Bioinformatics@Data Science A.Y. 2019-2020

Network Medicine Project 2

Vikranth Ale and Kentaro Kato

Group no. 7

Abstract

Acute myeloid leukemia (AML), M1 is a cancer of the myeloid tissue of blood cells i.e. on the bone marrow. Here M1 is a subtype which indicates minimal/without maturation. In this project we study possible genes that are involved in mutations by documenting their interactions with other genes with the interaction data collected from sources like BioGrid and IID, arranging this data into interactomes and use clustering methods to discover the disease modules, we also perform a detailed study which involves enrichment analysis of these putative disease modules.

Basic introduction about the disease/process

Acute myeloid leukemia (AML),M1 is a cancer of bone marrow that prevents blood cells from growing into healthy blood cells, bone marrow is a spongy tissue inside bones where blood cells are made, this cancer doesn't cause any tumors but the cells grow very rapidly in an abnormal way and become immature i.e. blast cells, they may even spread to other organs like liver and spleen. Early symptoms include feeling tiredness, pale skin, heartbeat irregularities, weight loss, headaches, shortness of breath, blurred vision etc. Since AML affects our blood cells, other complications include anemia, unusual bruising or bleeding, increased risk of infections and weakened immune system. It can be treated with chemotherapy, stem cell transplant, radiation and targeted therapy which is done in phases. Risk factors of AML include smoking, previous chemotherapy/radiation therapy, myelodysplastic syndrome and exposure to chemical benzene.

AML is the second most common type of leukemia and is seen mostly in adults all over the world, it makes up for 32% of all leukemia cases. The average age of diagnosis is 68. The 5-year survival rate for people less than 20 years is approximately 67% and that for 20 years and older is at 24%.

1.2 Seed genes

We downloaded the curated gene-disease associations dataset from DisGeNet database to get all the seed genes information along with their Gene Symbol, GeneID, UniProt entry and Protein names as shown in Table 1 below.

Summary on interaction data

For all the seed genes shown in Table 1, we got their corresponding approved symbols from HGNC website and collected their information from UniProt website, we manually collected interaction data from BioGrid and IID databases. We did some preprocessing to extract all genes data related to Human Beings with their approved symbols.

1.3 Arranging Interactions Data:

To arrange interactions from BioGrid and IID into specific interactomes into the combinations seed-seed, union and intersection interactomes, we have first created a mapping between all the Gene symbols and the UniProt entries for all human genes and then considered those interactions with proteins involved as per conditions.

Seed genes interactome: To create this interactome we have considered all the interactions from both BioGrid and IID database in which both the interacting proteins are seed genes. Total interactions found are 344 and the results are saved in the table as csv file. A sample is shown below.

Table 1.3.1. sample of seed genes interactome

Database	GeneA	GeneB	UniProtA	UniProtB
BioGrid	KMT2A	KMT2A	Q03164	Q03164
BioGrid	ASMTL	ASMTL	O95671	O95671
BioGrid	DNMT3A	MYC	Q9Y6K1	P01106

Union interactome: To create this interactome we have considered all the interactions from both Bio-Grid and IID databases with all proteins interacting with at least one seed gene, this union interactome has a total of 207034 interactions and the results are saved in the table as csv file.

Table 1.3.1. sample of union interactome

Database	GeneA	GeneB	UniProtA	UniProtB
BioGrid	GATA2	PML	P23769	P29590
BioGrid	RPA2	STAT3	P15927	P40763
BioGrid	ARF1	GGA3	P84077	Q9NZ52

Intersection interactome: To create this interactome we have considered all the interactions from both BioGrid and IID database with all proteins interacting with at least one seed gene confirmed, this intersection interactome has a total of 3943 interactions and the results are saved in the table as csv file. A sample is shown below.

Table 1.3.3. sample of intersection interactome

Database	GeneA	GeneB	UniProtA	UniProtB
IID	SGK1	TUFM	O00141	P49411
IID	SGK1	MAPT	O00141	P10636
IID	SGK1	CDKN1B	O00141	P46527

While doing the above interactomes we have encountered situations where the same interactions are present multiple times, and this is totally possible because a same protein may tend to interact on its own under different conditions in a lot of ways.

1.4 Enrichment Analysis:

We did enrichment analysis considering the seed genes from the above interactomes for both SG and Union by using a service called **Enrichr** to identify the classes of genes that are overrepresented in GO categories and pathways.

We have uploaded the list of seed genes into to the Enrichr website and extracted data of GO categories and Pathways to finds cases that occur more frequently than expected which can be linked to a particular biological process, molecular function, cellular components or KEGG 2019 Human for pathway analysis, all the relevant charts (Figure 1.4) and tables (Table 1.4) are saved in appendix.

2 DATA ANALYSIS

2.1 Calculation of Network Measures and Largely Connected Components [Local & Global]

For our 3 types of interactomes i.e. SGI (Seed-Seed Interactome), I (Intersection Interactome) and U (Union Interactome), we computed both global and local network measures using the tool cytoscape and betweenness centrality using a library in python called network. First, we computed global measures for SGI, I, and U (Table). Next, we extracted largest connected component for each I and U, called I-LCC and U-LCC as shown in the tables (Table 2.1).

2.2 Markov Clustering (MCL)

Once we got the networks of interactome, in this section, we identified the subnetworks (disease modules) in the LCC. These modules can be found because the genes found in some diseases could have a tendency of the same network in any disease. In order to identify the disease modules, we use a clustering method called Markov Clustering (MCL). It is implemented using cytoscape, additionally we applied hypergeometric test to find modules for which seed genes are statistically overrepresented. The modules that satisfy this condition are called **Putative disease modules** and shown in table 2.2

2.3 Enrichment Analysis for Disease Modules

We did enrichment analysis considering the seed genes from the above disease modules for both I_LCC and U_LCC networks by following the same procedure which was done in section 1.4 to identify the classes of genes that are overrepresented in GO categories and pathways.

These genes are then grouped to understand cases that occur more frequently than expected that can be linked to a particular biological process, molecular function, cellular components or KEGG 2019 Human for pathway analysis, all the relevant charts (Figure2.3.) and tables (Table 2.3.) are saved in appendix.

2.4 Putative disease genes using the DIAMOnD tool

We used DIAMOnD to identify the full disease module around a set of known disease genes that will help to compute the putative disease genes using a reference interactome of the whole Human genes from Bio Grid to collect PPIs with our seed gene list. We limited the putative disease proteins to 200 genes. Table describes the first 30 ranked genes of putative disease genes. Also, we find overrepresented GO categories and overrepresented pathways of such 200 newly found genes. (Table 2.4)

For the above 200 genes, we did enrichment analysis to identify the overrepresented GO categories and pathways following the same process as done in exercise 1.4. to group them into a relevant biological process, molecular function, Cellular components or KEGG 2019 Human for pathway analysis, all the relevant charts (Figure2.4) and tables (Table 2.4) are saved in appendix.

References:

[1] Integrative Biomedical Informatics (IBI) Group DisGeNet database
https://www.disgenet.org
[2] HGNC Team HuGO Gene Nomenclature Committee
https://www.genenames.org
[3] European Bioinformatics Institute (EMBLEBI), SIB Swiss Institute of Bioinformatics and the Protein Information Resource (PIR) Database for protein sequence and annotation data
https://www.uniprot.org
[4] BioGRID Team Database of genetic and protein interaction data
https://thebiogrid.org
[5] IID Group Database of detected and predicted protein-protein interactions
http://iid.ophid.utoronto.ca
[6] Ma'ayan Lab integrative web-based and mobile software application for ranking enriched terms
https://amp.pharm.mssm.edu/Enrichr/
[7]Cytoscape team Software platform for visualizing and studying complex networks
https://cytoscape.org/index.html
[8] NetworkX-Developers Python package for the creation and study of complex networks
https://networkx.github.io/documentation/stable/
[9] Susan Dina Ghiassian, J¨org Menche, Albert-L´aszl´o Barab´asi A DIseAse MOdule Detection (DIAMOnD) Algorithm Derived from a Systematic Analysis of Connectivity Patterns of Disease Proteins in the Human Interactome
https://journals.plos.org/ploscompbiol/article?id=10.1371/journal.pcbi.1004120
[10] WEBMD
https://www.webmd.com/cancer/lymphoma/acute-myeloid-leukemia-symptoms-treatments
[11] AMERICAN CANCER SOCIETY
https://www.cancer.org/cancer/acute-myeloid-leukemia/detection-diagnosis-staging/how-classified.html
[12] Acute myeloid leukemia
https://en.wikipedia.org/wiki/Acute_myeloid_leukemia

APPENDIX:

Table 1. Seed genes

GeneSymbol	GeneID	UniProtAC	ProteinName
FLT3	2322	P36888	fms related tyrosine kinase 3
NPM1	4869	P06748	nucleophosmin 1
IDH1	3417	O75874	isocitrate dehydrogenase (NADP(+)) 1, cytosolic
KMT2A	4297	Q03164	lysine methyltransferase 2A
IDH2	3418	P48735	isocitrate dehydrogenase (NADP(+)) 2, mitochondrial
MYH11	4629	P35749	myosin heavy chain 11
GATA2	2624	P23769	GATA binding protein 2
CEBPA	1050	P49715	CCAAT enhancer binding protein alpha
S100A10	6281	P60903	S100 calcium binding protein A10
SGK1	6446	O00141	serum/glucocorticoid regulated kinase 1
SPARC	6678	P09486	secreted protein acidic and cysteine rich
SPI1	6688	P17947	Spi-1 proto-oncogene
NF1	4763	P21359	neurofibromin 1
STAT3	6774	P40763	signal transducer and activator of transcription 3
SVIL	6840	O95425	supervillin
TCEA2	6919	Q15560	transcription elongation factor A2
TRH	7200	P20396	thyrotropin releasing hormone
S100A8	6279	P05109	S100 calcium binding protein A8
RGS2	5997	P41220	regulator of G protein 5ignalling 2
MN1	4330	Q10571	MN1 proto-oncogene, transcriptional regulator
MX1	4599	P20591	MX dynamin like GTPase 1
MYC	4609	P01106	MYC proto-oncogene, bHLH transcription factor
ASXL2	55252	Q76L83	ASXL transcriptional regulator 2
NRAS	4893	P01111	NRAS proto-oncogene, GTPase
NUP98	4928	P52948	nucleoporin 98
PDE4B	5142	Q07343	phosphodiesterase 4B
POU4F1	5457	Q01851	POU class 4 homeobox 1
PTPN11	5781	Q06124	protein tyrosine phosphatase, non-receptor type 11
TRIO	7204	O75962	trio Rho guanine nucleotide exchange factor
TSC2	7249	P49815	TSC complex subunit 2
VSIG4	11326	Q9Y279	V-set and immunoglobulin domain containing 4
EHD3	30845	Q9NZN3	EH domain containing 3
ZBTB7A	51341	O95365	zinc finger and BTB domain containing 7A
CHMP5	51510	Q9NZZ3	charged multivesicular body protein 5
FXYD6	53826	Q9H0Q3	FXYD domain containing ion transport regulator 6
ENAH	55740	Q8N8S7	ENAH, actin regulator
KMT2C	58508	Q8NEZ4	lysine methyltransferase 2C
BACH2	60468	Q9BYV9	BTB domain and CNC homolog 2
BAALC	79870	Q8WXS3	BAALC, MAP3K1 and KLF4 binding
VOPP1	81552	Q96AW1	VOPP1, WBP1/VOPP1 family member
SPRY4	81848	Q9C004	5ignal RTK 5ignalling antagonist 4
PSIP1	11168	O75475	PC4 and SFRS1 interacting protein 1
PIM2	11040	Q9P1W9	Pim-2 proto-oncogene, serine/threonine kinase

