجموعه داده اصلی

- نمونه‌ها و جمع‌آوری داده‌ها
- استخراج DNA از خون محیطی
- استخراج cDNA و تهیه RNA
- تحلیل بیان ژن MSCs
- یافتن سلول‌های بنیادی مزانشیمی (MSCs) با استفاده از داده‌های ریزآرایه موجود در پایگاه داده Gene Expression Omnibus (GEO)

مجموعه داده کمکی

- دیتاست GSE165870
- دیتاست GSE28974
- دیتاست GSE3807
- دیتاست GSE113033
- دیتاست GSE29105

نمایش داده خام

FeatureNum	Row	Col	accessions	ProbeUID	ControlType	ProbeName	GeneName	SystematicName	Description	...	rIsBGNonUnifOL	gIsFeatPopnOL	rIsFeatPopnOL	gIsBGPopnOL	rIsB
<int>	<int>	<int>	<chr>	<int>	<int>	<chr>	<chr>	<chr>	<chr>	...	<int>	<int>	<int>	<int>	<int>
1	1	1		0	1	GE_BrightCorner	GE_BrightCorner	GE_BrightCorner		...	0	1	1	1	1
2	1	2		1	1	DarkCorner	DarkCorner	DarkCorner		...	0	0	0	0	0
3	1	3		1	1	DarkCorner	DarkCorner	DarkCorner		...	0	1	0	0	0
4	1	4		1	1	DarkCorner	DarkCorner	DarkCorner		...	0	0	0	0	0
5	1	5		1	1	DarkCorner	DarkCorner	DarkCorner		...	0	0	0	0	0
6	1	6		1	1	DarkCorner	DarkCorner	DarkCorner		...	0	0	0	0	0

پیش پردازش داده

- مدیریت داده های تکراری
- مدیریت داده های گمشده
- حذف پروب های کنترلی

پیش پردازش داده

- تشکیل ماتریس بیان ژن

- مدیریت ژن های یکسان در پروپهای متفاوت

```
[ ] # Step 3: Merge the green and red data into a single dataframe by GeneName
final_data <- merge(green_data, red_data, by = "GeneName", suffixes = c("_green", "_red"))

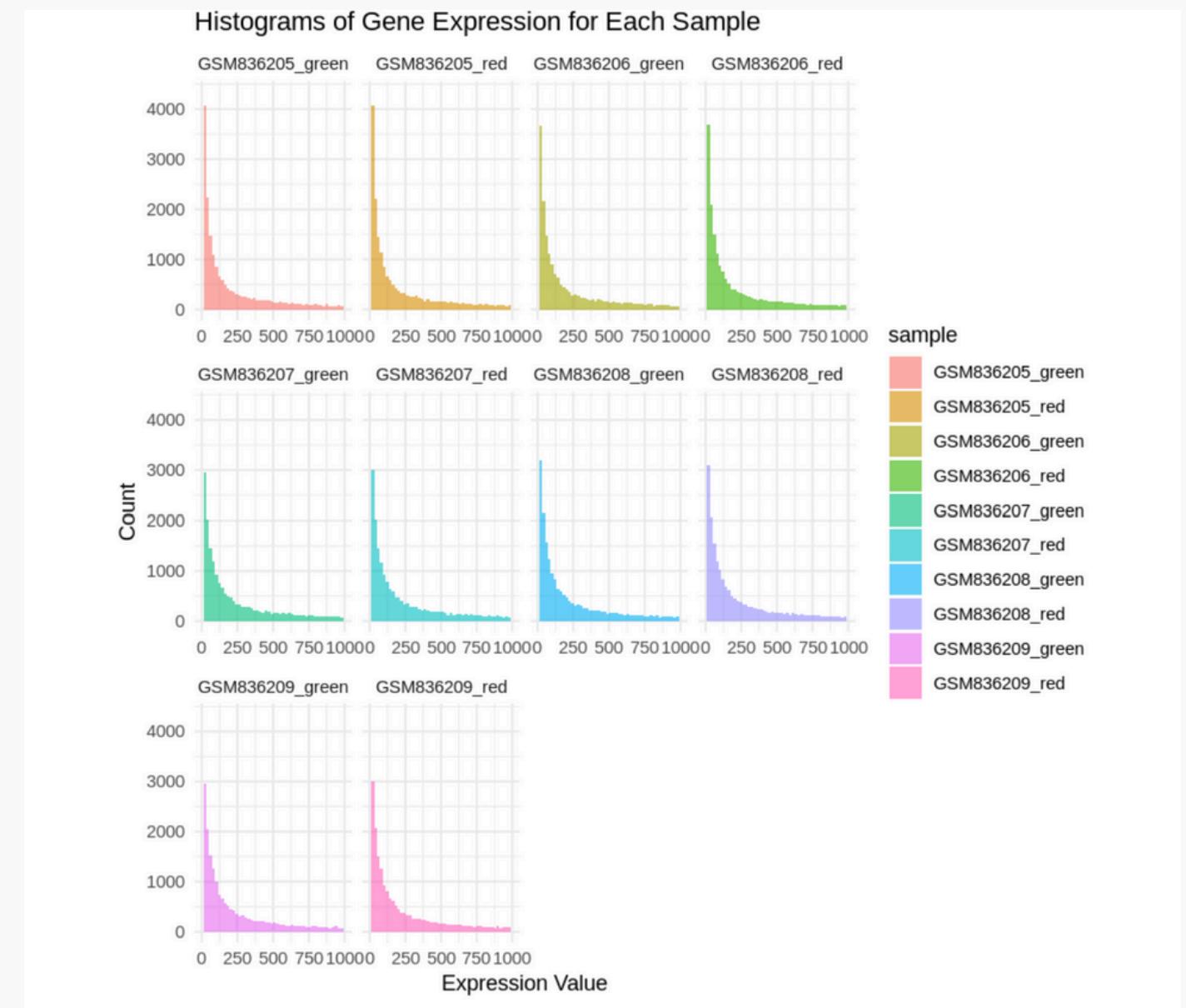
# View the final expression matrix
head(final_data)
```

	GeneName	GSM836205_green	GSM836206_green	GSM836207_green	GSM836208_green	GSM836209_green	GSM836205_red	GSM836206_red	GSM836207_red	GSM836208_red	GSM836209_red
	<chr>	<dbl>	<dbl>	<dbl>	<dbl>	<dbl>	<dbl>	<dbl>	<dbl>	<dbl>	<dbl>
1	15E1.2	412.1120	466.2163	472.5227	437.5120	445.8263	411.3263	568.1197	589.5123	407.9157	447.4650
2	2'-PDE	135.6863	160.4447	149.2900	139.8413	127.0240	141.4970	196.3577	170.4140	146.7380	145.0077
3	3.8-1	2.5500	2.9610	5.0560	3.8010	5.3220	7.8380	6.6710	4.6980	4.1320	4.5740
4	39508	16.4370	18.3620	12.2525	16.1835	19.0725	4.1725	33.8945	23.4145	3.8425	30.9545
5	39509	557.3825	648.1950	686.6245	627.3595	615.2060	902.8760	903.9380	1386.6660	1247.0575	838.3375
6	39510	168.1120	186.8980	184.4350	186.2810	160.7080	209.7900	565.9530	629.3900	407.8720	125.2710

