In the Name of God

Sharif University of Technology

Department of Computer Engineering

Introduction to Bioinformatics

Final Project

Parastoo Azizeddin

97106913

Instructor: Dr. Ali Sharifi Zarchi

Dr. Somayeh Koohi

Fall 1400

Content

1. Introduction

Leukemia is a group of blood cancers that usually begin in the bone marrow and result in high numbers of abnormal blood cells. These blood cells are not fully developed and are called blasts or leukemia cells. Symptoms may include bleeding and bruising, bone pain, fatigue, fever, and an increased risk of infections. These symptoms occur due to a lack of normal blood cells. Diagnosis is typically made by blood tests or bone marrow biopsy.

The exact cause of leukemia is unknown. A combination of genetic factors and environmental (non-inherited) factors are believed to play a role. Risk factors include smoking, ionizing radiation, some chemicals (such as benzene), prior chemotherapy, and Down syndrome. People with a family history of leukemia are also at higher risk. There are four main types of leukemia—acute lymphoblastic leukemia (ALL), acute myeloid leukemia (AML), chronic lymphocytic leukemia (CLL) and chronic myeloid leukemia (CML)—as well as a number of less common types. Leukemias and lymphomas both belong to a broader group of tumors that affect the blood, bone marrow, and lymphoid system, known as tumors of the hematopoietic and lymphoid tissues.

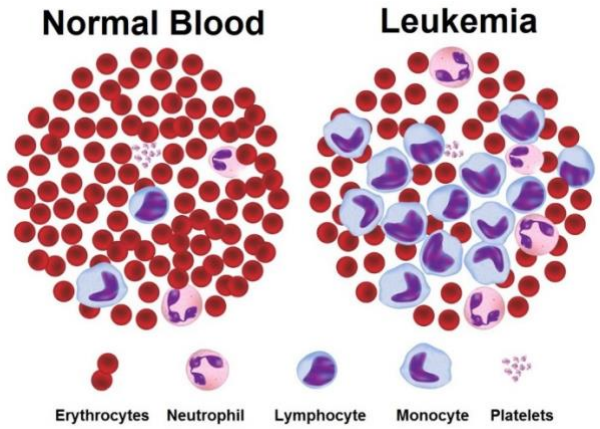


Figure 1:Normal Blood and Leukemia

The French-American-British (FAB) classification system divides AML into eight subtypes, M0 through to M7, based on the type of cell from which the leukemia developed and its degree of maturity. AML of types M0 to M2 may be called acute myeloblastic leukemia. Classification is done by examining the appearance of the malignant cells with light microscopy and/or by using cytogenetics to characterize any underlying chromosomal abnormalities. The subtypes have varying prognoses and responses to therapy.

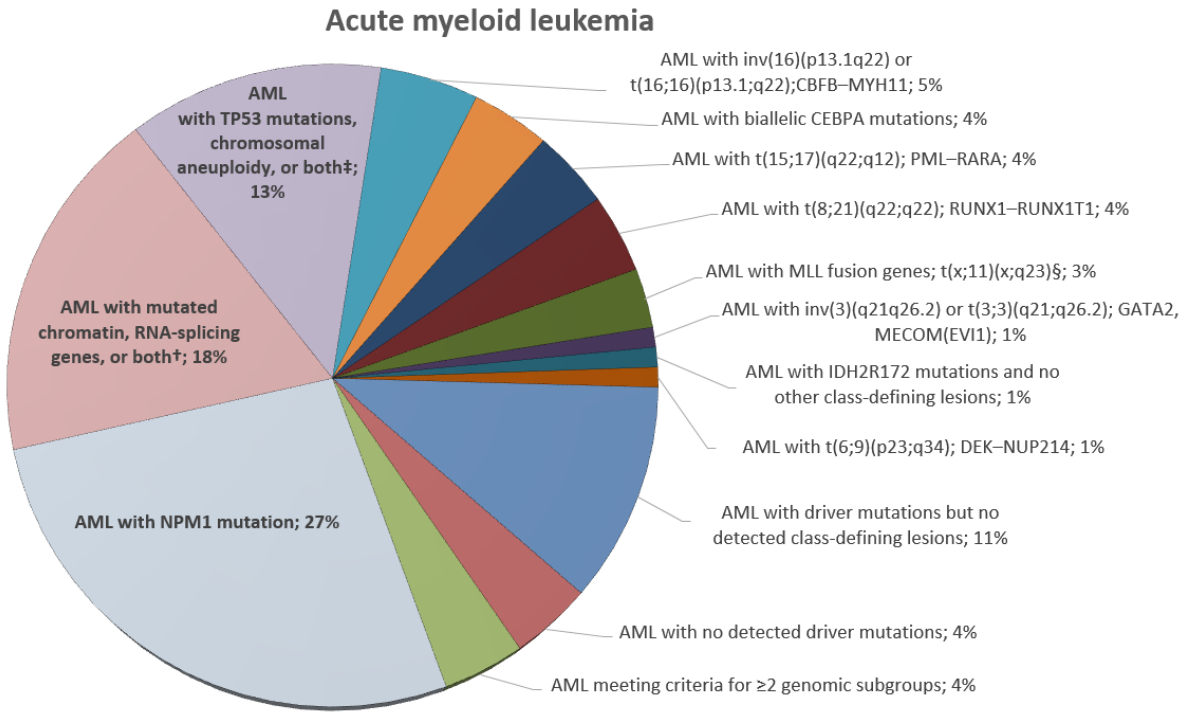


Figure 2: Relative incidence of acute myeloid leukemia subtypes by genetic changes

In this project we intend to analyze **GSE48558** data from GEO website and find genes which are responsible for this cancer.

At first we group the data into “Normal” and “AML” so this dataset has 121 Leukemia and 49 Normal samples.

2.Quality Control

Here for quality control, we have to plot boxplot which indicates that our data is normalized and it's not needed to use Quantile Normalization.

I have saved boxplot in 2 sizes in Results folder with names *boxplot.pdf* and *boxplot_zoomed.pdf.*

Also, the maximum value of the expression data is 13.76154 which shows that our values are in logarithmic unit.

Boxplot is shown below:

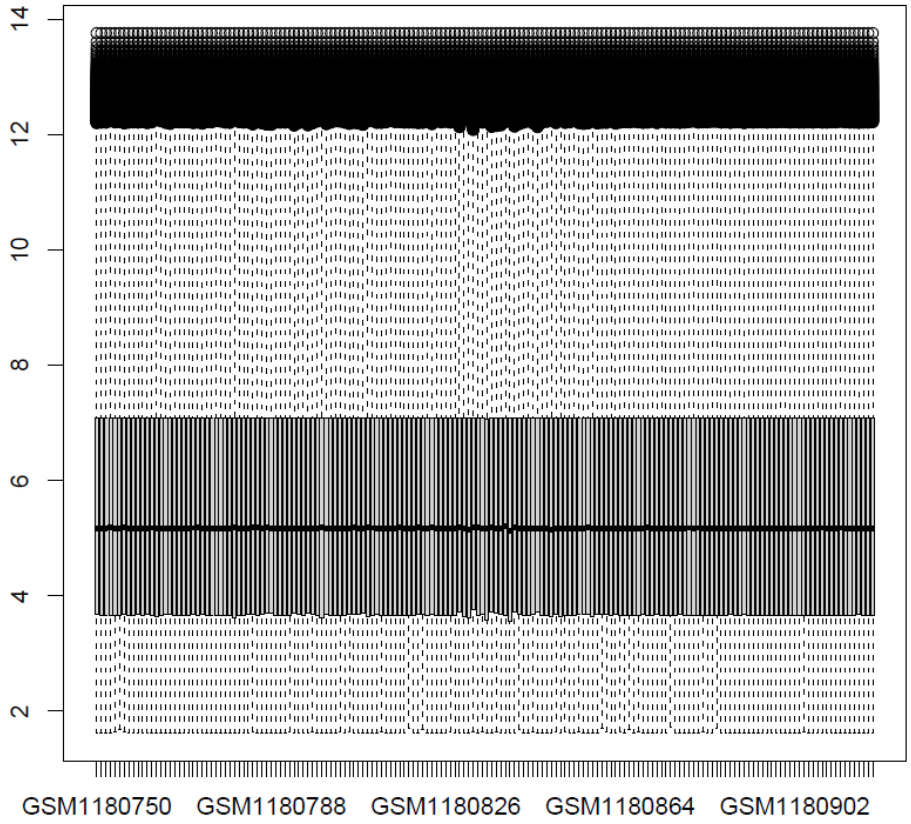


Figure 3 Boxplot

Also, we can do quality control by drawing heatmap of correlation.