EHMT2	10919	Q96KQ7	euchromatic histone lysine methyltransferase 2
TUBB2A	7280	Q13885	tubulin beta 2A class Iia
WT1	7490	P19544	Wilms tumor 1
PXDN	7837	Q92626	peroxidasin
ASMTL	8623	O95671	acetylserotonin O-methyltransferase like
TNFSF10	8743	P50591	TNF superfamily member 10
SYNGR1	9145	O43759	synaptogyrin 1
RASGRP1	10125	O95267	RAS guanyl releasing protein 1
IFI30	10437	P13284	IFI30, lysosomal thiol reductase
GAS2L1	10634	Q99501	growth arrest specific 2 like 1
SEPTIN9	10801	Q9UHD8	septin 9
AGRN	375790	O00468	agrin
ADCY7	113	P51828	adenylate cyclase 7
RUNX3	864	Q13761	runt related transcription factor 3
CBFB	865	Q13951	core-binding factor subunit beta
CCND2	894	P30279	cyclin D2
CD9	928	P21926	CD9 molecule
CD33	945	P20138	CD33 molecule
CD44	960	P16070	CD44 molecule (Indian blood group)
CDK6	1021	Q00534	cyclin dependent kinase 6
CEBPD	1052	P49716	CCAAT enhancer binding protein delta
CNR2	1269	P34972	cannabinoid receptor 2
CSF1R	1436	P07333	colony stimulating factor 1 receptor
RUNX1T1	862	Q06455	RUNX1 translocation partner 1
RUNX1	861	Q01196	runt related transcription factor 1
CASP7	840	P55210	caspase 7
ANXA2	302	P07355	annexin A2
ANXA4	307	P09525	annexin A4
ANXA5	308	P08758	annexin A5
ANXA6	309	P08133	annexin A6
FAS	355	P25445	Fas cell surface death receptor
AQP9	366	O43315	aquaporin 9
ATP1B1	481	P05026	ATPase Na+/K+ transporting subunit beta 1
BCL2	596	P10415	BCL2, apoptosis regulator
CAPG	822	P40121	capping actin protein, gelsolin like
CAPN2	824	P17655	calpain 2
CSF2	1437	P04141	colony stimulating factor 2
CSF3	1440	P09919	colony stimulating factor 3
CST3	1471	P01034	cystatin C
GTF2I	2969	P78347	general transcription factor Iii
H1-0	3005	P07305	H1 histone family member 0
H1-2	3006	P16403	histone cluster 1 H1 family member c
HGF	3082	P14210	hepatocyte growth factor
HOXA9	3205	P31269	homeobox A9
HSPB1	3315	P04792	heat shock protein family B (small) member 1
ID2	3398	Q02363	inhibitor of DNA binding 2
JAK2	3717	O60674	Janus kinase 2

KRAS	3845	P01116	KRAS proto-oncogene, GTPase
LYL1	4066	P12980	LYL1, basic helix-loop-helix family member
GFI1	2672	Q99684	growth factor independent 1 transcriptional repressor
FOXO1	2308	Q12778	forkhead box O1
FHL2	2274	Q14192	four and a half LIM domains 2
CTNNA1	1495	P35221	catenin alpha 1
CTSH	1512	P09668	cathepsin H
CTSZ	1522	Q9UBR2	cathepsin Z
DAPK1	1612	P53355	death associated protein kinase 1
DHX15	1665	O43143	DEAH-box helicase 15
DNMT3A	1788	Q9Y6K1	DNA methyltransferase 3 alpha
LPAR1	1902	Q92633	lysophosphatidic acid receptor 1
EIF4EBP1	1978	Q13541	eukaryotic translation initiation factor 4E binding protein 1
ENO2	2026	P09104	enolase 2
ERG	2078	P11308	ERG, ETS transcription factor
MET	4233	P08581	MET proto-oncogene, receptor tyrosine kinase

Table 1.2. Summary of interaction data

Data-base	No. of Seed Genes	No. of Interacting Proteins	Number of Interactions
BioGrid	100	17909	269952
IID	106	4759	8508

Table 1.2.1. Sample of BioGrid

GeneA	GeneB	UniProtA	UniProtB
KMT2A	KMT2A	Q03164	Q03164
ASMTL	ASMTL	O95671	O95671
DNMT3A	MYC	Q9Y6K1	P01106
MYC	DNMT3A	P01106	Q9Y6K1
CTNNA1	CTNNA1	P35221	P35221

Table 1.2.2. Sample of IID

GeneA	GeneB	UniProtA	UniProtB
SGK1	TUFM	O00141	P49411
SGK1	GANAB	O00141	Q14697
SGK1	MAP3K3	O00141	Q99759
SGK1	MAPT	O00141	P10636
SGK1	GAPDH	O00141	P04406

Table 1.4.1.1. GO Analysis - BP (seed genes)

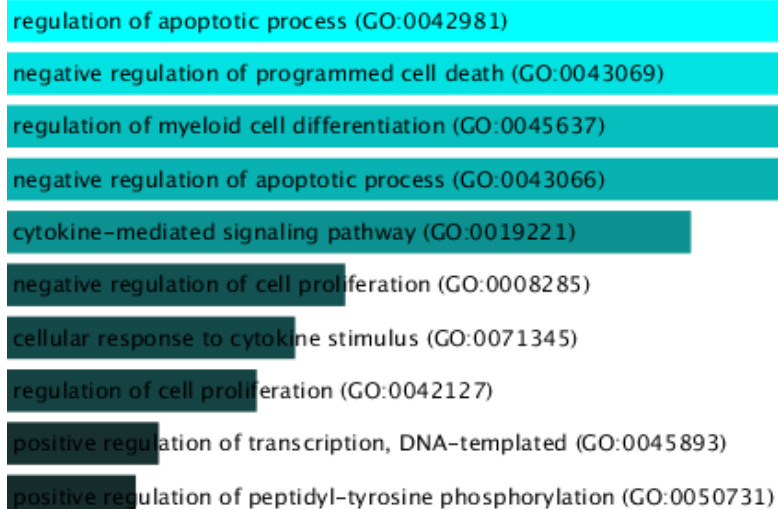| Term	Overlap	P-value	Adjusted P-value	Odds Ratio	Combined Score
regulation of apoptotic process (GO:0042981)	23/815	1.44E-13	7.33E-10	6.800206963	201.0957388
negative regulation of programmed cell death (GO:0043069)	17/408	7.84E-13	2.00E-09	10.04016064	279.8588695
regulation of myeloid cell differentiation (GO:0045637)	Sep-65	6.19E-12	1.05E-08	33.36422614	861.0859801
negative regulation of apoptotic process (GO:0043066)	17/485	1.23E-11	1.56E-08	8.446155757	212.2100292
cytokine-mediated signaling pathway (GO:0019221)	18/633	8.77E-11	8.95E-08	6.852052761	158.670495
negative regulation of cell proliferation (GO:0008285)	13/363	3.15E-09	2.68E-06	8.629559561	168.9301293
cellular response to cytokine stimulus (GO:0071345)	14/456	5.30E-09	3.86E-06	7.398013105	140.9786171
regulation of cell proliferation (GO:0042127)	17/740	7.89E-09	5.03E-06	5.535656138	103.285673
positive regulation of transcription, DNA-templated (GO:0045893)	20/1120	2.17E-08	1.23E-05	4.30292599	75.92216692
positive regulation of peptidyl-tyrosine phosphorylation (GO:0050731)	8/116	2.75E-08	1.41E-05	16.61819693	289.278629

Figure 1.4.1.1. GO Analysis - BP (seed genes)

Table 1.4.1.2. GO Analysis - MF (seed genes)