تحلیل آماری قبل از پیش پردازش

```
summary(final_data)
```

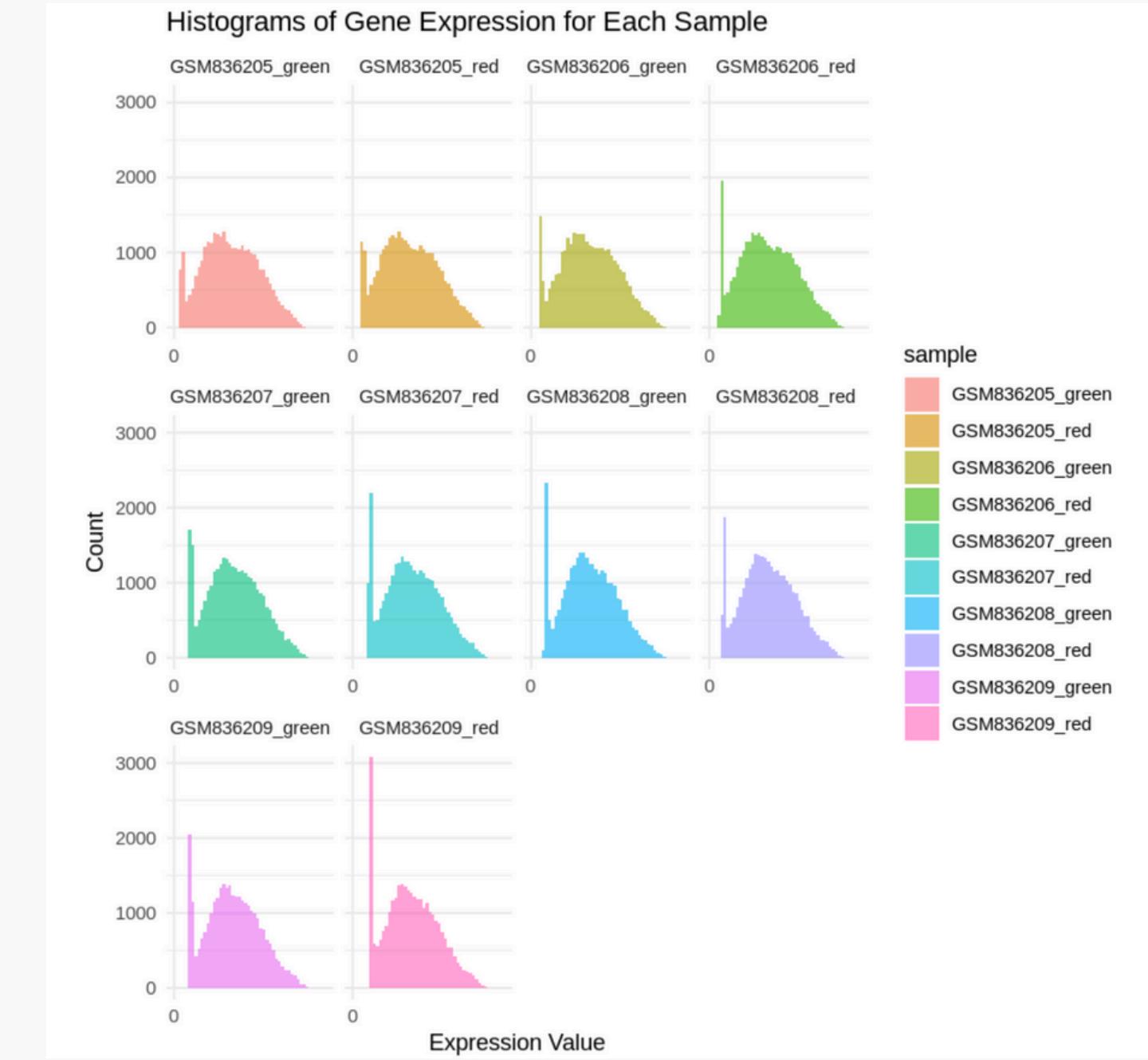
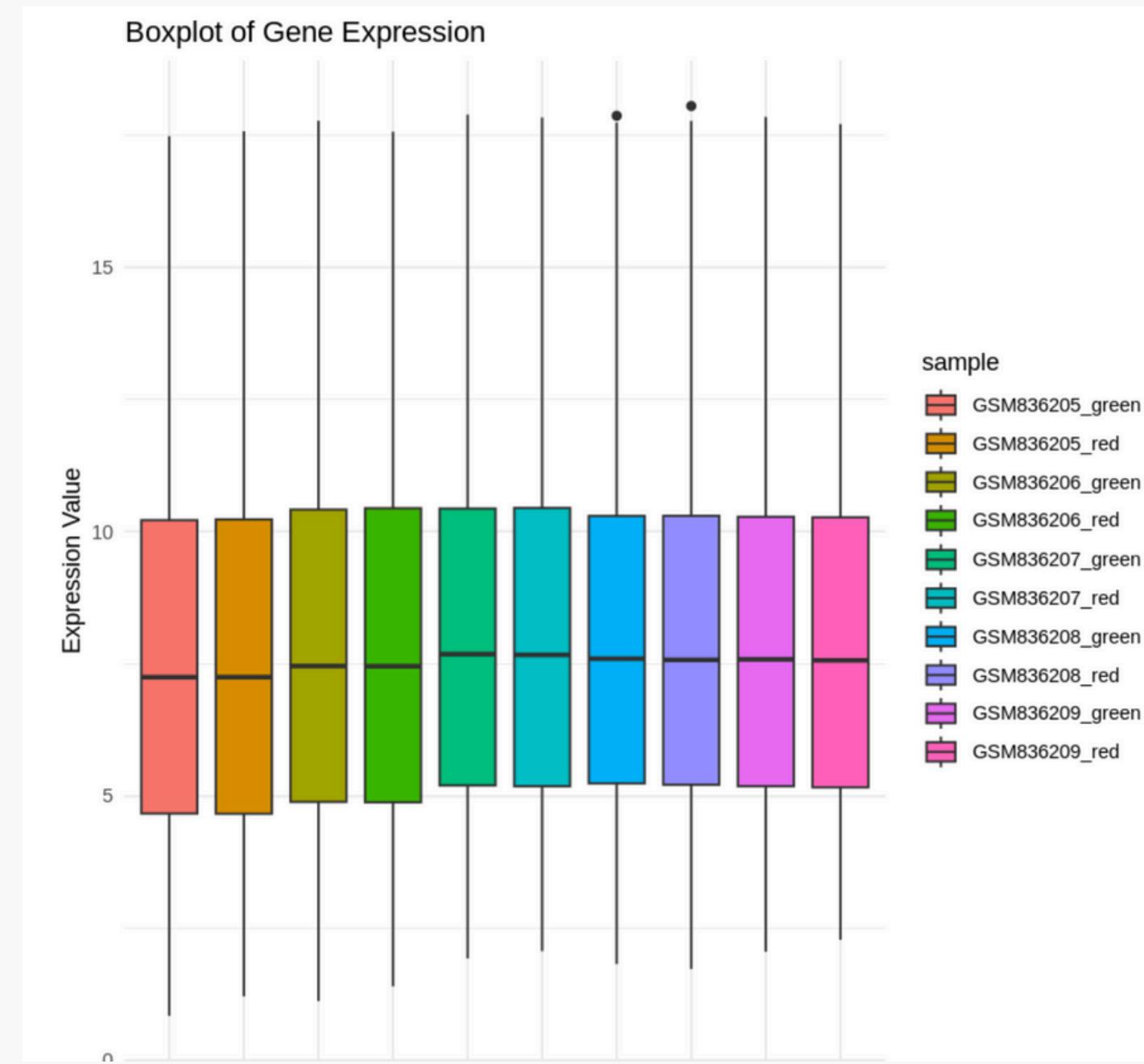
GeneName	GSM836205_green	GSM836206_green	GSM836207_green
Length:30587	Min. : 2.32	Min. : 2.64	Min. : 4.19
Class :character	1st Qu.: 25.42	1st Qu.: 29.56	1st Qu.: 36.46
Mode :character	Median : 152.71	Median : 176.19	Median : 204.53
	Mean : 2769.90	Mean : 3408.71	Mean : 3476.07
	3rd Qu.: 1197.67	3rd Qu.: 1386.28	3rd Qu.: 1391.04
	Max. :195145.20	Max. :193934.10	Max. :234258.80
GSM836208_green	GSM836209_green	GSM836205_red	
Min. : 3.32	Min. : 4.86	Min. : 1.79	
1st Qu.: 37.23	1st Qu.: 35.98	1st Qu.: 25.49	
Median : 191.41	Median : 190.24	Median : 152.55	
Mean : 3210.39	Mean : 3178.97	Mean : 2767.77	
3rd Qu.: 1254.45	3rd Qu.: 1229.62	3rd Qu.: 1184.76	
Max. :271853.20	Max. :214365.60	Max. :183058.70	
GSM836206_red	GSM836207_red	GSM836208_red	
Min. : 2.18	Min. : 3.81	Min. : 3.53	
1st Qu.: 29.72	1st Qu.: 36.85	1st Qu.: 37.86	
Median : 176.60	Median : 206.31	Median : 194.39	
Mean : 3344.29	Mean : 3433.10	Mean : 3224.97	
3rd Qu.: 1362.52	3rd Qu.: 1381.27	3rd Qu.: 1252.78	
Max. :224149.90	Max. :242766.20	Max. :239130.20	
GSM836209_red			
Min. : 4.16			
1st Qu.: 36.48			
Median : 192.70			
Mean : 3235.24			
3rd Qu.: 1239.64			
Max. :235419.80			