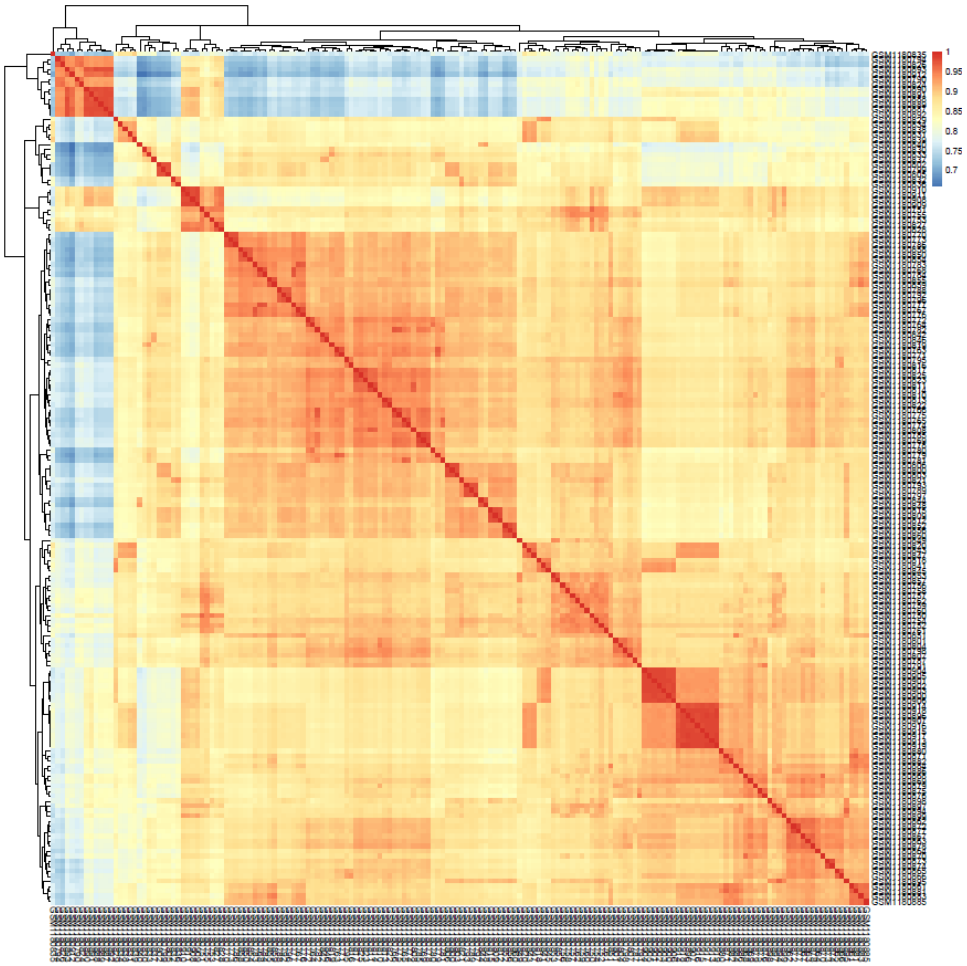


Figure 4: heatmap of correlation on expressions

3. Dimensionality Reduction

We use Principal Component Analysis for this purpose. The following figure shows PC plot which is saved in Results folder named *PC.pdf*.

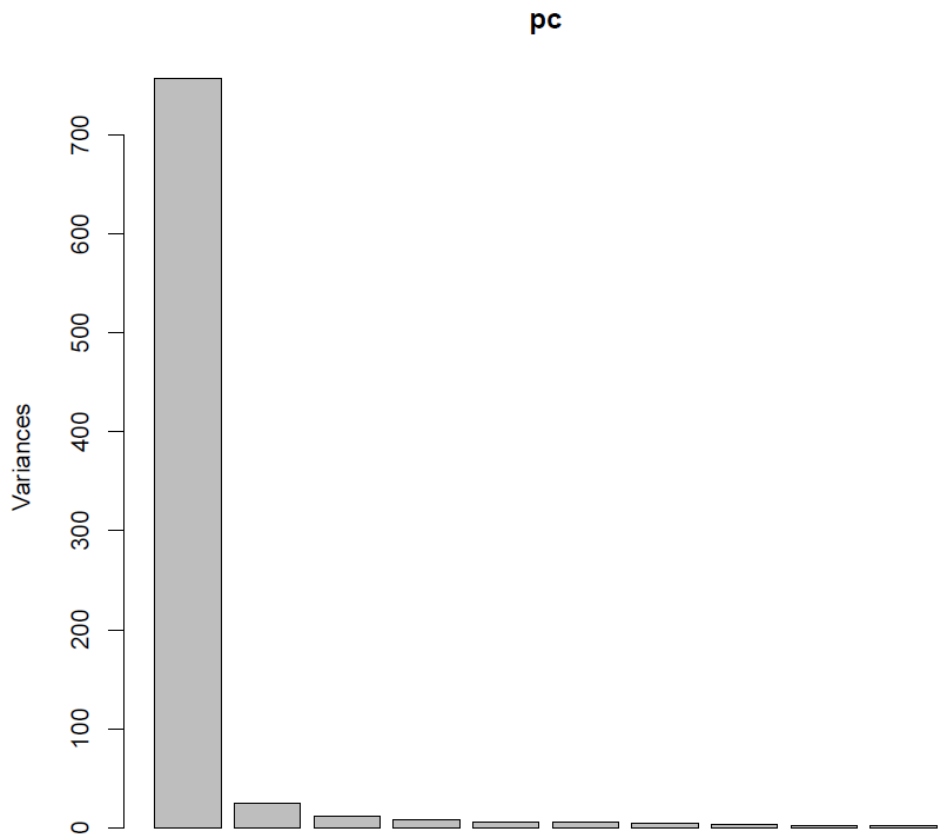


Figure 5: PC

For dimensionality reduction we have to use the features which have the most variance among data and in most cases, we use first 2 features PCA1, PCA2.
The figure below demonstrates data in pca1, pca2 space:

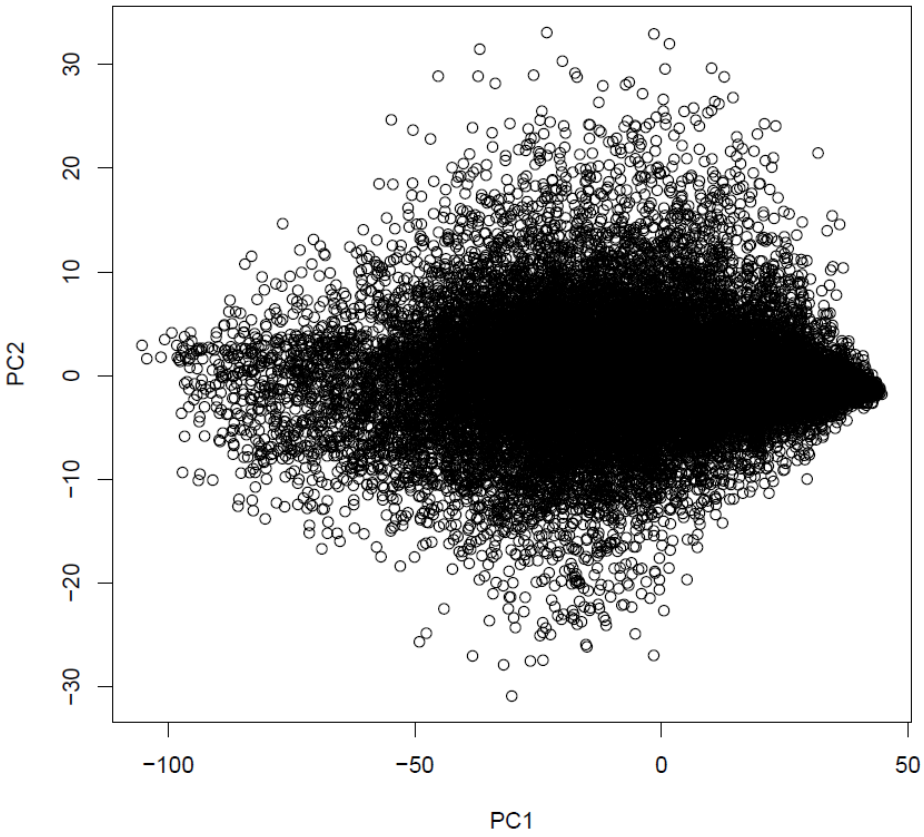


Figure 6: data in lower dimension

Each point in this plot stands for a gene. We see the most variation between pca1 and then pca2. If we pay attention to the plot we can observe that pca1 has separated housekeeping genes which are expressed very low and the ones which are expressed very high among all the samples. It's not useful for our purpose so we have to change it a bit.

A solution that comes to mind for solving this problem is to subtract the amount of gene expression in each sample from the average expression.

We can use *scale()* in R to get scaled gene expressions.

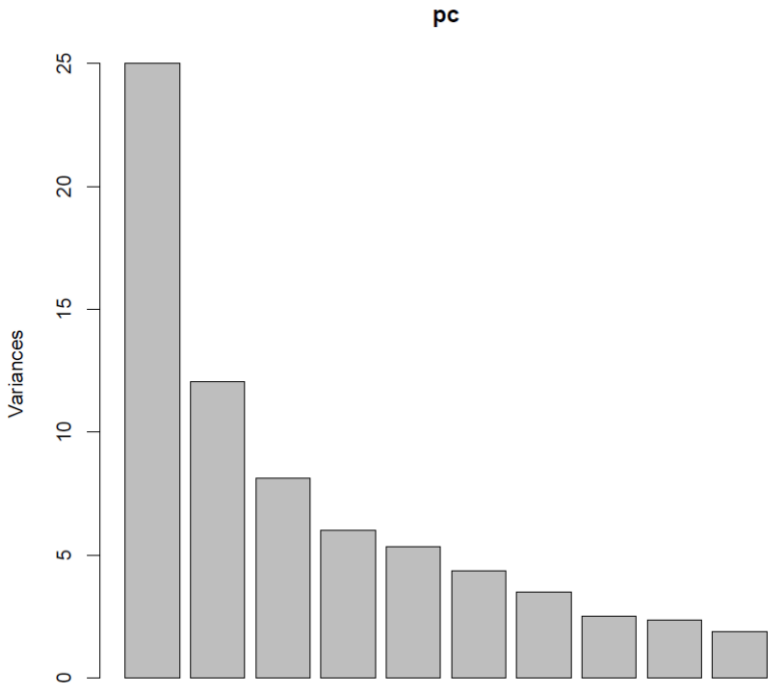


Figure 7: Scaled PC

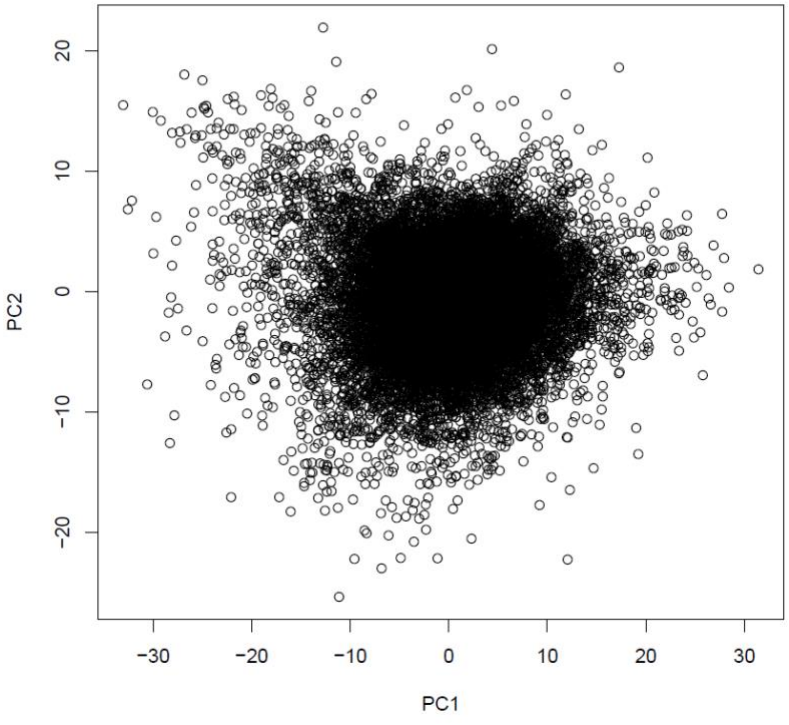


Figure 8: Scaled PC in pca1, pca2 space

Now we want to see the result of this dimensionality reduction on our samples.