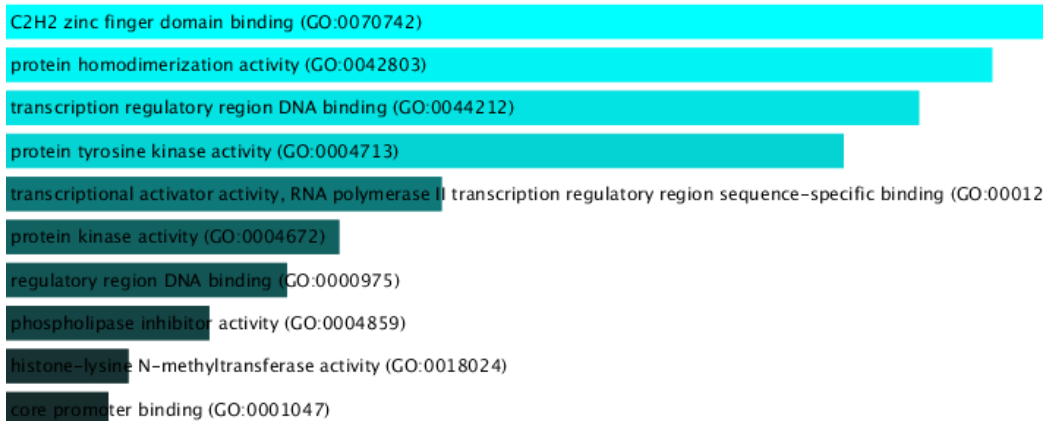| Term	Overlap	P-value	Adjusted P-value	Odds Ratio	Combined Score
C2H2 zinc finger domain binding (GO:0070742)	03-Dec	1.48E-05	0.016985818	60.24096386	670.1062012
protein homodimerization activity (GO:0042803)	12/664	1.82E-05	0.010479501	4.354768471	47.52608569
transcription regulatory region DNA binding (GO:0044212)	9/374	2.47E-05	0.009464721	5.798595451	61.52285485
protein tyrosine kinase activity (GO:0004713)	6/147	3.37E-05	0.009703809	9.835259405	101.2769
transcriptional activator activity, RNA polymerase II transcription regulatory region sequence-specific binding (GO:0001228)	7/284	0.000179	0.041146265	5.939249958	51.25317687
protein kinase activity (GO:0004672)	9/513	0.000273	0.052452193	4.22743606	34.68391489
regulatory region DNA binding (GO:0000975)	6/224	0.00034	0.055858119	6.454388985	51.553894
phospholipase inhibitor activity (GO:0004859)	02-Aug	0.000469	0.067444137	60.24096386	461.7711139
histone-lysine N-methyltransferase activity (GO:0018024)	Mar-41	0.000656	0.083847186	17.63150162	129.2375172
core promoter binding (GO:0001047)	Apr-97	0.000713	0.082019938	9.936653832	72.0068983

Figure 1.4.1.2. GO Analysis - MF (seed genes)

Table 1.4.1.3. GO Analysis - CC (seed genes)

Term	Overlap	P-value	Adjusted P-value	Odds Ratio	Combined Score
focal adhesion (GO:0005925)	15/356	1.73E-11	7.71E-09	10.15297144	251.599153
chromatin (GO:0000785)	9/296	3.80E-06	0.000847924	7.326603712	91.43522129
nuclear periphery (GO:0034399)	May-78	1.81E-05	0.002696647	15.44640099	168.6353523
nuclear matrix (GO:0016363)	Apr-59	0.000105479	0.011760878	16.33653257	149.5936501
membrane raft (GO:0045121)	5/119	0.000137408	0.012256765	10.12453174	90.03298888
nuclear chromatin (GO:0000790)	6/253	0.000646247	0.048037662	5.714557836	41.9695953
secretory granule lumen (GO:0034774)	6/317	0.002052377	0.130765716	4.560830071	28.22586804
platelet alpha granule membrane (GO:0031092)	Feb-17	0.002222324	0.123894553	28.34868887	173.187863
tertiary granule (GO:0070820)	4/164	0.004845662	0.240129492	5.877167205	31.32336952
nuclear chromosome part (GO:0044454)	6/392	0.005797451	0.258566335	3.688222277	18.99558715

Figure 1.4.1.3. GO Analysis - CC (seed genes)

Table 1.4.1.4. Pathway Analysis – KEGG 2019 (seed genes)

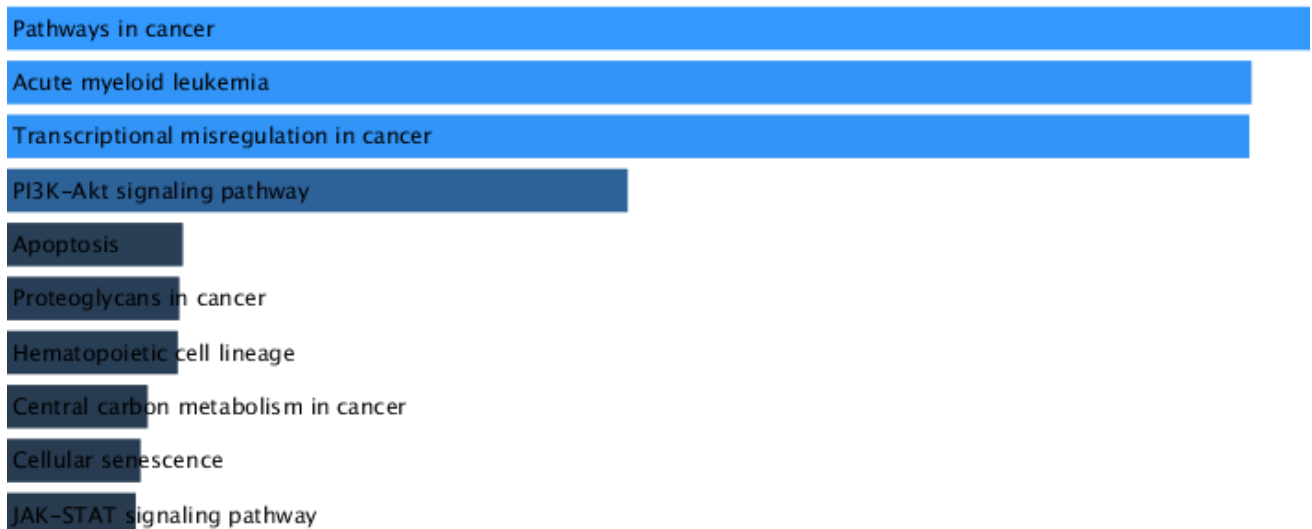| Term	Overlap	P-value	Adjusted P-value	Odds Ratio	Combined Score
Pathways in cancer	23/530	1.42E-17	4.36E-15	10.45692203	405.6821693
Acute myeloid leukemia	Dec-66	4.68E-17	7.21E-15	43.81161008	1647.333154
Transcriptional misregula-tion in cancer	16/186	4.89E-17	5.02E-15	20.72807358	778.4838544
PI3K-Akt signaling pathway	15/354	1.60E-11	1.23E-09	10.21033286	253.8349789
Apoptosis	8/143	1.41E-07	8.67E-06	13.48049541	212.668919
Proteoglycans in cancer	9/201	1.51E-07	7.76E-06	10.78942636	169.44368
Hematopoietic cell lineage	Jul-97	1.56E-07	6.85E-06	17.38914421	272.5729609
Central carbon metabolism in cancer	Jun-65	2.89E-07	1.11E-05	22.24281742	334.9216281
Cellular senescence	8/160	3.34E-07	1.14E-05	12.04819277	179.6707646
JAK-STAT signaling pathway	8/162	3.67E-07	1.13E-05	11.89944965	176.3229912

Figure 1.4.1.4. Pathway Analysis – KEGG 2019 (seed genes)

Table 1.4.2.1. GO Analysis - BP (union interactome genes)

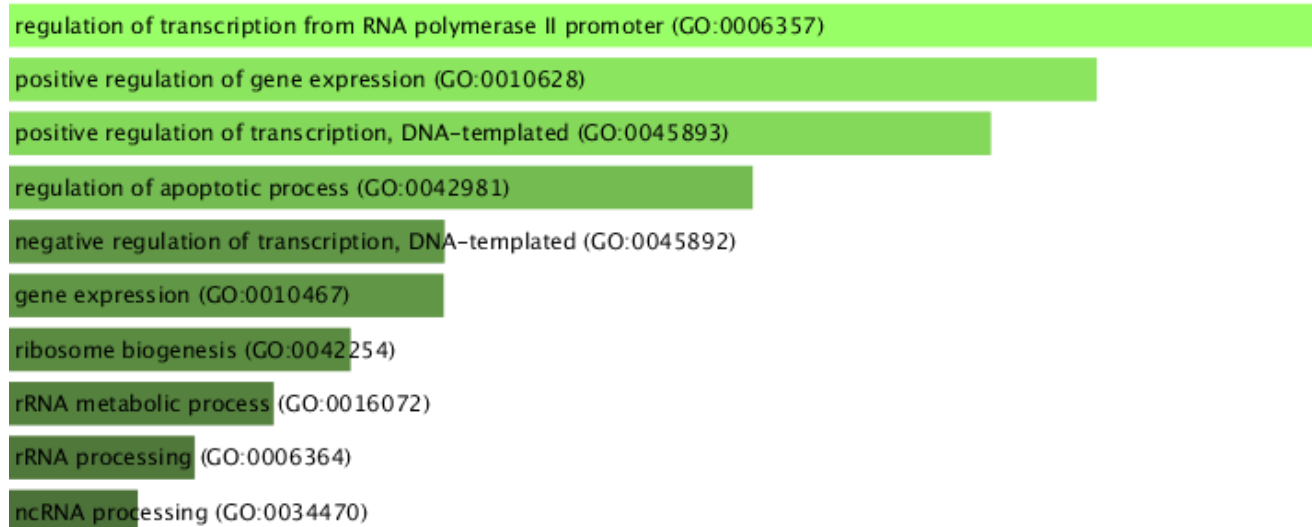| Term	Overlap	P-value	Adjusted P-value	Odds Ratio	Combined Score
regulation of transcription from RNA polymerase II promoter (GO:0006357)	812/1478	4.23E-72	2.16E-68	1.659541063	272.7337946
positive regulation of gene expression (GO:0010628)	488/771	1.77E-68	4.51E-65	1.911929401	298.2719494
positive regulation of transcription, DNA-templated (GO:0045893)	644/1120	1.07E-66	1.82E-63	1.73689775	263.8406433
regulation of apoptotic process (GO:0042981)	498/815	1.09E-62	1.38E-59	1.845772375	263.3521988
negative regulation of transcription, DNA-templated (GO:0045892)	487/813	1.67E-57	1.70E-54	1.809442653	236.5569461
gene expression (GO:0010467)	293/411	1.72E-57	1.46E-54	2.153437176	281.463617
ribosome biogenesis (GO:0042254)	189/226	6.36E-56	4.63E-53	2.52615371	321.062744
rRNA metabolic process (GO:0016072)	172/200	1.25E-54	7.96E-52	2.597794895	322.4320707
rRNA processing (GO:0006364)	172/202	2.67E-53	1.51E-50	2.572074153	311.3658049
ncRNA processing (GO:0034470)	186/227	2.41E-52	1.23E-49	2.475104244	294.1826411

Figure 1.4.2.1. GO Analysis - BP (union interactome genes)

