

# Pathway Analysis

# Pathway analysis – Scheme (20180828)

**Table S2**  
Zheng et al.,  
2017

List of signature  
genes in each  
cluster  
(C4-CD8-LAYN)

## Extract EntrezID (274 genes)

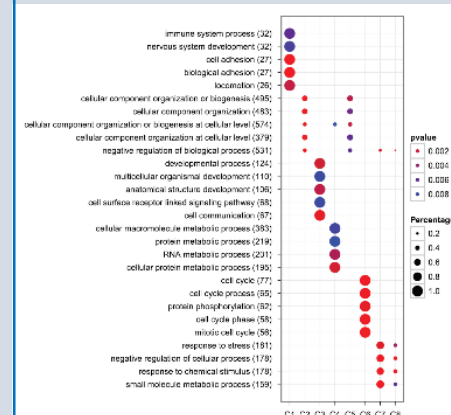
```
> head(test)
[1] 5996 84868 5133 1493 10663 201633
> test
      5996 84868 5133 1493 10663 201633 3604 22822
[1] 5996 84868 5133 1493 10663 201633 3604 22822
[9] 4647 6348 939 678655 1846 143903 3122
[17] 4643 5552 9495 84632 9760 1435 23406
[25] 1130 1757 2115 100287534 728558 3458 3805 55423
[33] 114614 9882 9452 10563 64759 116449 339789 79895
[41] 1390 64092 952 25861 3399 29851 51676 84969
[49] 406947 5740 3108 5997 2597 7412 8302 64231
[57] 60489 4345 639 116841 3560 7453 3682 54
[65] 3398 7133 3902 9901 26191 8638 7462 84159
[73] 27333 160365 27115 7292 2530 2731 51429 54602
[81] 55501 1509 285025 27350 4664 92906 3732 100528031
[89] 7128 5251 54438 5066 2280 677 3892 81
[97] 868 4046 3166 1890 80727 100616438 25897 200316
[105] 3516 5763 3162 10018 924 7227 5873 64333
[113] 64218 79890 4689 164118 94240 7188 6310 10447
[121] 22876 6503 3428 894 10797 29988 8611 160851
[129] 79413 4739 5778 22806 970 967 2820 130589
[137] 5900 23231 27334 8832 140564 83962 79713 84959
[145] 7167 2537 8565 23062 10538 9260 9619 57162
[153] 57205 100506190 10950 54806 9246 26031 5445 9585
[161] 684 9746 3987 5781 2650 8536 4061 100130093
[169] 54625 57157 23710 4522 403 54434 7979 5583
[177] 11315 3984 65094 5359 100506755 10791 7037 10906
[185] 10421 1495 115361 5496 3430 7867 7633 6709
[193] 23741 80219 134429 836 23517 9766 8027 57414
[201] 5315 2762 6774 79718 9997 25824 10589 2590
[209] 808 55320 81671 5573 4940 9144 5721 50615
[217] 4599 308 9051 55500 10982 10459 2969 10725
[225] 3660 2182 10379 2633 57835 29110 5606 405
[233] 1499 399 219285 9158 5214 2625 2634 11243
[241] 5993 29959 54790 55601 11262 55072 100527963 4170
[249] 79415 57819 9446 9049 5686 3939 2130 10956
[257] 10095 847 3385 55608 9205 830 8775 9784
[265] 11164 6717 132160 1508 10241 282991 80344 4678
[273] 1460 58496
```

## Over-representation test (e.g. 250 genes – depending on geneset)

```
# over-representation test
#
#...@organism Homo sapiens
#...@ontology DisGeNET
#...@keytype ENTREZID
#...@gene chr [1:274] "5996" "84868" "5133" "1493" "10663" "201633" "3604" "22822" "4647"
#...@p.adjusted "p.adjusted by 'BH' with cutoff <0.05"
#...@171 enriched terms found
'data.frame': 171 obs. of 9 variables:
 $ ID : chr "umls:c0021400" "umls:c0024138" "umls:c0026896" "umls:c0409974" ...
 $ Description: chr "Influenza" "Lupus Erythematosus, Discoid" "Myasthenia Gravis" "Lupus Erythe
 matorius" ...
 $ GeneRatio : chr "28/250" "21/250" "16/250" "22/250" ...
 $ BgRatio : chr "494/17381" "316/17381" "181/17381" "356/17381" ...
 $ pvalue : num 6.47e-10 6.39e-09 8.06e-09 1.04e-08 2.08e-08 ...
 $ p.adjust : num 1.40e-06 5.62e-06 5.62e-06 5.62e-06 8.99e-06 ...
 $ qvalue : num 1.07e-06 4.31e-06 4.31e-06 4.31e-06 6.90e-06 ...
 $ geneID : chr "1493/3604/6348/939/1757/3458/1390/2597/3682/8638/7292/25897/3162/2537/9246/
 5781/4061/5359/3430/836/9997/4940/45" | ____truncated____ "5133/1493/939/1435/3458/3805/10563/1390/298
 51/639/7453/7133/26191/356/3428/970/84959/10950/6774/50615/2625" "5133/1493/3122/1435/3458/10563/
 29851/406947/3560/26191/7128/5066/25897/9260/7979/836" "5133/1493/939/1435/3458/3805/10563/1390/9
 52/29851/3108/639/7133/26191/356/3428/970/84959/10950/6774/50615/2625" ...
 $ count : int 28 21 16 22 20 21 24 13 24 10 ...
#...@citation
Guangchuan Yu, Li-Gen Wang, Yanyan Han and Qing-Yu He.
clusterProfiler: an R package for comparing biological themes among
gene clusters. OMICS: A Journal of Integrative Biology
2012, 16(5):284-287
```

DOSE – DisGeNET, DO  
clusterProfiler – GO, KEGG

## Visualization



enrichPlot  
ggplot2

Limitation of the whole approach: ALL gene set analysis is knowledge-based. Novel genes and known genes with novel function will not be considered in the pathway analysis

# Pathway analysis: over-representation analysis (ORA)

## Over-representation analysis (ORA) - completed

- DisGenNET – Disease
- DO ontology - Disease
- KEGG - Pathway
- GO ontology – Cellular Component, Molecular Function, Biological Process

Limitations: ORA treats each gene equally! (No fold change or significance are considered for representation)

## Analysis requires fold-changes or p-values as input (FCS)

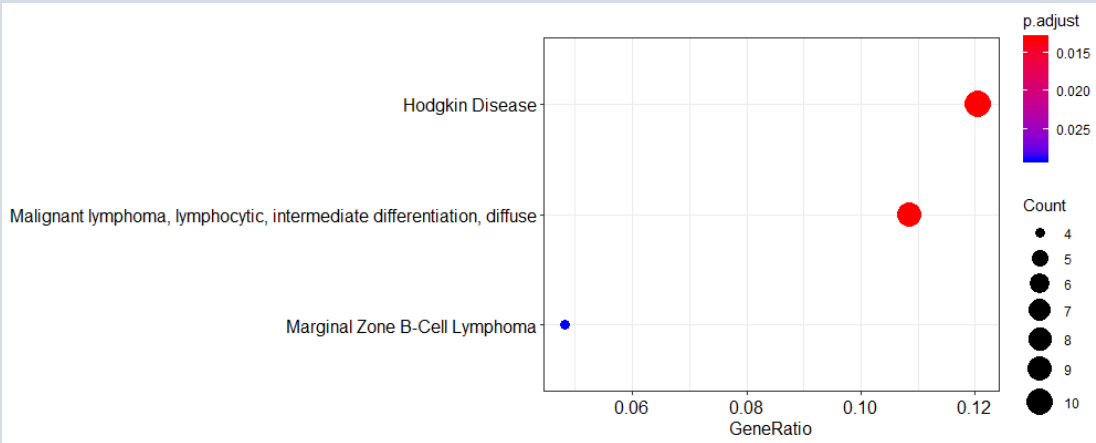
- **GSEA** and others

Comment: It does not require an arbitrary threshold for dividing expression data into significant and non-significant pool, but require whole gene expression dataset

