



پروژه درس بیوانفورماتیک

تحلیل دادگان ریزآرایه ی لوکمی حاد مغز استخوان

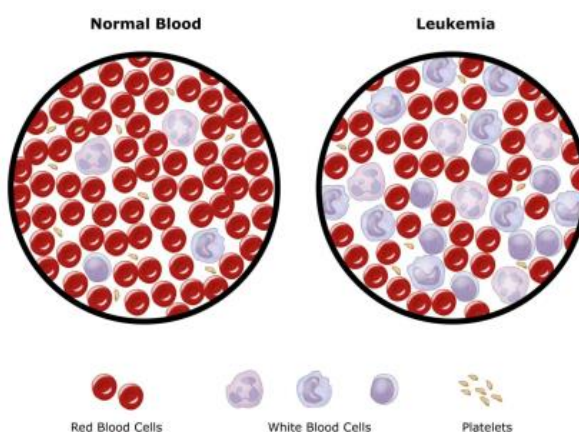
علیرضا کاظمی

۹۵۱۰۵۰۳۵

پائیز ۱۳۹۹

## ۱- مقدمه

لوکمیا یا سرطان خون، یکی از انواع سرطان است که معمولاً از مغز استخوان آغاز میشود و تعداد زیادی سلول خونی غیرعادی و نابالغ تولید می کند. این سلول های خونی به طور کامل تکامل نیافته اند و به آن ها سلول های لوکمی گفته می شود. لوکمی حاد مغز استخوان یا به اختصار **AML**، یکی از انواع لوکمیا است. این نوع از سرطان در اکثر مواقع در اثر بروز جهشی در دی ان ای سلول های پیشساز خون و یا جهش دیگری که موجب تقسیم و تکثیر غیر قابل کنترل سلول های خونی می شود، رخ میدهد. امروزه از آنالیز داده های میکروآرای برای تشخیص این بیماری و بیماری های از این دست که بر اثر تغییر و جهش در بیان ژن ها رخ میدهند، استفاده می شود.

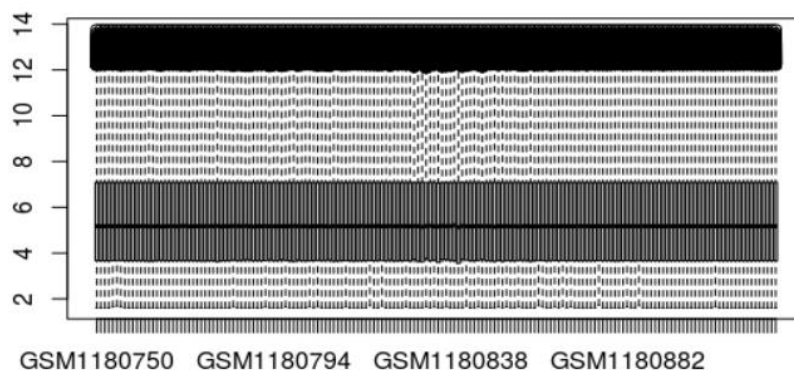


در این پروژه در قسمت ابتدایی با استفاده از داده های مربوط **GSE48558** و تحلیل آن، ژن هایی که در بروز این سرطان نقش موثر داشتند را پیدا کردیم. ابتدا باید داده ها را گروه بندی کرد که ۱۲۱ داده غیرسالم و ۴۹ داده سالم داریم.

1	GSM1180894	Primary AML sample 201	AML Patient	Leukemia
1	GSM1180895	Primary AML sample 201-bone marrow	AML Patient	Leukemia
1	GSM1180897	Primary AML sample 9	AML Patient	Leukemia
1	GSM1180898	Primary AML sample 9-spleen	AML Patient	Leukemia
1	GSM1180899	Primary AML sample 14-bone marrow	AML Patient	Leukemia
2	GSM1180790	Normal Granulocyte-1	Granulocytes	Normal
2	GSM1180794	Normal Granulocyte-2	Granulocytes	Normal
2	GSM1180818	Normal B-cell-1	B Cells	Normal
2	GSM1180820	Normal T cell-1	T Cells	Normal
2	GSM1180824	Normal Granulocyte-3	Granulocytes	Normal
2	GSM1180826	Normal Granulocyte-4	Granulocytes	Normal
2	GSM1180827	Normal Monocyte-1	Monocytes	Normal
2	GSM1180828	Normal Monocyte-2	Monocytes	Normal
2	GSM1180829	Normal B-cell-2	B Cells	Normal
2	GSM1180831	Normal T cell-2	T Cells	Normal
2	GSM1180834	Normal T cell-3	T Cells	Normal
2	GSM1180835	Normal T cell-4	T Cells	Normal

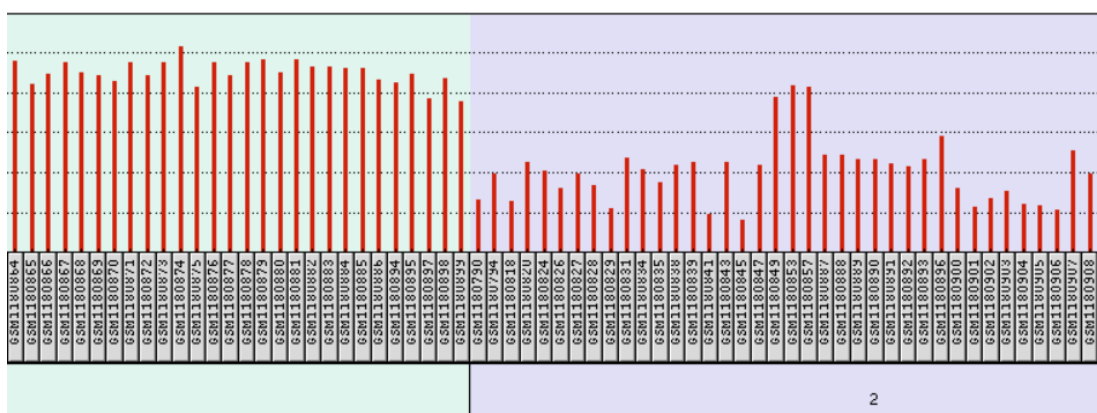
## ۲- کنترل کیفی

قبل از استنتاج داده باید نشان دهیم که داده های ما چه اندازه کیفیت دارند. باید ابتدا در نمودارهایی کیفیت آن را بررسی کنیم. با توجه به اینکه بزرگ ترین عدد که با دستور  $\max(ex)$  پیدا کردیم ۱۳/۷۶ بود، پس عددهای بیان ژن در مقیاس لگاریتمی آورده شده اند.