Table 1.4.2.2. GO Analysis - MF (union interactome genes)

Term	Overlap	P-value	Adjusted P-value	Odds Ratio	Combined Score
RNA binding (GO:0003723)	919/1387	2.20E-151	2.53E-148	2.001453286	694.3084747
cadherin binding (GO:0045296)	249/313	7.64E-65	4.40E-62	2.403042309	354.7723169
protein kinase binding (GO:0019901)	333/495	1.02E-55	3.91E-53	2.032101715	257.3092946
protein kinase activity (GO:0004672)	329/513	5.60E-48	1.61E-45	1.937246748	210.7763624
kinase binding (GO:0019900)	277/418	1.62E-44	3.74E-42	2.001750267	201.8329566
transcription regulatory region DNA binding (GO:0044212)	247/374	2.08E-39	3.99E-37	1.994948822	177.6854212
ubiquitin-like protein ligase binding (GO:0044389)	203/297	7.95E-36	1.31E-33	2.064647889	166.8637429
protein serine/threonine kinase activity (GO:0004674)	237/368	3.21E-35	4.62E-33	1.945391147	154.5122235
transcription coactivator activity (GO:0003713)	199/291	3.43E-35	4.39E-33	2.065696412	163.9297457
ubiquitin protein ligase binding (GO:0031625)	191/284	2.57E-32	2.96E-30	2.031521556	147.770996

RNA binding (GO:0003723)
cadherin binding (GO:0045296)
protein kinase binding (GO:0019901)
protein kinase activity (GO:0004672)
kinase binding (GO:0019900)
transcription regulatory region DNA binding (GO:0044212)
ubiquitin–like protein ligase binding (GO:0044389)
protein serine/threonine kinase activity (GO:0004674)
transcription coactivator activity (GO:0003713)
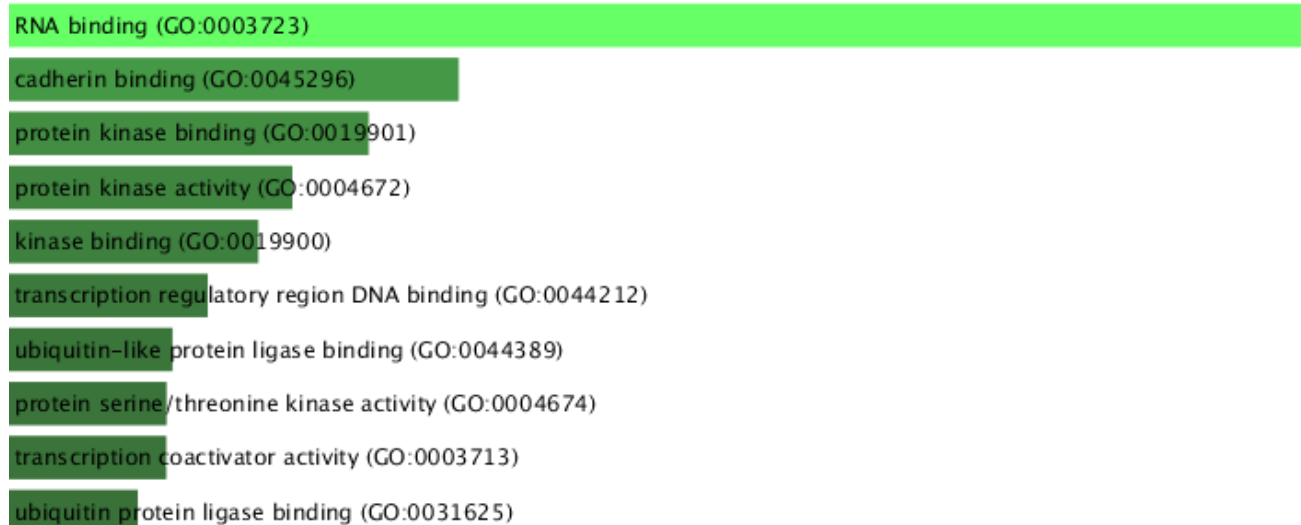ubiquitin protein ligase binding (GO:0031625)

Figure 1.4.2.2. GO Analysis - MF (union interactome genes)

Table 1.4.2.3. GO Analysis - CC (union interactome genes)

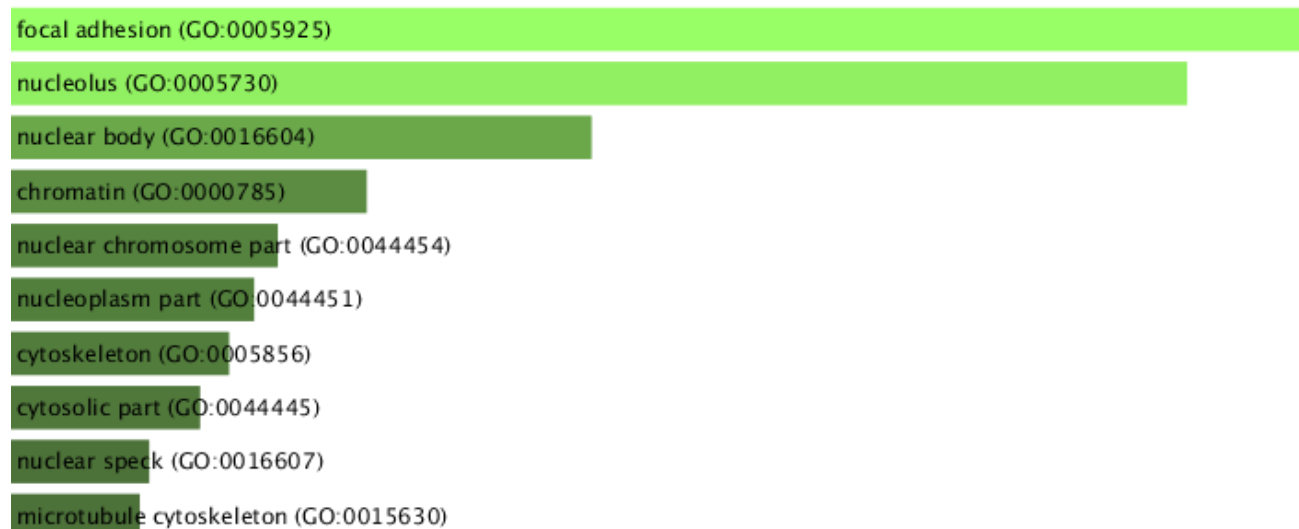| Term	Overlap	P-value	Adjusted P-value	Odds Ratio	Combined Score
focal adhesion (GO:0005925)	288/356	3.86E-78	1.72E-75	2.443705676	435.5908286
nucleolus (GO:0005730)	449/676	1.62E-72	3.61E-70	2.006347027	331.6582353
nuclear body (GO:0016604)	375/618	4.02E-46	5.98E-44	1.832944016	191.5921037
chromatin (GO:0000785)	203/296	3.72E-36	4.15E-34	2.07162305	169.000709
nuclear chromosome part (GO:0044454)	243/392	3.23E-32	2.88E-30	1.872520644	135.7759773
nucleoplasm part (GO:0044451)	248/407	3.61E-31	2.69E-29	1.840618062	129.0194665
cytoskeleton (GO:0005856)	297/520	4.71E-30	3.00E-28	1.725279704	116.5060193
cytosolic part (GO:0044445)	121/159	9.04E-29	5.04E-27	2.298765411	148.4378577
nuclear speck (GO:0016607)	187/296	1.62E-26	8.02E-25	1.908342416	113.3278281
microtubule cytoskeleton (GO:0015630)	229/388	4.21E-26	1.88E-24	1.782830949	104.1723179

Figure 1.4.2.3. GO Analysis - CC (union interactome genes)

Table 1.4.2.4. Pathway Analysis – KEGG 2019 (union interactome genes)

| Term	Overlap	P-value	Adjusted P-value	Odds Ratio	Combined Score
Pathways in cancer	332/530	4.04E-45	1.24E-42	1.8922069	193.4205436
Human T-cell leukemia virus 1 infection	163/219	3.17E-36	4.88E-34	2.248277413	183.7737783
Cellular senescence	126/160	1.57E-32	1.61E-30	2.378794744	174.201718
Apoptosis	116/143	2.70E-32	2.08E-30	2.45035134	178.1136803
Cell cycle	104/124	1.60E-31	9.84E-30	2.533483394	179.6539006
Proteoglycans in cancer	143/201	1.85E-28	9.50E-27	2.149049346	137.2323054
Epstein-Barr virus infection	142/201	9.27E-28	4.08E-26	2.134021029	132.8323524
Hepatitis B	121/163	4.77E-27	1.84E-25	2.24235399	135.9029857
Chronic myeloid leukemia	69/76	8.28E-26	2.83E-24	2.742470131	158.387624
Ubiquitin mediated proteolysis	105/137	1.39E-25	4.27E-24	2.315128705	132.5125902
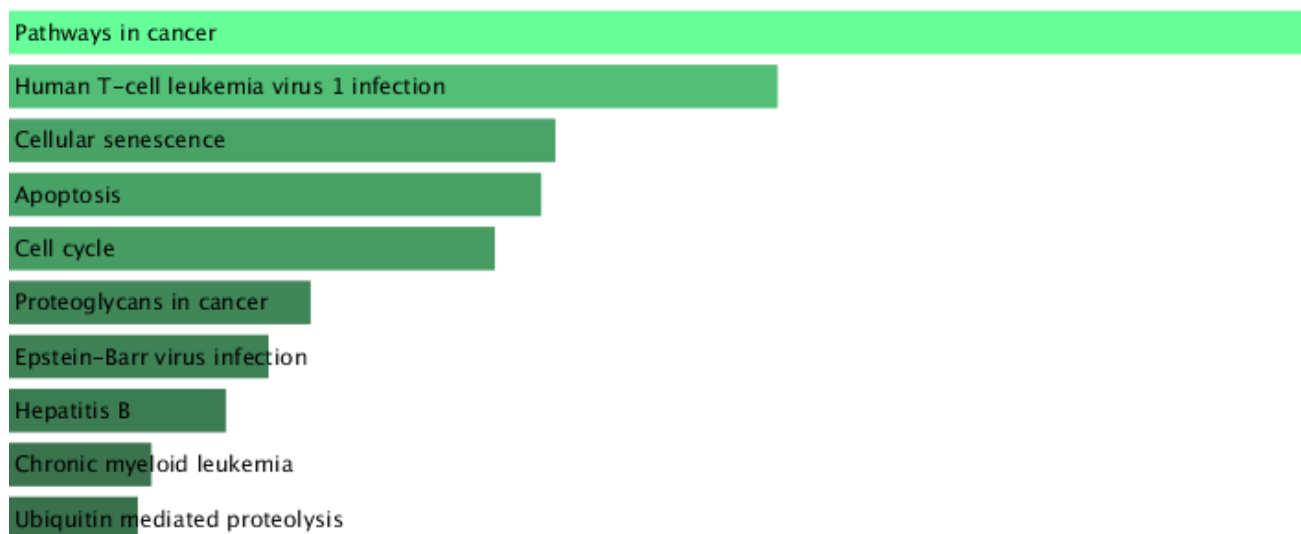
Figure 1.4.2.4. Pathway Analysis – KEGG 2019 (union interactome genes)