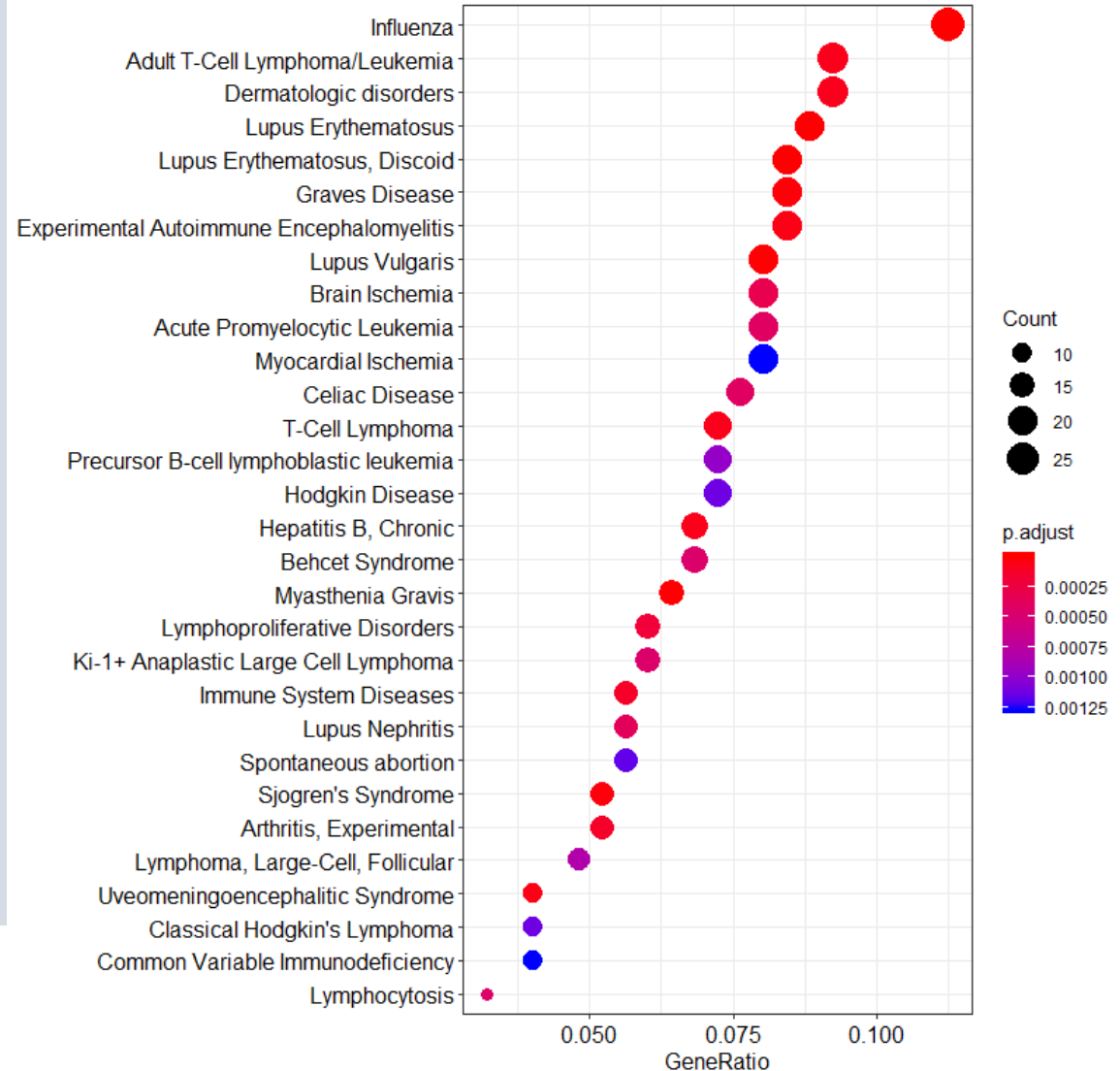
# Analysis: Over-representation test

## DisGeNET Ontology

### C1-CD8-LEF1 (Input: 101 genes)

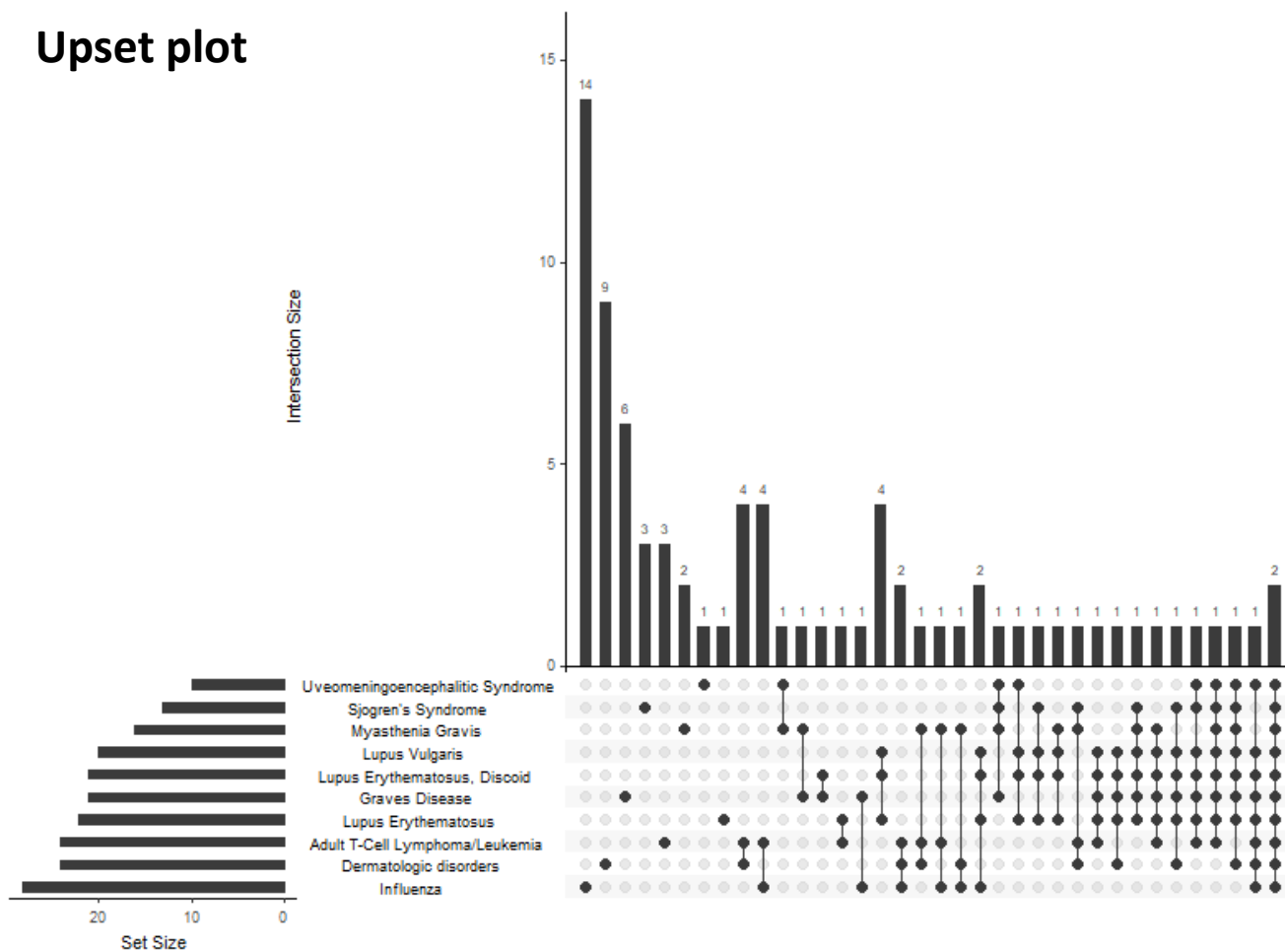


### C4-CD8-LAYN (Input: 274 genes)

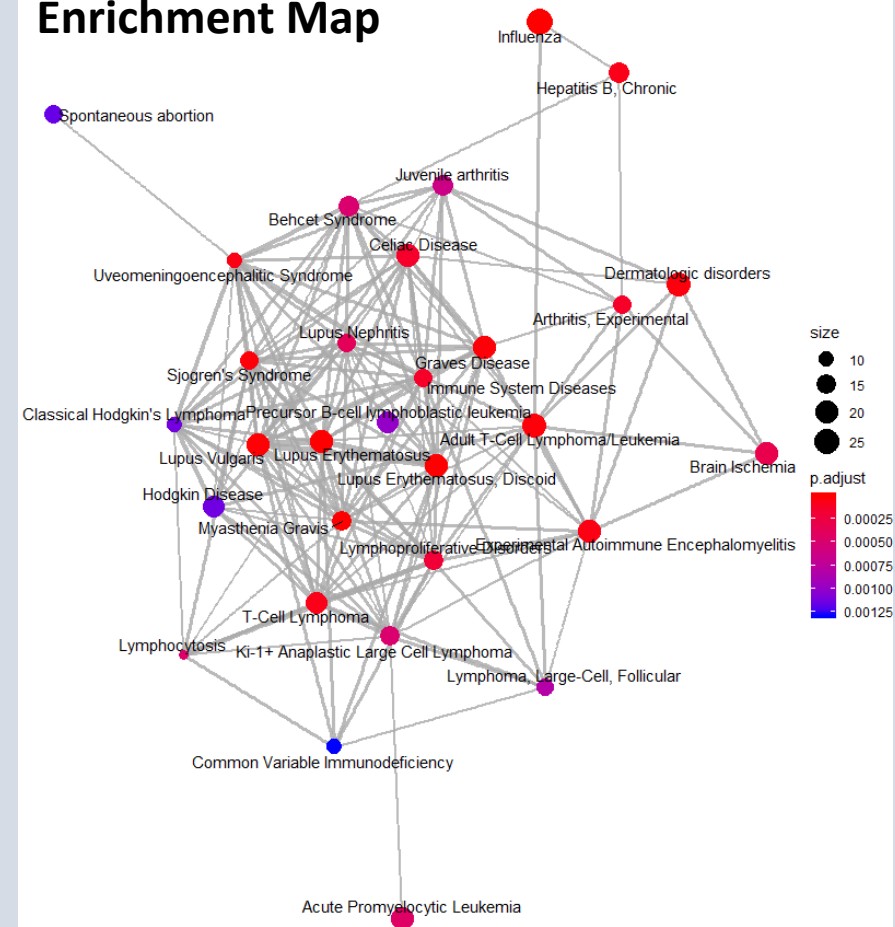


# Analysis: Visualization

## Upset plot



## Enrichment Map





# Analysis: Top Gene List (DisGeNet)

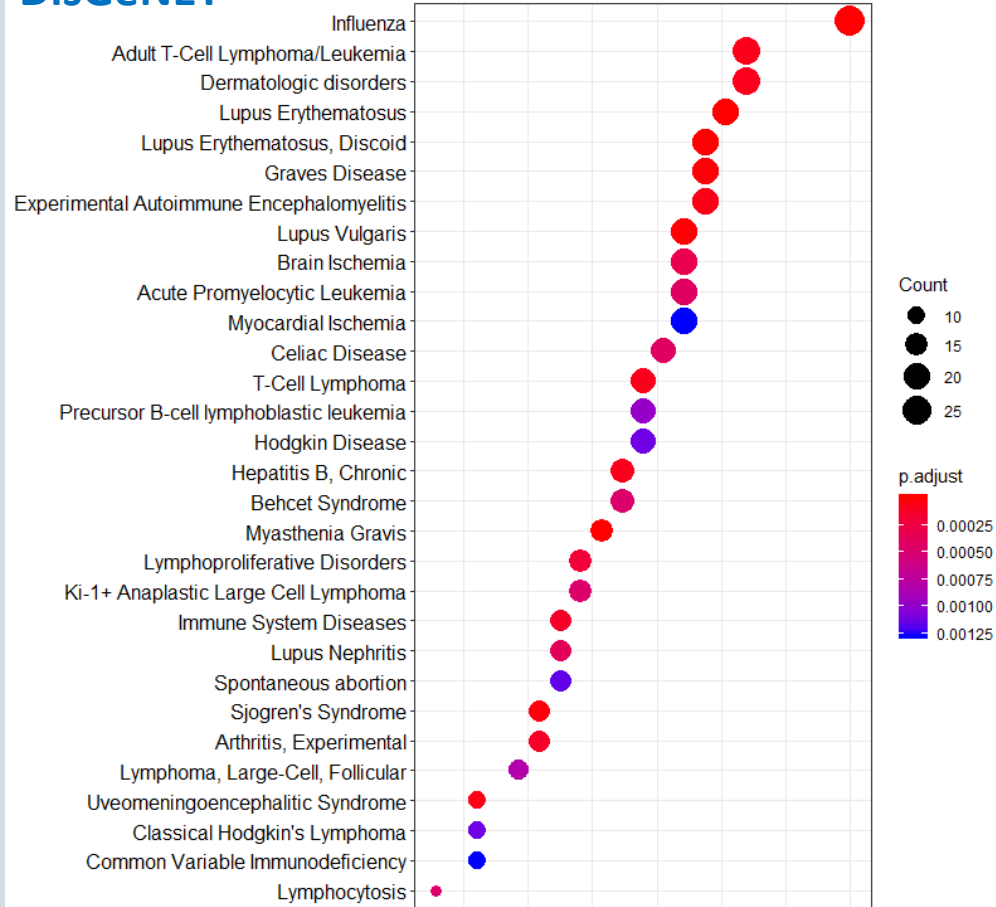
Description	GeneRatio	pvalue	geneID
Influenza	28/250	6.47E-10	CTLA4/TNFRSF9/CCL3/CD27/SARDH/IFNG/CREM/GAPDH/ITGAE/OASL/TNFSF4/RNF19A/HMOX1/IFI6/UBE2L6/PTPN11/LY6E/PLSCR1/IFI35/CASP3/SCO2/OAS3/MX1/IRF9/GBP1/GATA3/DDX60/CAT
Lupus Erythematosus, Discoid	21/250	6.39E-09	PDCD1/CTLA4/CD27/CSF1/IFNG/KIR2DL4/CXCL13/CREM/ICOS/PRDM1/WARS/TNFRSF1B/PTPN22/FASLG/IFI16/CD70/UBASH3B/BTG3/STAT3/IL21R/GATA3
Myasthenia Gravis	16/250	8.06E-09	PDCD1/CTLA4/HLA-DRA/CSF1/IFNG/CXCL13/ICOS/MIR155/IL2RB/PTPN22/TNFAIP3/PAM/RNF19A/PDLIM7/SEM1/CASP3
Lupus Erythematosus	22/250	1.04E-08	PDCD1/CTLA4/CD27/CSF1/IFNG/KIR2DL4/CXCL13/CREM/CD38/ICOS/HLA-DMA/PRDM1/TNFRSF1B/PTPN22/FASLG/IFI16/CD70/UBASH3B/BTG3/STAT3/IL21R/GATA3
Lupus Vulgaris	20/250	2.08E-08	PDCD1/CTLA4/CD27/CSF1/IFNG/KIR2DL4/CXCL13/CREM/ICOS/PRDM1/TNFRSF1B/PTPN22/FASLG/IFI16/CD70/UBASH3B/BTG3/STAT3/IL21R/GATA3
Graves Disease	21/250	4.83E-08	PDCD1/CTLA4/CXCR6/IFNG/ITM2A/CXCL13/ICOS/WARS/ITGAE/TNFRSF1B/PTPN22/ARID5B/TNFAIP3/CBLB/FASLG/SEM1/STAT3/PRDX5/IL21R/RHOH/GATA3
Adult T-Cell Lymphoma/Leukemia	24/250	8.09E-08	RGS1/PDCD1/CTLA4/CCL3/TOX/IFNG/CD38/ICOS/MIR155/VCAM1/IL2RB/TNFSF4/HMOX1/AHI1/CASP3/STAT3/OAS3/IL21R/IRF9/GATA3/MCL1/GSTO1/CAT/CSNK2B
Sjogren's Syndrome	13/250	8.33E-08	PDCD1/CTLA4/IFNG/CXCL13/ID3/MIR155/PRDM1/PTPN22/TNFAIP3/FASLG/SPTAN1/STAT3/GTF2I
Dermatologic disorders	24/250	1.20E-07	RGS1/CTLA4/CCL3/NR4A2/TOX/IFNG/MIR155/VCAM1/IL2RB/ID2/TNFRSF1B/PTPN22/FKBP1A/FASLG/RNF19A/HMOX1/IGFLR1/PELI1/PON2/MTHFD1/LIMK1/IRF2/GATA3/GSTO1
Uveomeningoencephalitic Syndrome	10/250	2.16E-07	PDCD1/CTLA4/HLA-DRA/IFNG/KIR2DL4/PTPN22/TNFAIP3/TRAFF5/STAT3/GATA3
Experimental Autoimmune Encephalomyelitis	21/250	2.25E-07	HAVCR2/CTLA4/TNFRSF9/CCL3/ENTPD1/NR4A2/IFNG/CXCL13/MIR155/VCAM1/IL2RB/TNFRSF1B/CBLB/RNF19A/SEMA4A/CASP3/STAT3/PRDX5/GATA3/GSTO1/CAT
Hepatitis B, Chronic	17/250	3.41E-07	HAVCR2/PDCD1/CTLA4/TNFRSF9/IFNG/KIR2DL4/KLR4/APOBEC3G/WARS/ACP5/OASL/FASLG/OAS3/MX1/IRF2/IRF9/CAT
T-Cell Lymphoma	18/250	3.68E-07	PDCD1/CTLA4/CD27/IFNG/CXCL13/ID3/MIR155/PRDM1/IL2RB/ID2/BCL2L11/CCND2/CD70/PDLIM7/STAT3/CTNNA1/GATA3/TET2
Immune System Diseases	14/250	8.60E-07	HAVCR2/PDCD1/CTLA4/ENTPD1/IFNG/ICOS/PTPN22/TNFAIP3/RAB27A/IKZF3/BATF/STAT3/GATA3/TET2
Arthritis, Experimental	13/250	9.07E-07	HAVCR2/CTLA4/CCL3/NR4A2/IFNG/TNFRSF1B/PAM/CBLB/FASLG/STAT3/ACSL4/CAT/CTSB
Celiac Disease	20/250	9.88E-07	RGS1/PDCD1/CTLA4/TNFRSF9/HLA-DRA/IFNG/KIR2DL4/CD38/ICOS/IL2RB/PTPN22/TNFSF4/TNFAIP3/ZFP36L1/FASLG/PON2/PARK7/TFRC/GATA3/CTSB
Lymphoproliferative Disorders	15/250	1.56E-06	CTLA4/CD27/IFNG/CXCL13/CD38/MIR155/PRDM1/IL2RB/TNFRSF1B/TNFAIP3/CBLB/FASLG/BCL2L11/PDLIM7/STAT3
Brain Ischemia	20/250	2.56E-06	CTLA4/CCL3/TOX/PTGIS/VCAM1/ITGAE/TNFRSF1B/FASLG/HMOX1/PON2/MTHFD1/PRKCH/LIMK1/CASP3/STAT3/GALNT2/GSTO1/CAT/CTSB/LY6G5B
Lupus Nephritis	14/250	3.25E-06	PDCD1/CSF1/IFNG/CXCL13/VCAM1/TNFSF4/TNFAIP3/FASLG/UBASH3B/LY6E/CASP3/STAT3/MX1/GATA3
Acute Promyelocytic Leukemia	20/250	4.08E-06	SRGN/CSF1/IFNG/CD38/ASB2/ID2/FASLG/ATXN1/UBE2L6/PLSCR1/TFRC/CASP3/STAT3/TBL1XR1/PRDX5/PRKAR1A/IRF9/TET2/SP140/MCL1
Lymphocytosis	8/250	4.58E-06	CTLA4/CD27/IFNG/CD38/MIR155/ITGAE/FASLG/PDLIM7
Behcet Syndrome	17/250	5.05E-06	PDCD1/CTLA4/IFNG/KIR2DL4/KLR4/TNFRSF1B/PTPN22/TNFAIP3/HMOX1/TRAFF5/IFI16/UBASH3B/PHTF2/STAT3/PSTPIP1/GATA3/CAT
Ki-14 Anaplastic Large Cell Lymphoma	15/250	5.10E-06	IFNG/MIR155/PRDM1/ID3/TNFRSF1B/FASLG/BCL2L11/PDLIM7/PTPN11/CACD2/STAT3/CALM2/IL21R/GATA3/MCL1

**Caution 1 :** Only “upregulated genes” were defined as signature genes to start with, but loss of function mutation etc could define select gene as “geneset”

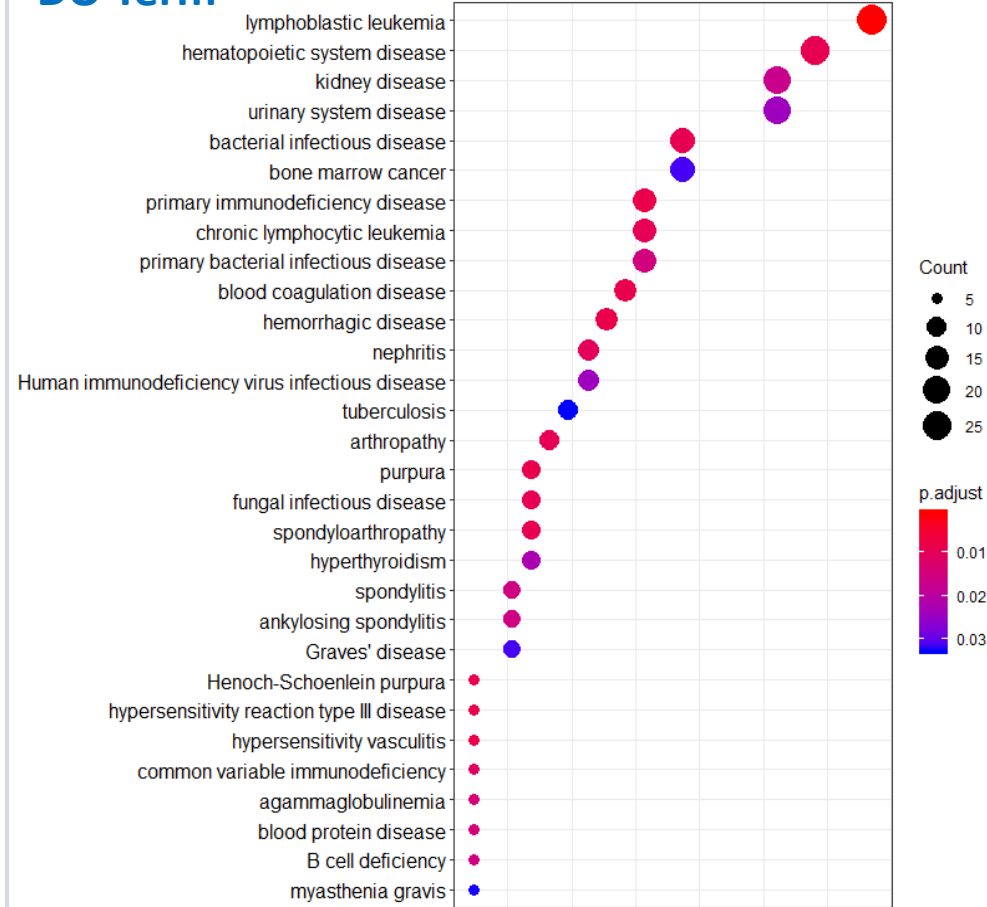
**Caution 2 :** Each pathway was treated independently – a gene can function in multiple pathway so identified pathways were frequently overlapped → Obscure true source of signals