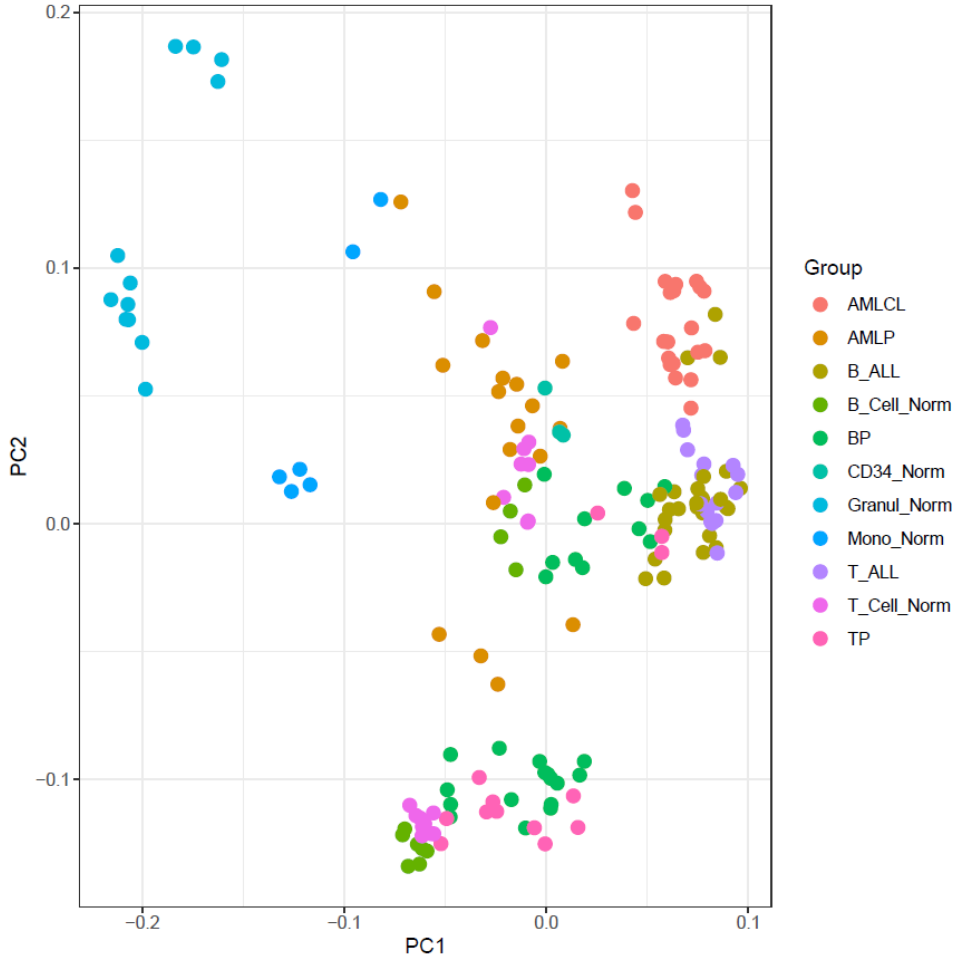


Figure 9: PCA on Samples

As we analyze figure.9 we understand that the 2 groups AMLP and Normal group which is made of Granul_Norm and Mono_Norm are separated with PCA1 well. But different kinds of Leukemia are not separated well and they are mixed so it's better to find another way to do dimensionality reduction on our data.

****Bonus****

In this part I introduce some techniques for dimensionality reduction:

Method	Description
PCA	Principal component analysis
CA, COA	Correspondence analysis
NSC	Nonsymmetric correspondence analysis
PCoA, MDS	Principal co-ordinate analysis/multiple dimensional scaling
NMF	Nonnegative matrix factorization
nmMDS	Nonmetric multidimensional scaling
sPCA, nsPCA, pPCA	Sparse PCA, nonnegative sparse PCA, penalized PCA. (PCA with feature selection)
NIPALS PCA	Nonlinear iterative partial least squares analysis (PCA on data with missing values)
pPCA, bPCA	Probabilistic PCA, Bayesian PCA
MCA	Multiple correspondence analysis
ICA	Independent component analysis
sIPCA	Sparse independent PCA (combines sPCA and ICA)
plots	Graphical resources

Figure 10: Table of Dimensionality Reduction Methods

There are many dimension reduction approaches related to PCA, including principal co-ordinate analysis (PCoA), correspondence analysis (CA) and nonsymmetrical correspondence analysis (NSCA). These may be computed by SVD, but differ in how the data are transformed before decomposition, and therefore, each is optimized for specific data properties. PCoA (also known as Classical Multidimensional

Scaling) is versatile, as it is a SVD of a distance matrix that can be applied to decompose distance matrices of binary, count or continuous data. It is frequently applied in the analysis of microbiome data.

PCA is designed for the analysis of multi-normal distributed data. If data are strongly skewed or extreme outliers are present, the first few axes may only separate those objects with extreme values instead of displaying the main axes of variation. If data are unimodal or display nonlinear trends, one may see distortions or artifacts in the resulting plots, in which the second axis is an arched function of the first axis. Both nonmetric Multi-Dimensional Scaling (MDS) and CA perform better than PCA in these cases. Unlike PCA, CA can be applied to sparse count data with many zeros. Although designed for contingency tables of nonnegative count data, CA and NSCA, decompose a chi-squared matrix, but have been successfully applied to continuous data including gene expression and protein profiles. Gene and protein expression can be seen as an approximation of the number of corresponding molecules present in the cell during a certain measured condition. These two arguments support the suitability of CA and NSCA as analysis methods for omics data. While CA investigates symmetric associations between two variables, NSCA captures asymmetric relations between variables. **Spectral map analysis is related to CA, and performs comparably with CA, each outperforming PCA in the identification of clusters of leukemia gene expression profiles.**

Nonnegative matrix factorization (NMF) forces a positive or nonnegative constraint on the resulting data matrices and, similar to Independent Component Analysis (ICA), there is no requirement for orthogonality or independence in the components. The nonnegative constraint guarantees that only the additive combinations of latent variables are allowed. This may be more intuitive in biology where many biological measurements (e.g., protein concentrations, count data) are represented by positive

values. ICA was recently applied to molecular subtype discovery in bladder cancer. ICA successfully decomposed and extracted multiple layers of signal from the data. The first two components were associated with batch effects but other components revealed new biology about tumor cells and the tumor microenvironment. They also applied ICA to non-bladder cancers and, by comparing the correlation between components, were able to identify a set of bladder cancer-specific components and their associated genes.

As omics data sets tend to have high dimensionality ($p \gg n$) it is often useful to reduce the number of variables. Several recent extensions of PCA include variable selection, often via a regularization step or L-1 penalization (e.g., Least Absolute Shrinkage and Selection Operator, LASSO). The NIPALS algorithm uses an iterative regression approach to calculate the components and loadings, which is easily extended to have a sparse operator that can be included during regression on the component. A cross-validation approach can be used to determine the level of sparsity. Sparse, penalized and regularized extensions of PCA and related methods have been described recently.

Independent Component Analysis (ICA)

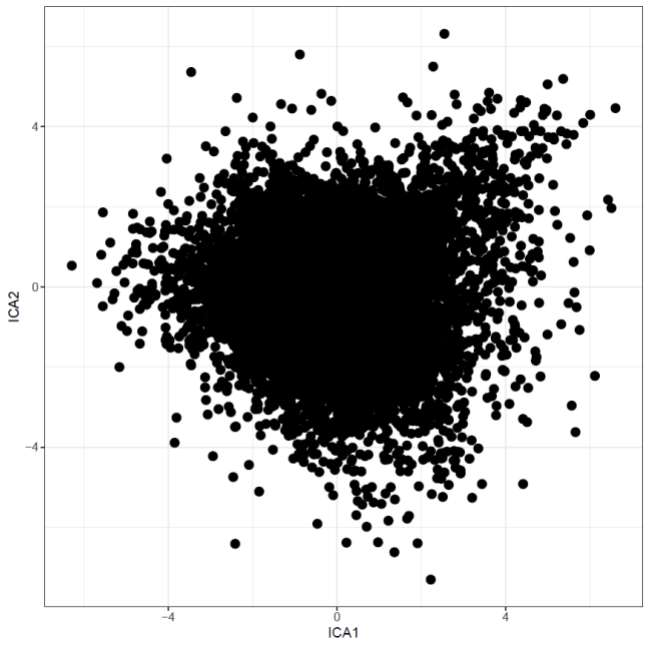


Figure 11: ICA

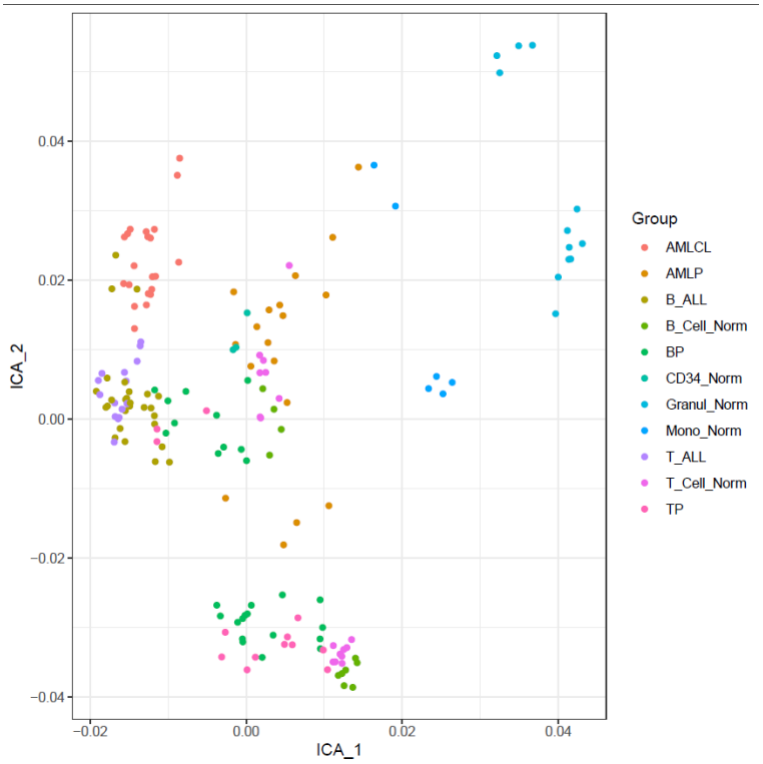


Figure 12: ICA on samples

As we can observe this method doesn't make classification much better so we try another method.