Table 2.1.1. Global measure of SGI, I, U, I-LCC, U-LCC

| interac-tome	nodes	self-loops	links	con-nected compo-nents	iso-lated nodes	aver-age path length	aver-age clus-tering coeffi-cient	aver-age de-gree	diam-eter	ra-dius	centrali-zation
SGI	83	60	256	14	12	3.872	0.125	3.566	9	1	0.00346
I	2677	38	3943	4	0	3.58	0.014	2.917	9	1	7.00E-05
U	6621	1638	203091	1	0	3.073	0.098	55.148	7	1	5.00E-05
I-LCC	2664	0	3895	1	0	3.58	0.014	2.923	9	1	7.00E-05
U-LCC	6621	1638	203091	1	0	3.073	0.098	55.148	7	1	5.00E-05
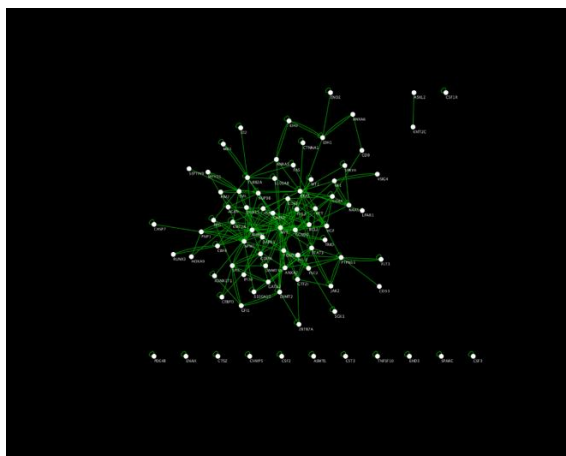
Figure 2.1. Graph of SGI
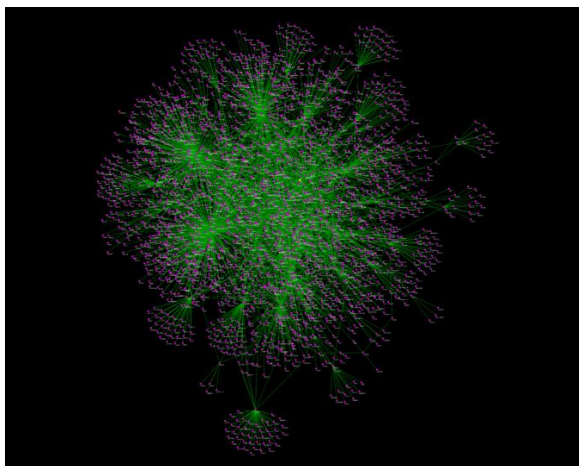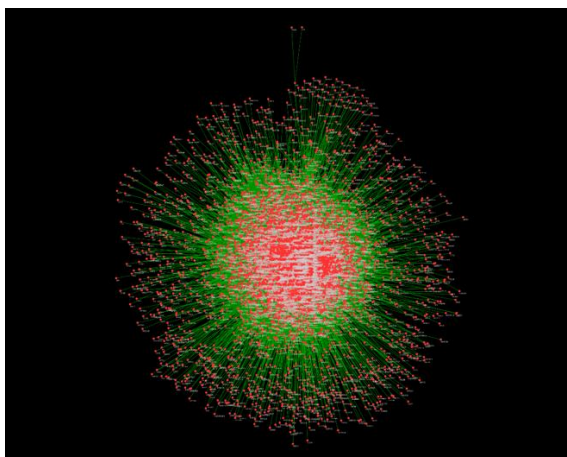

Figure 2.1. Graph of I-LCC

Figure 2.1. Graph of U-L

Table 2.1.2. Global measure of I-LCC

Rank	name	In de-gree	Out de-gree	Between-ness Cen-trality	Eigenvector Centrality	Closeness Centrality	ratio Between-ness/Nodedegree
1	STAT3	74	84	0.00642521	0.06063617	0.3408521	4.07E-05
2	NPM1	27	383	0.0048705	0.31375241	0.4408421	1.19E-05
3	MYC	16	508	0.00376588	0.59990604	0.4464785	7.19E-06
4	HSPB1	17	250	0.00288502	0.14126747	0.367162	1.08E-05
5	CEBPA	15	44	0.00236639	0.05110214	0.3540462	4.01E-05
6	CDK6	22	62	0.00226934	0.05041978	0.4052419	2.70E-05
7	PTPN11	71	40	0.0021788	0.0210828	0.5362319	1.96E-05
8	MAPK8	3	2	0.0019503	0.03056941	0.2802617	0.00039006
9	RUNX1	25	24	0.00176623	0.06256403	0.4378882	3.60E-05
10	FHL2	47	29	0.00169521	0.04454231	0.5257732	2.23E-05
11	PPP2CA	5	1	0.00126786	0.04326851	0.2893983	0.00021131
12	BCL2	15	43	0.00126334	0.02480463	0.2996214	2.18E-05
13	EHMT2	51	18	0.00116422	0.022917	0.875	1.69E-05
14	MET	9	23	0.00115482	0.01188002	0.3065611	3.61E-05
15	KMT2A	24	31	0.00103737	0.05839002	0.5480769	1.89E-05
16	TCEA2	16	27	0.00102971	0.00582536	1	2.39E-05
17	HSP90AA1	5	5	0.00102545	0.05068803	0.3203787	0.000102545
18	EP300	6	3	0.00093109	0.04712255	0.3786408	0.000103454
19	CD44	10	54	0.00083703	0.01955834	0.2981771	1.31E-05
20	RUNX1T1	14	27	0.00081405	0.01501295	0.5714286	1.99E-05

Table 2.1.2. Global measure of U-LCC

Rank	name	In degree	Out degree	Betweenness Centrality	Eigenvector Centrality	ClosenessCentrality	ratio Betweenness/Nodedegree
1	MYC	266	2190	0.07340315	0.13962367	0.568414	2.99E-05
2	APP	830	200	0.04558901	0.05041435	0.43362	4.43E-05
3	KRAS	60	1547	0.02627012	0.06796302	0.524812	1.63E-05
4	TP53	453	635	0.02446928	0.07439483	0.489375	2.25E-05
5	NPM1	280	629	0.01916379	0.07641053	0.465607	2.11E-05
6	HSPB1	209	436	0.01656969	0.03856582	0.454877	2.57E-05
7	ELAVL1	123	828	0.01600121	0.06004126	0.502967	1.68E-05
8	STAT3	250	255	0.01523638	0.02140088	0.440136	3.02E-05
9	TRIM25	92	1199	0.0148058	0.09051068	0.518984	1.15E-05
10	PTPN11	221	173	0.01430834	0.01553258	0.419621	3.63E-05
11	HSP90AA1	326	395	0.014196	0.05975502	0.466975	1.97E-05
12	UBC	305	510	0.01283605	0.06339625	0.461952	1.57E-05
13	BRCA1	190	777	0.01246467	0.08355762	0.49769	1.29E-05
14	EGFR	151	732	0.01208748	0.05878169	0.495363	1.37E-05
15	VCP	238	459	0.01130163	0.06268199	0.474505	1.62E-05
16	CTNNB1	244	414	0.01082945	0.06313504	0.485673	1.65E-05
17	XPO1	165	672	0.01045885	0.06040625	0.46984	1.25E-05
18	HNRNPL	109	593	0.00942863	0.04736486	0.457562	1.34E-05
19	LPAR1	97	144	0.00942781	0.00171766	0.335533	3.91E-05
20	FLT3	20	137	0.00919752	0.00355299	0.373333	5.86E-05

Table 2.2. Putative diseases modules with MCL algorithm

interactome	algorithm	module ID	no. of seed genes	total no. of genes	ratio seed/total
I-LCC	Marcov	1	54	904	0.05973
I-LCC	Marcov	4	3	36	0.08333
I-LCC	Marcov	6	2	22	0.09091
I-LCC	Marcov	13	1	10	0.1
I-LCC	Marcov	14	1	10	0.1
I-LCC	Marcov	15	1	10	0.1
U-LCC	Marcov	15	2	35	0.05714
U-LCC	Marcov	24	2	27	0.07407

Table 2.3.1.1. GO Analysis - BP (I-LCC)