# Analysis: DO term analysis

## DisGeNET



## DO Term



**Network of Cancer Gene (NCG) Database**  
– top 30 pathway, only one as significant

## Commonly Enriched Genesets:

Infectious disease, Leukemia/Lymphoma, Autoimmune disease, Graves Disease/hyperthyroidism, Blood Vessle (Ischemia  $\leftrightarrow$  Purpura/hemorrhagic/Behcet Syndrome)



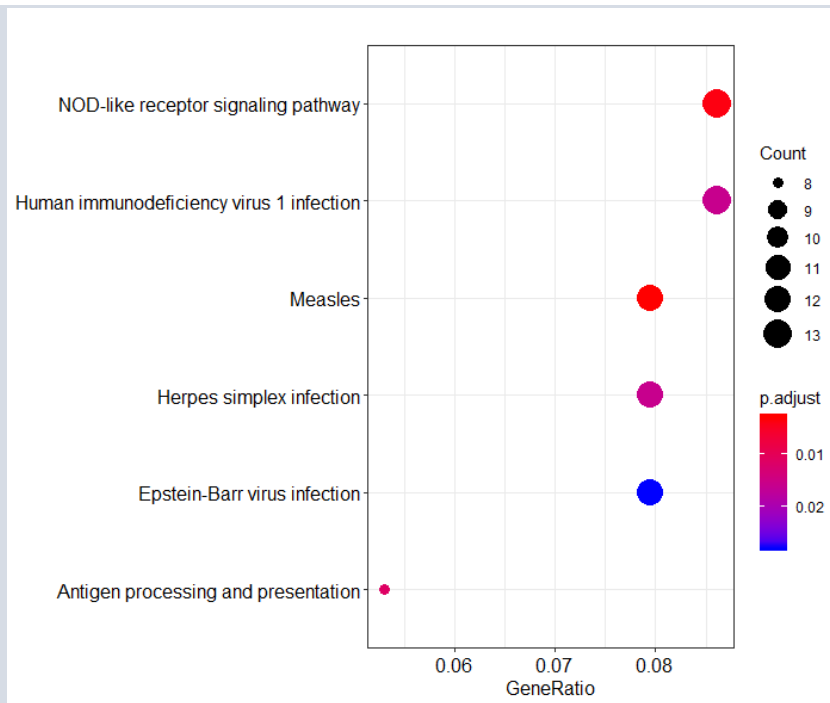
# Analysis: Top Gene List (DO term)

ID	Description	GeneRatio	p.adjust	geneID
DOID:1037	lymphoblastic leukemia	26/169	0.001016	CCL3/IFNG/CXCL13/CD38/IL2RB/LAT2/ARID5B/CLECL1/PDE7B/TNFAIP3/ZFP36L1/FASLG/BCL2L11/CCND2/IKZF3/CD70/PTPN11/TFRC/PKM/TBL1XR1/IL21R/ARNT/RHOH/MCL1/LDHA/CSNK2B
DOID:612	primary immunodeficiency disease	14/169	0.008202	PDCD1/CTLA4/IFNG/ICOS/APOBEC3G/TNFRSF1B/FASLG/APOBEC3F/NCF4/CD70/CD63/BST2/PRKAR1A/NFAT5
DOID:11123	Henoch-Schoenlein purpura	5/169	0.008202	CTLA4/PTPN22/TNFSF4/HMOX1/CAT
DOID:1557	hypersensitivity reaction type III disease	5/169	0.008202	CTLA4/PTPN22/TNFSF4/HMOX1/CAT
DOID:9809	hypersensitivity vasculitis	5/169	0.008202	CTLA4/PTPN22/TNFSF4/HMOX1/CAT
DOID:2213	hemorrhagic disease	12/169	0.008202	CTLA4/HLA-DRA/IFNG/PTPN22/TNFSF4/FASLG/HMOX1/CASP3/CTNNB1/TET2/LDHA/CAT
DOID:1247	blood coagulation disease	13/169	0.008473	CTLA4/HLA-DRA/IFNG/PTPN22/TNFSF4/FASLG/HMOX1/PTPN11/CASP3/CTNNB1/TET2/LDHA/CAT
DOID:3326	purpura	8/169	0.008473	CTLA4/IFNG/PTPN22/TNFSF4/FASLG/HMOX1/CASP3/CAT
DOID:74	hematopoietic system disease	23/169	0.009264	CTLA4/HLA-DRA/IFNG/CD38/ICOS/VCAM1/TNFRSF1B/PTPN22/TNFSF4/FASLG/HMOX1/CD70/PTPN11/TFRC/SPTAN1/CASP3/PRKAR1A/NFAT5/CTNNB1/GATA3/TET2/LDHA/CAT
DOID:104	bacterial infectious disease	16/169	0.009264	PDCD1/CTLA4/CXCR6/CCL3/CD27/IFNG/CD38/ICOS/PRDM1/TNFRSF1B/PTPN22/CBLB/FASLG/BCL2L11/STAT3/MCL1
DOID:1564	fungal infectious disease	8/169	0.009264	PDCD1/CTLA4/CXCR6/IFNG/TNFRSF1B/PTPN22/STAT3/ICAM3
DOID:381	arthropathy	9/169	0.009264	RGS1/PDCD1/CTLA4/CSF1/IFNG/ICOS/PTGIS/TNFRSF1B/STAT3
DOID:1123	spondyloarthropathy	8/169	0.009264	RGS1/PDCD1/CTLA4/IFNG/ICOS/PTGIS/TNFRSF1B/STAT3
DOID:1040	chronic lymphocytic leukemia	14/169	0.00968	CCL3/IFNG/CXCL13/CD38/PDE7B/TNFAIP3/ZFP36L1/FASLG/BCL2L11/TFRC/IL21R/RHOH/MCL1/CSNK2B
DOID:10952	nephritis	11/169	0.010447	PDCD1/MYO1E/CSF1/IFNG/CXCL13/VCAM1/TNFSF4/ACTN4/FASLG/CASP3/GATA3
DOID:12177	common variable immunodeficiency	5/169	0.011504	CTLA4/ICOS/TNFRSF1B/CD70/PRKAR1A
DOID:2583	agammaglobulinemia	5/169	0.01414	CTLA4/ICOS/TNFRSF1B/CD70/PRKAR1A
DOID:620	blood protein disease	5/169	0.01414	CTLA4/ICOS/TNFRSF1B/CD70/PRKAR1A
DOID:0050338	primary bacterial infectious disease	14/169	0.015208	PDCD1/CTLA4/CCL3/CD27/IFNG/CD38/ICOS/PRDM1/TNFRSF1B/PTPN22/CBLB/FASLG/BCL2L11/MCL1
DOID:6590	spondylitis	7/169	0.015915	RGS1/PDCD1/CTLA4/IFNG/PTGIS/TNFRSF1B/STAT3
DOID:2115	B cell deficiency	5/169	0.015915	CTLA4/ICOS/TNFRSF1B/CD70/PRKAR1A
DOID:7147	ankylosing spondylitis	7/169	0.015915	RGS1/PDCD1/CTLA4/IFNG/PTGIS/TNFRSF1B/STAT3
DOID:557	kidney disease	21/169	0.017219	PDCD1/MYO1E/CSF1/IFNG/CXCL13/RGS2/VCAM1/ACP5/TNFRSF1B/TNFSF4/ACTN4/FASLG/RBPJ/HMOX1/AHI1/LIMS1/TFRC/CASP3/STAT3/GATA3/ICAM3
DOID:7998	hyperthyroidism	8/169	0.022529	CTLA4/CXCR6/TNFRSF9/IFNG/TNFRSF1B/PTPN22/FASLG/PRDX5
DOID:526	Human immunodeficiency virus infection	11/169	0.024254	PDCD1/CXCR6/CCL3/CD27/CSF1/IFNG/CXCL13/VCAM1/APOBEC3G/IL2RB/FASLG
DOID:18	urinary system disease	21/169	0.024254	PDCD1/MYO1E/CSF1/IFNG/CXCL13/RGS2/VCAM1/ACP5/TNFRSF1B/TNFSF4/ACTN4/FASLG/RBPJ/HMOX1/AHI1/LIMS1/TFRC/CASP3/STAT3/GATA3/ICAM3
DOID:12361	Graves' disease	7/169	0.031455	CTLA4/CXCR6/TNFRSF9/TNFRSF1B/PTPN22/FASLG/PRDX5
DOID:4960	bone marrow cancer	16/169	0.031469	CCL3/SRGN/IFNG/VCAM1/PRDM1/ACP5/LAG3/FASLG/BCL2L11/CCND2/PTPN11/STAT3/CTNNB1/TET2/MCL1/LDHA
DOID:437	myasthenia gravis	5/169	0.032656	PDCD1/CTLA4/HLA-DRA/CXCL13/PTPN22



# Analysis: KEGG analysis

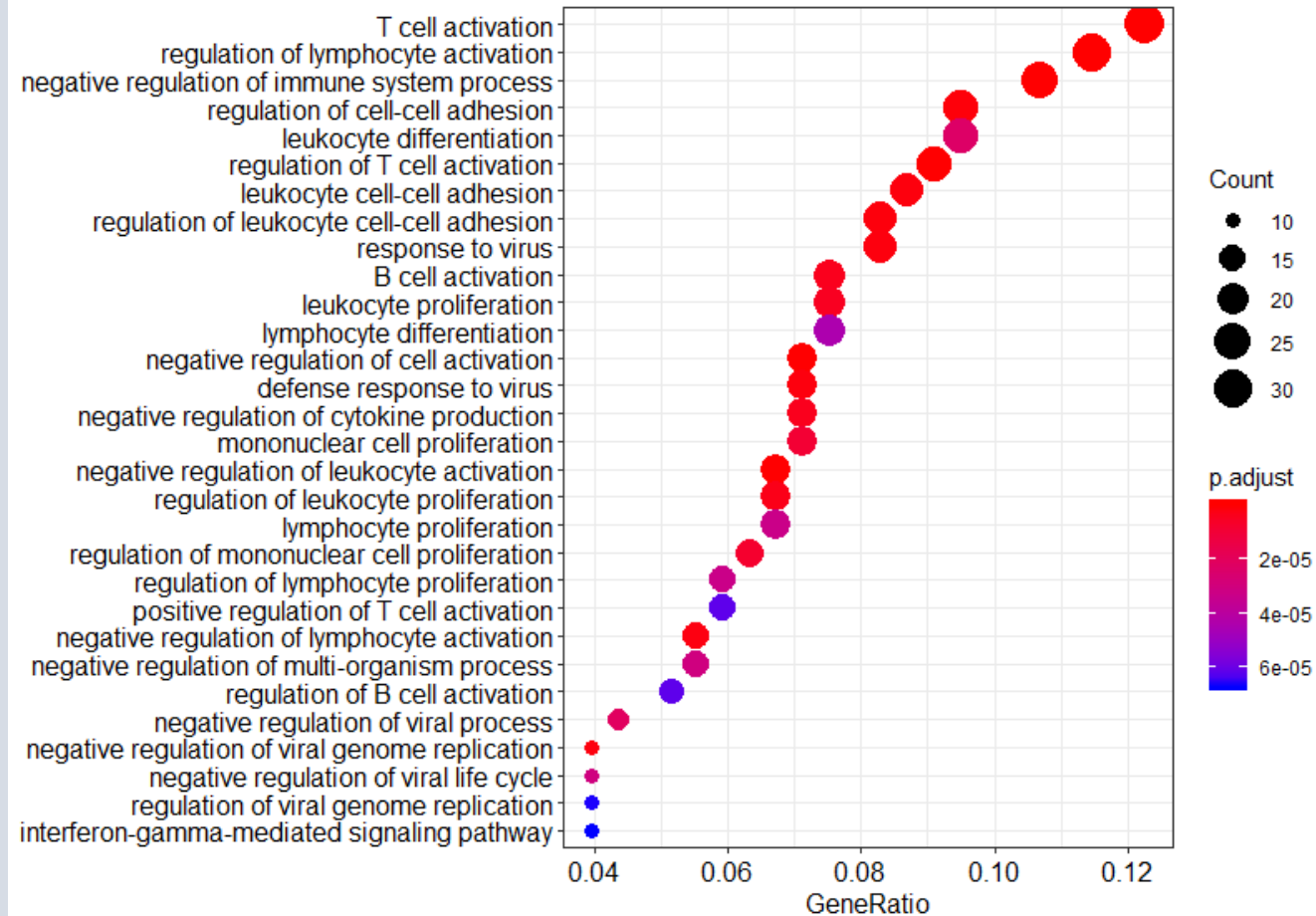
KEGG



ID	Description	GeneRatio	p.adjust	qvalue	geneID	Count
hsa05162	Measles	12/151	0.003005	0.002636	IFNG/IL2RB/TNFAIP3/CBLB/FASLG/CCND2/STAT3/OAS3/MX1/IRF9/TBK1/CSNK2B	12
hsa04621	NOD-like receptor signaling pathway	13/151	0.003707	0.003251	TNFAIP3/TRAFF5/IFI16/GABARAPL1/GBP4/OAS3/PSTPIP1/IRF9/GBP1/TBK1/GBP2/RNF31/CTSB	13
hsa04612	Antigen processing and presentation	8/151	0.011375	0.009978	HLA-DRA/IFNG/KIR2DL4/HLA-DMA/KLRK4/PSME2/RFX5/CTSB	8
hsa05170	Human immunodeficiency virus 1 infection	13/151	0.015951	0.013992	APOBEC3G/TNFRSF1B/APOBEC3C/FASLG/APOBEC3F/TRAFF5/APOBEC3D/BST2/LIMK1/CASP3/CALM3/TBK1/MAP2K3	13
hsa05168	Herpes simplex infection	12/151	0.015951	0.013992	HLA-DRA/IFNG/HLA-DMA/FASLG/TRAFF5/PTPN11/CASP3/OAS3/GTF2I/IRF9/TBK1/CSNK2B	12
hsa05169	Epstein-Barr virus infection	12/151	0.028047	0.024603	ENTPD1/HLA-DRA/IFNG/CD38/TNFAIP3/RBPJ/TRAFF5/SEM1/STAT3/TBK1/MAP2K3/CSNK2B	12