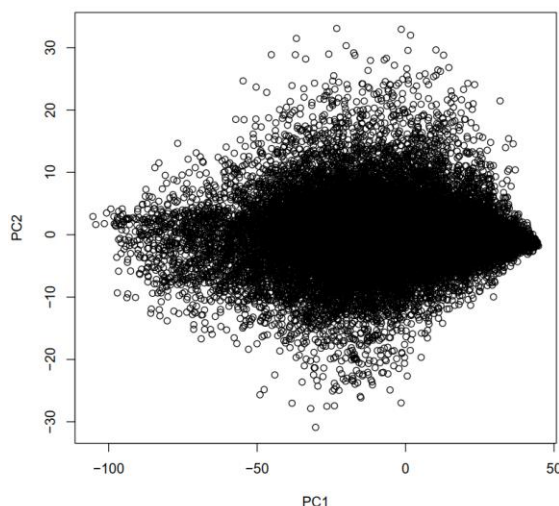
در تصویر بالا باکس پلات داده ها را رسم کردیم، همانطور که مشاهده می شود داده ها عملاً نرمالایز شده اند و نیازی به کار اضافه ای نیست.

## ۳- کاهش ابعاد داده ها

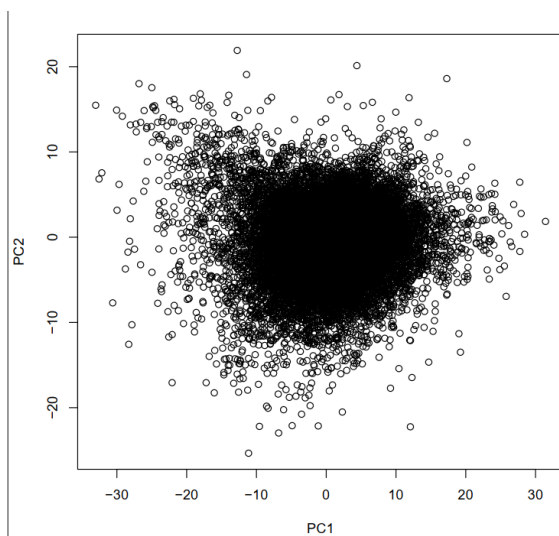


در اینجا بیان یک ژن خاص در یک سری از نمونه ها آمده است. حال در این قسمت می خواهیم ژن هایی که در تمامی نمونه ها میزان بیان تقریباً یکسانی دارند را پیدا کنیم و ضریب تاثیر آنها را کم کنیم چون عملاً نمی توانند اطلاعات جدیدی به ما در جهت تمایز ژن ها بدهند. پس ژن هایی برای ما مناسب هستند که پراکندگی بیشتری دارند. حال با استفاده از روش  $pca$  به

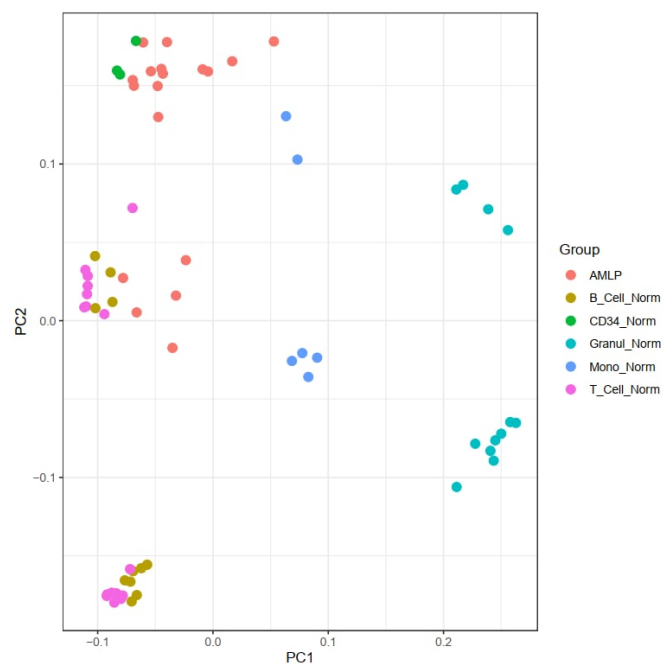
دنبال فیچرهایی می‌گردیم که ضریب تاثیر ژن‌هایی با پراکندگی بیشتر، در آن بیشتر است. معمولاً با دو یا سه فیچر اول جدید در این روش می‌توان قسمت خوبی از واریانس را کسب کرد.



همانطور که مشاهده می‌شود توزیع داده‌های ما در فضای برداری جدید به صورت بالاست که فقط دو فیچر  $pc1$  و  $pc2$  دارد. چیزی که در شکل بالا ملموس است، اینه که هنوز نتوانستیم ژن‌هایی که پراکندگی کمی دارند رو جدا کنیم. به نوعی با توجه به میانگین بیان ژن این‌ها جدا شده‌اند، پس طبق روش دکتر که در ویدئوها گفتند می‌آییم از میانگین بیان ژنشان کم می‌کنیم و سپس دوباره کاهش ابعاد می‌دهیم.