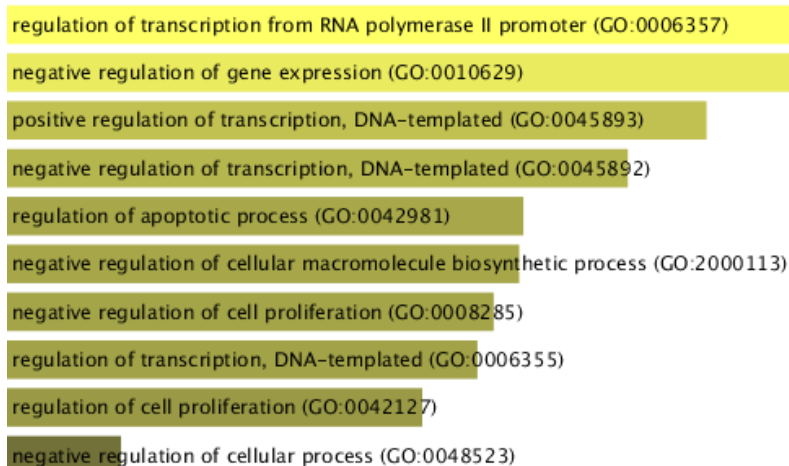| Term	Overlap	P-value	Adjusted P-value	Odds Ratio	Combined Score
regulation of transcription from RNA polymerase II promoter (GO:0006357)	161/1478	1.77E-26	9.03E-23	2.409977607	142.9021733
negative regulation of gene expression (GO:0010629)	94/618	1.30E-25	3.32E-22	3.365123006	192.8276118
positive regulation of transcription, DNA-templated (GO:0045893)	130/1120	8.99E-24	1.53E-20	2.56795196	136.2698904
negative regulation of transcription, DNA-templated (GO:0045892)	106/813	2.65E-23	3.38E-20	2.88454212	149.9518865
regulation of apoptotic process (GO:0042981)	105/815	1.11E-22	1.14E-19	2.850317607	144.0848532
negative regulation of cellular macromolecule biosynthetic process (GO:2000113)	80/512	1.18E-22	1.00E-19	3.456858407	174.5437989
negative regulation of cell proliferation (GO:0008285)	66/363	1.67E-22	1.22E-19	4.022526146	201.6976769
regulation of transcription, DNA-templated (GO:0006355)	160/1598	2.09E-22	1.33E-19	2.215158328	110.5786581
regulation of cell proliferation (GO:0042127)	98/740	4.46E-22	2.53E-19	2.929921072	144.0415394
negative regulation of cellular process (GO:0048523)	78/534	2.79E-20	1.43E-17	3.231579994	145.4989685

Figure 2.3.1.1. GO Analysis - BP (I-LCC)

Table 2.3.1.2. GO Analysis - MF (I-LCC)

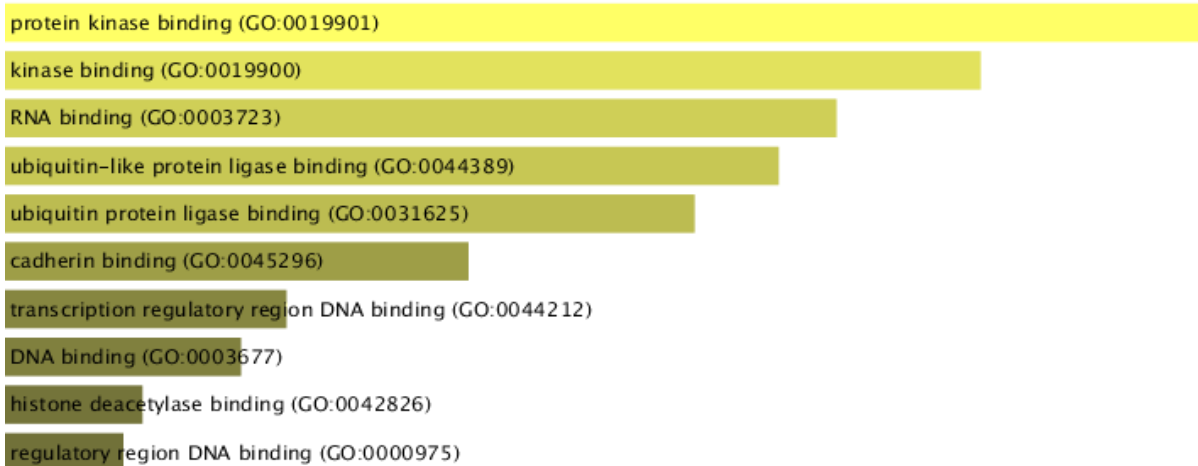| Term	Overlap	P-value	Adjusted P-value	Odds Ratio	Combined Score
protein kinase bind-ing (GO:0019901)	85/495	6.32E-27	7.28E-24	3.799052472	229.1796111
kinase binding (GO:0019900)	73/418	1.09E-23	6.26E-21	3.86374222	204.2959746
RNA binding (GO:0003723)	144/1387	1.35E-21	5.20E-19	2.296929133	110.3688191
ubiquitin-like protein ligase binding (GO:0044389)	57/297	9.40E-21	2.70E-18	4.245999821	195.8003225
ubiquitin protein lig-ase binding (GO:0031625)	54/284	1.57E-19	3.62E-17	4.206655864	182.1332419
cadherin binding (GO:0045296)	52/313	3.06E-16	5.87E-14	3.675535073	131.3025078
transcription regula-tory region DNA binding (GO:0044212)	53/374	1.36E-13	2.24E-11	3.135204202	92.88163263
DNA binding (GO:0003677)	90/893	6.25E-13	8.99E-11	2.229731738	62.65841454
histone deacetylase binding (GO:0042826)	22/85	1.69E-11	2.17E-09	5.726184279	142.0168834
regulatory region DNA binding (GO:0000975)	36/224	3.20E-11	3.69E-09	3.55562579	85.91724084

Figure 2.3.1.2. GO Analysis - MF (I-LCC)

Table 2.3.1.3. GO Analysis - CC (I-LCC)

Term	Overlap	P-value	Adjusted P-value	Odds Ratio	Combined Score
focal adhesion (GO:0005925)	64/356	1.40E-21	6.24E-19	3.977329223	190.9854602
microtubule (GO:0005874)	35/210	2.11E-11	4.70E-09	3.687315634	90.64578286
nucleolus (GO:0005730)	70/676	8.15E-11	1.21E-08	2.290935749	53.21962076
nuclear body (GO:0016604)	65/618	2.07E-10	2.31E-08	2.326946759	51.88467578
chromatin (GO:0000785)	40/296	5.96E-10	5.32E-08	2.989715379	63.50277064
nuclear chromosome part (GO:0044454)	47/392	1.11E-09	8.23E-08	2.652609716	54.69969064
secretory granule lumen (GO:0034774)	41/317	1.38E-09	8.81E-08	2.861449987	58.37232791
cytoskeleton (GO:0005856)	56/520	1.63E-09	9.09E-08	2.382573179	48.20976893
microtubule cytoskeleton (GO:0015630)	46/388	2.40E-09	1.19E-07	2.622935864	52.06287474
polymeric cytoskeletal fiber (GO:0099513)	32/221	5.73E-09	2.56E-07	3.203459737	60.79345459

Figure 2.3.1.3. GO Analysis - CC (I-LCC)

Table 2.3.1.4. Pathway Analysis – KEGG 2019 (I-LCC)

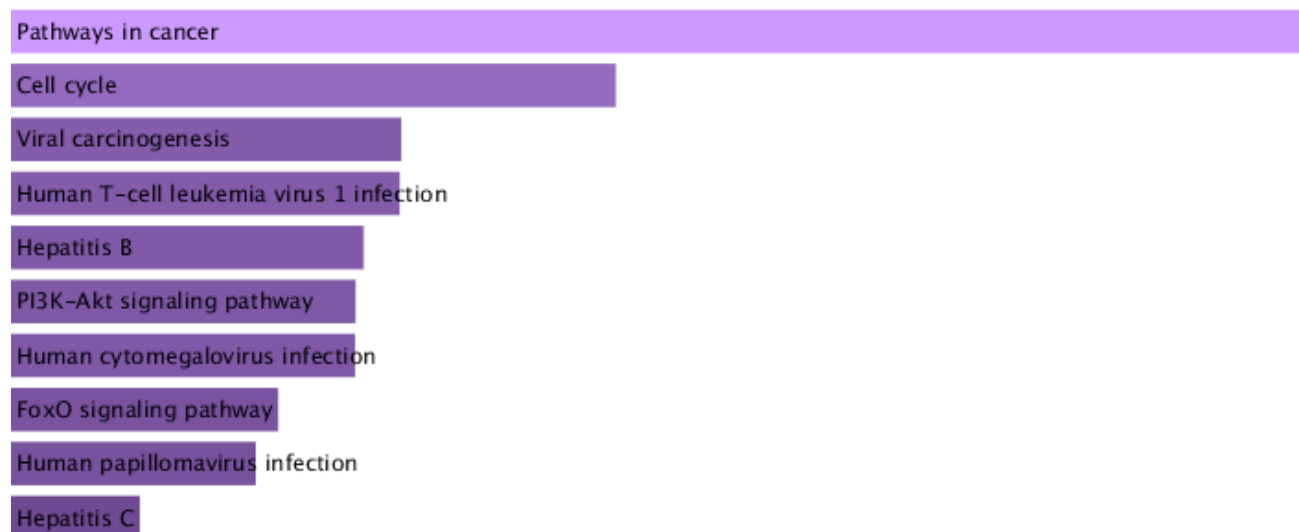| Term	Overlap	P-value	Adjusted P-value	Odds Ratio	Combined Score
Pathways in cancer	94/530	7.48E-31	2.30E-28	3.923860411	272.1882341
Cell cycle	39/124	1.07E-22	1.64E-20	6.958321439	352.0326053
Viral carcinogenesis	46/201	3.40E-20	3.49E-18	5.063179677	226.9717032
Human T-cell leukemia virus 1 infection	48/219	3.56E-20	2.74E-18	4.849072615	217.1554085
Hepatitis B	41/163	9.40E-20	5.79E-18	5.564905804	243.8041918
PI3K-Akt signaling pathway	61/354	1.16E-19	5.98E-18	3.812309385	166.2045677
Human cytomegalovirus infection	48/225	1.18E-19	5.19E-18	4.719764012	205.7041836
FoxO signaling pathway	36/132	9.33E-19	3.59E-17	6.03378922	250.4975517
Human papillomavirus infection	57/330	1.70E-18	5.82E-17	3.821399839	156.353153
Hepatitis C	37/155	3.84E-17	1.18E-15	5.281187554	199.6183108

Figure 2.3.1.4. Pathway Analysis – KEGG 2019 (I-LCC)

Table 2.3.2.1. GO Analysis - BP (U-LCC)