پیش پردازش داده

- اصلاح پس زمینه
- نرمالسازی
- تبدیل لگاریتمی

تحلیل آماری پس از پیش پردازش

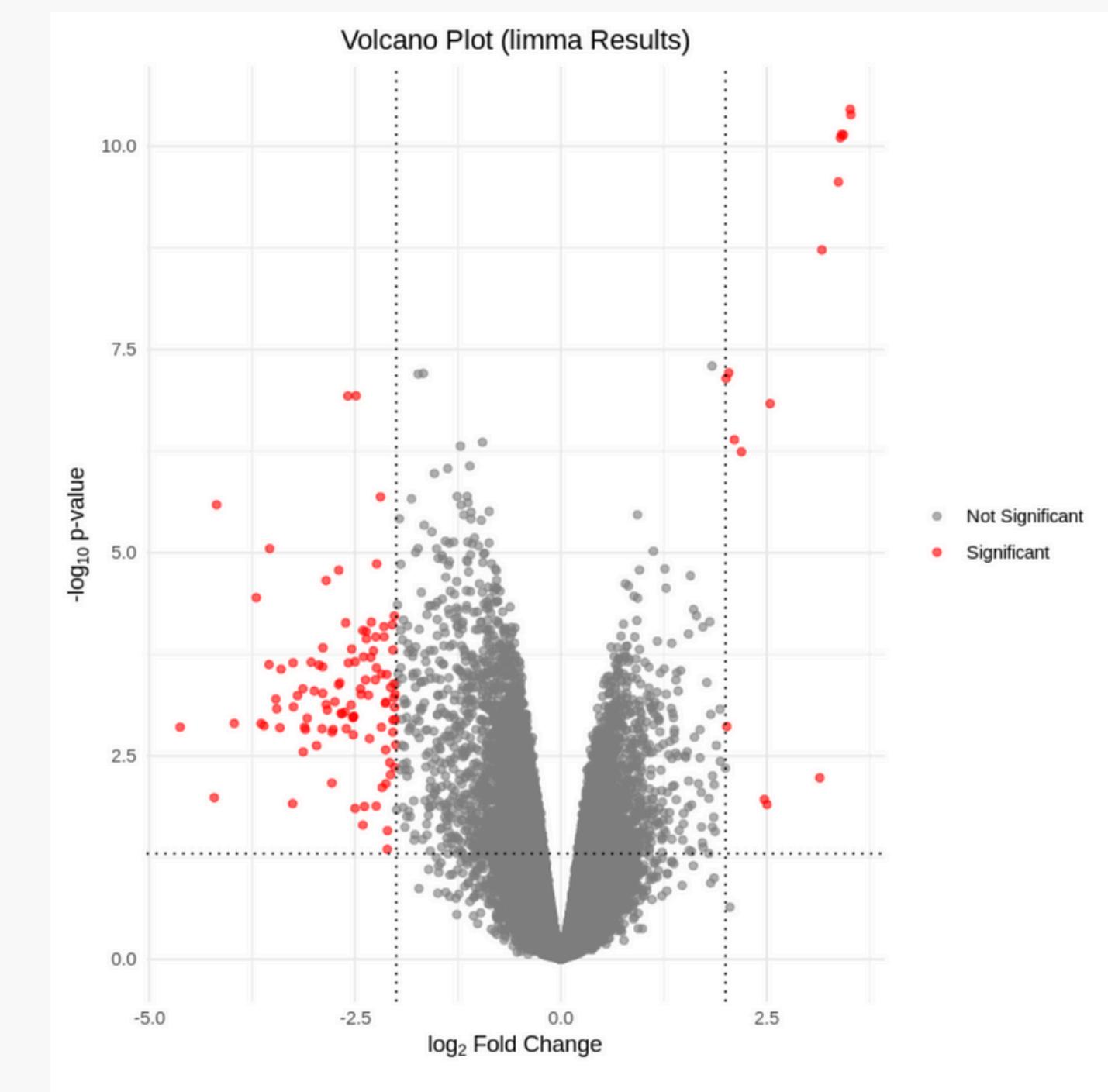


افتراق بیان ژنی

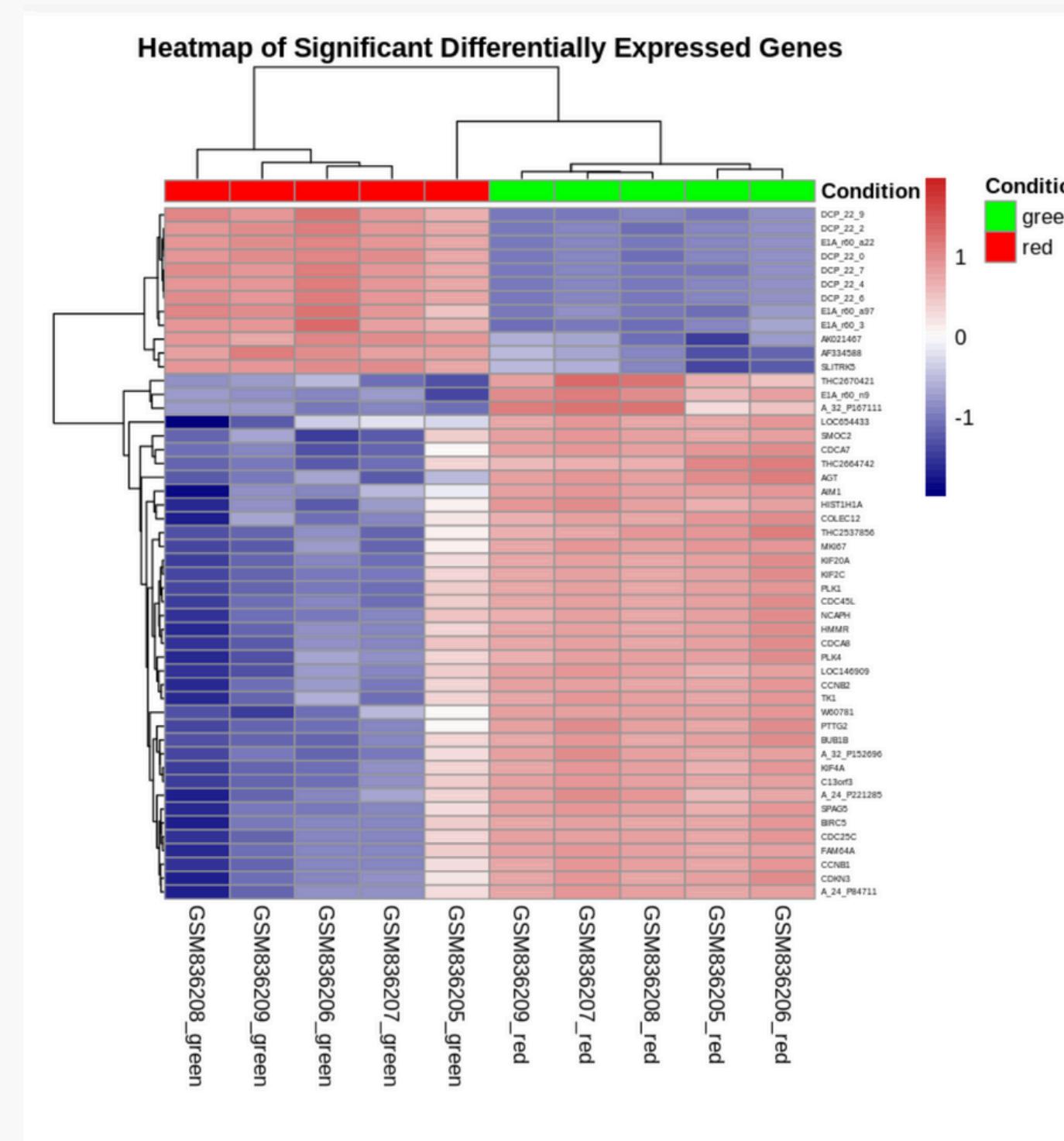
A data.frame: 6 × 6						
	logFC	AveExpr	t	P.Value	adj.P.Val	B
	<dbl>	<dbl>	<dbl>	<dbl>	<dbl>	<dbl>
E1A_r60_a22	3.514793	15.24751	28.37765	3.501205e-11	4.792656e-07	12.37786
DCP_22_0	3.522298	15.19104	27.95262	4.087718e-11	4.792656e-07	12.31968
DCP_22_4	3.408842	14.40768	26.47121	7.144605e-11	4.792656e-07	12.10036
DCP_22_2	3.434484	14.95367	26.44633	7.213764e-11	4.792656e-07	12.09644
DCP_22_6	3.396627	13.67475	26.23413	7.834466e-11	4.792656e-07	12.06269
DCP_22_7	3.370446	12.60764	23.21671	2.733109e-10	1.393293e-06	11.51015