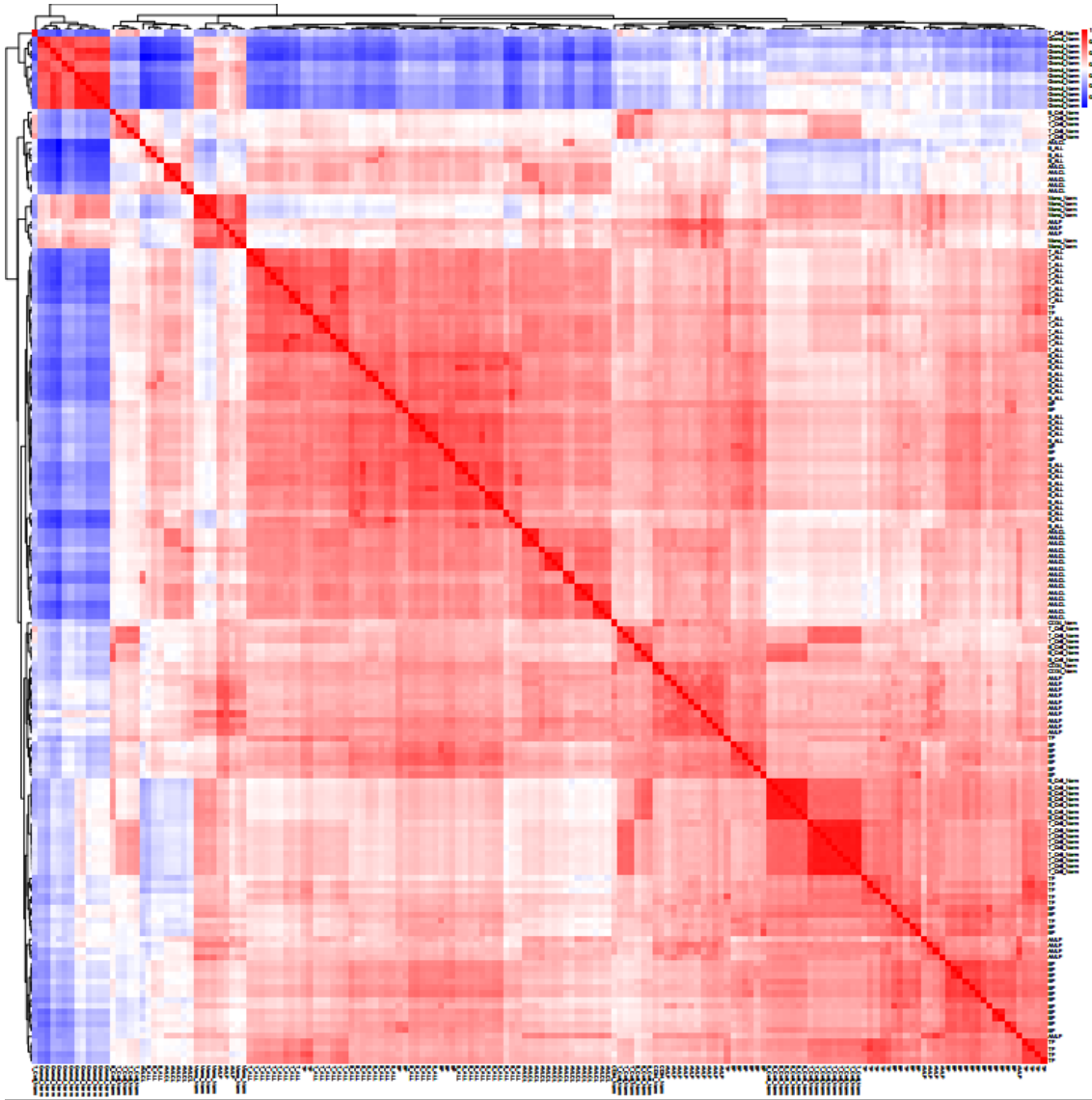


با توجه به شکل بالا می‌بینیم که توزیع فیچرها به صورت یک توزیع مستقل درآمده‌اند، یعنی با این فیچرهای جدید  $pc1$  و  $pc2$  احتمالاً بتوان به جدایی پذیری بهتری دست پیدا کرد، حال می‌آییم اثر اینگونه کاهش بعد را بر روی دیتاها مشاهده می‌کنیم.

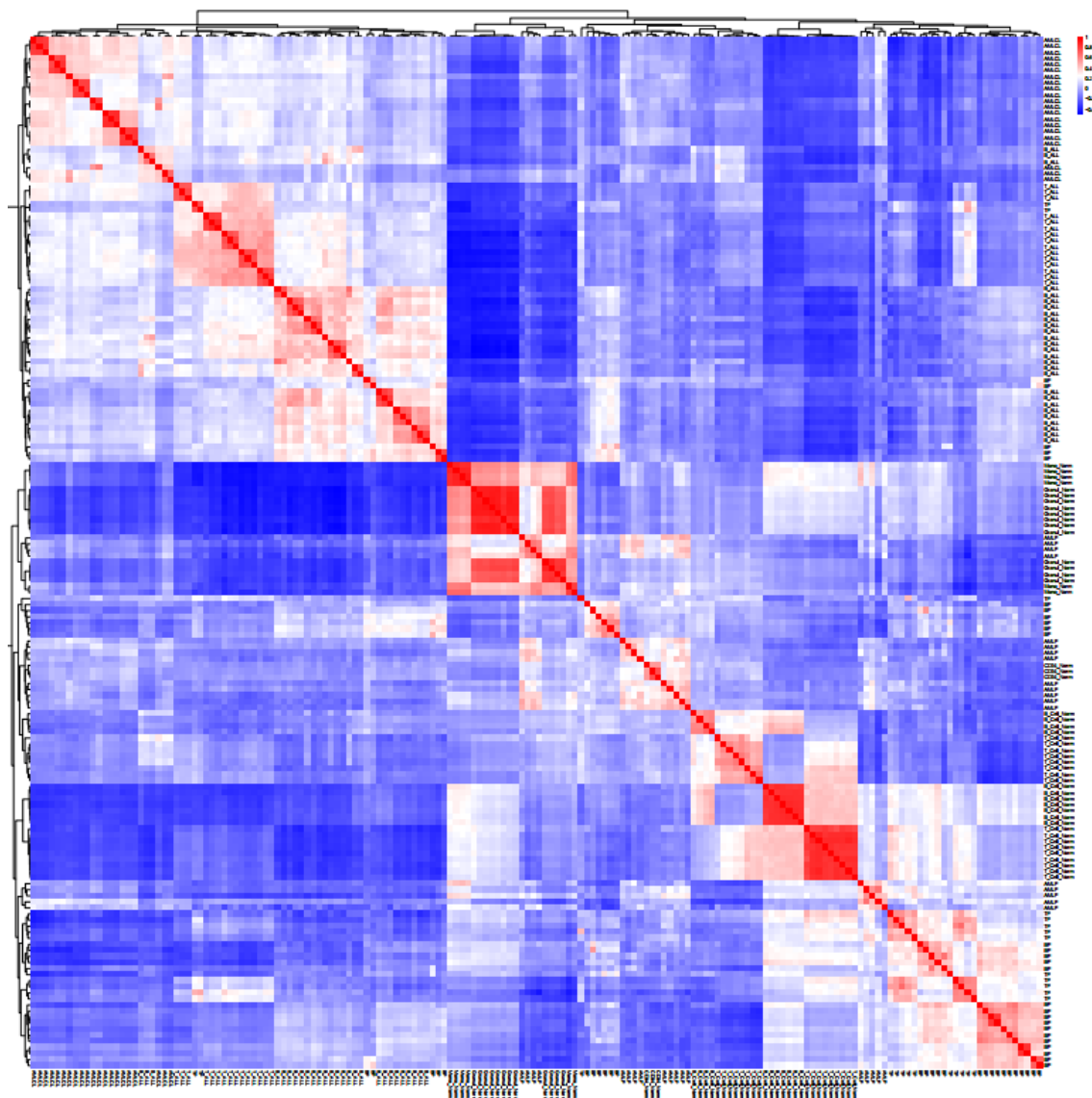


گروه های سالم به طور مناسبی از ناسالم ها جدا شده اند و فقط مقداری از آن با AMLP غیر قابل تشخیص است. هم چنین در ناسالم ها، AMLCL و TP هم به طور خوبی جدا شده اند. B\_ALL نیز در دو خوشه جدا آمده است. هم چنین T\_ALL از B\_ALL نیز قابل تشخیص است. پس به طور کلی، دیتا به خوبی توانسته است از یکدیگر جدا شود.