Term	Overlap	P-value	Adjusted P-value	Odds Ratio	Combined Score
hexose biosynthetic process (GO:0019319)	Jun-44	1.23E-10	6.28E-07	77.92207792	1778.084752
gluconeogenesis (GO:0006094)	May-41	8.72E-09	2.23E-05	69.68641115	1293.18349
glucose metabolic process (GO:0006006)	May-64	8.62E-08	0.000146702	44.64285714	726.1641222
glycolytic process through glucose-6-phosphate (GO:0061620)	Apr-25	9.68E-08	0.000123511	91.42857143	1476.614347
canonical glycolysis (GO:0061621)	Apr-25	9.68E-08	9.88E-05	91.42857143	1476.614347
glucose catabolic process to pyruvate (GO:0061718)	Apr-25	9.68E-08	8.23E-05	91.42857143	1476.614347
2-oxoglutarate metabolic process (GO:0006103)	Mar-17	3.28E-06	0.002392983	100.8403361	1273.299674
glycolytic process (GO:0006096)	Mar-23	8.49E-06	0.005414206	74.53416149	870.3258117
ATP generation from ADP (GO:0006757)	Mar-24	9.69E-06	0.005493558	71.42857143	824.6098854
carbohydrate catabolic process (GO:0016052)	Mar-35	3.09E-05	0.015778643	48.97959184	508.608821

Figure 2.3.2.1. GO Analysis - BP (U-LCC)

Table 2.3.2.2. GO Analysis - MF (U-LCC)

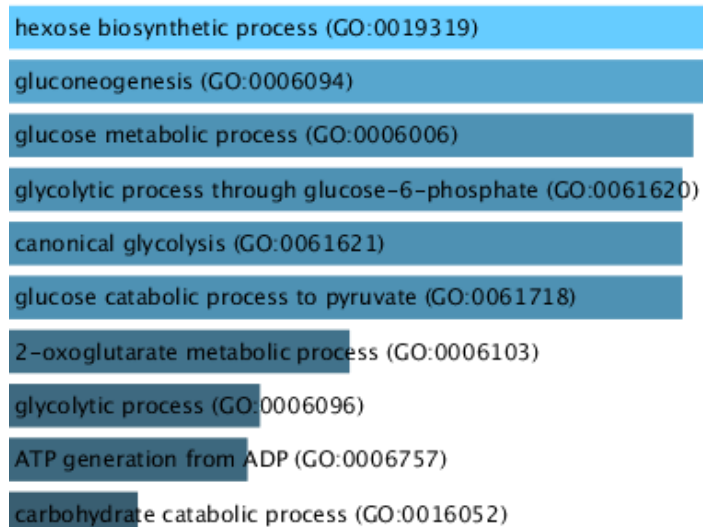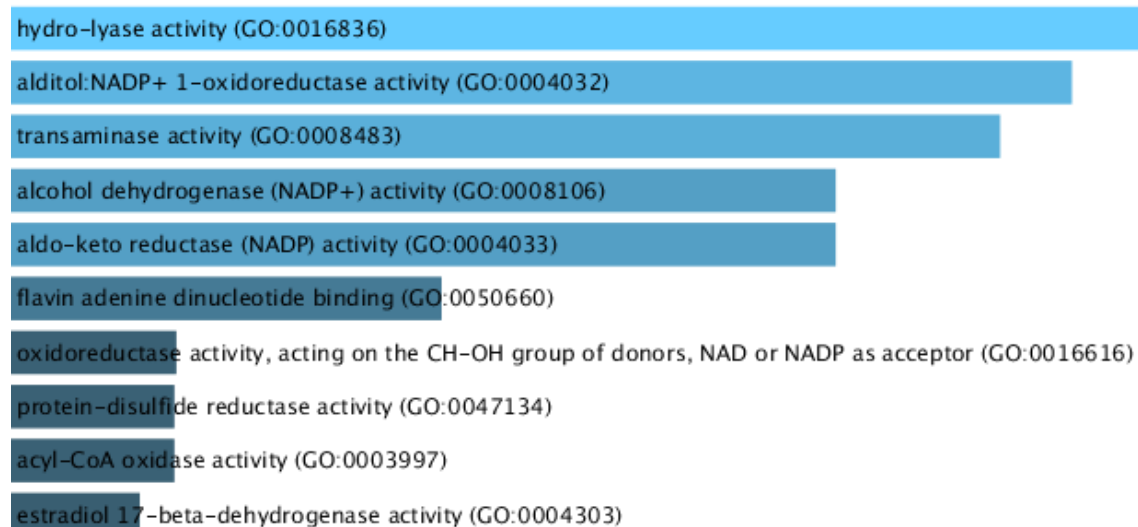| Term	Overlap	P-value	Adjusted P-value	Odds Ratio	Combined Score
hydro-lyase activity (GO:0016836)	Mar-45	6.62E-05	0.07624082	38.0952381	366.5617019
alditol:NADP+ 1-oxidore-ductase activity (GO:0004032)	02-Dec	0.000194204	0.111764655	95.23809524	813.9618229
transaminase activity (GO:0008483)	Feb-14	0.000267179	0.102507752	81.63265306	671.6400797
alcohol dehydrogenase (NADP+) activity (GO:0008106)	Feb-20	0.000554183	0.159466141	57.14285714	428.4580405
aldo-keto reductase (NADP) activity (GO:0004033)	Feb-20	0.000554183	0.127572913	57.14285714	428.4580405
flavin adenine dinucleotide binding (GO:0050660)	Feb-48	0.00319059	0.612061455	23.80952381	136.8464172
oxidoreductase activity, acting on the CH-OH group of donors, NAD or NADP as acceptor (GO:0016616)	Feb-88	0.01036436	1	12.98701299	59.34262643
protein-disulfide reductase activity (GO:0047134)	01-Jun	0.010455365	1	95.23809524	434.3466677
acyl-CoA oxidase activity (GO:0003997)	01-Jun	0.010455365	1	95.23809524	434.3466677
estradiol 17-beta-dehydro-genase activity (GO:0004303)	01-Jul	0.012187578	1	81.63265306	359.7826969

Figure 2.3.2.2. GO Analysis - MF (U-LCC)

Table 2.3.2.3. GO Analysis - CC (U-LCC)

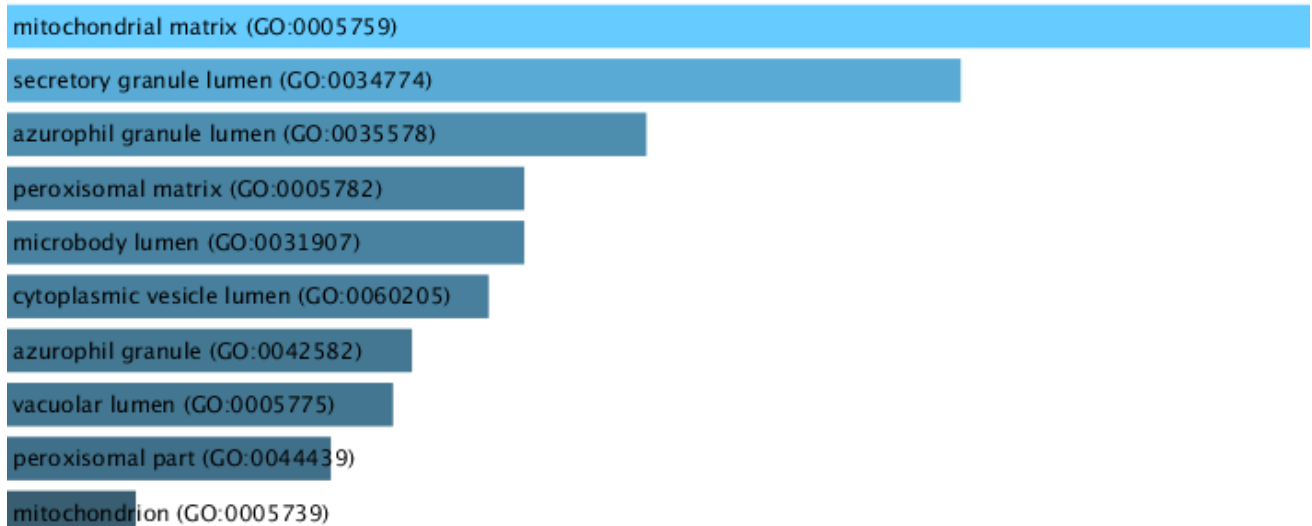| Term	Overlap	P-value	Adjusted P-value	Odds Ratio	Combined Score
mitochondrial matrix (GO:0005759)	8/308	4.74E-08	2.11E-05	14.84230056	250.321719
secretory granule lumen (GO:0034774)	7/317	1.08E-06	0.000240941	12.61829653	173.3517901
azurophil granule lumen (GO:0035578)	Apr-90	1.80E-05	0.00268227	25.3968254	277.4044049
peroxisomal matrix (GO:0005782)	Mar-42	5.38E-05	0.005996638	40.81632653	401.248198
microbody lumen (GO:0031907)	Mar-42	5.38E-05	0.004797311	40.81632653	401.248198
cytoplasmic vesicle lumen (GO:0060205)	4/129	7.41E-05	0.005505832	17.71871539	168.5139656
azurophil granule (GO:0042582)	4/154	0.000146976	0.0093645	14.84230056	130.986837
vacuolar lumen (GO:0005775)	4/161	0.000174361	0.009720635	14.19698314	122.8661117
peroxisomal part (GO:0044439)	Mar-75	0.000304081	0.015068925	22.85714286	185.1020574
mitochondrion (GO:0005739)	7/1026	0.001741843	0.077686184	3.898635478	24.76729709

Figure 2.3.2.3. GO Analysis - CC (U-LCC)

Table 2.3.2.4. Pathway Analysis – KEGG 2019 (U-LCC)