نمایش ژن ها با بیان معنادار

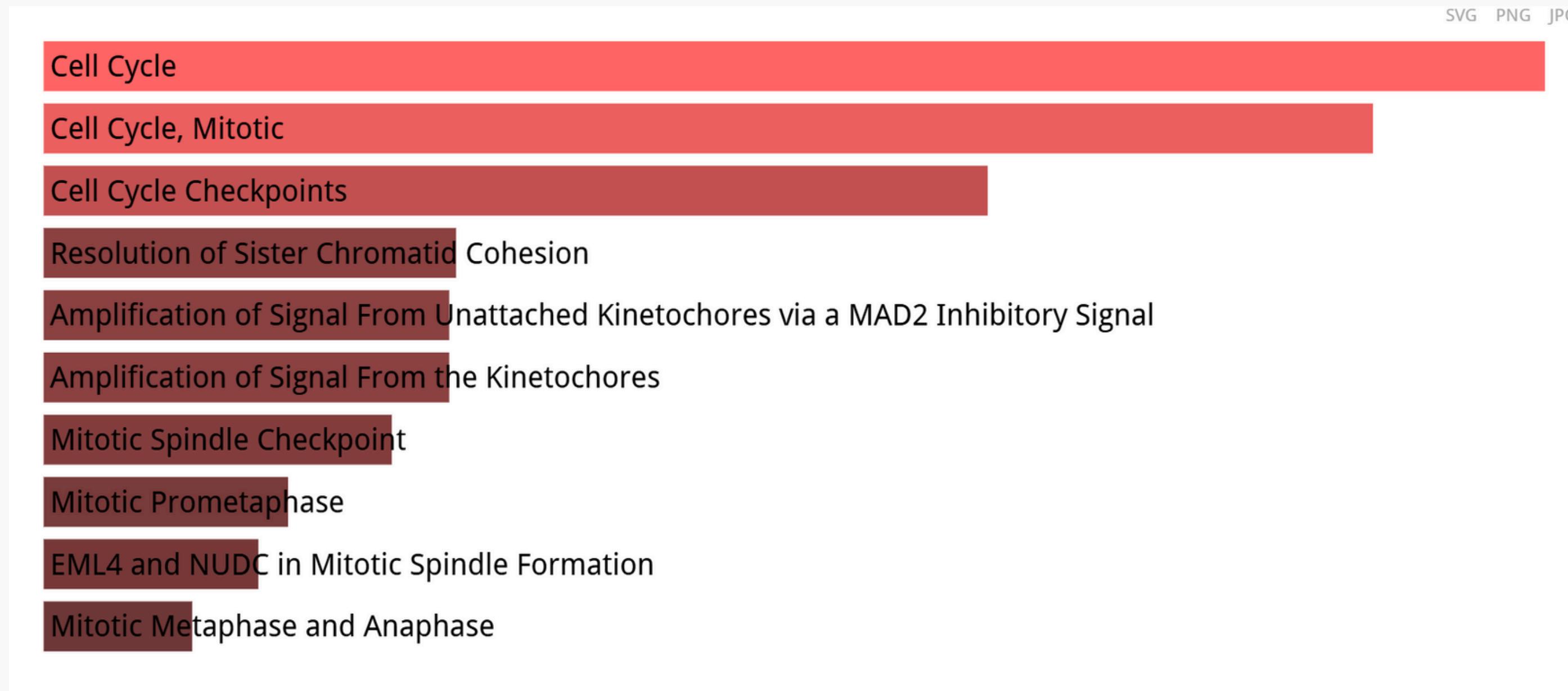
- adj.pvalue < 0.05
- $|LogFoldChange| < 2$



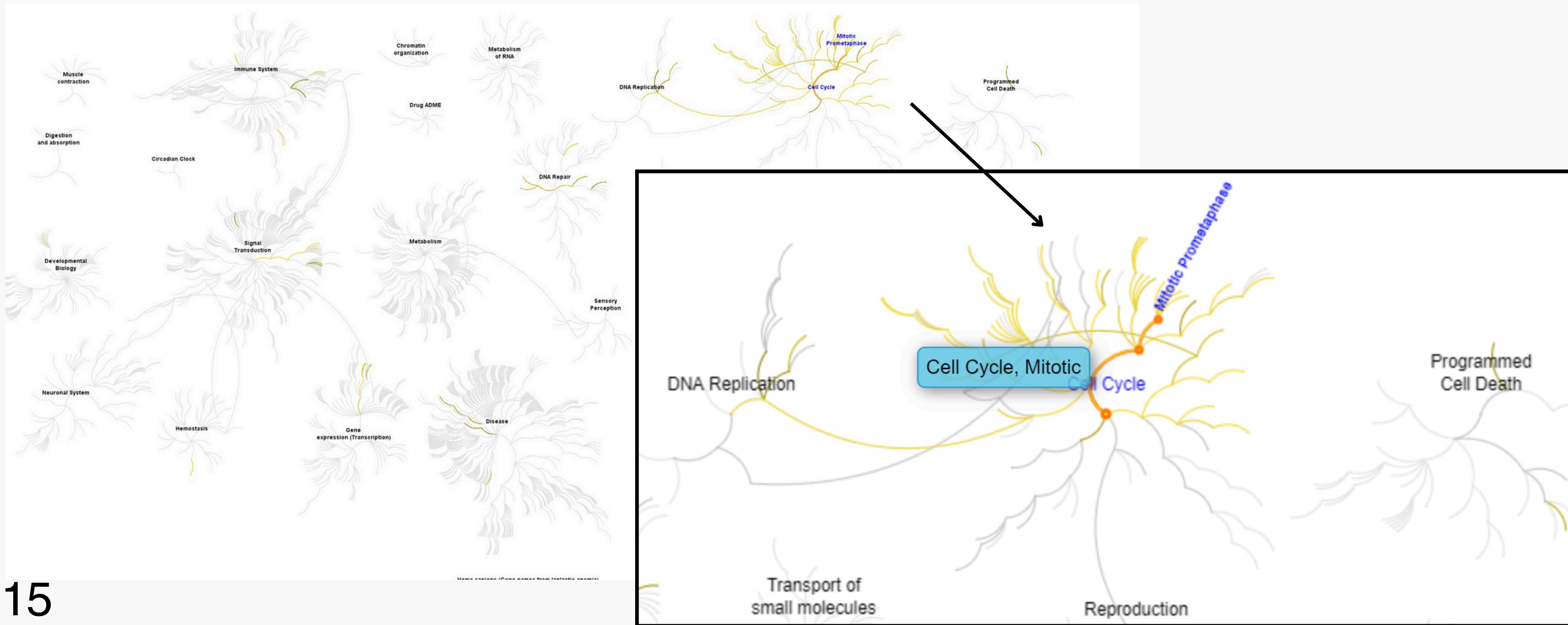
نمایش شباهت نمونه ها براساس ژن های معنادار



تعیین مسیرهای زیستی



نمایش اهمیت ژن ها در تعیین مسیرهای زیستس



مقایسه با نتایج مقاله

• بیان ژن های معنادار AGT و GAS2L3, MK167

Table 3

Differentially downregulated genes among paediatric patients with aplastic anaemia compared to healthy donors.