#### ۴- بررسی همبستگی

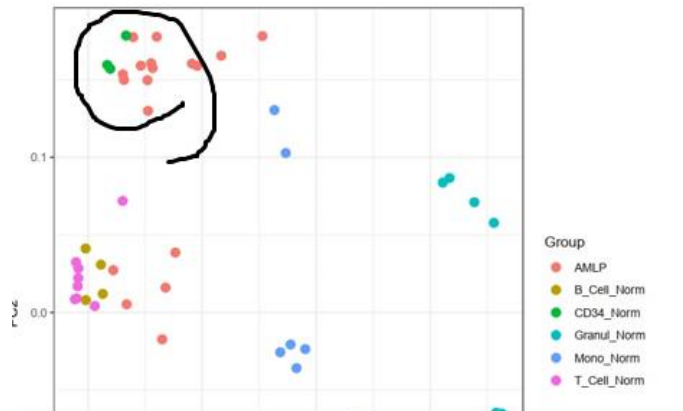


فایل همبستگی بین داده ها را بر اساس سورس نیم ها ابتدا بر روی کل داده ها قبل از کاستن میانگین از آنها و سپس بر روی دیتای اسکیل شده انجام دادم و خروجی آن را ذخیره کردم. در ابتدا می توان فهمید که خود داده های نرمال به چند دسته تقسیم می شوند. بعضی از این زیرگروه ها باز چند دسته ای هستند. دسته نرمال B Cell و T Cell همانطور که در نمودار پی سی ای قسمت قبل، قابل مشاهده بود، اینجا نیز داری همبستگی بالایی هستند. که این پدیده نیز باز در دو دسته رخ داده است. دسته Granul نیز با اکثر داده های غیر نرمال همبستگی منفی دارد.



پس از کسر میانگین متوجه می شویم که نمونه های ناسالم با خودشان همبستگی زیادی ندارند ولی اما با داده های سالم همبستگی منفی زیادی دارند پس یعنی میتوان آن ها را از یکدیگر تشخیص داد.

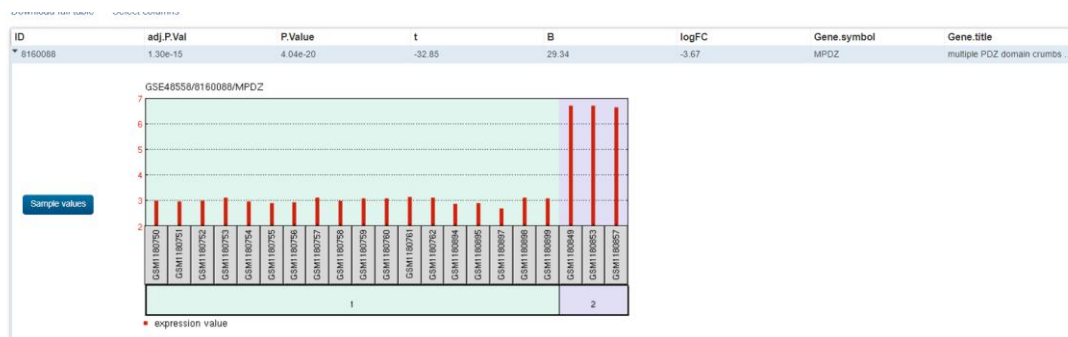
## 5- بررسی تمایز بین ژن ها



همانطور که در بخش سوم بررسی کردیم نمونه های سالم CD34 به خوشه AMLP بسیار نزدیک است و بقیه نمونه های نرمال دورتر هستند. پس این یعنی همبستگی بیشتری دارند. حال در سایت نمونه های AMLP را در گروه یک و نمونه های CD34 را در گروه دو قرار می دهیم.

ID	adj.P.Val	P.Value	t	B	logFC	Gene.symbol	Gene.title
* 8160088	1.30e-15	4.04e-20	-32.85	29.34	-3.67	MPDZ	multiple PDZ domain crumbs ...
* 8101084	1.30e-11	8.54e-16	-20.54	23.57	-3.875	PRKG2	protein kinase, cGMP-depend...
* 8173805	8.80e-11	8.17e-15	-18.42	21.96	-2.286	RPS6KA6	ribosomal protein S6 kinase A6
* 7910915	4.01e-10	4.95e-14	-16.87	20.6	-4.379	CHRM3	cholinergic receptor muscarini...
* 8160168	1.22e-09	1.88e-13	-15.8	19.55	-3.41	FREM1	FRAS1 related extracellular ...
* 7902127	2.15e-09	3.99e-13	-15.22	18.95	-2.267	SGIP1	SH3 domain GRB2 like endo...
* 8121814	1.34e-08	3.26e-12	-13.69	17.22	-4.673	NKAIN2	Na+/K+ transporting ATPase l...
* 8173917	1.34e-08	3.31e-12	-13.68	17.21	-4.121	NAP1L3	nucleosome assembly protein...
* 8149438	1.52e-08	4.23e-12	-13.51	17	-1.877	SGCZ	sarcoglycan zeta
* 8054439	1.91e-08	5.92e-12	-13.28	16.72	-2.071	ST6GAL2	ST6 beta-galactoside alpha-2...
* 7927606	3.45e-08	1.18e-11	-12.83	16.13	-1.975	PRKG1	protein kinase, cGMP-depend...
* 7906954	4.73e-08	1.76e-11	-12.56	15.79	-4.196	PBX1	PBX homeobox 1
* 8172708	1.03e-07	4.15e-11	-12.01	15.04	-3.173	NUDT11	nudix hydrolase 11
* 8077356	2.28e-07	9.89e-11	-11.48	14.28	-2.611	LRRN1	leucine rich repeat neuronal 1
* 8174610	2.34e-07	1.08e-10	-11.42	14.2	-2.054	LRCH2	leucine rich repeats and calpo...
* 8060997	3.39e-07	1.68e-10	-11.16	13.81	-3.436	PLCB4	phospholipase C beta 4
* 7943376	3.45e-07	1.81e-10	-11.11	13.74	-2.78	CEP126	centrosomal protein 126
* 8008088	4.31e-07	2.49e-10	-10.93	13.46	-4.14	HLF	HLF, PAR bZIP transcription f...
* 8021668	4.31e-07	2.53e-10	-10.92	13.44	-1.718	CDH7	cadherin 7
* 7932118	4.41e-07	2.73e-10	-10.87	13.38	-2.016	BEND7	BEN domain containing 7
* 8097288	8.43e-07	5.48e-10	-10.47	12.75	-1.728	FAT4	FAT atypical cadherin 4
* 8102938	1.17e-06	7.98e-10	-10.26	12.41	-2.08	RNF150	ring finger protein 150