Some of the codes for methods are attached in my code but didn't have time to run them and get results cause they needed python tensorflow.keras library.

These methods are implemented:

UMAP,AUTOENCODER,TSNE.

4. Correlation Analysis

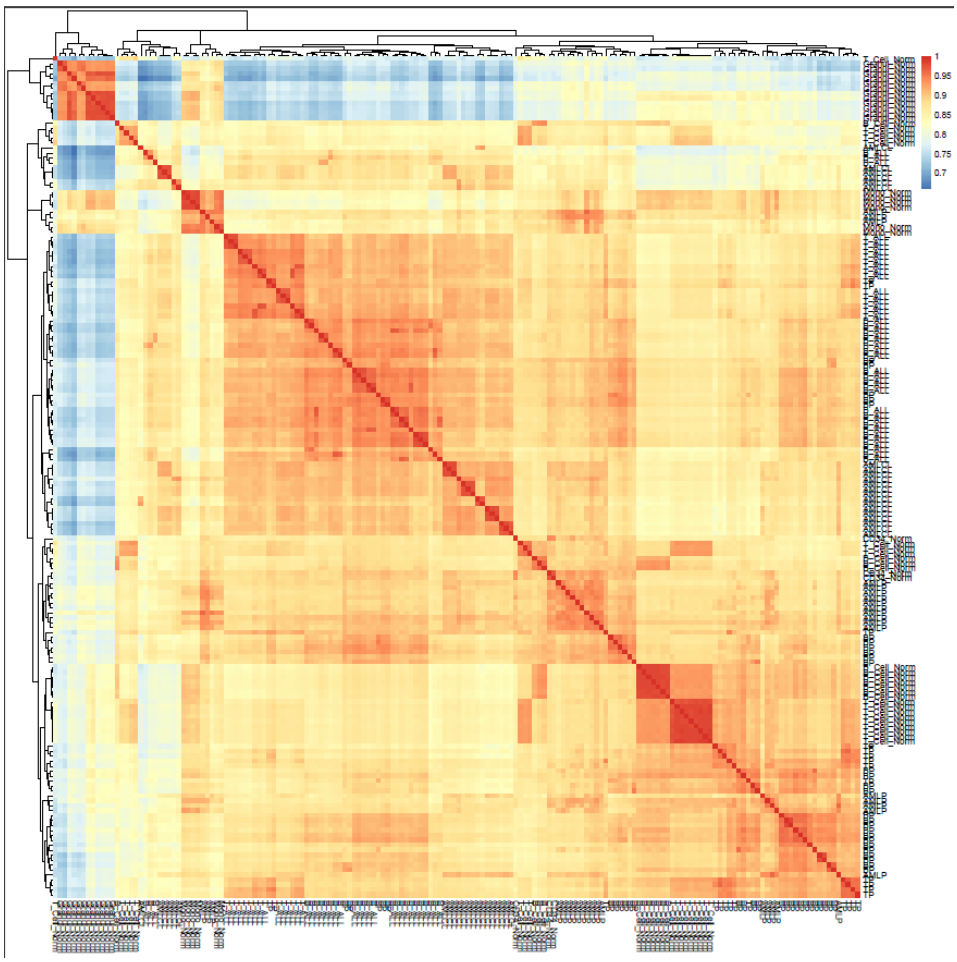


Figure 13: Correlation Heatmap

As we observe in figure.11, it's the result of heatmap on expression of genes before reducing the mean of them. At first we should consider that Normal data is separated into 2 groups of Granul_Norm and Mono_Norm. B-ALL and T-ALL have high correlation. And also, we can observe that Granul_Norm has negative correlation with most types.

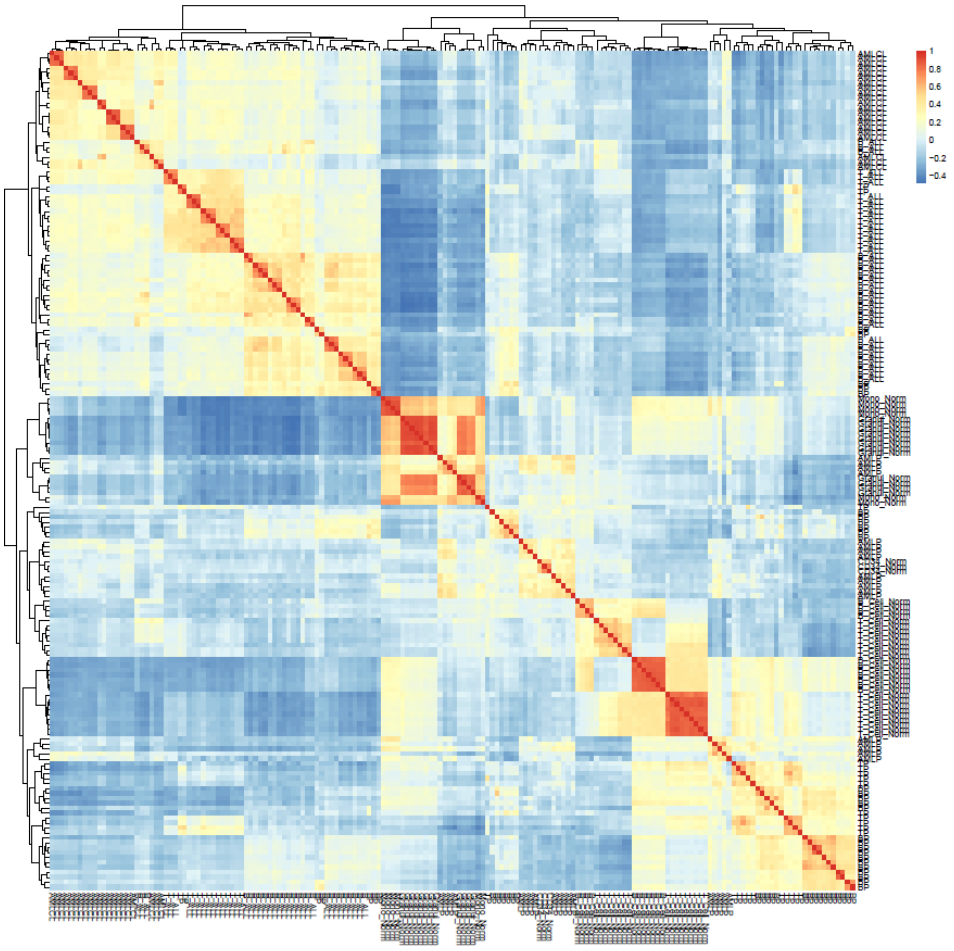


Figure 14: Correlation Heatmap after removing mean

Here is the result after reducing mean of data and got scaled expression of data.

We can conclude that leukemia different types don't have correlation with each other but have negative correlation with Normal types so we can distinguish the leukemia and the normal data from each other.

5. Cell Differentiation Analysis

As we observed in figure.9, AMLP and CD34-Normal data are very close to each other and not separated and far from other normal samples. It means that these 2 groups has more correlation. We can use the website to define 2 groups for AMLP and CD34-Normal and get the results from there and also in my code there exists a part for "Differential Expression Analysis".

In the following table we can see genes that are ordered from largest adj.P.value to lowest adj.P.value. Also, I have downloaded the full table and attached it to my files.

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 ...
8101284	1.38e-11	8.54e-16	20.54	23.57	5.875	PRKG2	protein kinase, cGMP-depend...
8173825	8.80e-11	8.17e-15	18.42	21.96	2.286	RPS6KA6	ribosomal protein S6 kinase A6
7910915	4.01e-10	4.96e-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 m...
7902127	2.15e-09	3.99e-13	15.22	18.95	2.267	SGIP1	SH3 domain GRB2 like endop...
8121814	1.34e-08	3.26e-12	13.69	17.22	4.673	NKAIN2	Na+/K+ transporting ATPase i...
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
8077366	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...
8060897	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
8008588	4.31e-07	2.49e-10	10.93	13.46	4.14	HLF	HLF, PAR bZIP transcription fa...
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
8169158	1.43e-06	1.02e-09	10.12	12.19	2.153	CXorf57	chromosome X open reading f...
8106418	2.29e-06	1.70e-09	9.84	11.72	4.412	CRHBP	corticotropin releasing hormon...
8139933	2.40e-06	1.86e-09	9.79	11.64	2.676		
8173600	3.07e-06	2.47e-09	9.64	11.38	2.635	NAP1L2	nucleosome assembly protein ...