Patient 1_log fold change	Patient 2_log fold change	Patient 3_log fold change	Patient4_log fold change	Patient 5_log fold change	Gene	Gene name	Remarks
-2.144	-2.524	-2.97	-2.841	-2.535	AGT	angiotensinogen	Universally downregulated among all patients
-1.048	-2.275	-2.343	-4.192	-2.003	GAS2L3	growth arrest specific 2 like 3	Differentially downregulated among patients with immunosuppressive therapy
-1.281	-2.809	-3.336	-3.586	-3.243	MKI67	marker of proliferation Ki-67	Differentially downregulated among patients with immunosuppressive therapy
-1.437	-4.012	-3.921	-3.828	-3.173	TMSB15A	thymosin beta 15a	Differentially downregulated among patients with immunosuppressive therapy

چالش کمبود داده های اصلی مقاله

- برای انجام تحلیل های جامع تر مجموعه داده های مشابه پیدا کنیم.
- هر کدام را جدا و بعد در کنار هم تحلیل کنیم!

1. GSE165870 (<https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE165870>)

Bulk RNA-seq data of Lin-CD34+ hematopoietic stem and progenitor cells derived from bone marrow of healthy donors and untreated aplastic anemia patients

2. GSE28974 (<https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE28974>)

Genome-wide gene expression profiling of mesenchymal stem cells

3. GSE3807 (<https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE3807>)

CD3+ T cells derived from bone marrow or peripheral blood of aplastic anemia patients

4. GSE29105 (<https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE29105>)

Isolation of single human hematopoietic stem cells capable of long-term multilineage engraftment

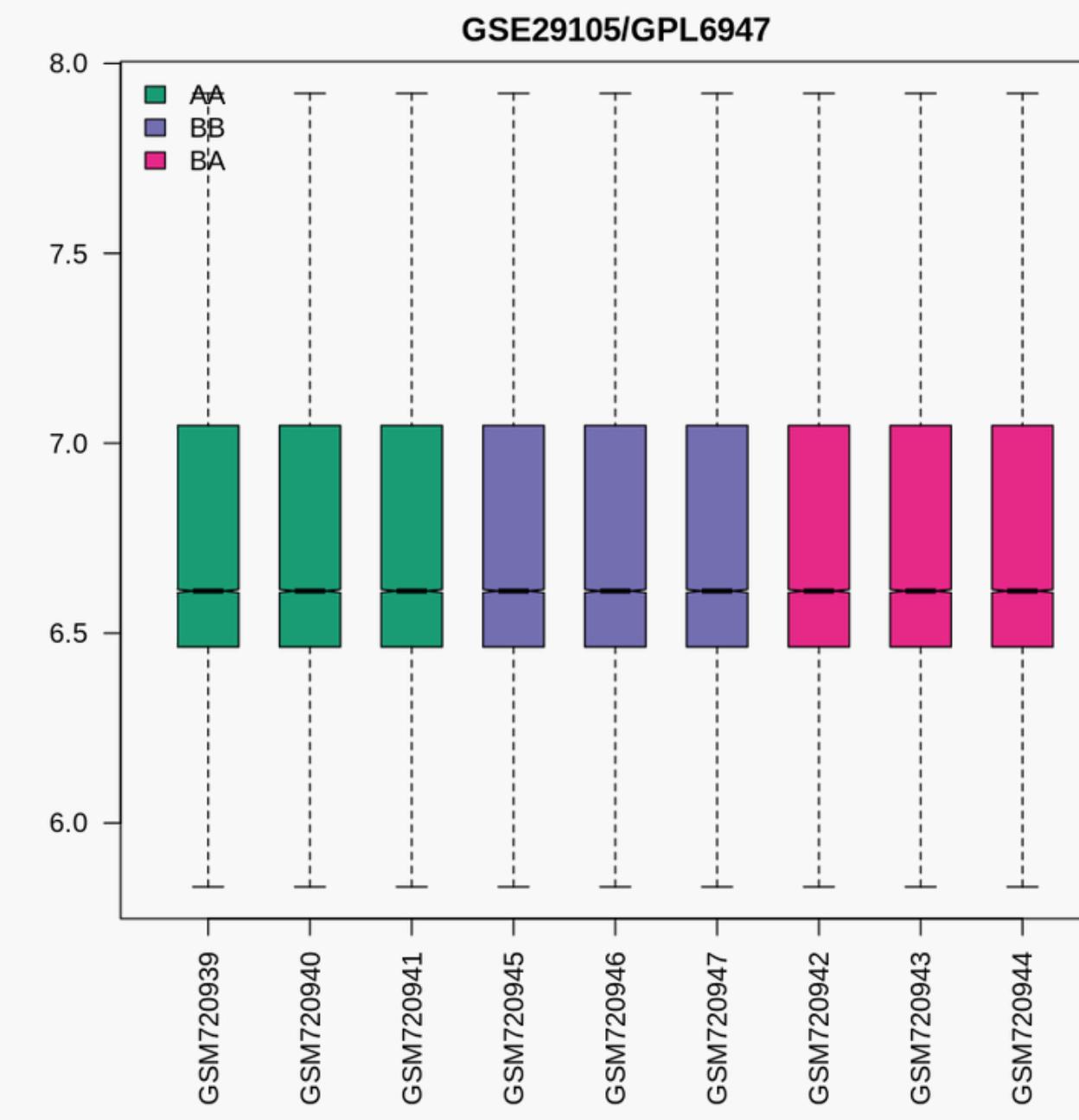
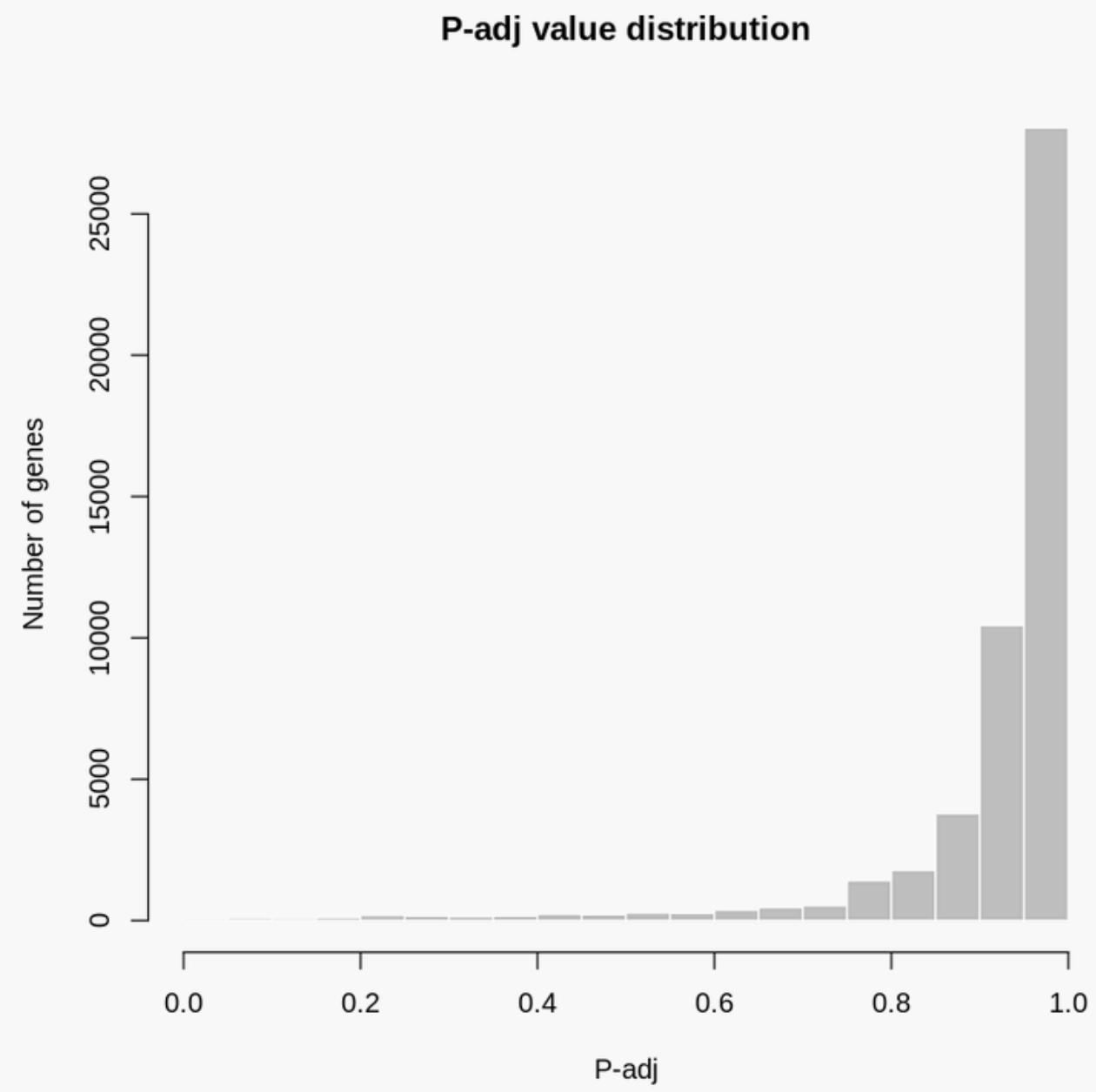
5. GSE113033 (<https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE113033>)