C4-CD8-LAYN (Input: 274 genes)

# Analysis: GO – Biological Process



C4-CD8-LAYN (Input: 274 genes)

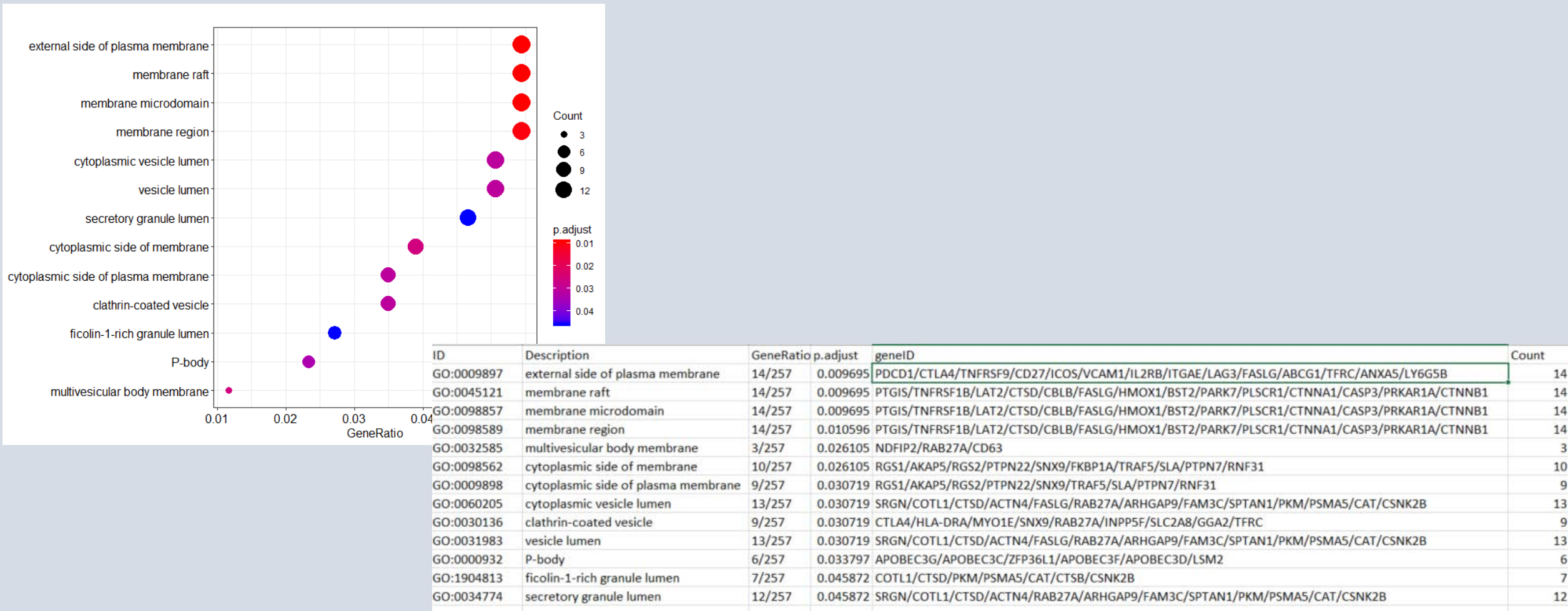
# Analysis: GO – Biological Process

ID	Description	GeneRatio	p.adjust	geneID
GO:0042110	T cell activation	31/253	3.49E-09	HAVCR2/PDCD1/CTLA4/TIGIT/CD27/HLA-DRA/IFNG/SIRPG/ICOS/MIR155/VCAM1/PRDM1/LAG3/PTPN22/CLECL1/TNFSF4/FKBP1A/ZFP36L1/CD7/RAB27A/SEMA4A/BATF/PELI1/PTPN11/TFRC/CASP3/STAT3/PRKAR1A/CTNNB1/RHOH/GATA3
GO:0051249	regulation of lymphocyte activation	29/253	3.69E-08	HAVCR2/PDCD1/CTLA4/TIGIT/CD27/HLA-DRA/IFNG/SIRPG/SAMSN1/CD38/ICOS/MIR155/VCAM1/PRDM1/ID2/LAG3/PTPN22/CLECL1/TNFSF4/TNFAIP3/ZFP36L1/IKZF3/PELI1/PTPN11/TFRC/CASP3/PRKAR1A/CTNNB1/GATA3
GO:0050866	negative regulation of cell activation	18/253	7.36E-08	HAVCR2/CTLA4/TIGIT/SAMSN1/MIR155/CD200/PRDM1/ID2/LAG3/PTPN22/TNFSF4/TNFAIP3/HMOX1/CD84/UBASH3B/PELI1/CASP3/PRKAR1A
GO:0002695	negative regulation of leukocyte activation	17/253	7.45E-08	HAVCR2/CTLA4/TIGIT/SAMSN1/MIR155/CD200/PRDM1/ID2/LAG3/PTPN22/TNFSF4/TNFAIP3/HMOX1/CD84/PELI1/CASP3/PRKAR1A
GO:0002683	negative regulation of immune system process	27/253	7.87E-08	HAVCR2/PDCD1/CTLA4/TIGIT/CCL3/SAMSN1/MIR155/CD200/PRDM1/ID2/LAG3/PTPN22/TNFSF4/TNFAIP3/ZFP36L1/HMOX1/IFI16/CD84/UBASH3B/PELI1/BST2/PARP14/TRAFF1/CASP3/PRKAR1A/GBP1/CTNNB1
GO:0050863	regulation of T cell activation	23/253	1.57E-07	HAVCR2/PDCD1/CTLA4/TIGIT/CD27/HLA-DRA/IFNG/SIRPG/ICOS/MIR155/VCAM1/PRDM1/LAG3/PTPN22/CLECL1/TNFSF4/PELI1/PTPN11/TFRC/CASP3/PRKAR1A/CTNNB1/GATA3
GO:0022407	regulation of cell-cell adhesion	24/253	1.06E-06	HAVCR2/PDCD1/CTLA4/TIGIT/CD27/HLA-DRA/IFNG/SIRPG/CXCL13/ICOS/MIR155/VCAM1/LAG3/PTPN22/CLECL1/TNFSF4/UBASH3B/PELI1/PTPN11/TFRC/CASP3/PRKAR1A/MAD2L2/GATA3
GO:1903037	regulation of leukocyte cell-cell adhesion	21/253	1.26E-06	HAVCR2/PDCD1/CTLA4/TIGIT/CD27/HLA-DRA/IFNG/SIRPG/ICOS/MIR155/VCAM1/LAG3/PTPN22/CLECL1/TNFSF4/PELI1/PTPN11/TFRC/CASP3/PRKAR1A/GATA3
GO:0007159	leukocyte cell-cell adhesion	22/253	1.33E-06	HAVCR2/PDCD1/CTLA4/TIGIT/CD27/HLA-DRA/IFNG/SIRPG/ICOS/MIR155/VCAM1/LAG3/PTPN22/CLECL1/TNFSF4/PELI1/PTPN11/GCNT1/TFRC/CASP3/PRKAR1A/GATA3
GO:0051607	defense response to virus	18/253	1.46E-06	LYST/IFNG/APOBEC3G/OASL/APOBEC3C/TNFAIP3/APOBEC3F/IFI16/APOBEC3D/BST2/PLSCR1/OAS3/MX1/IRF2/IRF9/GBP1/TBK1/DDX60
GO:0009615	response to virus	21/253	1.46E-06	LYST/IFNG/APOBEC3G/OASL/TNFSF4/APOBEC3C/TNFAIP3/APOBEC3F/BCL2L11/IFI16/APOBEC3D/BST2/PLSCR1/OAS3/MX1/IRF2/IRF9/GBP1/TBK1/GATA3/DDX60
GO:0045071	negative regulation of viral genome replication	10/253	1.58E-06	APOBEC3G/OASL/APOBEC3C/APOBEC3F/IFI16/APOBEC3D/BST2/PLSCR1/OAS3/MX1
GO:0051250	negative regulation of lymphocyte activation	14/253	1.88E-06	HAVCR2/CTLA4/TIGIT/SAMSN1/MIR155/PRDM1/ID2/LAG3/PTPN22/TNFSF4/TNFAIP3/PELI1/CASP3/PRKAR1A
GO:0070663	regulation of leukocyte proliferation	17/253	3.01E-06	HAVCR2/CTLA4/CSF1/CD38/MIR155/VCAM1/PRDM1/PTPN22/CLECL1/TNFSF4/TNFAIP3/IKZF3/PELI1/TFRC/CASP3/PRKAR1A/CTNNB1
GO:0042113	B cell activation	19/253	3.81E-06	CTLA4/CD27/ITM2A/SAMSN1/CD38/VCAM1/PRDM1/ID2/LAT2/TNFSF4/TNFAIP3/ZFP36L1/RBP1/IKZF3/BATF/PELI1/BST2/TFRC/CASP3
GO:0001818	negative regulation of cytokine production	18/253	3.81E-06	HAVCR2/TIGIT/TNFRSF9/SRGN/IFNG/MIR155/ACPD/LAG3/PTPN22/TNFSF4/TNFAIP3/HMOX1/CD84/UBE2L6/BST2/GBP1/TBK1/GATA3
GO:0070661	leukocyte proliferation	19/253	4.59E-06	HAVCR2/CTLA4/CSF1/CD38/MIR155/VCAM1/PRDM1/PTPN22/CLECL1/TNFSF4/TNFAIP3/FKBP1A/IKZF3/PELI1/TFRC/CASP3/PRKAR1A/TBK1/CTNNB1
GO:0032944	regulation of mononuclear cell proliferation	16/253	7.60E-06	HAVCR2/CTLA4/CSF1/CD38/MIR155/VCAM1/PRDM1/PTPN22/CLECL1/TNFSF4/IKZF3/PELI1/TFRC/CASP3/PRKAR1A/CTNNB1
GO:0032943	mononuclear cell proliferation	18/253	9.03E-06	HAVCR2/CTLA4/CSF1/CD38/MIR155/VCAM1/PRDM1/PTPN22/CLECL1/TNFSF4/FKBP1A/IKZF3/PELI1/TFRC/CASP3/PRKAR1A/TBK1/CTNNB1
GO:0048525	negative regulation of viral process	11/253	2.12E-05	CCL3/APOBEC3G/OASL/APOBEC3C/APOBEC3F/IFI16/APOBEC3D/BST2/PLSCR1/OAS3/MX1
GO:0002521	leukocyte differentiation	24/253	2.33E-05	CTLA4/CCL3/CD27/CSF1/IFNG/ITM2A/MIR155/VCAM1/PRDM1/ID2/PTPN22/TNFSF4/ZFP36L1/RBP1/SEMA4A/IFI16/IKZF3/UBASH3B/BATF/TFRC/STAT3/CTNNB1/RHOH/GATA3
GO:1903901	negative regulation of viral life cycle	10/253	3.08E-05	APOBEC3G/OASL/APOBEC3C/APOBEC3F/IFI16/APOBEC3D/BST2/PLSCR1/OAS3/MX1
GO:0043901	negative regulation of multi-organism process	14/253	3.08E-05	HAVCR2/CCL3/APOBEC3G/PRDM1/OASL/APOBEC3C/TNFAIP3/APOBEC3F/IFI16/APOBEC3D/BST2/PLSCR1/OAS3/MX1
GO:0050670	regulation of lymphocyte proliferation	15/253	3.33E-05	HAVCR2/CTLA4/CD38/MIR155/VCAM1/PRDM1/PTPN22/CLECL1/TNFSF4/IKZF3/PELI1/TFRC/CASP3/PRKAR1A/CTNNB1
GO:0046651	lymphocyte proliferation	17/253	3.33E-05	HAVCR2/CTLA4/CD38/MIR155/VCAM1/PRDM1/PTPN22/CLECL1/TNFSF4/FKBP1A/IKZF3/PELI1/TFRC/CASP3/PRKAR1A/TBK1/CTNNB1
GO:0030098	lymphocyte differentiation	19/253	4.45E-05	CTLA4/CD27/IFNG/ITM2A/MIR155/VCAM1/PRDM1/ID2/PTPN22/TNFSF4/ZFP36L1/RBP1/SEMA4A/IKZF3/BATF/STAT3/CTNNB1/RHOH/GATA3
GO:0050870	positive regulation of T cell activation	15/253	6.17E-05	HAVCR2/PDCD1/CTLA4/CD27/HLA-DRA/IFNG/SIRPG/ICOS/MIR155/VCAM1/CLECL1/TNFSF4/PTPN11/TFRC/GATA3
GO:0050864	regulation of B cell activation	13/253	6.17E-05	CTLA4/CD27/SAMSN1/CD38/PRDM1/ID2/TNFSF4/TNFAIP3/ZFP36L1/IKZF3/PELI1/TFRC/CASP3
GO:0045069	regulation of viral genome replication	10/253	6.66E-05	APOBEC3G/OASL/APOBEC3C/APOBEC3F/IFI16/APOBEC3D/BST2/PLSCR1/OAS3/MX1
GO:0060333	interferon-gamma-mediated signaling pathway	10/253	6.71E-05	HLA-DRA/IFNG/VCAM1/OASL/PARP14/OAS3/IRF2/IRF9/GBP1/GBP2
GO:0060337	type I interferon signaling pathway	10/253	6.71E-05	OASL/IFI6/BST2/PTPN11/IFI35/OAS3/MX1/IRF2/IRF9/GBP2
GO:0071357	cellular response to type I interferon	10/253	6.71E-05	OASL/IFI6/BST2/PTPN11/IFI35/OAS3/MX1/IRF2/IRF9/GBP2
GO:0034341	response to interferon-gamma	14/253	7.74E-05	CCL3/HLA-DRA/IFNG/GAPDH/VCAM1/OASL/BST2/PARP14/OAS3/IRF2/IRF9/GBP1/GBP2/CALCOCO2
GO:0051251	positive regulation of lymphocyte activation	18/253	7.84E-05	HAVCR2/PDCD1/CTLA4/CD27/HLA-DRA/IFNG/SIRPG/CD38/ICOS/MIR155/VCAM1/PRDM1/CLECL1/TNFSF4/PELI1/PTPN11/TFRC/GATA3
GO:0019079	viral genome replication	11/253	8.59E-05	CXCR6/APOBEC3G/OASL/APOBEC3C/APOBEC3F/IFI16/APOBEC3D/BST2/PLSCR1/OAS3/MX1

C4-CD8-LAYN (Input: 274 genes)