تصویر ژن هایی که کمترین پی ویو را دارند در جدول بالا به تصویر کشیده ایم. هم چنین جدول کلی را نیز دانلود کرده ام و در پیوست قرار داده ام. مثلاً در پایین ژن MPDZ که بیشترین تفاوت آماری را در دو گروه دارد، بیان آن در نمونه ها به صورت زیر است:





حال همان کدی که سایت تولید کرده را کپی کردم و همانند ویدئو آنهایی که بیان ژن زیاد (AMLN\_Normal\_up4.txt) و یا کم (AMLN\_Normal\_down4.txt) دارند را از یکدیگر جدا و در فایل تکست ذخیر کردم. به طور مثال چند خروجی در در جدول زیر آورده ام:

GENE	Adj P-Value	Log FC	Expression in Patient
MPDZ	1.3 e-15	-3.67	Low
PRKG2	1.3 e-11	-5.875	Low
FYB	4.4e-05	4.382	High
DPP9	5.5e-05	1.678	High

## ۶- آنالیز Pathway

Pathway هایی که در این سایت پیدا میکنیم، به این معنی است که با ژن هایی که ما پیدا کردیم اشتراک زیادی دارند. باید فایل ژن هایی که پیدا کرده ایم را در این سایت آپلود کنیم. حال ابتدا ژن های کاهشی را در این سایت پیدا می کنیم. مثلاً در اولین مورد که در زیر آمده است:

**WikiPathways 2019 Human** Bar Graph **Table** Clustergram Appyter ⚙️ ⓘ

Hover each row to see the overlapping genes.

10 entries per page Search:

Index	Name	P-value	Adjusted p-value	Odds Ratio	Combined score
1	Type II interferon signaling (IFNG) WP619	6.911e-8	0.00002253	9.82	161.96
2	T-Cell Receptor and Co-stimulatory Signaling WP2583	4.526e-7	0.00004919	10.75	157.06
3	The human immune response to tuberculosis WP4197	0.000006123	0.0003327	10.87	130.52
4	T-Cell antigen Receptor (TCR) pathway during Staphylococcus aureus infection WP3863	8.432e-7	0.00005497	5.97	83.56
5	T-Cell antigen Receptor (TCR) Signaling Pathway WP69	1.696e-7	0.00002764	5.14	80.08
6	Allograft Rejection WP2328	7.390e-7	0.00005497	4.85	68.42
7	ncRNAs involved in STAT3 signaling in hepatocellular carcinoma WP4337	0.002484	0.06230	9.03	54.13
8	Selective expression of chemokine receptors during T-cell polarization WP4494	0.0003154	0.01285	6.48	52.22
9	Inflammatory Response Pathway WP453	0.0003948	0.01287	6.20	48.56
10	IL-7 Signaling Pathway WP205	0.0008800	0.02608	6.42	45.19

Showing 1 to 10 of 326 entries | [Export entries to table](#) Previous Next

Terms marked with an \* have an overlap of less than 5



در مورد دوم T-Cell مربوط به سیستم ایمنی است طبق سرچی که کردم.

### T cell receptor signalling and the immune synapse

The spatial organization of co-signalling receptors on naive T cells is thought to be somewhat random; therefore specific events involving the reorganization of T cell surface molecules are required for optimal functional interactions to occur. The formation of the immune synapse is the primary reorganizing event that enables productive T cell receptor (TCR) signalling and co-signalling<sup>2</sup>. The immune synapse is composed of the central, peripheral and distal supra-molecular activation

مورد سوم هم مشخصاً مربوط به سیستم ایمنی است. مورد چهار و پنج هم مانند مورد دوم است. با توجه به نوع بیماری ای که در پروژه کار کردیم، به نظر منطقی می آید که بیان ژن هایی که مرتبط با پاسخ ایمنی است کاهش می یابد.

## KEGG 2019 Human

Bar Graph **Table** Clustergram Appyter  

Hover each row to see the overlapping genes.

10 entries per page

Search:

Index	Name	P-value	Adjusted p-value	Odds Ratio	Combined score
1	Primary immunodeficiency	5.323e-10	1.299e-7	12.49	266.61
2	T cell receptor signaling pathway	1.360e-9	1.659e-7	5.74	117.29
3	Th17 cell differentiation	4.293e-9	3.492e-7	5.34	102.84
4	NF-kappa B signaling pathway	7.738e-8	0.000003147	5.14	84.19
5	Th1 and Th2 cell differentiation	2.404e-7	0.000008380	5.00	76.17
6	Epstein-Barr virus infection	2.117e-8	0.000001292	3.63	64.20
7	Antigen processing and presentation	0.000002540	0.00006199	4.96	63.87
8	Allograft rejection	0.00005204	0.0007055	6.33	62.43
9	Human T-cell leukemia virus 1 infection	4.357e-8	0.000002126	3.42	57.89
10	Measles	5.017e-7	0.00001530	3.90	56.64

Showing 1 to 10 of 244 entries | [Export entries to table](#)

[Previous](#) [Next](#)



Terms marked with an \* have an overlap of less than 5

در اینجا هم نیز T-Cell مانند قبل در بعضی موارد دیده می شود.

پس در پایان جمع بندی ژن های کاهشی، به نظر می آید ژن هایی که مرتبط با پاسخ سیستم ایمنی است بیان آنها کاهش می یابد.

حال ژن های افزایشی را می خواهیم بررسی کنیم.

## WikiPathways 2019 Human

Bar Graph **Table** Clustergram Appyter  

Hover each row to see the overlapping genes.