Gene expression profiling of novel hematopoietic stem cell subpopulations defined by CD35 expression and progenitor populations in human adult hematopoiesis

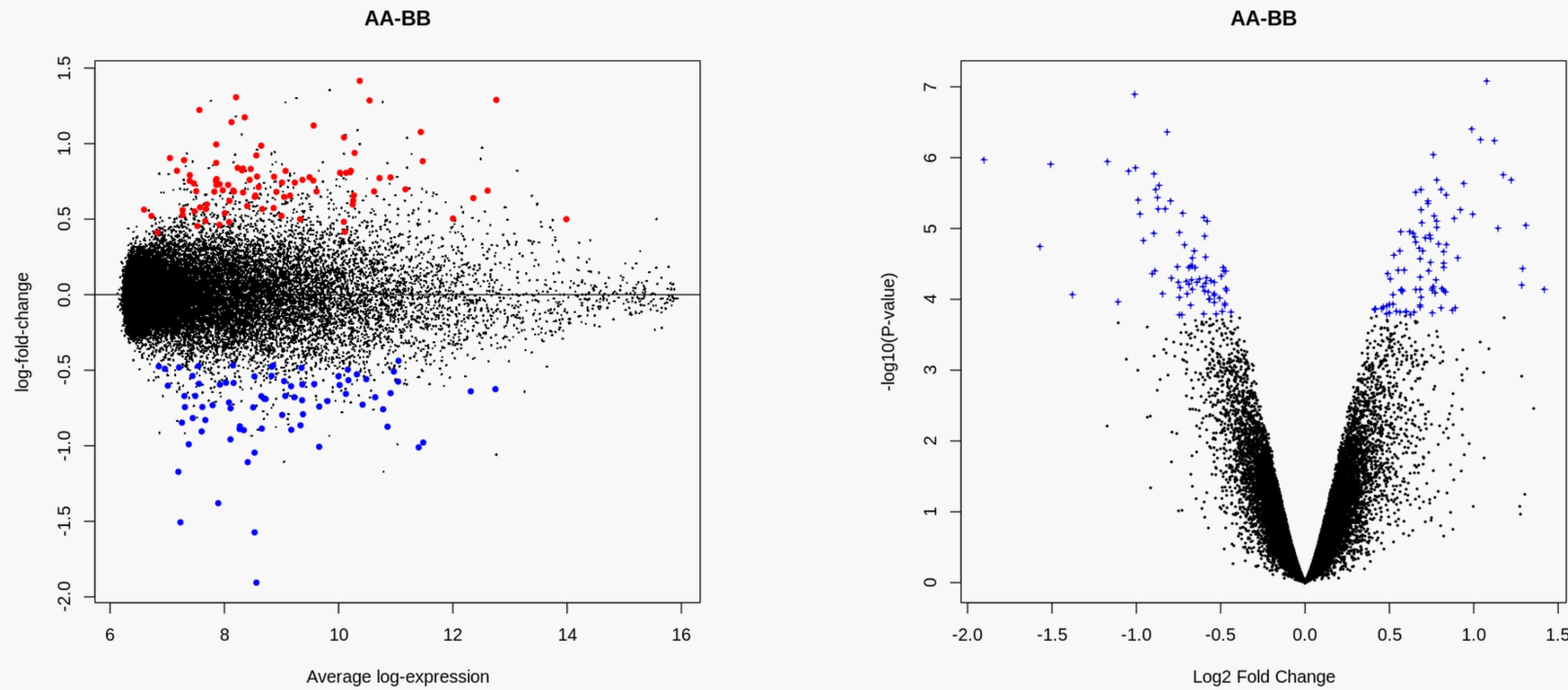
تحلیل جدا جدای یک نمونه از دیتاست ها

- باید GSE29105 را تحلیل کنیم.
- نصب package و کتابخانه
- دانلود و گروه بندی
- نرمال سازی
- به دست آوردن آماره ها و تحلیل