Figure 15: Statistics for AMLP and CD34-Normal

As an example, here is the most different gene in these 2 cases which is **MPDZ**

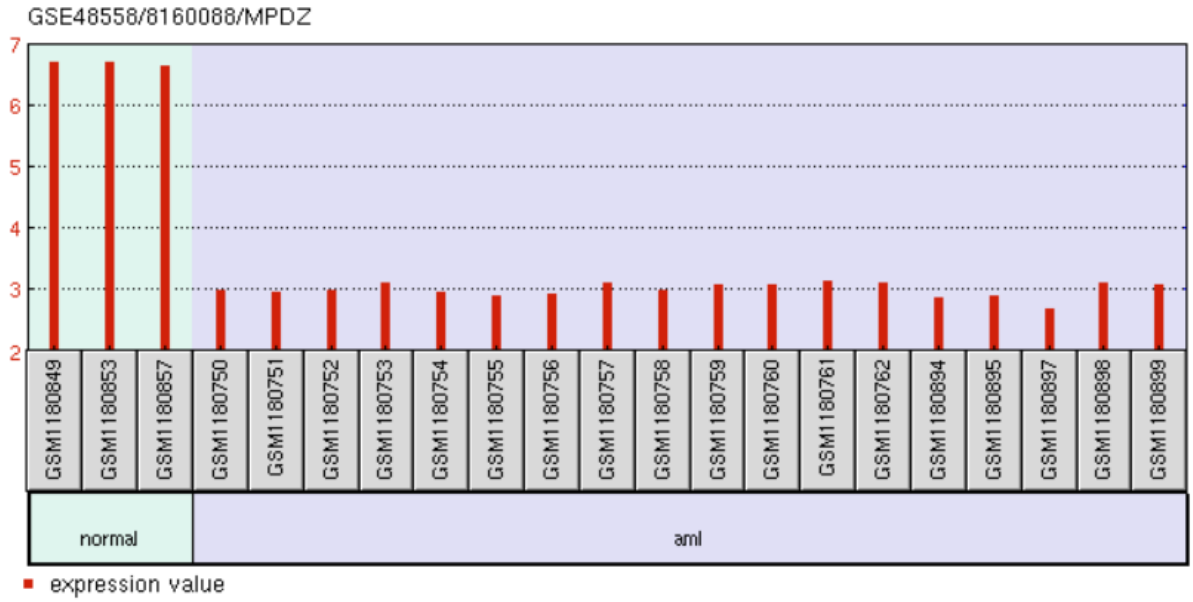


Figure 16: MPDZ gene difference between 2 groups

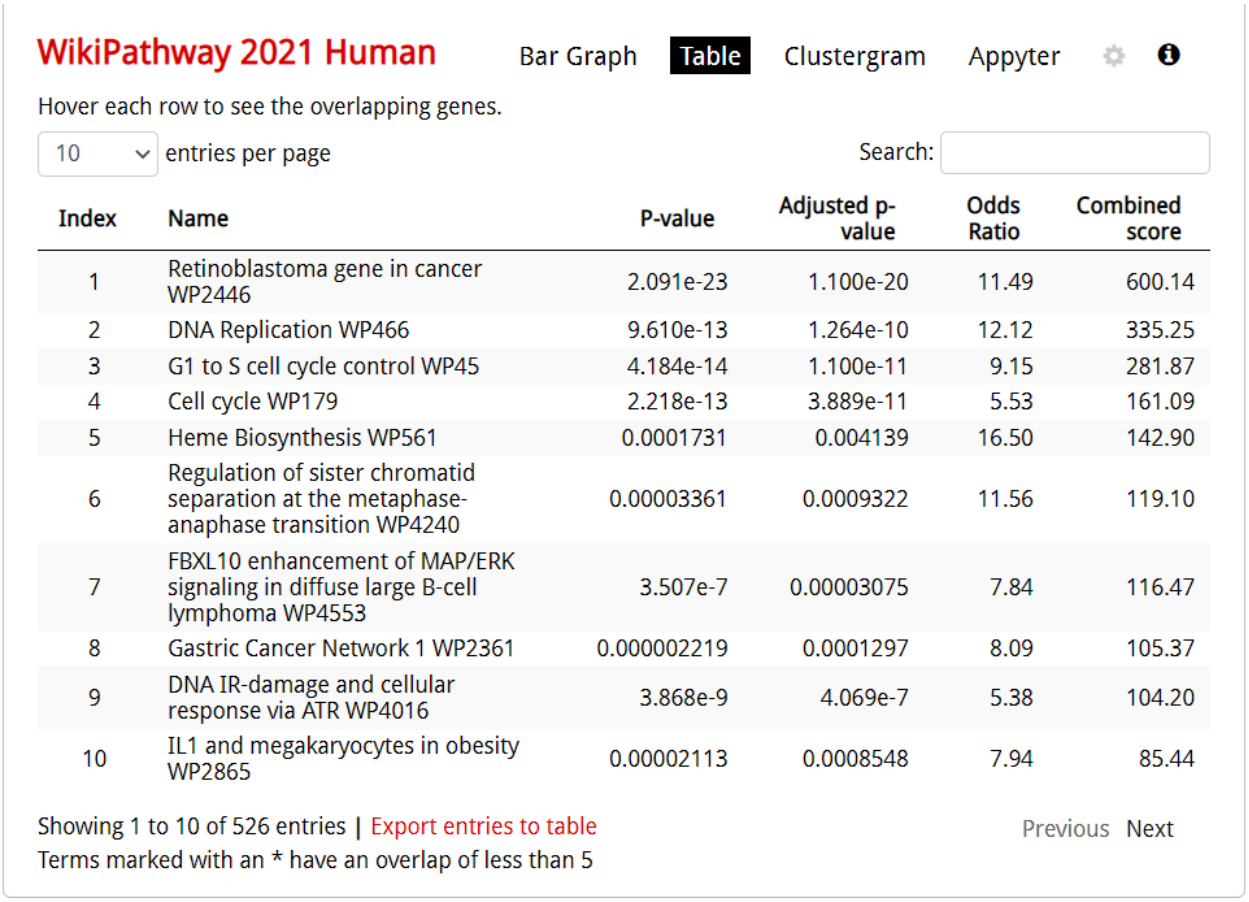
Also I have coded this part in R using a design matrix and fitting a line to the data. Also I found genes that are responsible for high expression and low expression in AMLP and are saved in Results folder files: "*AMLP_Normal_up.txt"* and "*AMLP_Normal_down.txt"*.

Here I have brought some of the results of my code for this part:

Gene	Adj.P.Val	LogFC	Expression in AMLP
MPDZ	1.30e-15	-3.67	Low
PRKG2	1.38e-11	-20.54	Low
FYB	4.4e-05	4.382	High
DPP9	5.5e-05	1.678	High

6. Pathway Analysis

For pathway analysis we use Enrichr website and enter our set of genes there to obtain genes



WikiPathway 2021 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	1.100e-20	11.49	600.14
2	DNA Replication WP466	9.610e-13	1.264e-10	12.12	335.25
3	G1 to S cell cycle control WP45	4.184e-14	1.100e-11	9.15	281.87
4	Cell cycle WP179	2.218e-13	3.889e-11	5.53	161.09
5	Heme Biosynthesis WP561	0.0001731	0.004139	16.50	142.90
6	Regulation of sister chromatid separation at the metaphase-anaphase transition WP4240	0.00003361	0.0009322	11.56	119.10
7	FBXL10 enhancement of MAP/ERK signaling in diffuse large B-cell lymphoma WP4553	3.507e-7	0.00003075	7.84	116.47
8	Gastric Cancer Network 1 WP2361	0.000002219	0.0001297	8.09	105.37
9	DNA IR-damage and cellular response via ATR WP4016	3.868e-9	4.069e-7	5.38	104.20
10	IL1 and megakaryocytes in obesity WP2865	0.00002113	0.0008548	7.94	85.44

Showing 1 to 10 of 526 entries | Export entries to table Previous Next

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

Figure 17:Pathway for AML Up Gene

KEGG 2021 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	Cell cycle	2.071e-14	6.399e-12	5.72	180.16
2	DNA replication	5.135e-7	0.00007934	7.49	108.54
3	One carbon pool by folate	0.0002958	0.007030	7.11	57.80
4	Base excision repair	0.00006170	0.002118	5.75	55.76
5	Valine, leucine and isoleucine degradation	0.00001929	0.001192	4.92	53.43
6	Propanoate metabolism	0.00008191	0.002531	5.51	51.88
7	Malaria	0.00003108	0.001372	4.66	48.32
8	Pyrimidine metabolism	0.00002516	0.001296	4.42	46.79
9	p53 signaling pathway	0.000009835	0.0009913	4.03	46.46
10	Fatty acid elongation	0.0004010	0.008850	5.56	43.52

Showing 1 to 10 of 309 entries | Export entries to table

Previous Next

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

Figure 18: Pathway for AML Up Gene

Here we encounter genes that are related to cell cycle or expression regulation and transcription rate regulation.

So, we conclude that AML Up genes cause increase in cell division and decrease apoptosis of cells.

For AML Down genes we can do the same steps and get the following results:

WikiPathway 2021 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	Modulators of TCR signaling and T cell activation WP5072	2.354e-12	1.001e-9	10.06	269.39
2	Type II interferon signaling (IFNG) WP619	6.911e-8	0.00001469	9.82	161.96
3	T-Cell Receptor and Co-stimulatory Signaling WP2583	4.526e-7	0.00004809	10.75	157.06
4	Cancer immunotherapy by PD-1 blockade WP4585	0.000006123	0.0003253	10.87	130.52
5	Immune response to tuberculosis WP4197	0.000006123	0.0003253	10.87	130.52
6	Pathogenesis of SARS-CoV-2 Mediated by nsp9-nsp10 Complex WP4884	0.00003269	0.001263	10.18	105.18
7	Interactions between immune cells and microRNAs in tumor microenvironment WP4559	0.00003146	0.001263	8.15	84.52
8	T-Cell antigen Receptor (TCR) pathway during Staphylococcus aureus infection WP3863	8.432e-7	0.00005972	5.97	83.56
9	T-cell receptor (TCR) signaling pathway WP69	1.696e-7	0.00002402	5.14	80.08
10	FOXP3 in COVID-19 WP5063	0.0004636	0.01300	10.16	78.03

Showing 1 to 10 of 425 entries | Export entries to table Previous Next

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

Figure 19:Pathway for AML Down Gene

For example, I searched about gene T-cell and here is the result:

T cell, also called T lymphocyte, type of leukocyte (white blood cell) that is an essential part of the immune system. T cells are one of two primary types of lymphocytes—B cells being the second type—that determine the specificity of immune response to antigens (foreign substances) in the body.

So we can conclude that AML down genes can cause problems in our immune system.