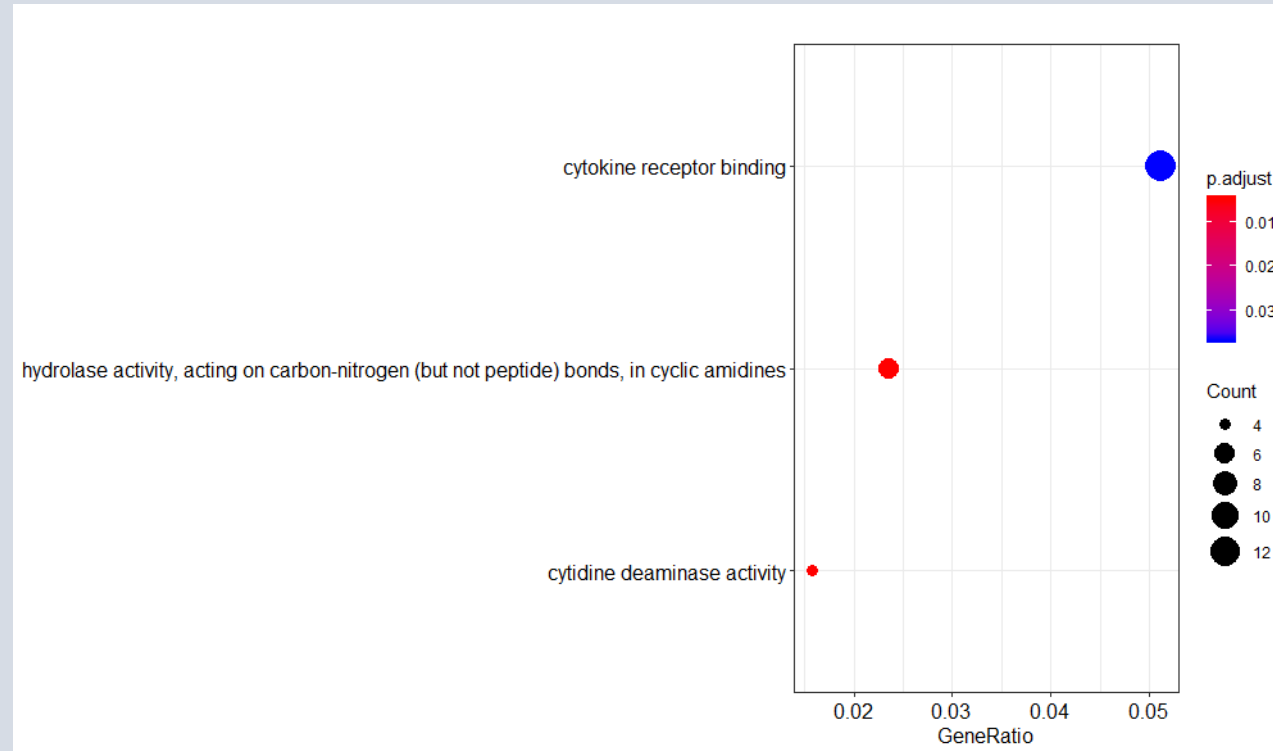
# Analysis: GO – Cellular component



C4-CD8-LAYN (Input: 274 genes)



# Analysis: GO – molecular function



ID	Description	GeneRatio	p.adjust	geneID
GO:0016814	hydrolase activity, acting on carbon-nitrogen (but not peptide) bonds, in cyclic amidines	6/254	0.00492	APOBEC3G/APOBEC3C/APOBEC3F/MTHFD2/APOBEC3D/MTHFD1
GO:0004126	cytidine deaminase activity	4/254	0.00492	APOBEC3G/APOBEC3C/APOBEC3F/APOBEC3D
GO:0005126	cytokine receptor binding	13/254	0.036655	CCL3/CSF1/IFNG/CXCL13/TNFSF4/FKBP1A/FASLG/TRAFF5/CD70/YARS/CASP3/STAT3/GATA3

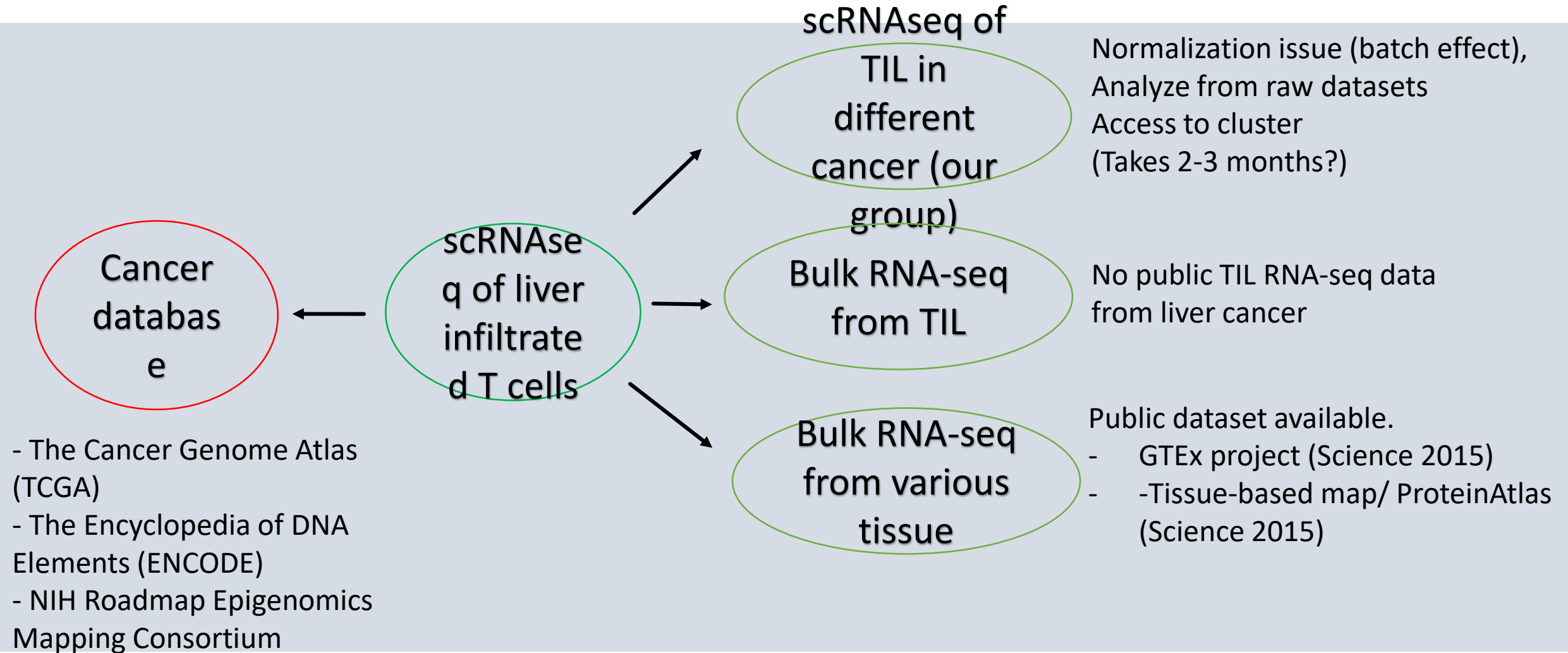
C4-CD8-LAYN (Input: 274 genes)

# Ongoing pathway Analysis (Future plan)

- Advanced analysis with fold changes & p-values inclusion
- Modified GSEA analysis using partial list of gene set might be doable
  - Without triplicate input? – Maybe I can manually add calculated statistics.
  - I have only partial list of gene expression – How rank is determined?
- Sorting genes expressed on cell surfaces
- Came up with potential target to follow

# Tissue enrichment analysis and candidate selection

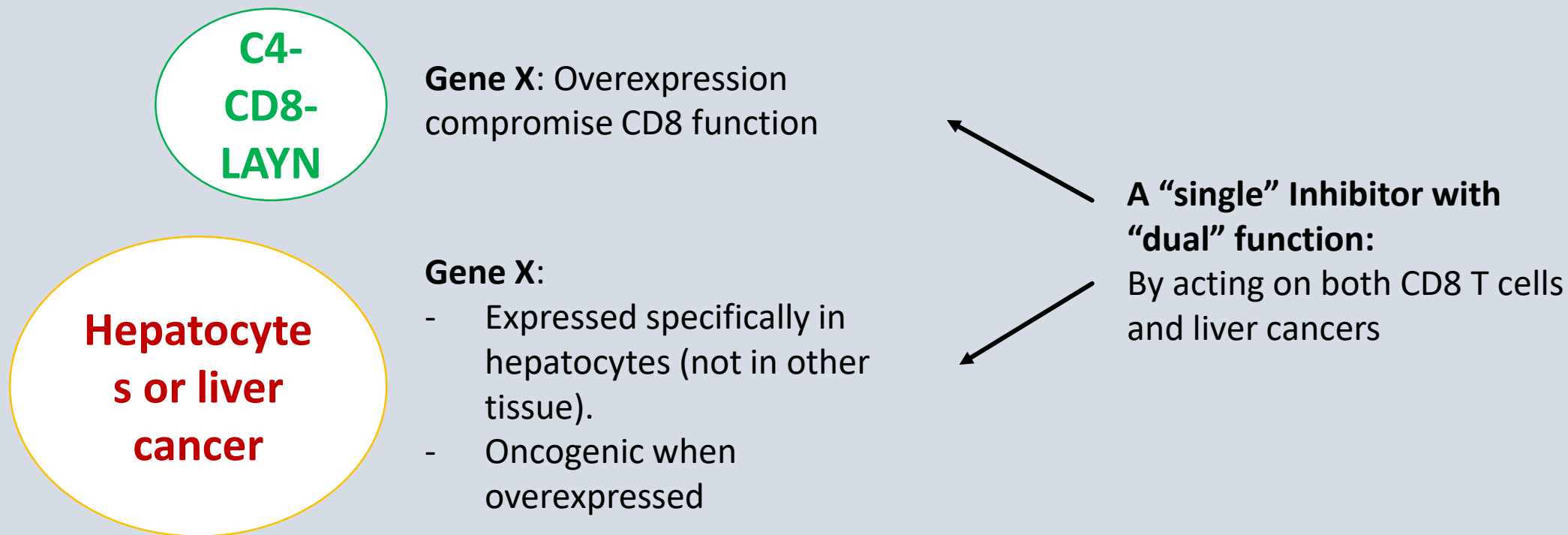
# How to extract useful information from Public datasets?





# IDEA / Experimental Design

Looking for single targets both expressed in Liver cancer and C4-CD8-LAYN





**Table S2**  
Zheng et al.,  
2017

List of signature genes in each cluster  
(C4-CD8-LAYN)

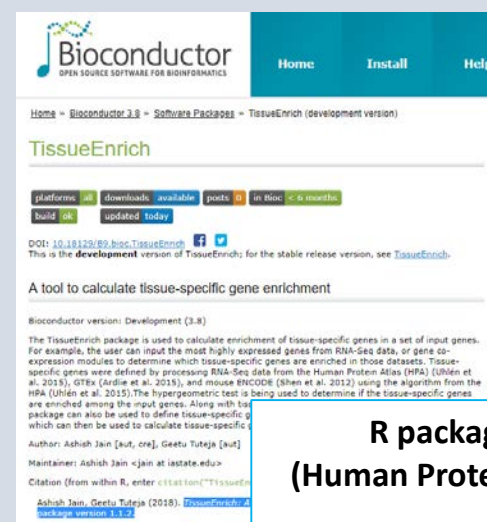
## Extract GeneID (272 genes)

2 entrez ID were not assigned with gene name ("NA")

[illegible]

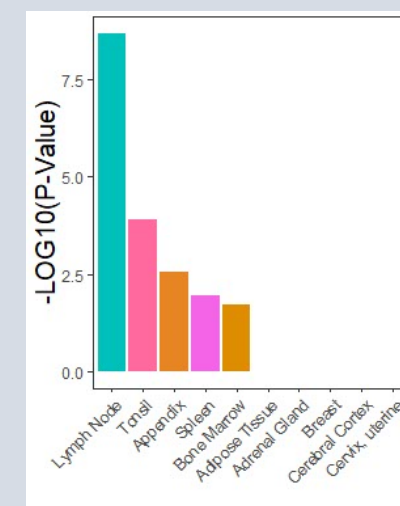
## Tissue Enrichment Analysis

## Human RNA-seq dataset



## R package – TissueEnrich (Human Protein Atlas / GTEx / ENCODE)

## Visualization

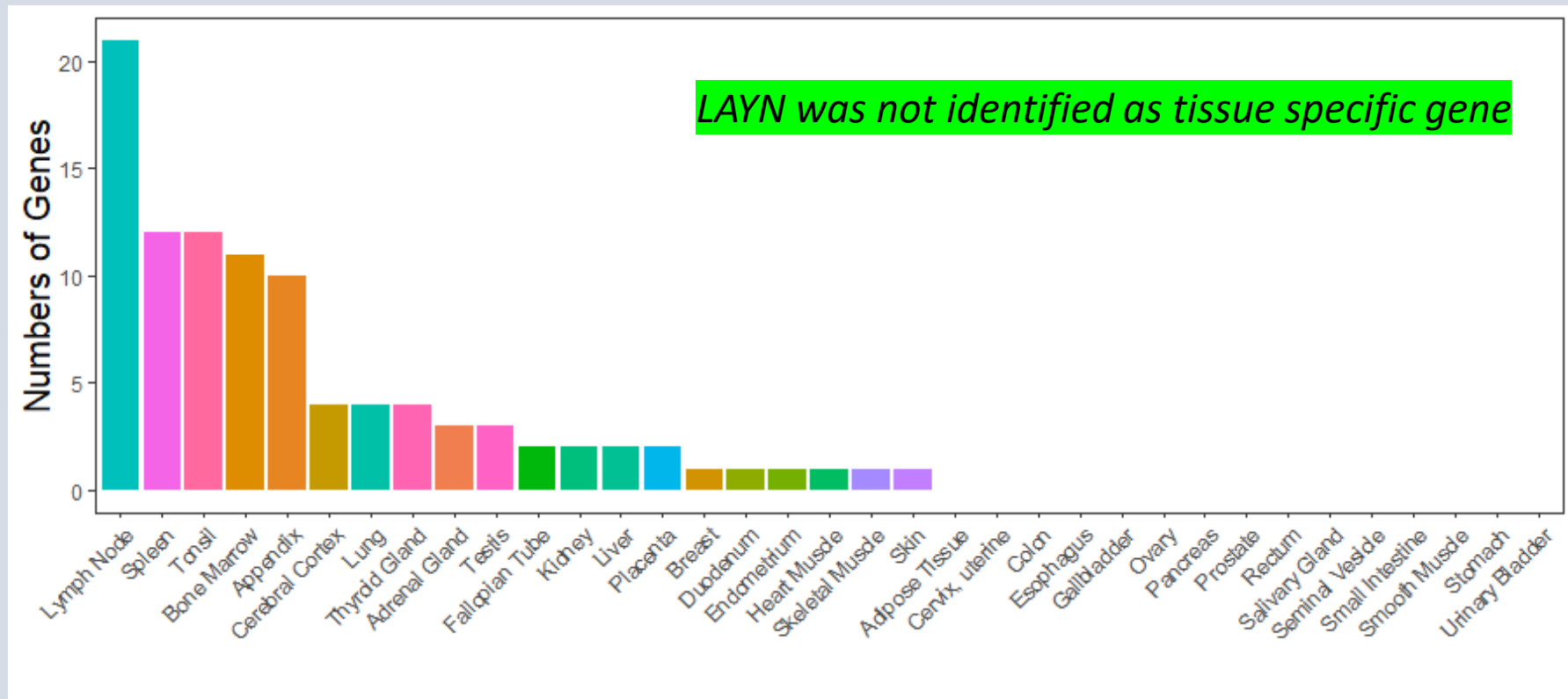


enrichPlot  
ggplot2

**Question: From 274 genes identified in C4-CD8-LAYN cluster, are there any common genes that are enriched in primary liver or liver cancers?**