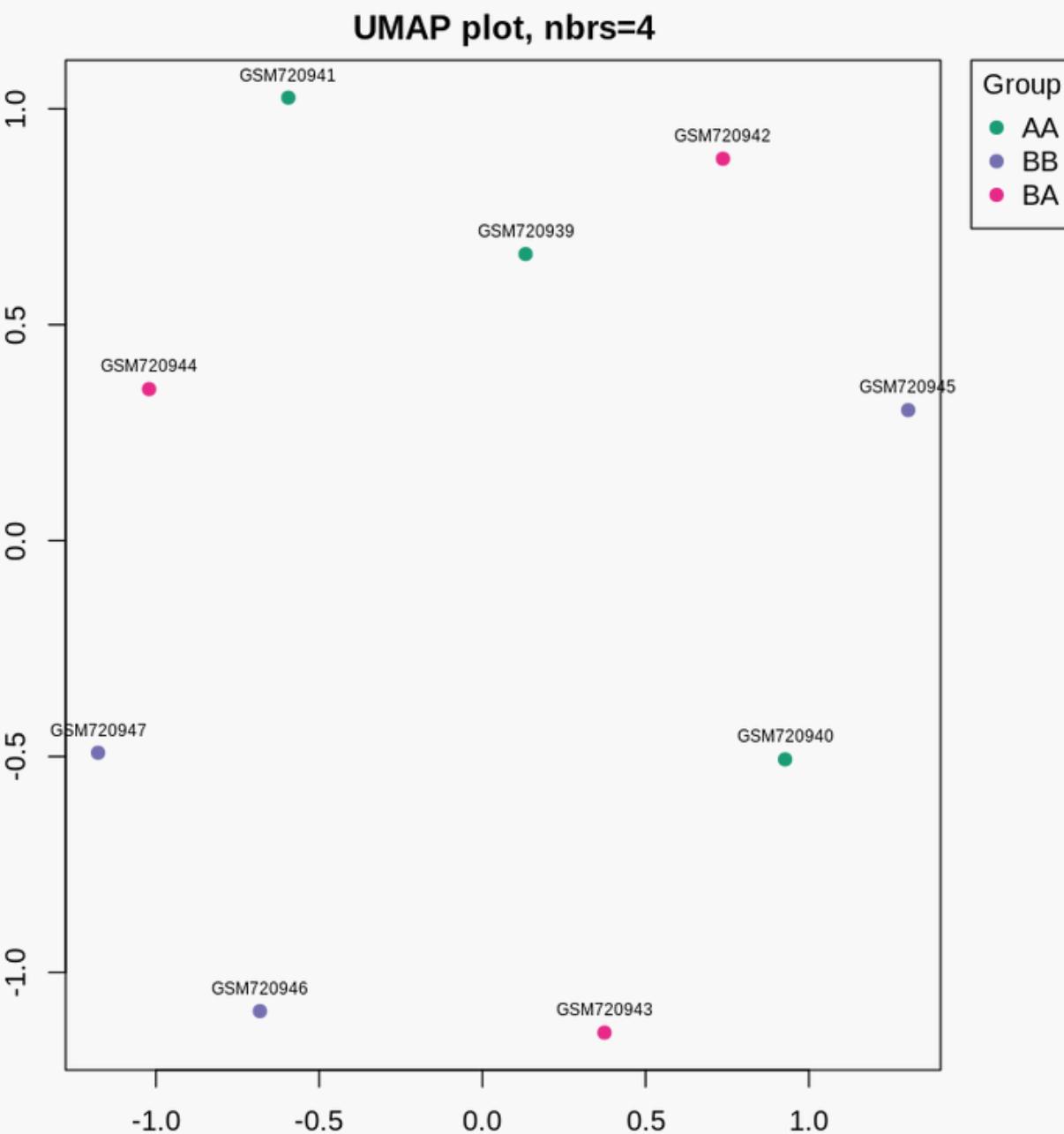
تحالیل GSE29105



GSE29105 تحلیل



تحليل GSE29105



تحلیل همزمان مجموعه داده ها

- برای ارزیابی شباهت با دیتاست اصلی
- چالش فرمت و تکنولوژی های مختلف برای ترکیب دیتاست ها
- اشتراک ژن های یکسان
- کاهش ابعاد
- خوش بندی

تحلیل همزمان مجموعه داده ها

• ترکیب دیتاست ها

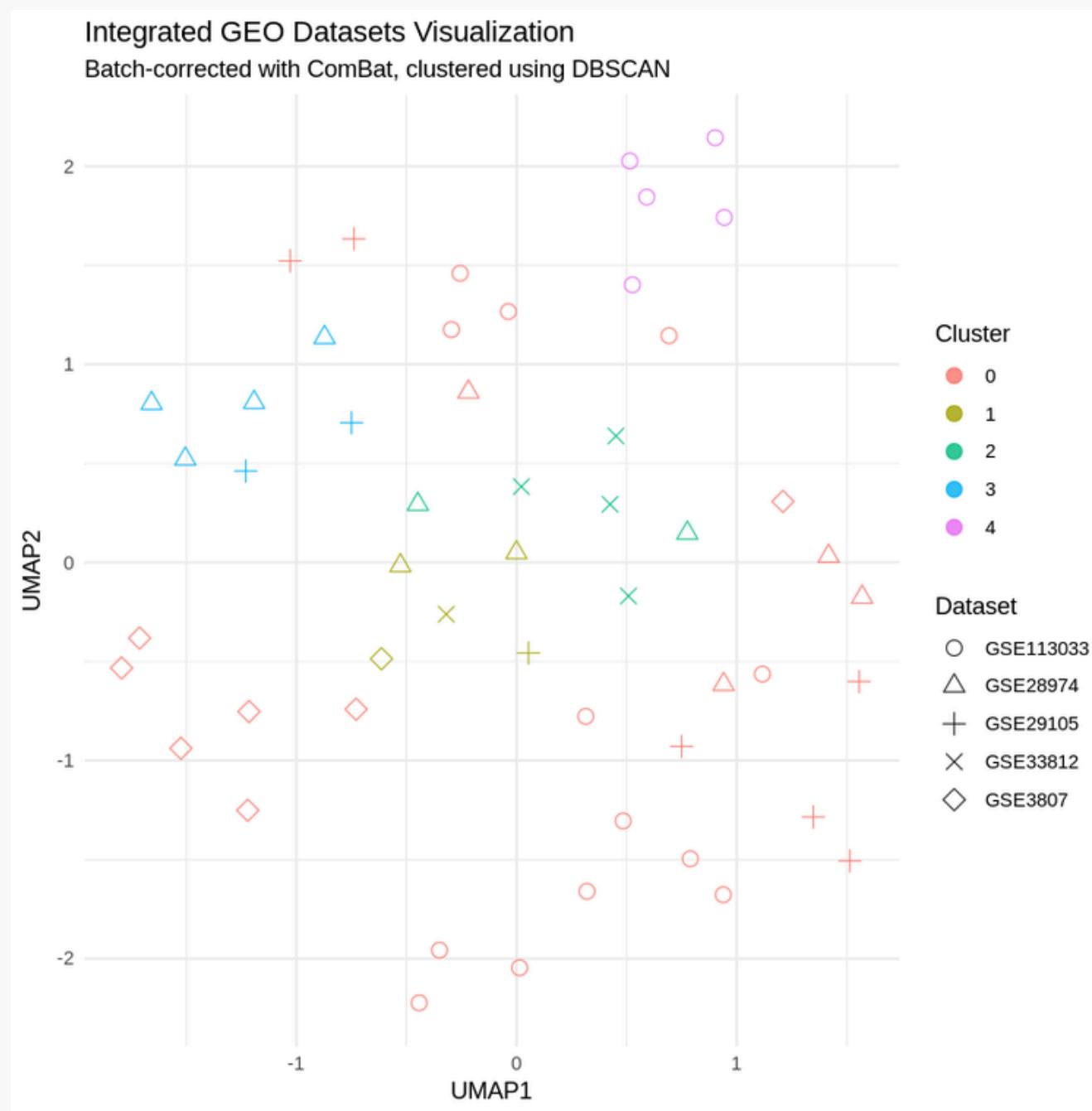
1 head(combat_expr)

A matrix: 6 × 52 of type dbl

GSM836205	GSM836206	GSM836207	GSM836208	GSM836209	GSM717846	GSM717847	GSM717848	GSM717849	GSM717850	...	GSM3094192	GSM3094193	GSM3094194	GSM3094195	GSM3094196	GSM3094197
54.51187	2.146381	27.56052	30.515183	45.08888	29.04906	47.27927	25.88757	41.43310	30.51736	...	39.59476	28.49316	20.79424	15.71909	34.14419	49.53370
28.82338	26.417119	24.93053	9.774833	25.78721	17.58702	31.07903	33.65275	27.44207	7.56443	...	26.89500	24.84303	13.24175	38.12516	23.49731	27.56444
23.78404	26.341856	28.86062	23.022553	26.36138	28.31288	21.20725	23.85619	26.05447	26.66121	...	22.48672	25.85532	25.63612	26.15854	23.54987	23.64850
28.53135	29.577987	30.18504	29.829181	30.51996	45.28547	34.84323	35.34772	25.96273	29.41819	...	21.00886	30.17767	35.53235	37.41993	18.01415	22.72584
53.99718	68.721868	62.68516	56.730021	62.11412	57.96213	65.37473	56.33859	53.22416	61.20798	...	52.22896	62.27357	60.85675	60.89102	41.99847	54.15801
13.99478	13.782799	17.82968	18.773957	20.12292	19.63477	17.44086	18.00990	16.84043	17.13710	...	19.60460	20.49751	19.89568	30.94676	10.43767	16.63061