7.Gene Anthology Analysis

For AML Up Genes we have:

GO Biological Process 2021 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	positive regulation of superoxide anion generation (GO:0032930)	2.806e-7	0.00005886	99.53	1501.53
2	regulation of hippocampal neuron apoptotic process (GO:0110089)	0.0001729	0.004395	161.89	1402.48
3	regulation of microglial cell mediated cytotoxicity (GO:1904149)	0.0001729	0.004395	161.89	1402.48
4	regulation of neutrophil activation (GO:1902563)	0.0001729	0.004395	161.89	1402.48
5	regulation of superoxide anion generation (GO:0032928)	3.815e-7	0.00006401	90.48	1337.19
6	immunoglobulin mediated immune response (GO:0016064)	0.000008396	0.0009656	105.34	1231.18
7	B cell mediated immunity (GO:0019724)	0.00001151	0.0009656	92.17	1048.15
8	negative regulation of B cell proliferation (GO:0030889)	0.00001530	0.001110	81.92	908.33
9	complement activation, alternative pathway (GO:0006957)	0.0003611	0.007768	97.13	769.87
10	negative regulation of leukocyte activation (GO:0002695)	0.00003135	0.001754	61.43	637.06

Showing 1 to 10 of 839 entries | Export entries to table Previous Next

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

Figure 20: GO AML Up Gene Biological

GO Molecular Function 2021

Bar Graph | **Table** | Clustergram | Appyter

Hover each row to see the overlapping genes.

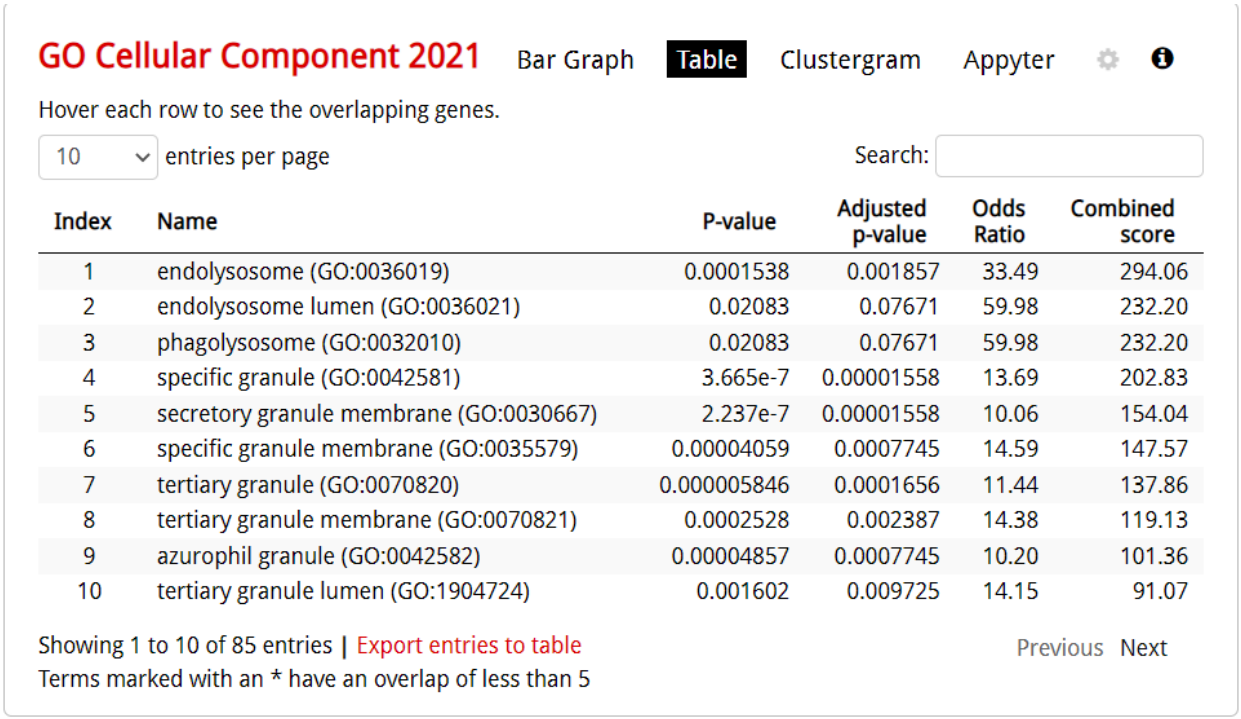
10 entries per page Search:

Index	Name	P-value	Adjusted p-value	Odds Ratio	Combined score
1	IgG binding (GO:0019864)	0.0001729	0.01936	161.89	1402.48
2	complement component C3b binding (GO:0001851)	0.02083	0.1343	59.98	232.20
3	inositol 1,3,4,5 tetrakisphosphate binding (GO:0043533)	0.02083	0.1343	59.98	232.20
4	endogenous lipid antigen binding (GO:0030883)	0.02083	0.1343	59.98	232.20
5	phosphatidylserine flippase activity (GO:0140346)	0.02083	0.1343	59.98	232.20
6	exogenous lipid antigen binding (GO:0030884)	0.02083	0.1343	59.98	232.20
7	cobalamin binding (GO:0031419)	0.02494	0.1343	47.98	177.10
8	collagen receptor activity (GO:0038064)	0.02494	0.1343	47.98	177.10
9	oncostatin-M receptor activity (GO:0004924)	0.02494	0.1343	47.98	177.10
10	phosphatidylethanolamine flippase activity (GO:0090555)	0.02494	0.1343	47.98	177.10

Showing 1 to 10 of 112 entries | Export entries to table Previous Next

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

Figure 21: GO AML Up Gene Molecular Function



GO Cellular Component 2021 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	endolysosome (GO:0036019)	0.0001538	0.001857	33.49	294.06
2	endolysosome lumen (GO:0036021)	0.02083	0.07671	59.98	232.20
3	phagolysosome (GO:0032010)	0.02083	0.07671	59.98	232.20
4	specific granule (GO:0042581)	3.665e-7	0.00001558	13.69	202.83
5	secretory granule membrane (GO:0030667)	2.237e-7	0.00001558	10.06	154.04
6	specific granule membrane (GO:0035579)	0.00004059	0.0007745	14.59	147.57
7	tertiary granule (GO:0070820)	0.000005846	0.0001656	11.44	137.86
8	tertiary granule membrane (GO:0070821)	0.0002528	0.002387	14.38	119.13
9	azurophil granule (GO:0042582)	0.00004857	0.0007745	10.20	101.36
10	tertiary granule lumen (GO:1904724)	0.001602	0.009725	14.15	91.07

Showing 1 to 10 of 85 entries | Export entries to table Previous Next

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

Figure 22: GO AML Up Gene Cellular Component

From this analysis we can get that AML Up gene are correlated with genes responsible for cell division and cell apoptosis.

Here we analyze AML Down genes:

GO Biological Process 2021

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	antigen processing and presentation of endogenous peptide antigen via MHC class I via ER pathway (GO:0002484)	0.000004450	0.0005184	50.84	626.53
2	antigen processing and presentation of endogenous peptide antigen via MHC class I via ER pathway, TAP-independent (GO:0002486)	0.000004450	0.0005184	50.84	626.53
3	cellular response to type I interferon (GO:0071357)	6.702e-16	1.054e-12	12.13	423.68
4	type I interferon signaling pathway (GO:0060337)	6.702e-16	1.054e-12	12.13	423.68
5	negative regulation of humoral immune response mediated by circulating immunoglobulin (GO:0002924)	0.000001917	0.0002809	30.54	401.99
6	protection from natural killer cell mediated cytotoxicity (GO:0042270)	0.00006797	0.004649	40.63	389.91
7	antigen processing and presentation of exogenous peptide antigen via MHC class I, TAP-independent (GO:0002480)	0.00001140	0.001121	33.89	385.76
8	cellular response to interleukin-21 (GO:0098757)	0.00001140	0.001121	33.89	385.76
9	interleukin-21-mediated signaling pathway (GO:0038114)	0.00001140	0.001121	33.89	385.76
10	negative regulation of complement activation, classical pathway (GO:0045959)	0.00001140	0.001121	33.89	385.76

Showing 1 to 10 of 3,146 entries | Export entries to table Previous Next

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

Figure 23:GO AML Down Gene Biological

GO Molecular Function 2021

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	TAP1 binding (GO:0046978)	0.00002354	0.001636	81.27	866.05
2	C-C chemokine binding (GO:0019957)	5.497e-8	0.00001019	14.60	243.98
3	C-C chemokine receptor activity (GO:0016493)	4.936e-7	0.00006861	13.12	190.55
4	chemokine receptor activity (GO:0004950)	0.000001661	0.0001539	10.80	143.79
5	chemokine binding (GO:0019956)	0.000001281	0.0001425	9.28	125.96
6	G protein-coupled chemoattractant receptor activity (GO:0001637)	0.00008497	0.004575	11.10	104.04
7	MHC class I protein binding (GO:0042288)	0.00008497	0.004575	11.10	104.04
8	peptide transmembrane transporter activity (GO:1904680)	0.003160	0.08993	15.22	87.62
9	GTPase activator activity (GO:0005096)	2.564e-10	1.426e-7	3.24	71.45
10	GTPase regulator activity (GO:0030695)	4.437e-9	0.000001234	3.55	68.25

Showing 1 to 10 of 556 entries | Export entries to table Previous Next

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

Figure 24:GO AML Down Gene Molecular Function

GO Cellular Component 2021

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	MHC class I protein complex (GO:0042612)	0.00006797	0.003104	40.63	389.91
2	T cell receptor complex (GO:0042101)	0.000007775	0.001065	20.35	239.47
3	alpha-beta T cell receptor complex (GO:0042105)	0.001871	0.03205	20.29	127.46
4	MHC protein complex (GO:0042611)	0.00002272	0.002075	10.97	117.27
5	X chromosome (GO:0000805)	0.003160	0.05094	15.22	87.62
6	phagocytic vesicle membrane (GO:0030670)	0.00003630	0.002290	5.83	59.62
7	integral component of lumenal side of endoplasmic reticulum membrane (GO:0071556)	0.0002493	0.006832	6.79	56.31
8	lumenal side of endoplasmic reticulum membrane (GO:0098553)	0.0002493	0.006832	6.79	56.31
9	Barr body (GO:0001740)	0.02015	0.1725	13.51	52.77
10	MPP7-DLG1-LIN7 complex (GO:0097025)	0.02015	0.1725	13.51	52.77

Showing 1 to 10 of 274 entries | Export entries to table Previous Next

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

Figure 25: GO AML Down Gene Cellular Component

Interleukins (IL) are a type of cytokine first thought to be expressed by leukocytes alone but have later been found to be produced by many other body cells. They play essential roles in the activation and differentiation of immune cells, as well as proliferation, maturation, migration, and adhesion. They also have pro-inflammatory and anti-inflammatory properties. The primary function of interleukins is, therefore, to modulate growth, differentiation, and activation during inflammatory and immune responses.