# Tissue specific expression of 274 genes enriched in C4-CD8-LAYN

20+ genes were specific to lymph node, 2 to liver and 200+ genes were relatively universal, which are probably not good drug targets



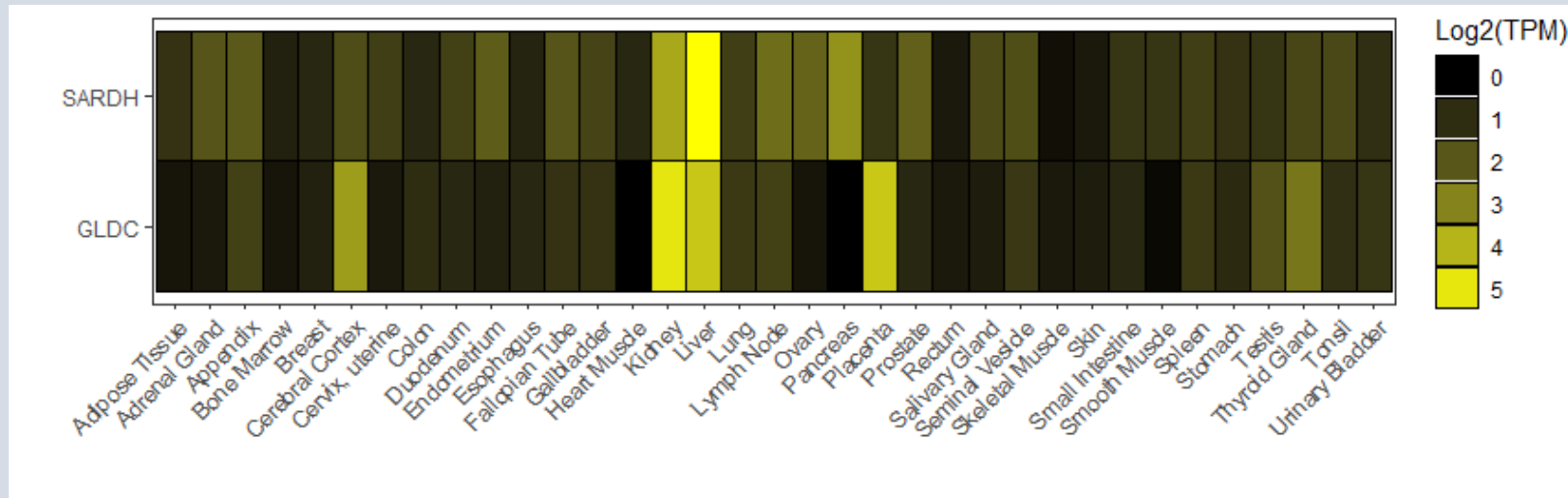




# Two liver specific genes

Liver specific genes that were commonly expressed in C4-CD8-LAYN cluster:

SARDH and GLDC



Plan: Validate this results using mouse tissue with qRT-PCR

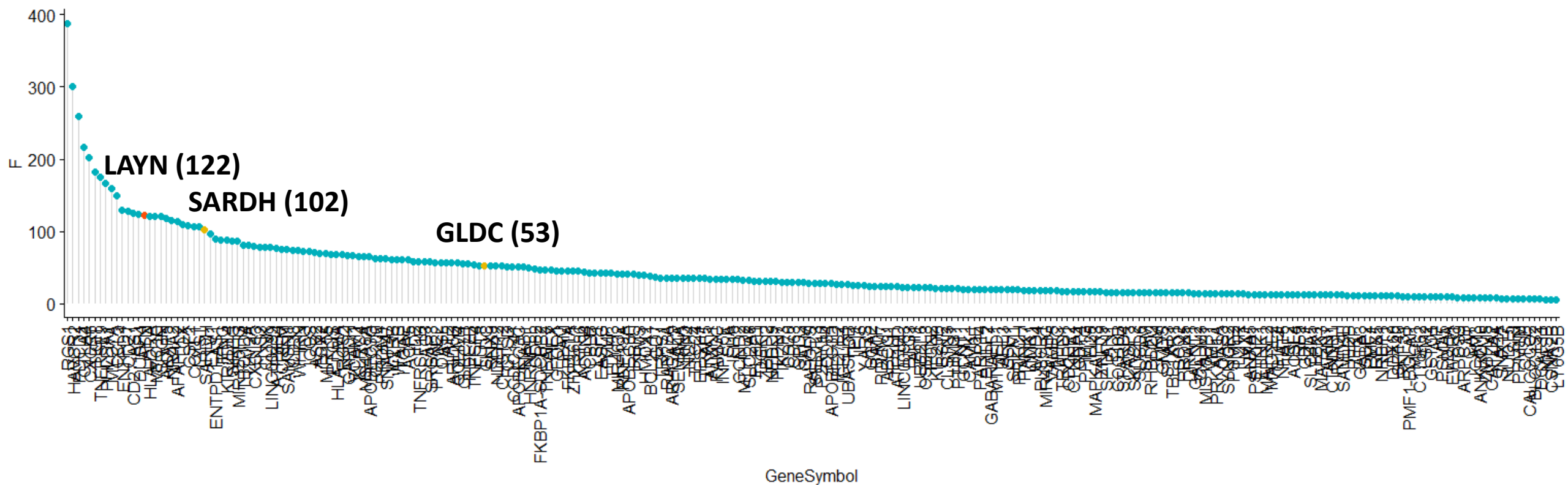
(Perfusion → liver, lung, heart, lymph node, spleen, bone marrow, Thymus, and Kidney) :

Do I need animal protocol for this?

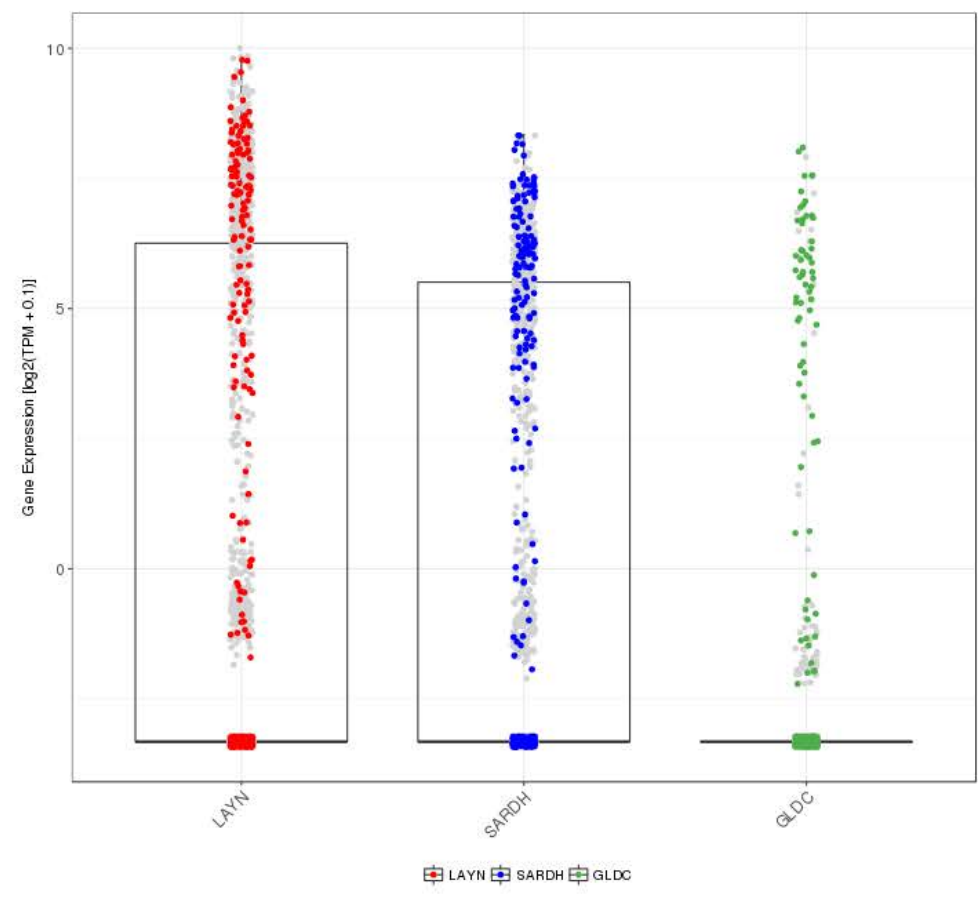
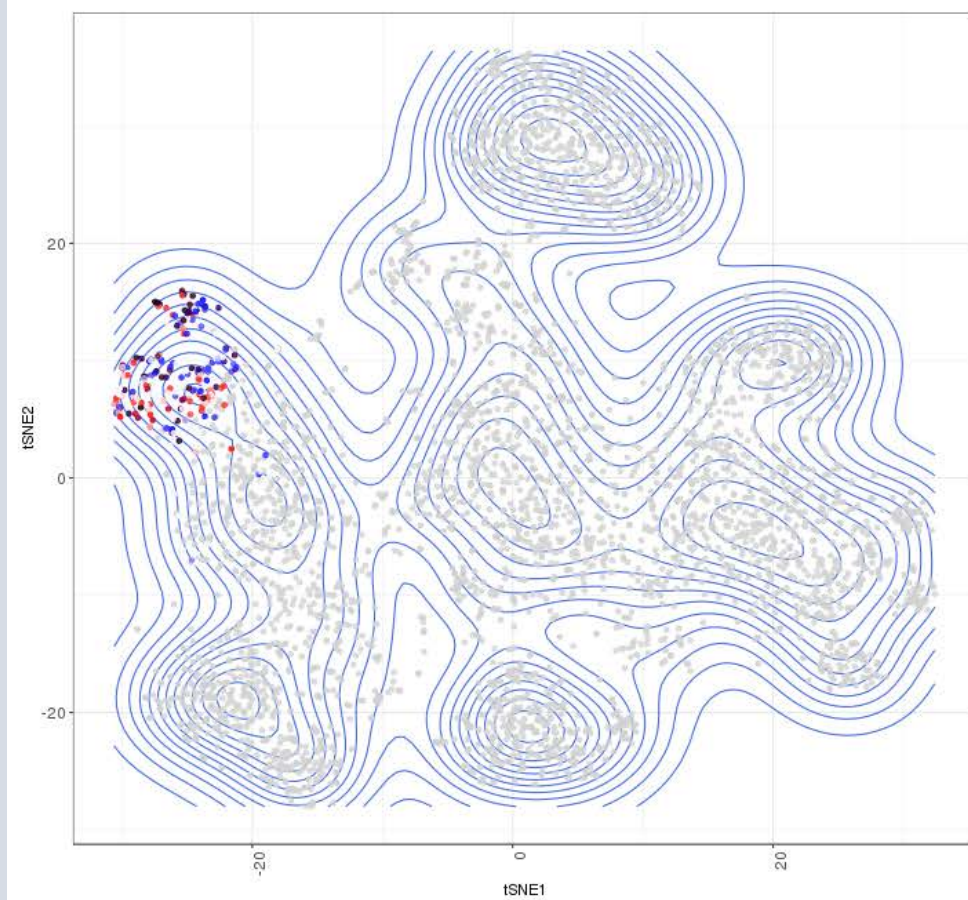
# Fold Changes of Targets of interests

### C4-CD8-LAYN (272 genes, two NAs were excluded)

## What is F value?

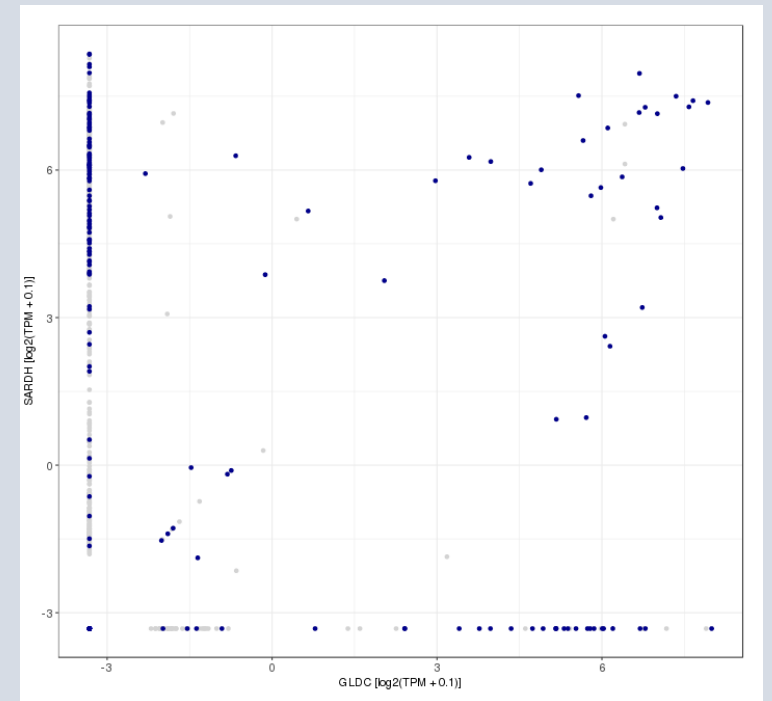
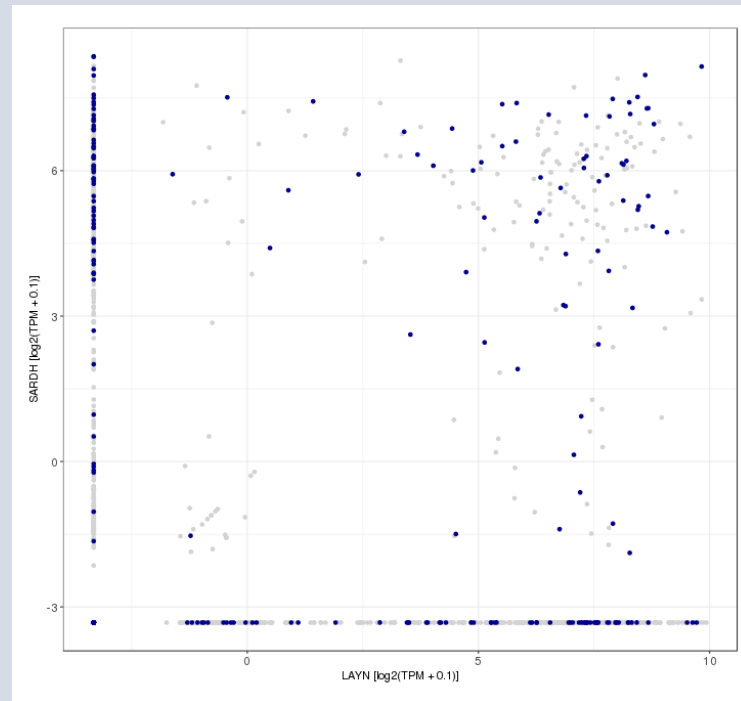
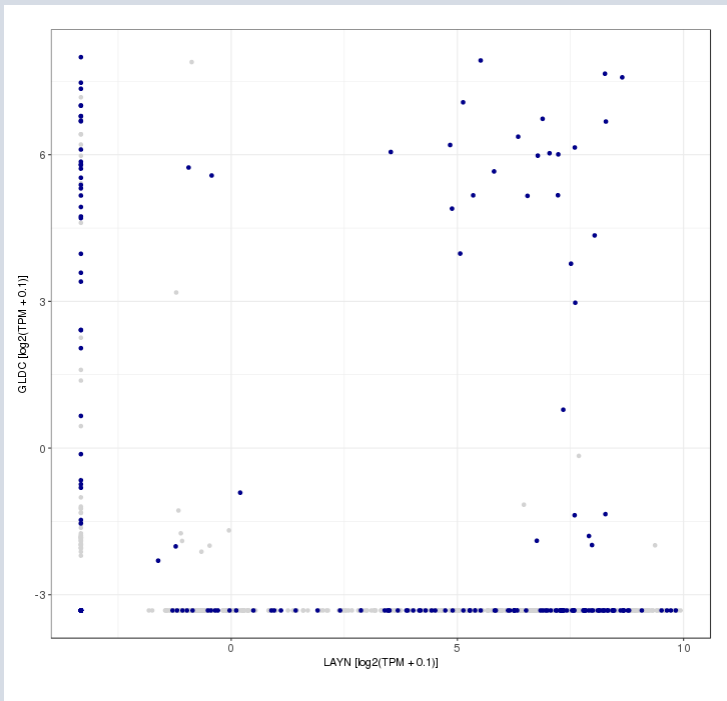