10 entries per page

Search:

Index	Name	P-value	Adjusted p-value	Odds Ratio	Combined score
1	Retinoblastoma Gene in Cancer WP2446	2.091e-23	8.342e-21	11.49	600.14
2	DNA Replication WP466	9.610e-13	9.586e-11	12.12	335.25
3	G1 to S cell cycle control WP45	4.184e-14	8.348e-12	9.15	281.87
4	Cell Cycle WP179	2.218e-13	2.950e-11	5.53	161.09
5	Heme Biosynthesis WP561	0.0001731	0.003837	16.50	142.90
6	Regulation of sister chromatid separation at the metaphase-anaphase transition WP4240	0.00003361	0.0008957	11.56	119.10
7	Gastric Cancer Network 1 WP2361	0.000002219	0.0001265	8.09	105.37
8	DNA IR-damage and cellular response via ATR WP4016	3.868e-9	3.087e-7	5.38	104.20
9	IL1 and megakaryocytes in obesity WP2865	0.00002113	0.0007663	7.94	85.44
10	Fluoropyrimidine Activity WP1601	0.000009539	0.0004229	6.62	76.54

Index	Name	P-value	Adjusted p-value	Odds Ratio	Combined score
1	Systemic lupus erythematosus	2.385e-21	7.083e-19	7.40	351.48
2	Cell cycle	2.071e-14	3.075e-12	5.72	180.16
3	Alcoholism	1.993e-13	1.973e-11	4.36	127.52
4	DNA replication	5.135e-7	0.00002542	7.49	108.54
5	Propanoate metabolism	0.000006791	0.0002881	6.94	82.54
6	Transcriptional misregulation in cancer	6.397e-10	4.750e-8	3.56	75.38
7	One carbon pool by folate	0.0002958	0.005167	7.11	57.80
8	Base excision repair	0.00006170	0.001410	5.75	55.76
9	Valine, leucine and isoleucine degradation	0.00001929	0.0006367	4.92	53.43
10	Malaria	0.00002457	0.0007299	4.78	50.79

به طور کلی، در اینجا موارد بالا مرتبط با cell cycle یا تقسیم سلولی هستند که به نوعی با افزایش آن در بیماری مورد بررسی در ارتباط است.

## ۷- آنالیز Gene Antology

ابتدا برای ژن های کاهشی آنالیز را انجام میدهیم.

**GO Cellular Component 2018** Bar Graph **Table** Clustergram Appyter ⚙️ ⓘ

Hover each row to see the overlapping genes.

10 entries per page Search:

Index	Name	P-value	Adjusted p-value	Odds Ratio	Combined score
1	T cell receptor complex (GO:0042101)	9.319e-9	0.000002283	26.25	485.45
2	MHC protein complex (GO:0042611)	0.0001224	0.01174	10.17	91.65
3	integral component of luminal side of endoplasmic reticulum membrane (GO:0071556)	0.0003154	0.01932	6.48	52.22
4	AP-1 adaptor complex (GO:0030121)	0.02929	0.2990	10.14	35.78
5	early endosome membrane (GO:0031901)	0.0001438	0.01174	4.02	35.55
6	phagocytic vesicle membrane (GO:0030670)	0.001763	0.05400	4.60	29.14
7	phagocytic vesicle (GO:0045335)	0.0004923	0.02285	3.45	26.28
8	actin cortical patch (GO:0030479)	0.03974	0.3441	8.11	26.15
9	AIM2 inflammasome complex (GO:0097169)	0.03974	0.3441	8.11	26.15
10	endocytic patch (GO:0061645)	0.03974	0.3441	8.11	26.15

**GO Molecular Function 2018** Bar Graph **Table** Clustergram Appyter ⚙️ ⓘ

Hover each row to see the overlapping genes.

10 entries per page Search:

Index	Name	P-value	Adjusted p-value	Odds Ratio	Combined score
1	C-C chemokine binding (GO:0019957)	0.0002939	0.01608	20.31	165.19
2	chemokine receptor activity (GO:0004950)	0.000002765	0.0005041	12.55	160.60
3	G-protein coupled chemoattractant receptor activity (GO:0001637)	0.00003727	0.003398	13.57	138.35
4	T cell receptor binding (GO:0042608)	0.001871	0.06398	20.29	127.46
5	superoxide-generating NADPH oxidase activator activity (GO:0016176)	0.007066	0.1390	10.14	50.24
6	phosphatidylinositol-3,4-bisphosphate binding (GO:0043325)	0.0006963	0.03461	6.78	49.29
7	transmembrane receptor protein serine/threonine kinase activity (GO:0004675)	0.003350	0.08420	8.12	46.29
8	MHC class I protein binding (GO:0042288)	0.003350	0.08420	8.12	46.29
9	cytokine receptor activity (GO:0004896)	0.00003224	0.003398	4.43	45.83
10	SH3/SH2 adaptor activity (GO:0005070)	0.0001125	0.008793	4.98	45.25

**GO Biological Process 2018** Bar Graph **Table** Clustergram Appyter ⚙️ ⓘ

Hover each row to see the overlapping genes.

10 entries per page Search:

Index	Name	P-value	Adjusted p-value	Odds Ratio	Combined score
1	cellular response to interleukin-21 (GO:0098757)	0.00001140	0.001894	33.89	385.76
2	interleukin-21-mediated signaling pathway (GO:0038114)	0.00001140	0.001894	33.89	385.76
3	cellular response to type I interferon (GO:0071357)	7.972e-15	1.125e-11	11.33	367.85
4	type I interferon signaling pathway (GO:0060337)	7.972e-15	1.125e-11	11.33	367.85
5	antigen processing and presentation of exogenous peptide antigen via MHC class I, TAP-independent (GO:0002480)	0.00002466	0.003315	25.42	269.70
6	regulation of lymphocyte activation (GO:0051249)	0.00001386	0.002060	17.45	195.16
7	T-helper 17 cell differentiation (GO:0072539)	0.0002939	0.02128	20.31	165.19
8	lymphocyte mediated immunity (GO:0002449)	0.00008353	0.007738	16.94	159.11
9	T cell differentiation (GO:0030217)	9.670e-8	0.00003900	9.45	152.55
10	regulation of T cell receptor signaling pathway (GO:0050856)	0.000001661	0.0003907	10.80	143.79

در همه موارد که پی ویوی زیر ۰,۰۵ دارند بعضی موارد مشخصاً مربوط به سیستم ایمنی و هم چنین t cell که مربوط به سیستم ایمنی میشود، موجود است. هم چنین مورد اول آن مربوط به interleukin است که طبق سرچی که کردم مربوط به تنظیم رشد سلولی است:

**Interleukin (IL)**, any of a group of naturally occurring proteins that mediate communication between cells. **Interleukins** regulate cell growth, differentiation, and motility. They are particularly important in stimulating immune responses, such as inflammation.

موارد دیگه نیز شامل interferon است که طبق سرچ زیر مربوط به سیستم ایمنی است:

**Interferons (IFNs, /ˌɪntərˈfɛərən/)** are a group of signaling proteins made and released by host cells in response to the presence of several viruses. In a typical scenario, a virus-infected cell will release **interferons** causing nearby cells to heighten their anti-viral defenses.

حال این آنالیز را نیز دوباره برای ژن های افزایشی انجام می دهیم:

#### GO Molecular Function 2018

Bar Graph

Table

Clustergram

Appyter



Hover each row to see the overlapping genes.