Term	Overlap	P-value	Adjusted P-value	Odds Ratio	Combined Score
Glycolysis / Gluconeogenesis	Apr-68	5.91E-06	0.001820072	33.61344538	404.6715577
HIF-1 signaling pathway	4/100	2.73E-05	0.004211393	22.85714286	240.1580462
Glutathione metabolism	Mar-56	0.000127703	0.013110796	30.6122449	274.4634731
Butanoate metabolism	Feb-28	0.001092892	0.084152699	40.81632653	278.3235801
Citrate cycle (TCA cycle)	Feb-30	0.001254936	0.077304069	38.0952381	254.5017355
Glyoxylate and dicarboxylate metabolism	Feb-30	0.001254936	0.064420057	38.0952381	254.5017355
Glycine, serine and threonine metabolism	Feb-40	0.002225686	0.097930165	28.57142857	174.5054368
Tryptophan metabolism	Feb-42	0.002451434	0.094380218	27.21088435	163.5668577
Fatty acid degradation	Feb-44	0.002687548	0.091973881	25.97402597	153.7435286
Valine, leucine and isoleucine degradation	Feb-48	0.00319059	0.098270162	23.80952381	136.8464172
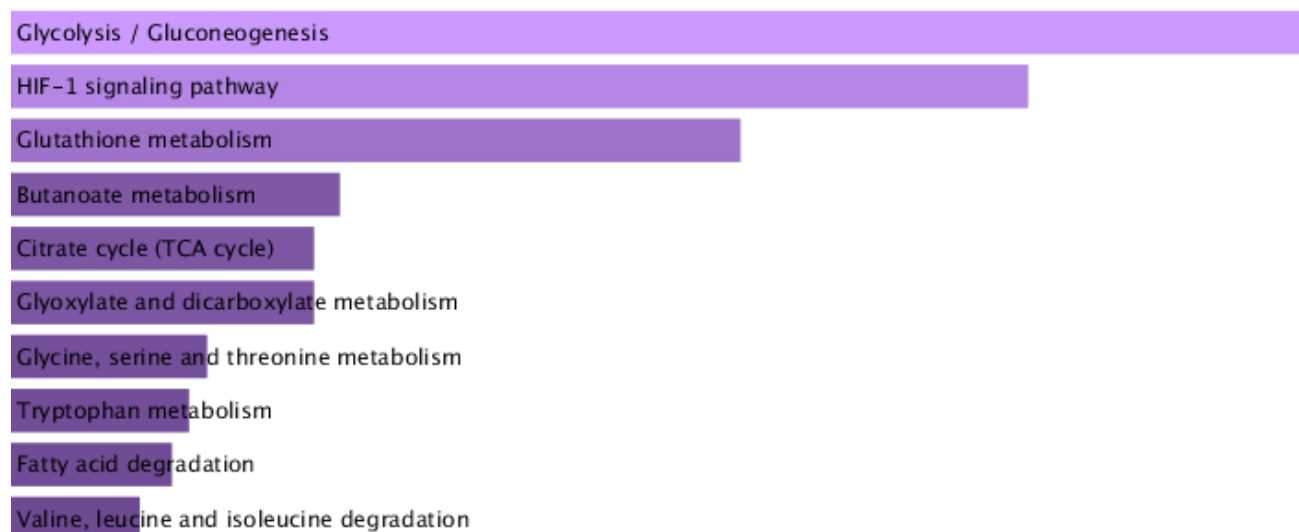
Figure 2.3.2.4. Pathway Analysis – KEGG 2019 (U-LCC)

Table. 2.4. 30 genes identified by the DIAMOnD

Rank	DIAMOnD_node	Gene
1	5925	RB1
2	5894	RAF1
3	5371	PML
4	84662	GLIS2
5	1969	EPHA2
6	405	ARNT
7	1030	CDKN2B
8	5607	MAP2K5
9	2264	FGFR4
10	1019	CDK4
11	5156	PDGFRA
12	5606	MAP2K3
13	2064	ERBB2
14	6794	STK11
15	2911	GRM1
16	1031	CDKN2C
17	11014	KDELR2
18	4771	NF2
19	26524	LATS2
20	8678	BECN1
21	207	AKT1
22	7015	TERT
23	11186	RASSF1
24	51343	FZR1
25	8463	TEAD2
26	898	CCNE1
27	1029	CDKN2A
28	369	ARAF
29	4217	MAP3K5
30	23624	CBLC

transcription regulatory region DNA binding (GO:0044212)

transcription coactivator activity (GO:0003713)

RNA polymerase II regulatory region sequence-specific DNA binding (GO:0000977)

DNA binding (GO:0003677)

regulatory region DNA binding (GO:0000975)

transcription regulatory region sequence-specific DNA binding (GO:0000976)

core promoter proximal region sequence-specific DNA binding (GO:0000987)

RNA polymerase II core promoter proximal region sequence-specific DNA binding (GO:0000978)

activating transcription factor binding (GO:0033613)

transcription factor activity, RNA polymerase II core promoter proximal region sequence-specific binding (GO:0000982)

Figure 2.4.1. GO Analysis - BP (genes from DIAMOnD)

nuclear chromosome part (GO:0044454)

nuclear chromatin (GO:0000790)

chromatin (GO:0000785)

RNA polymerase II transcription factor complex (GO:0090575)

nuclear body (GO:0016604)

nuclear transcription factor complex (GO:0044798)

PML body (GO:0016605)

nucleoplasm part (GO:0044451)

chromosome, telomeric region (GO:0000781)

SWI/SNF complex (GO:0016514)
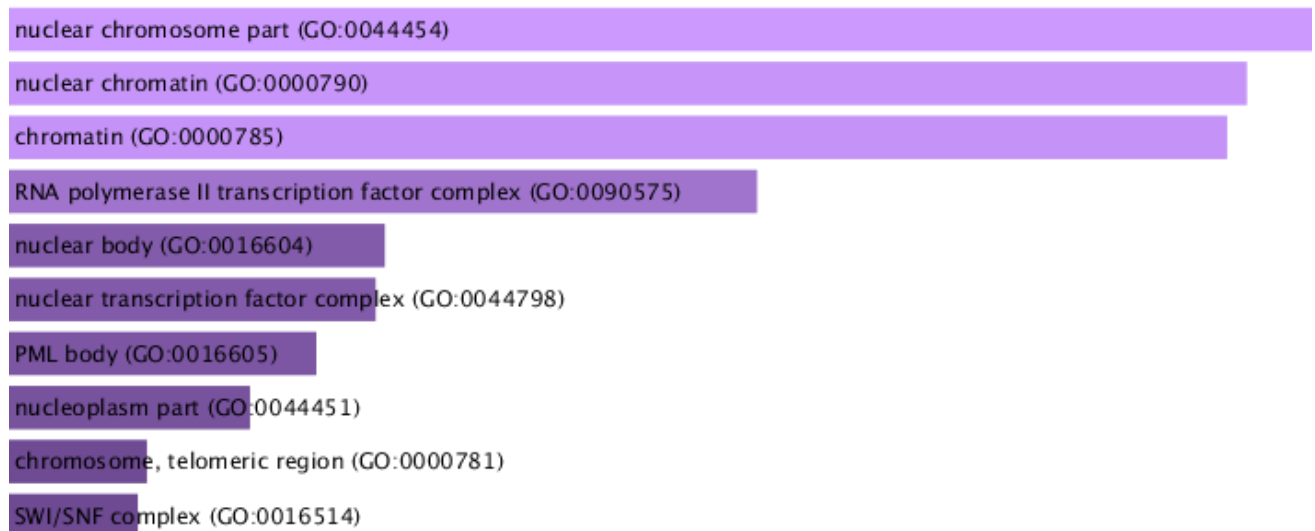
Figure 2.4.2. GO Analysis - MF (genes from DIAMOnD)

positive regulation of transcription, DNA-templated (GO:0045893)

regulation of transcription from RNA polymerase II promoter (GO:0006357)

negative regulation of transcription, DNA-templated (GO:0045892)

positive regulation of transcription from RNA polymerase II promoter (GO:0045944)

regulation of transcription, DNA-templated (GO:0006355)

positive regulation of gene expression (GO:0010628)

negative regulation of gene expression (GO:0010629)

positive regulation of nucleic acid-templated transcription (GO:1903508)

negative regulation of transcription from RNA polymerase II promoter (GO:0000122)

negative regulation of cellular macromolecule biosynthetic process (GO:2000113)

Figure 2.4.3. GO Analysis - CC (genes from DIAMOnD)

Pathways in cancer

Cell cycle

Viral carcinogenesis

Cellular senescence

Human T-cell leukemia virus 1 infection

Prostate cancer

Hepatitis B

Thyroid hormone signaling pathway
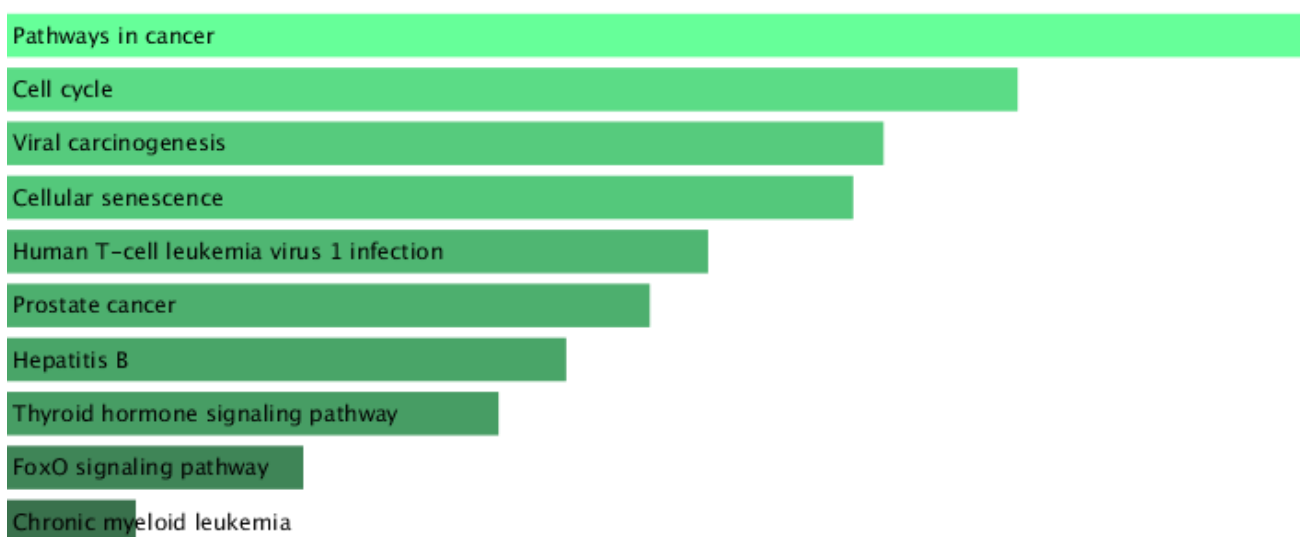
FoxO signaling pathway

Chronic myeloid leukemia

Figure 2.4.4. Pathway Analysis – KEGG 2019 (genes from DIAMOnD)