# Three gene expression in C4-CD8-LAYN cluster (Liver)



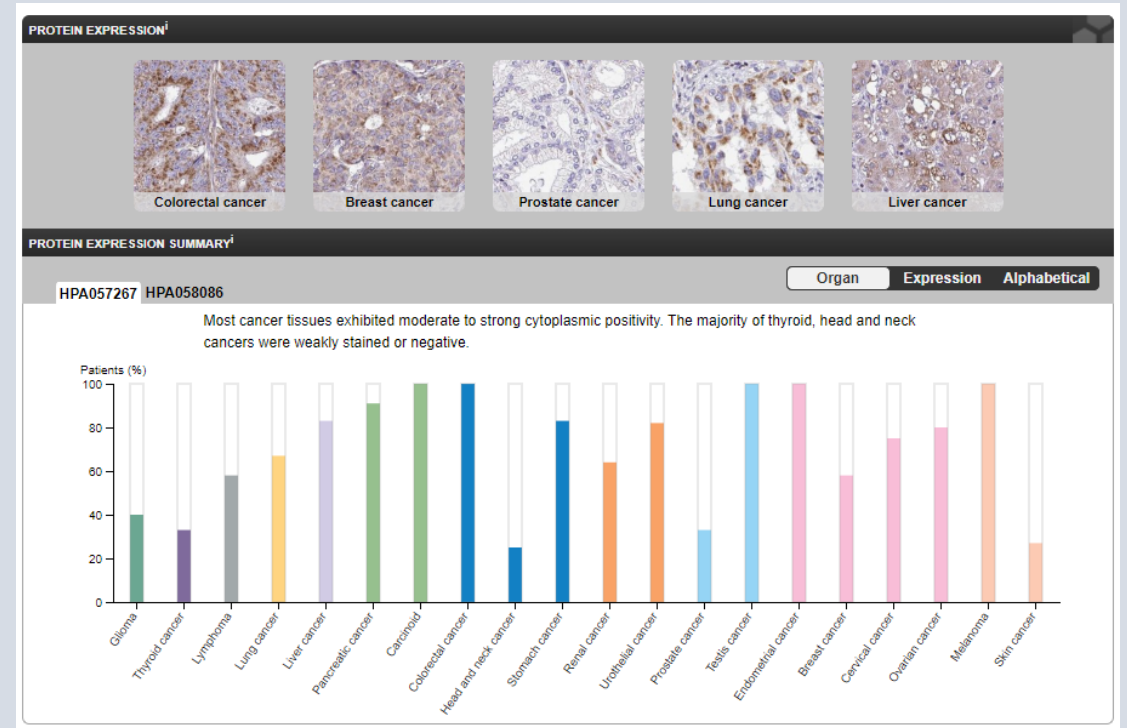
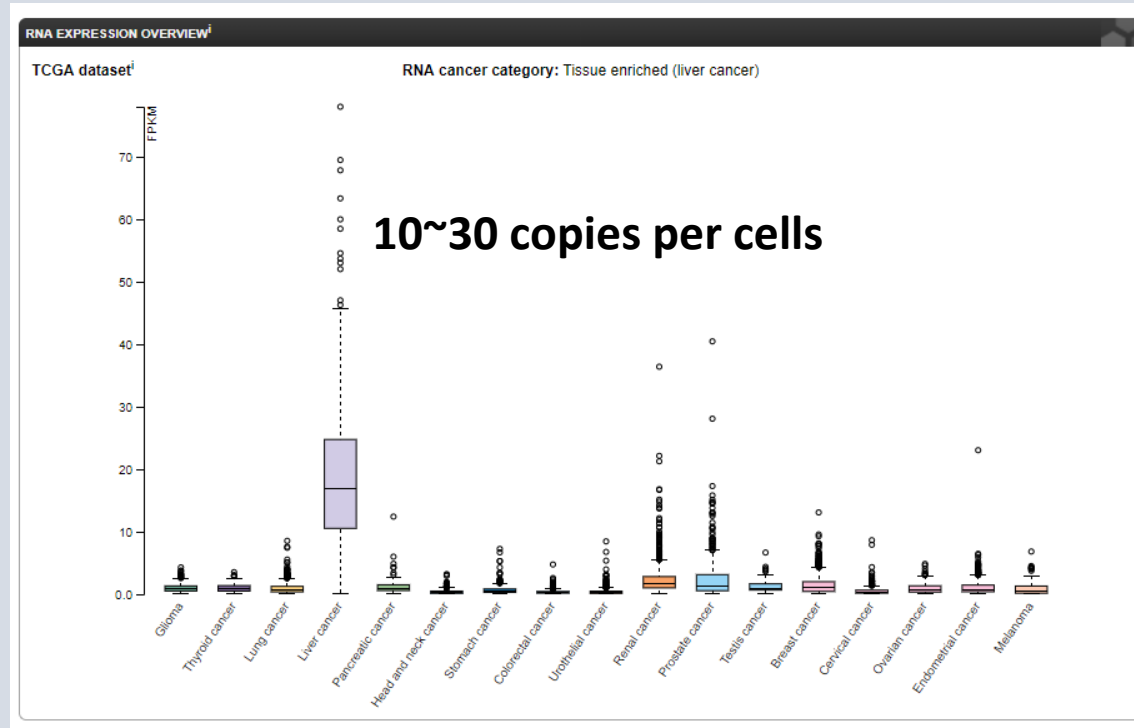
# Co-expression profile of LAYN, SARdh and gldc

Poor co-expression? Detection issue?

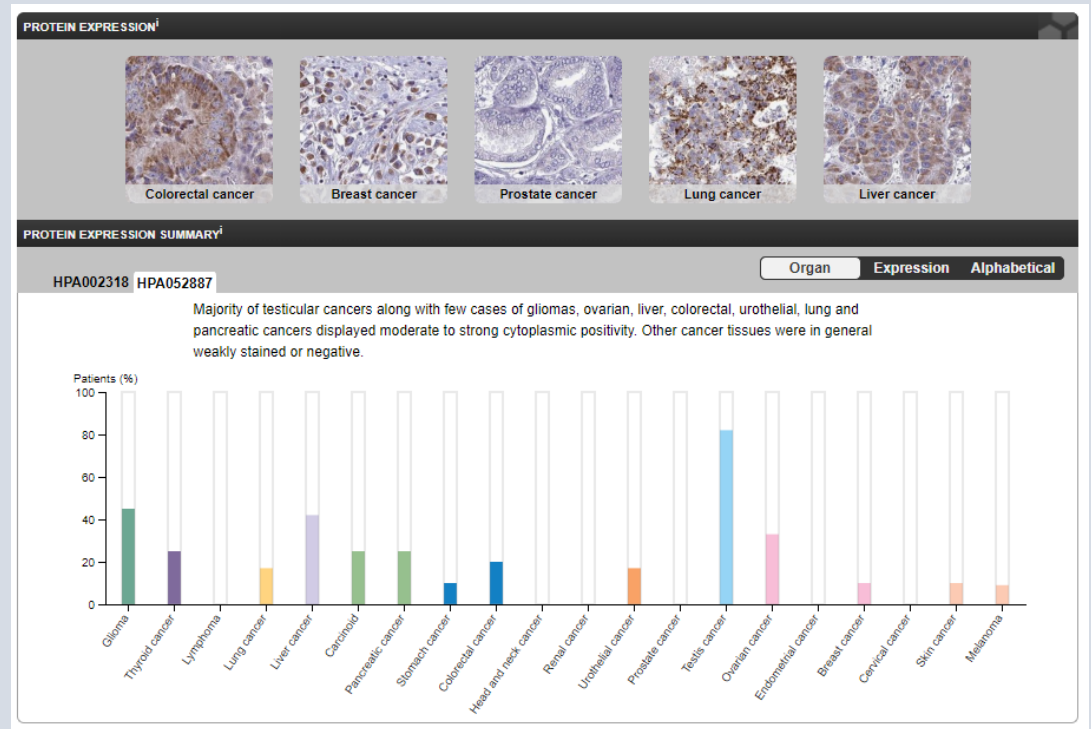
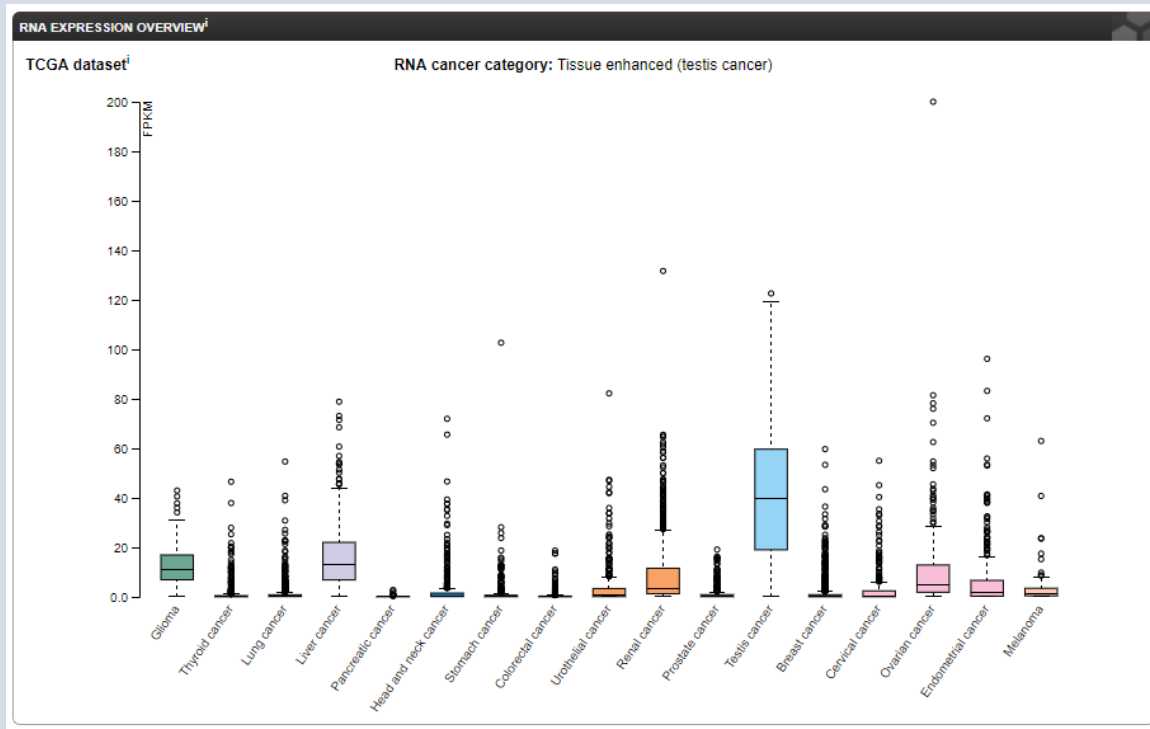




# SARDH expression in Cancer (The Human Protein Atlas)



# GLDC expression in Cancer (The Human Protein Atlas)



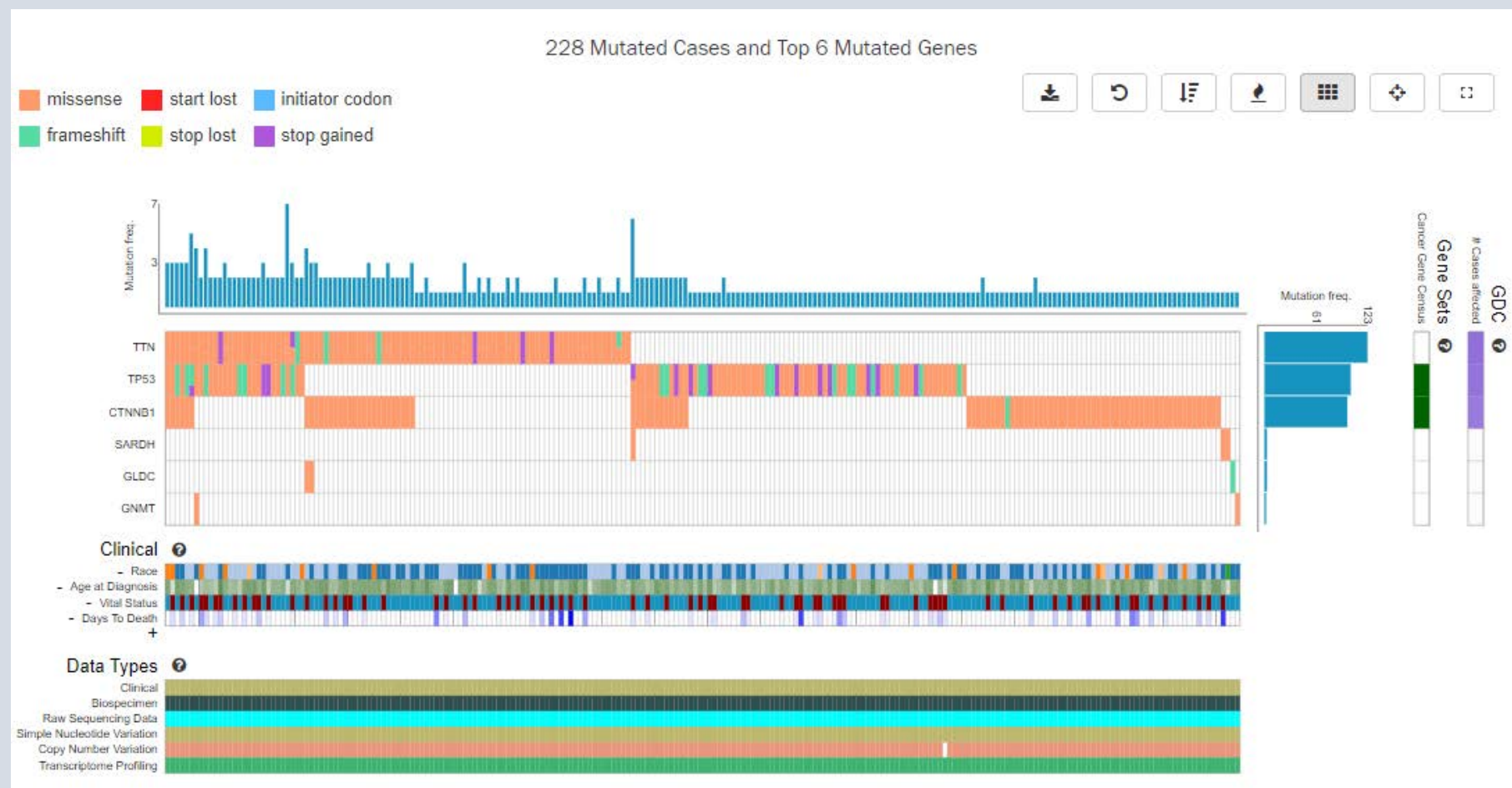
Without comparison with adjacent normal tissue, this inter-tissue comparison does not mean much. We can just naively say these two genes are at least expressed in liver cancer, and relatively liver

# SARDH and GLDC mutation in liver cancer (TCGA)

TTN, TP53, CTNNb1 are the top 3 genes frequently mutated genes in liver.