10 entries per page

Search:

Index	Name	P-value	Adjusted p-value	Odds Ratio	Combined score
1	sphingolipid transporter activity (GO:0046624)	0.0003315	0.01167	26.38	211.36
2	histone kinase activity (GO:0035173)	0.0001731	0.008248	16.50	142.90
3	DNA polymerase binding (GO:0070182)	0.00005610	0.004131	10.28	100.59
4	four-way junction DNA binding (GO:0000400)	0.00005610	0.004131	10.28	100.59
5	lipoprotein particle receptor activity (GO:0030228)	0.0003531	0.01192	8.80	69.97
6	Y-form DNA binding (GO:0000403)	0.005984	0.08976	13.18	67.47
7	arylesterase activity (GO:0004064)	0.005984	0.08976	13.18	67.47
8	glucosyltransferase activity (GO:0046527)	0.005984	0.08976	13.18	67.47
9	DNA insertion or deletion binding (GO:0032135)	0.005984	0.08976	13.18	67.47
10	DNA ligase (ATP) activity (GO:0003910)	0.005984	0.08976	13.18	67.47

#### GO Cellular Component 2018

Bar Graph

Table

Clustergram

Appyter



Hover each row to see the overlapping genes.

10 entries per page

Search:

Index	Name	P-value	Adjusted p-value	Odds Ratio	Combined score
1	MutLalpha complex (GO:0032389)	0.000008584	0.00008888	26.42	308.16
2	chromosome, centromeric region (GO:0000775)	3.275e-11	3.504e-9	9.19	221.81
3	Golgi cisterna membrane (GO:0032580)	1.549e-7	0.000006217	9.94	155.89
4	Golgi cis cisterna (GO:0000137)	1.549e-7	0.000006217	9.94	155.89
5	condensed chromosome, centromeric region (GO:0000779)	6.038e-7	0.00001385	9.71	139.11
6	platelet alpha granule membrane (GO:0031092)	0.000008305	0.00008887	11.75	137.50
7	spindle (GO:0005819)	2.846e-12	4.567e-10	4.05	107.62
8	spindle microtubule (GO:0005876)	1.271e-7	0.000006217	6.44	102.25
9	nuclear chromosome part (GO:0044454)	1.617e-14	5.191e-12	3.12	99.07
10	spindle midzone (GO:0051233)	0.000003280	0.00004049	7.67	96.82

## GO Biological Process 2018

Bar Graph

**Table**

Clustergram

Appyter



Hover each row to see the overlapping genes.

10 entries per page

Search:

Index	Name	P-value	Adjusted p-value	Odds Ratio	Combined score
1	DNA strand elongation involved in DNA replication (GO:0006271)	1.622e-8	0.000003021	22.07	395.92
2	regulation of transcription involved in G1/S transition of mitotic cell cycle (GO:0000083)	3.124e-10	9.694e-8	15.49	339.03
3	positive regulation of establishment of protein localization to telomere (GO:1904851)	0.000008584	0.0006524	26.42	308.16
4	positive regulation of protein localization to chromosome, telomeric region (GO:1904816)	0.000002219	0.0002233	23.13	301.11
5	G1/S transition of mitotic cell cycle (GO:0000082)	4.725e-17	8.797e-14	7.32	275.28
6	centromere complex assembly (GO:0034508)	6.686e-10	1.915e-7	10.63	224.62
7	regulation of establishment of protein localization to telomere (GO:0070203)	0.00002017	0.001252	19.81	214.19
8	regulation of attachment of spindle microtubules to kinetochore (GO:0051988)	0.00002017	0.001252	19.81	214.19
9	lagging strand elongation (GO:0006273)	0.00008170	0.003710	22.00	207.06
10	regulation of protein localization to Cajal body (GO:1904869)	0.00008170	0.003710	22.00	207.06

در اینجا نیز با ژن هایی که مرتبط با cell cycle و یا تنظیم بیان و تنظیم نرخ رونویسی است، برخوردیم. و یا elongation که به معنای زیر است:

1 : the state of being elongated or lengthened also : the process of growing or increasing in length chain **elongation** in DNA synthesis the **elongation** of a muscle under tension. 2 : something that is elongated.

## ۸- موارد اضافی

### COVID-19 Related Gene Sets

Bar Graph

**Table**

Clustergram

Appyter



Hover each row to see the overlapping genes.

10 entries per page

Search:

Index	Name	P-value	Adjusted p-value	Odds Ratio	Combined score
1	SARS perturbation Down Genes PBMCs GDS1028:GPL201	1.112e-27	2.530e-26	7.65	474.97
2	Up-regulated by SARS-CoV-2 infection of Calu3 cells	7.052e-36	6.417e-34	5.54	448.39
3	Up-regulated by SARS-CoV-2 2 MOI in Calu-3 from GSE147507	7.052e-36	6.417e-34	5.54	448.39
4	Up-regulated by SARS-CoV-2 in Calu-3 24hr from GSE148729	1.791e-34	1.086e-32	5.40	419.79
5	Up-regulated by SARS-CoV-1 in Calu-3 from GSE148729	4.345e-33	1.977e-31	5.27	392.41
6	SARS-CoV perturbation Up Genes bronchial epithelial 2B4 from GSE17400:GPL570:6	2.183e-29	5.677e-28	5.49	362.40
7	Healthy lung biopsy vs. COVID-19 infected lung series 15 from GSE147507 up genes	8.404e-31	2.549e-29	5.02	347.63
8	Up-regulated by SARS-CoV-2 in lung tissue from GSE147507	8.404e-31	2.549e-29	5.02	347.63
9	SARS-CoV perturbation Up Genes bronchial epithelial 2B4 from GSE17400:GPL570:5	1.315e-19	2.393e-18	4.88	212.02
10	SARS-CoV perturbation Up Genes bronchial epithelial 2B4 from GSE17400:GPL570:3	1.481e-18	2.450e-17	5.12	210.22

ژن های کاهشی ای که در پروژه به دست آوردیم در رابطه با ژن هایی هستند که ویروس کووید بر روی آنان اثر گذار است که با توجه نتیجه قبلی که آن را بر روی سیستم ایمنی موثر میدانستیم در اینجا هم احتمالاً همانند این سرطان، کووید ۱۹ روی کاهش بیان ژن های مرتبط به پاسخ سیستم ایمنی تاثیر می گذارد.

ClinVar 2019

Bar Graph **Table** Clustergram Appyter

Hover each row to see the overlapping genes.