Interferons are proteins that are part of your natural defenses. They tell your immune system that germs or cancer cells are in your body. And they trigger killer immune cells to fight those invaders.

Interferons got their name because they "interfere" with viruses and keep them from multiplying.

From this analysis we can get that AML Down gene are correlated with genes responsible for immune system.

**** Bonus ****

Acute myeloid leukemia (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. The mutation causes the stem cells to produce many more white blood cells than are needed. The white blood cells produced are still immature, so they do not have the infection-fighting properties of fully developed white blood cells. As the number of immature cells increases, the amount of healthy red blood

cells and platelets decrease, and it's this fall that causes many of the symptoms of leukemia.

According to the report above, the main reason is the excessive increase of red blood cells and the involvement of white blood cells. As we have seen in the above analyzes, the decreasing genes we obtained were mostly associated with P values less than 0.05 with the genetic mechanisms associated with the immune system, and the incremental genes we obtained were mostly associated with P values less than 0.05 were associated with genetic mechanisms related to cell division and cell cycle.

- Adult acute myeloid leukemia (AML) is a type of cancer in which the bone marrow makes a large number of abnormal blood cells.
- Leukemia may affect red blood cells, white blood cells, and platelets.
- There are different subtypes of AML.
- Smoking, previous chemotherapy treatment, and exposure to radiation may affect the risk of AML.
- Signs and symptoms of AML include fever, feeling tired, and easy bruising or bleeding.
- Tests that examine the blood and bone marrow are used to diagnose AML.
- Certain factors affect prognosis (chance of recovery) and treatment options.

Acute myeloid leukemia (AML) is a cancer of the blood and bone marrow. It is the most common type of acute leukemia in adults. This type of cancer usually gets

worse quickly if it is not treated. AML is also called acute myelogenous leukemia and acute nonlymphocytic leukemia.

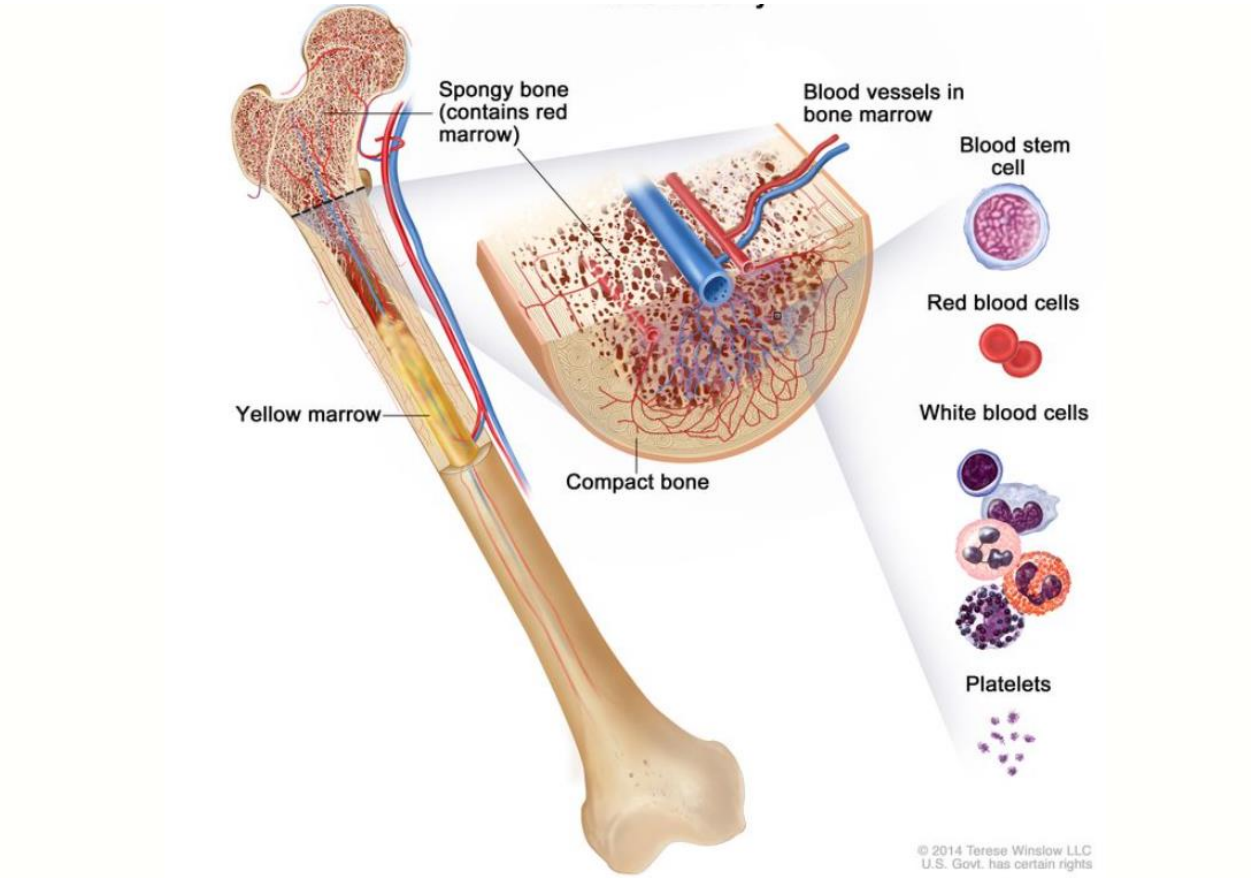


Figure 26:Anatomy of the bone. The bone is made up of compact bone, spongy bone, and bone marrow. Compact bone makes up the outer layer of the bone. Spongy bone is found mostly at the ends of bones and contains red marrow. Bone marrow is found in the center of most bones and has many blood vessels. There are two types of bone marrow: red and yellow. Red marrow contains blood stem cells that can become red blood cells, white blood cells, or platelets. Yellow marrow is made mostly of fat.

Leukemia may affect red blood cells, white blood cells, and platelets.

Normally, the bone marrow makes blood stem cells (immature cells) that become mature blood cells over time. A blood stem cell may become a myeloid stem cell or a lymphoid stem cell. A lymphoid stem cell becomes a white blood cell.

A myeloid stem cell becomes one of three types of mature blood cells:

- Red blood cells that carry oxygen and other substances to all tissues of the body.
- White blood cells that fight infection and disease.
- Platelets that form blood clots to stop bleeding.

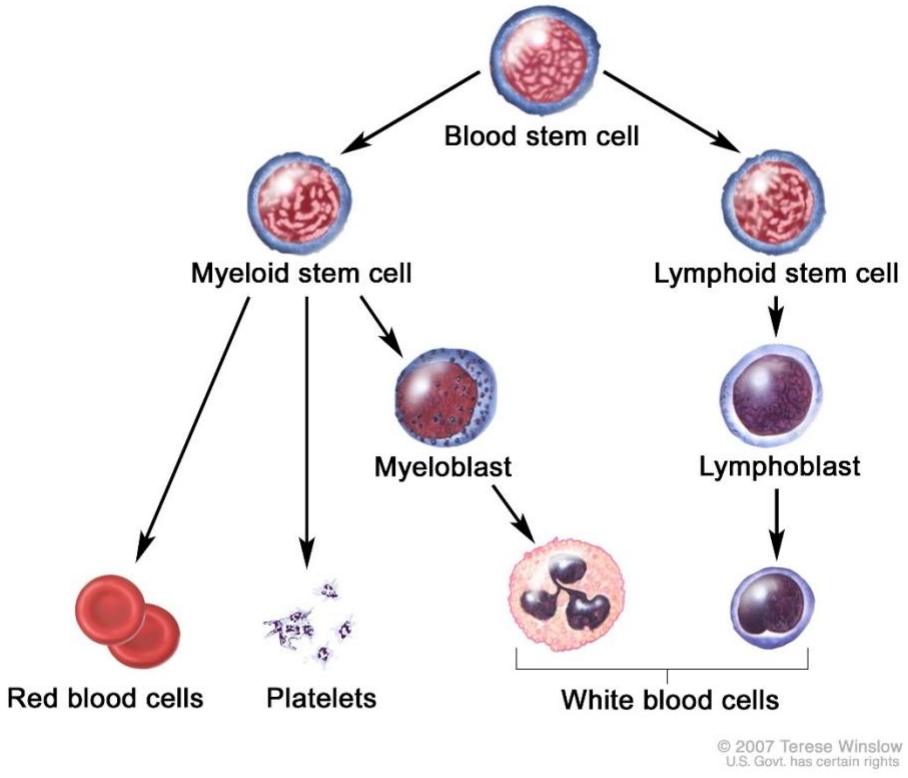


Figure 27:Blood cell development. A blood stem cell goes through several steps to become a red blood cell, platelet, or white blood cell.

In AML, the myeloid stem cells usually become a type of immature white blood cell called myeloblasts (or myeloid blasts). The myeloblasts in AML are abnormal and do not become healthy white blood cells. Sometimes in AML, too many stem cells become abnormal red blood cells or platelets. These abnormal white blood cells, red blood cells, or platelets are also called leukemia cells or blasts. Leukemia cells can build up in the bone marrow and blood so there is less room for healthy white blood cells, red blood cells, and platelets. When this happens, infection, anemia, or easy bleeding may occur.

The leukemia cells can spread outside the blood to other parts of the body, including the central nervous system (brain and spinal cord), skin, and gums. Sometimes leukemia cells form a solid tumor called a myeloid sarcoma. Myeloid sarcoma is also called extramedullary myeloid tumor, granulocytic sarcoma, or chloroma.

Smoking, previous chemotherapy treatment, and exposure to radiation may affect the risk of AML.