SARDH, GLDC and GNMT are rarely mutated in liver cancer.

- SARDH P383T (two cases)
- SARDH P118L
- GLDC A454S
- GLDC V322E
- GLDC L207Hfs\*25(frameshift)
- GNMT R157Q
- GNMT A230D



# Advanced TCGA analysis (Future Plan)

IDEA: Phenotype could be from “expression” not by “mutations” in CDS

TCGA datasets:

- Level 1 – Raw Data
- Level 2 – Processed Data
- Level 3 – Segmented or Interpreted Data
- Level 4 – Region of Interest Data
- Data access: TCGA data portal (Level 1-3, [TCGAbiolinks in R](#)), Firehose (Level 3 and 4, [RTGCAtoolbox in R](#))
- DNA Methylation and RNA-seq data is available from TCGA database but not accessible from Web

# Advanced TCGA Analysis – Methylation and expression datasets

- Download: Liver cancer RNA-seq dataset → I need to normalize by myself
- Download: Methylation dataset
- Integrated analysis of methylation and expression → “Starburst plot” with gene name. Does SARDH and GLDC is overexpressed? Correlated with methylation status?
- Survival analysis → Which categories? Does it really meaningful?
- Question: What should I compare with? Different subtypes of HCC? Compare stages (IV vs I)? I need something to compare to

# SARDh in literature

- Only 41 literatures in Pubmed.
- (poorly) associated with sarcosinemia and prostate cancer.

The screenshot shows the PubMed website interface. At the top, the PubMed logo and 'US National Library of Medicine National Institutes of Health' are visible. A search bar contains the term 'sardh'. Below the search bar, there are links for 'Create RSS', 'Create alert', and 'Advanced'. On the left side, there are filters for 'Article types' (Clinical Trial, Review, Customize ...), 'Text availability' (Abstract, Free full text, Full text), 'Publication dates' (5 years, 10 years, Custom range...), and 'Species' (Humans, Other Animals). There are also links for 'Clear all' and 'Show additional filters'. The main content area displays 'Search results' for 'Items: 1 to 20 of 50'. The results are sorted by 'Most Recent' and shown in 'Summary' format. The first four results are listed, each with a checkbox, a title link, authors, journal information, PMID, and a 'Free PMC Article' link. The results are: 1. 'Recurrent attacks of acute hepatic porphyria: major role of the chronic inflammatory response in the liver.' by Schmitt C, Lenglet H, Yu A, Delaby C, Benecke A, Lefebvre T, Letteron P, Paradis V, Wahlin S, Sandberg S, Harper P, SARDH E, Sandvik AK, Hov JR, Aarsand AK, Chiche L, Bazille C, Scoazec JY, To-Figueras J, Carrascal M, Abian J, Mirmiran A, Karim Z, Deybach JC, Puy H, Peoc'h K, Manceau H, Gouya L. J Intern Med. 2018 Jul;284(1):78-91. doi: 10.1111/joim.12750. Epub 2018 Mar 26. PMID: 29498764. 2. 'Differential expression of genes identified by suppression subtractive hybridization in liver and adipose tissue of gerbils with diabetes.' by Gong J, Du X, Li Z, Li X, Guo M, Lu J, Wang Y, Chen Z, Li C. PLoS One. 2018 Feb 2;13(2):e0191212. doi: 10.1371/journal.pone.0191212. eCollection 2018. PMID: 29394254. 3. 'Gene-centric analysis implicates nuclear encoded mitochondrial protein gene variants in migraine susceptibility.' by Stuart S, Benton MC, Eccles DA, Sutherland HG, Haupt LM, Lea RA, Griffiths LR. Mol Genet Genomic Med. 2017 Jan 17;5(2):157-163. doi: 10.1002/mgg3.270. eCollection 2017 Mar. PMID: 28361102. 4. 'Dataset of liver proteins of eu- and hypothyroid rats affected in abundance by any of three factors: in vivo exposure to hexabromocyclododecane (HBCD), thyroid status, gender differences.' by Miller I, Renaut J, Cambier S, Murk AJ, Gutleb AC, Serchi T. Data Brief. 2016 Aug 5;8:1344-7. doi: 10.1016/j.dib.2016.07.063. eCollection 2016 Sep. PMID: 27579339.



# SARDH-GNMT homeostasis

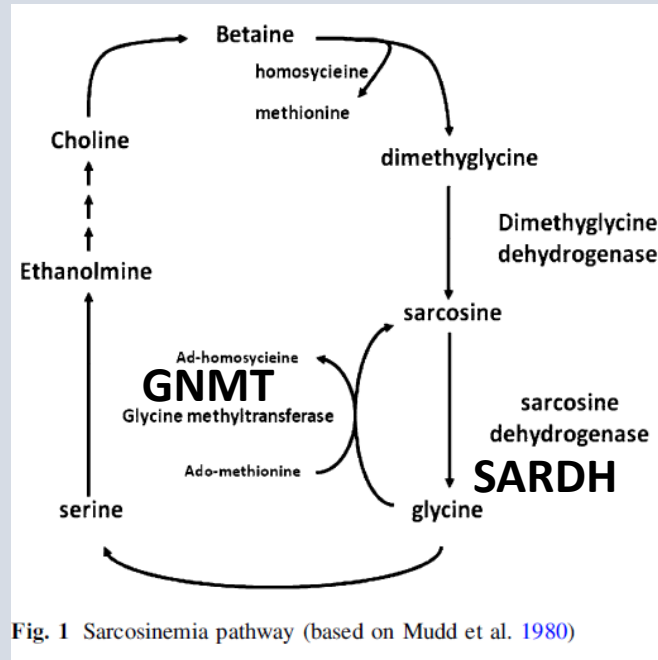
Hum Genet (2012) 131:1805–1810  
DOI 10.1007/s00439-012-1207-x

## ORIGINAL INVESTIGATION

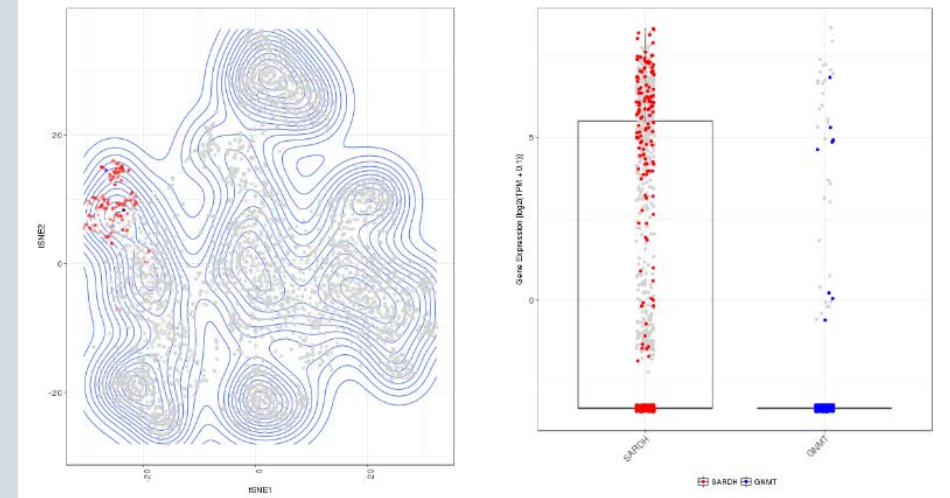
### Mutations in the sarcosine dehydrogenase gene in patients with sarcosinemia

Ifat Bar-joseph · Elon Pras · Haïke Reznik-Wolf · Dina Marek-Yagel · Almogit Abu-Horvitz · Maya Dushnitsky · Nurit Goldstein · Shlomit Rienstein · Michal Dekel · Ben Pode-Shakked · Joseph Zlotnik · Anelia Benarrosh · Philippe Gillery · Niklaus Hoffiger · Christiane Auray-Blais · Roselyne Garnotel · Yair Anikster

Among 6 family with Sarcosinemia (Increased level of sarcosine) investigated, four families contains four different mutations (**P287L, V71F, R723X, R514X**) or a uniparental disomy in the region of SARDH gene. **None of them were matched with mutations found in TCGA.**



## C4-CD8-LAYN in liver dataset



Caution: It could be simple detection problem

**Prediction:** Cells in C4-CD8-LAYN → Glycine (or downstream such as serine) accumulation and sarcosine depletion

# GNMT knock-out in HCC and T cells

## Characterization of a glycine N-methyltransferase gene knockout mouse model for hepatocellular carcinoma: Implications of the gender disparity in liver cancer susceptibility

Yi-Jen Liao<sup>1</sup>, Shih-Ping Liu<sup>2,3</sup>, Cheng-Ming Lee<sup>2</sup>, Chia-Hung Yen<sup>1</sup>, Pei-Chun Chuang<sup>1</sup>, Chia-Yen Chen<sup>2</sup>, Ting-Fen Tsai<sup>4</sup>, Shiu-Feng Huang<sup>5</sup>, Yan-Hwa Wu Lee<sup>6</sup> and Yi-Ming Arthur Chen<sup>1,2,7\*</sup>

<sup>1</sup>Molecular Medicine Program, Institute of Public Health, School of Medicine, National Yang-Ming University, Taipei, Taiwan

<sup>2</sup>AIDS Prevention and Research Center, National Yang-Ming University, Taipei, Taiwan

<sup>3</sup>Center for Neuropsychiatry, China Medical University and Hospital, Taichung, Taiwan

<sup>4</sup>Faculty of Life Sciences and Institute of Genome Sciences, National Yang-Ming University, Taipei, Taiwan

<sup>5</sup>Division of Molecular and Genomic Medicine, National Health Research Institute, Miaoli, Taiwan

<sup>6</sup>Institute of Biochemistry and Molecular Biology, School of Life Sciences, National Yang-Ming University, Taipei, Taiwan

<sup>7</sup>Department of Microbiology, School of Medicine, National Yang-Ming University, Taipei, Taiwan

## Role of Glycine N-Methyltransferase in the Regulation of T-Cell Responses in Experimental Autoimmune Encephalomyelitis

Chung-Hsien Li,<sup>1,2</sup> Ming-Hong Lin,<sup>3</sup> Shih-Han Chu,<sup>1,2</sup> Pang-Hsien Tu,<sup>4</sup> Cheng-Chieh Fang,<sup>1,2</sup> Chia-Hung Yen,<sup>2,5</sup> Peir-In Liang,<sup>6</sup> Jason C Huang,<sup>7</sup> Yu-Chia Su,<sup>8</sup> Huey-Kang Sytwu,<sup>3</sup> and Yi-Ming Arthur Chen<sup>2,9,10</sup>

<sup>1</sup>Institute of Microbiology and Immunology, National Yang-Ming University, Taipei, Taiwan; <sup>2</sup>Center for Infectious Disease and Cancer Research (CICAR), Kaohsiung Medical University, Kaohsiung, Taiwan; <sup>3</sup>Department and Graduate Institute of Microbiology and Immunology, National Defense Medical Center, Taipei, Taiwan; <sup>4</sup>Institute of Biomedical Sciences, Academia Sinica, Taipei, Taiwan; <sup>5</sup>Graduate Institute of Natural Products, College of Pharmacy, Kaohsiung Medical University, Kaohsiung, Taiwan;

<sup>6</sup>Department of Pathology, Kaohsiung Medical University Hospital, Kaohsiung Medical University, Kaohsiung, Taiwan; <sup>7</sup>Department of Biotechnology and Laboratory Science in Medicine, National Yang-Ming University, Taipei, Taiwan; <sup>8</sup>National Laboratory Animal Center, National Applied Research Laboratories; <sup>9</sup>Institute of Biomedical Sciences, National Sun Yat-sen University, Kaohsiung, Taiwan; and <sup>10</sup>Department of Microbiology and Immunology, Institute of Medical Research and Institute of Clinical Medicine, College of Medicine, Kaohsiung Medical University, Kaohsiung, Taiwan

GNMT -/- presumably phenocopies SARDH overexpression

GNMT -/- caused spontaneous HCC.

GNMT -/- defective T cells response in EAE, lower IFN $\gamma$  and IL-17A production, MOG induced Th17 was inhibited but Treg numbers were increased.

Proposed mechanism: GNMT  $\rightarrow$  mTORC1  $\rightarrow$  T cell activation

# GLDC in literature

**GLDC promotes cellular transformation in NSCLC,**

Via glycolysis and glycine/serine metabolism  
→ pyrimidine metabolism to regulate cancer proliferation

GLDC is generally overexpressed in many different cancers including liver

Cell

## **Glycine Decarboxylase Activity Drives Non-Small Cell Lung Cancer Tumor-Initiating Cells and Tumorigenesis**

Wen Cai Zhang,<sup>1,3</sup> Ng Shyh-Chang,<sup>1</sup> He Yang,<sup>5</sup> Amit Rai,<sup>6</sup> Shivshankar Umashankar,<sup>6,7</sup> Siming Ma,<sup>1</sup> Boon Seng Soh,<sup>1</sup> Li Li Sun,<sup>1</sup> Bee Choo Tai,<sup>11</sup> Min En Nga,<sup>9</sup> Kishore Kumar Bhakoo,<sup>12</sup> Senthil Raja Jayapal,<sup>13</sup> Massimo Nichane,<sup>1</sup> Qiang Yu,<sup>2</sup> Dokeu A. Ahmed,<sup>4</sup> Christie Tan,<sup>4</sup> Wong Poo Sing,<sup>10</sup> John Tam,<sup>10</sup> Agasthian Thirugananam,<sup>14</sup> Monireh Soroush Noghabi,<sup>1</sup> Yin Huei Pang,<sup>9</sup> Haw Siang Ang,<sup>5</sup> Wayne Mitchell,<sup>16,17</sup> Paul Robson,<sup>1</sup> Philipp Kaldis,<sup>13</sup> Ross Andrew Soo,<sup>5,8</sup> Sanjay Swarup,<sup>6,7</sup> Elaine Hsuen Lim,<sup>3,8,15,\*</sup> and Bing Lim<sup>1,18,\*</sup>

# Future Plan

- Advanced TCGA analysis
- A few more candidates as backup
- Amgen Body Map analysis
- Read literatures