10 entries per page

Search:

Index	Name	P-value	Adjusted p-value	Odds Ratio	Combined score
1	common variable immunodeficiency 2	0.0009697	0.05139	30.44	211.22
2	ischemic stroke	0.02015	0.2670	13.51	52.77
3	human immunodeficiency virus type 1, susceptibility to	0.003350	0.08876	8.12	46.29
4	systemic lupus erythematosus	0.01293	0.2285	7.61	33.08
5	autism spectrum disorder	0.05136	0.3403	6.76	20.06
6	familial hypercholesterolemias	0.05136	0.3403	6.76	20.06
7	susceptibility to malaria	0.03686	0.3403	4.68	15.45
8	familial colorectal cancer	0.04323	0.3403	4.35	13.65
9	mycobacterium tuberculosis, susceptibility to	0.07759	0.4569	5.07	12.95
10	prader-willi syndrome	0.1070	0.5672	4.05	9.06

Showing 1 to 10 of 53 entries | [Export entries to table](#)

Terms marked with an \* have an overlap of less than 5

Previous Next

در اینجا نیز می بینیم ژن های کاهشی این سرطان با ژن های تحت تاثیر این بیماری های عصبی مرتبط است.

## جمع بندی آنالیزهای نهایی:

**Acute myeloid leukaemia (AML) is caused by a DNA mutation in the stem cells in your bone marrow that produce red blood cells, platelets and infection-fighting white blood cells.<sup>1</sup>**

طبق گزارش بالا علت اصلی افزایش بی رویه سلول های گلبول قرمز و درگیری گلبول های سفید است. همانطور که در آنالیزهای بالا دیدیم، ژن های کاهشی ای که به دست آوردیم اکثراً با پی ویوی کمتر از ۰,۰۵ با ساز و کارهای ژنی ای که مرتبط با سیستم ایمنی بودند ارتباط داشتند و هم چنین ژن های افزایشی ای که به دست آوردیم اکثراً با پی ویوی کمتر از ۰,۰۵ با ساز و کارهای ژنی ای که مرتبط به تقسیم سلولی و چرخه سلولی بودند ارتباط داشتند.

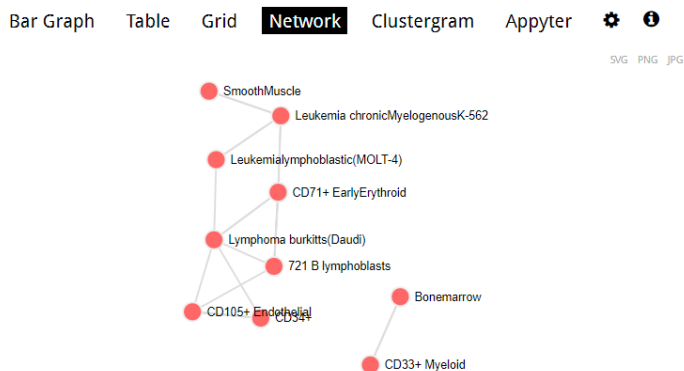
یکی از مواردی که با این بیماری اشتراکات زیادی در بیان ژن بود، T Cell بود. حال کمی به بررسی آن می پردازیم. **لنفوسیت تی** از انواع سلول های دستگاه ایمنی بدن است. دفاع اختصاصی بدن شامل ایمنی هومورال و ایمنی سلولی است. در ایمنی سلولی لنفوسیت های (T سلول های T) فعالیت دارند. لنفوسیت های T پس از اتصال به پادگن (آنتی ژن) تکثیر پیدا می کنند و انواعی از سلول های T از جمله تعدادی T کشنده و تعدادی T خاطره به وجود می آورند. سلول های T کشنده به طور مستقیم به سلول های آلوده به ویروس و سلول های سرطانی حمله می کنند و با تولید پروتئین خاص به نام پرفورین منافذی در این سلول ها به وجود می آورند و موجب مرگ آن ها می شوند. به همین علت این نوع از پاسخ ایمنی به ایمنی سلولی معروف است. انواع سلول های T عبارت است از T کشنده، T خاطره، یاری کننده (h.c) و مهارکننده.<sup>۲</sup>

<sup>1</sup> [Acute myeloid leukaemia - Causes - NHS \(www.nhs.uk\)](https://www.nhs.uk/conditions/acute-myeloid-leukaemia/causes/)

<sup>2</sup> [لنفوسیت تی - ویکی پدیا، دانشنامه آزاد \(wikipedia.org\)](https://en.wikipedia.org/wiki/Lymphocyte)

در یکی از مقاله هایی که در رابطه با AML است، بر روی یکی از بیمارهایی که از ای ام ال نجات یافته است به این نتیجه رسیده است که تعداد سلول های لنفوسیتی زیاد شده است.<sup>۳</sup>

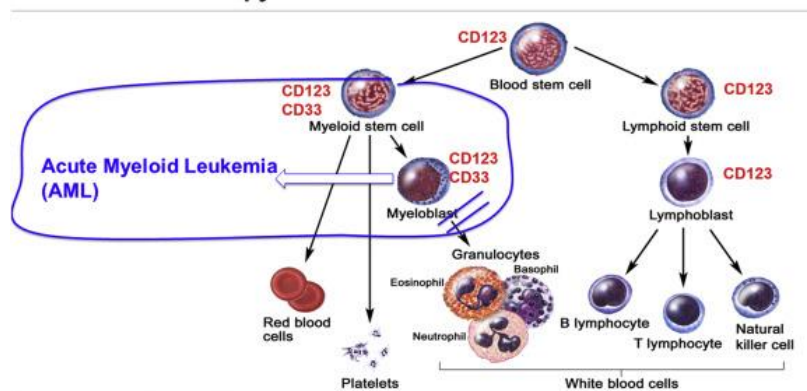
### Human Gene Atlas



هم چنین در این مقاله بالا نیز به افزایش بیش از حد CD34+ شده است که با توجه به سرچی که با استفاده از ژن های به دست آمده در پروژه در قسمت cell type در enricher زدم، در شبکه این پایگاه که human gene atlas بود که جزو سلول های خونی بود.

حتی یک نوع درمان بر اساس همین نوع سلول ها برای درمان این نوع سرطان شده است.<sup>۴</sup>

### CAR T-cell therapy of AML



<sup>3</sup> [Peripheral blood T cells in acute myeloid leukemia \(AML\) patients at diagnosis have abnormal phenotype and genotype and form defective immune synapses with AML blasts | Blood | American Society of Hematology \(ashpublications.org\)](https://ashpublications.org/blood/article/127/12/3542/3542)

<sup>4</sup> [How close are we to CAR T-cell therapy for AML? - ScienceDirect](https://www.sciencedirect.com/science/article/pii/S1521692619300011)

علل ایجاد این بیماری در گزارش ها در دسته های زیر آمده است:<sup>۵</sup>

پیری - مصرف دخانیات - اضافه وزن - تابش های پر انرژی (یونیزه کننده) - بنزن - شرایط ارثی - شیمیوتراپی - بیماری های  
خونی - بیماری های ایمنی

---

<sup>5</sup> [Risks and causes | Acute myeloid leukaemia | Cancer Research UK](#)