Anything that increases your risk of getting a disease is called a risk factor. Having a risk factor does not mean that you will get cancer; not having risk factors doesn't mean that you will not get cancer. Talk with your doctor if you think you may be at risk. Possible risk factors for AML include the following:

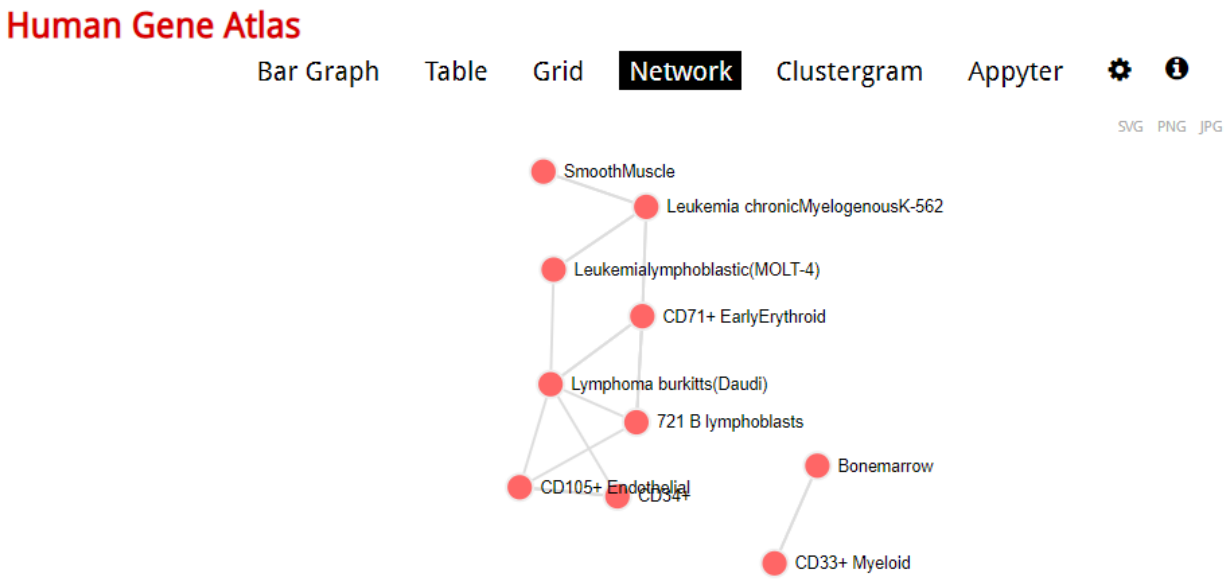
- Being male.
- Older age.
- Smoking.
- Having had treatment with chemotherapy or radiation therapy in the past.

- Being exposed to radiation in the environment (such as nuclear radiation) or to the chemical benzene.
- Having a personal history of a blood disorder such as myelodysplastic syndrome.
- Having certain syndromes or inherited disorders.

In one of the papers I read was mentioned that the results show a significant increase in the absolute number of total T cells in the PB of AML patients compared with healthy controls.

In an article on AML, one patient who survived AML concluded that the number of lymphocytes had increased.

Also, in this article above, CD33 + has been increased too much, which according to the search I did in the enricher section of the cell type using the genes obtained in the project, in the network of this database, which was human gene atlas, which is part of It was blood cells.



Also, I found a paper which introduces a method for treating AML:

Chimeric antigen receptor (CAR) T-cell therapy for acute myeloid leukemia (AML) has thus far been elusive, in part owing to the absence of truly AML-specific surface antigens, making AML difficult to target. However, progress has been made toward the use of CAR T-cell therapy in this disease, prompting the topic of this paper. Discussion and clinical examples of potential solutions to creating a safe and effective CAR T cell for AML include: (1) Decreasing the potency or activity of CAR T cells to enhance the therapeutic window; (2) Using transient or depletable CAR T cells as part of pre-transplant conditioning; and (3) Using a gene-edited allogeneic donor hematopoietic stem cell transplant in order to allow safe and protracted anti-AML CAR T-cell function.

Targeted therapy

Targeted therapy is a type of treatment that uses drugs or other substances to identify and attack specific cancer cells. Targeted therapies usually cause less harm to normal cells than chemotherapy or radiation therapy do. There are different types of targeted therapy:

Monoclonal antibodies: Monoclonal antibodies are immune system proteins made in the laboratory to treat many diseases, including cancer. As a cancer treatment, these antibodies can attach to a specific target on cancer cells or other cells that may help cancer cells grow. The antibodies are able to then kill the cancer cells, block their growth, or keep them from spreading. Monoclonal antibodies are given by infusion. They may be used alone or to carry drugs, toxins, or radioactive material directly to cancer cells. Gemtuzumab ozogamicin is a type

of antibody-drug conjugate used to treat patients with newly diagnosed or relapsed AML. It contains a monoclonal antibody that binds to CD33, which is found on some leukemia cells, and also contains a toxic substance, which may help kill cancer cells.

Other targeted therapies include:

- Midostaurin, a protein kinase inhibitor used with certain types of chemotherapy to treat newly diagnosed patients with AML that has a mutation in the *FLT3 gene*.
- Gilteritinib, a tyrosine kinase inhibitor that may be used to treat patients with AML that has come back or did not get better with other treatment and has a mutation in the *FLT3* gene.
- Glasdegib, ivosidenib, and enasidenib, which may be used as less intensive treatments in older or frail patients who cannot receive standard treatment.

Arsenic trioxide and all-trans retinoic acid (ATRA) are anticancer drugs that kill leukemia cells, stop the leukemia cells from dividing, or help the leukemia cells mature into white blood cells. These drugs are used in the treatment of a subtype of AML called acute promyelocytic leukemia.

Drugs Approved for Acute Myeloid Leukemia (AML)

- Arsenic Trioxide

  Arsenic trioxide causes morphological changes and DNA fragmentation characteristic of apoptosis in NB4 human promyelocytic leukemia cells *in vitro*. Arsenic trioxide also causes damage or degradation of the fusion protein PML/RAR-alpha.

- Cyclophosphamide

Cyclophosphamide (CP), also known as cytophosphane among other names, is a medication used as chemotherapy and to suppress the immune system. As chemotherapy it is used to treat lymphoma, multiple myeloma, leukemia, ovarian cancer, breast cancer, small cell lung cancer, neuroblastoma, and sarcoma. As an immune suppressor it is used in nephrotic syndrome, granulomatosis with polyangiitis, and following organ transplant, among other conditions. It is taken by mouth or injection into a vein.

- Daunorubicin Hydrochloride

The sugar moiety of cytarabine hinders the rotation of the molecule within the DNA. The DNA replication ceases, specifically during the S phase of the cell cycle, making it a specific drug for rapidly dividing cells, such as those seen in cancer.

- Gemtuzumab Ozogamicin

Gemtuzumab ozogamicin is a recombinant humanized IgG4 kappa antibody which is conjugated with calicheamicin derivative, a cytotoxic antitumor antibiotic isolated from fermentation of Micromonospora echinospora ssp. calichensis. Gemtuzumab ozogamicin has approximately 50% of the antibody loaded with 4-6 moles calicheamicin per mole of antibody [Label]. The antibody is specifically directed against the CD33 antigen present on leukemic myeloblasts in most patients with acute myeloid leukemia (AML). By binding to the CD33 antigen on tumors, the cytotoxic agent blocks the growth of cancerous cells and causes cell death.

So we can conclude that all these drugs affect cell cycle and cell division speed or the immune system so the results I analyzed for the gene anthology and pathways match the results from the papers in this field.

Some Amazing Findings

When analyzing AML Down Genes, I checked the tab Disease/Drug and found a dataset for COVID-19 Genes 2021 that the genes had very low P values that shows AML Down Genes are responsible for immune system just like genes correlated for covid-19 disease.

COVID-19 Related Gene Sets 2021

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	500 genes up-regulated by SARS-CoV-2 in Calu-3 cells from GSE148729	6.582e-36	7.487e-34	6.09	493.04
2	SARS perturbation; 220 Down Genes from GEN3VA; Human PBMCs; Accession: GDS1028 Platform: GPL201	1.053e-27	3.686e-26	7.85	487.92
3	500 genes up regulated by SARS-COV-2 infection of Calu3 cells	2.814e-36	4.267e-34	5.74	469.71
4	500 genes up-regulated by SARS-CoV-2 (2 MOI) in human Calu-3 cells from GSE147507	2.814e-36	4.267e-34	5.74	469.71
5	500 genes up-regulated by SARS-CoV-2 in human Calu3 cells from GSE147507	2.383e-36	4.267e-34	5.63	461.47
6	500 genes up-regulated by SARS-CoV-2 in human Calu3 cells at 24h from GSE148729 s1 polyA	2.292e-35	2.086e-33	5.51	439.16
7	500 genes up-regulated by SARS-CoV-1 in Calu-3 cells from GSE148729 24h	2.824e-33	1.835e-31	5.68	425.65
8	500 genes up-regulated by SARS-CoV-2 in human Calu3 cells at 24h from GSE148729 totalRNA	4.837e-34	3.668e-32	5.38	412.86
9	500 genes up-regulated by SARS-CoV-2 in human Organoids cells from GSE154613	2.924e-31	1.663e-29	5.28	371.02

References

[1] Acute Myeloid Leukemia Treatment (PDQ®)–Patient Version - National Cancer Institute

[2] 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)

[4] How close are we to CAR T-cell therapy for AML? - ScienceDirect

[5] Drugs Approved for Leukemia - National Cancer Institute

[6] Genomics of Acute Myeloid Leukemia Diagnosis and Pathways (ascopubs.org)

[7] Acute Myeloid Leukemia (AML): Causes, Symptoms, Diagnosis, and Treatment (webmd.com)

[8] Acute myeloid leukaemia - Causes - NHS (www.nhs.uk)

[9] Arsenic trioxide: Uses, Interactions, Mechanism of Action | DrugBank Online

Gemtuzumab ozogamicin: Uses, Interactions, Mechanism of Action | DrugBank Online

Gilteritinib: Uses, Interactions, Mechanism of Action | DrugBank Online

Midostaurin: Uses, Interactions, Mechanism of Action | DrugBank Online

Prednisone: Uses, Interactions, Mechanism of Action | DrugBank Online

Azacitidine - an overview | ScienceDirect Topics