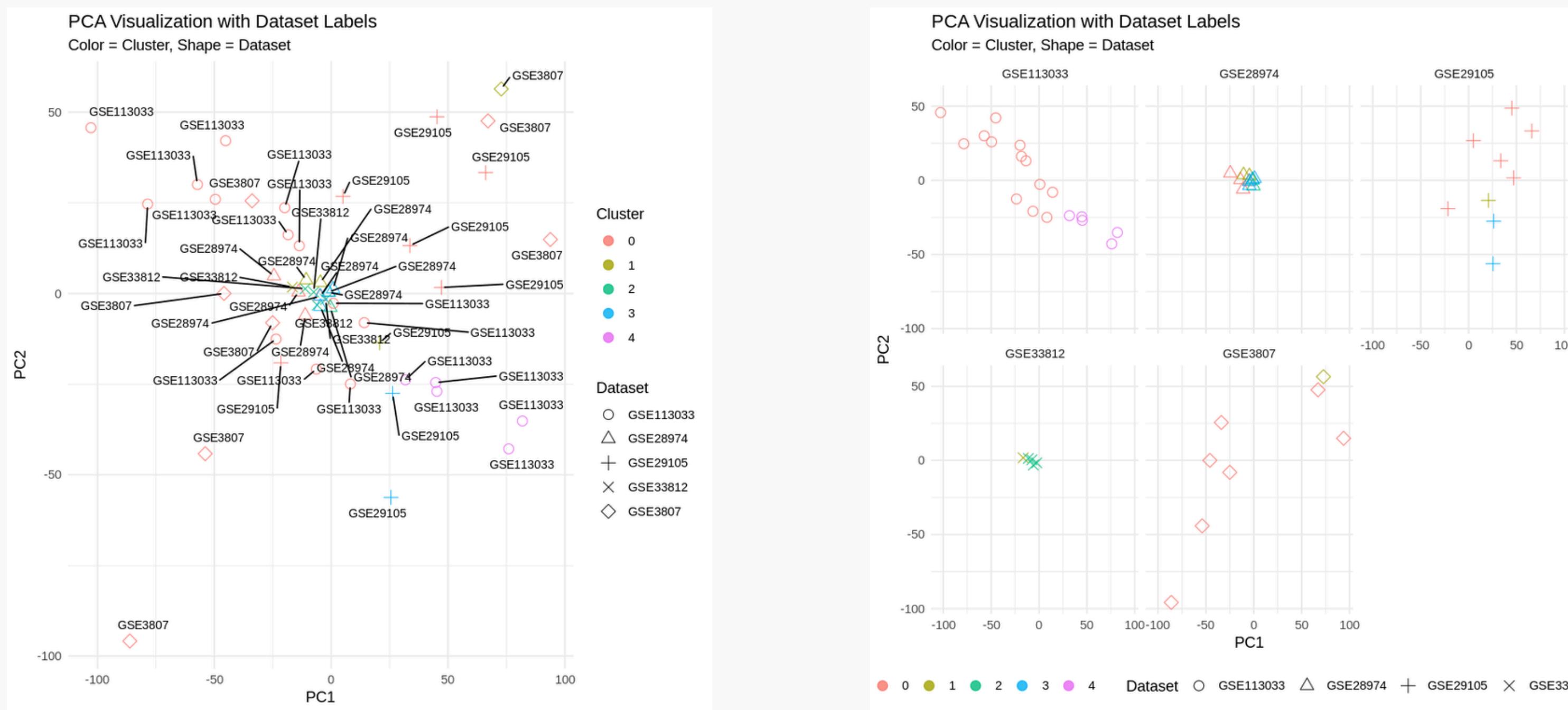
تحلیل همزمان مجموعه داده ها

• کاهش ابعاد و خوش بندی با UMAP



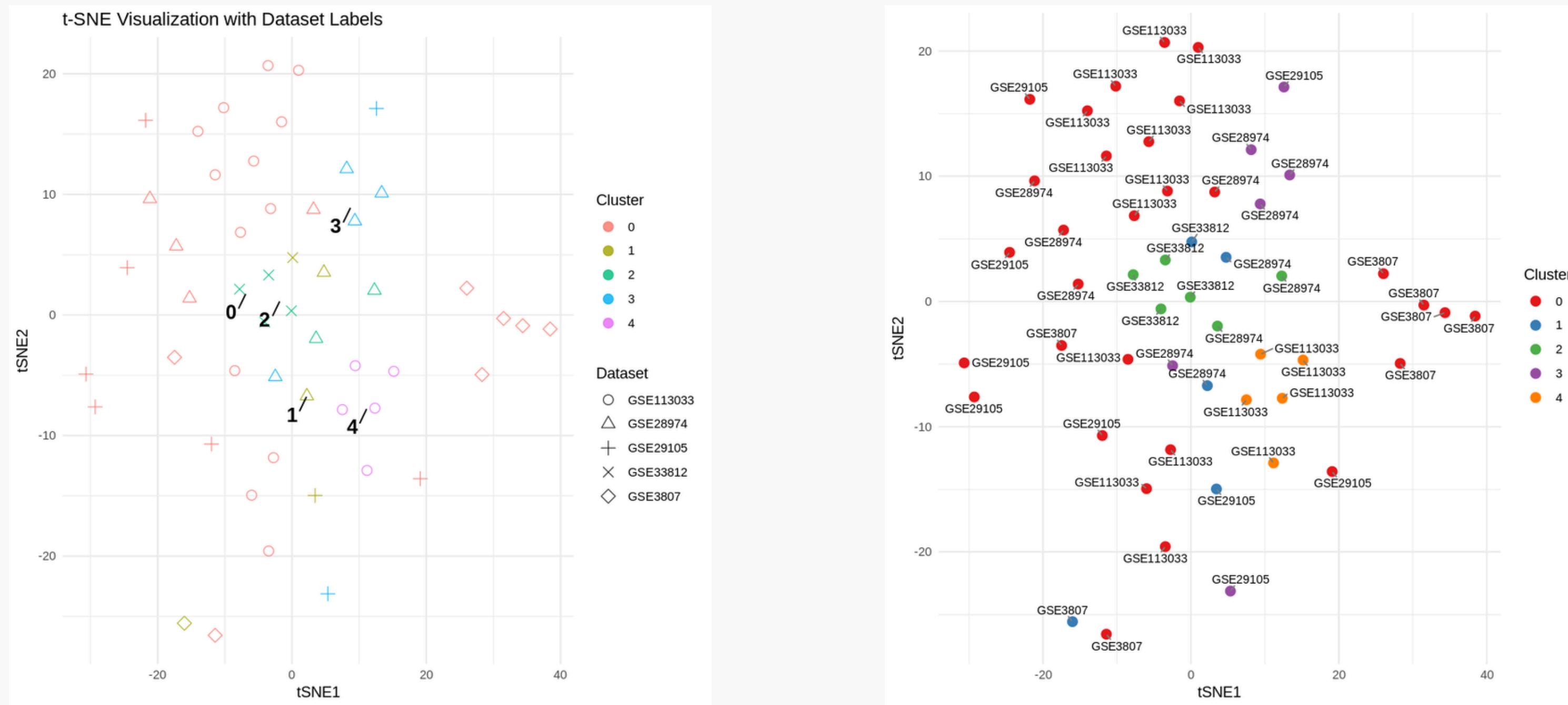
تحلیل همزمان مجموعه داده ها

- کاهش ابعاد و خوش بندی با PCA



تحلیل همزمان مجموعه داده ها

• کاهش ابعاد و خوش بندی با t-SNE



ممنون از حسن توجه شما