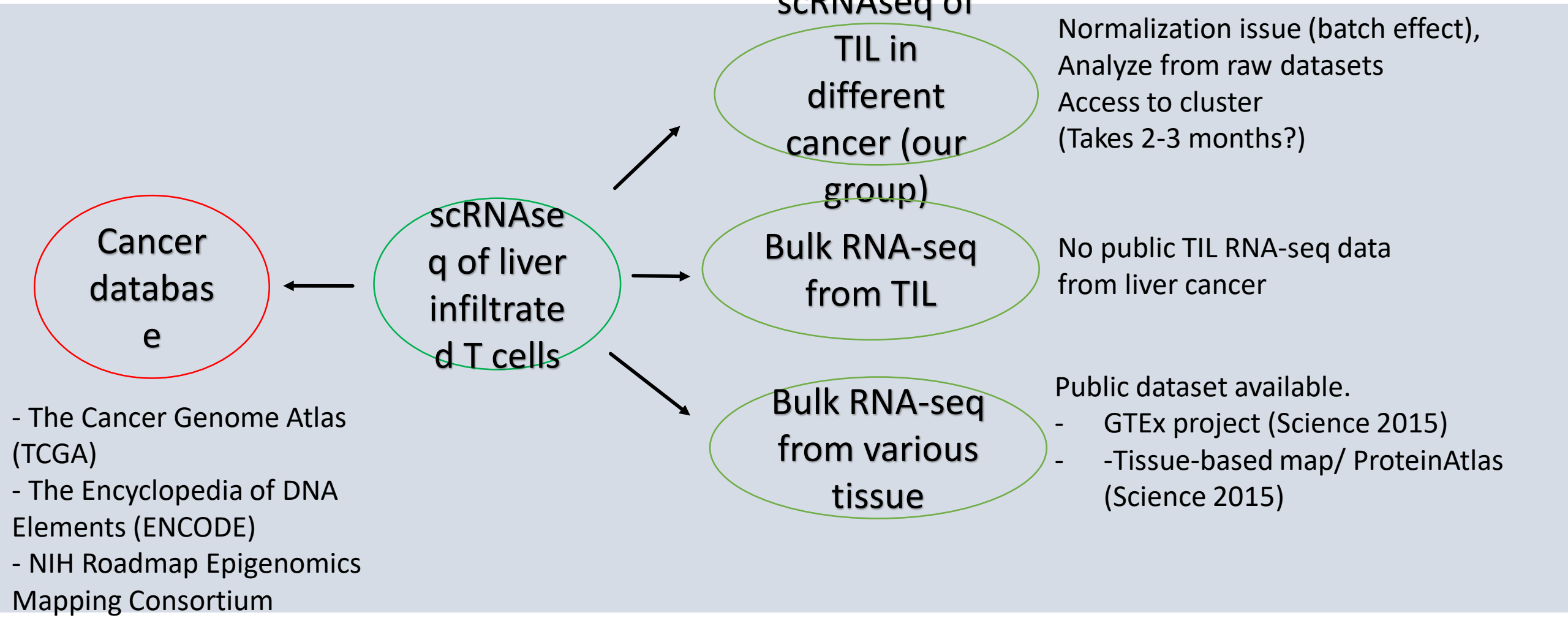
Tissue enrichment analysis and candidate selection

Zheng et al. 2017 , T cell scRNAseq of Liver cancer. C4-CD8-LAYN cluster

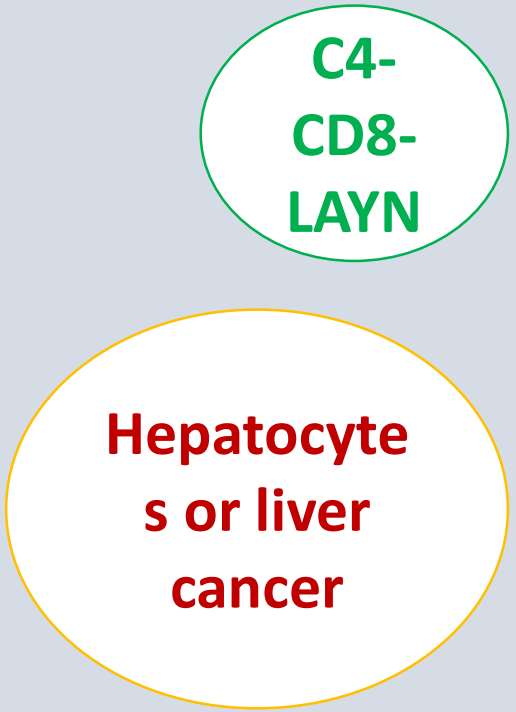
20180905

How to extract useful information from Public datasets?



IDEA / Experimental Design

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



Gene X: Overexpression compromise CD8 function

Gene X:

- Expressed specifically in hepatocytes (not in other tissue).
- Oncogenic when overexpressed

A "single" Inhibitor with "dual" function:
By acting on both CD8 T cells and liver cancers

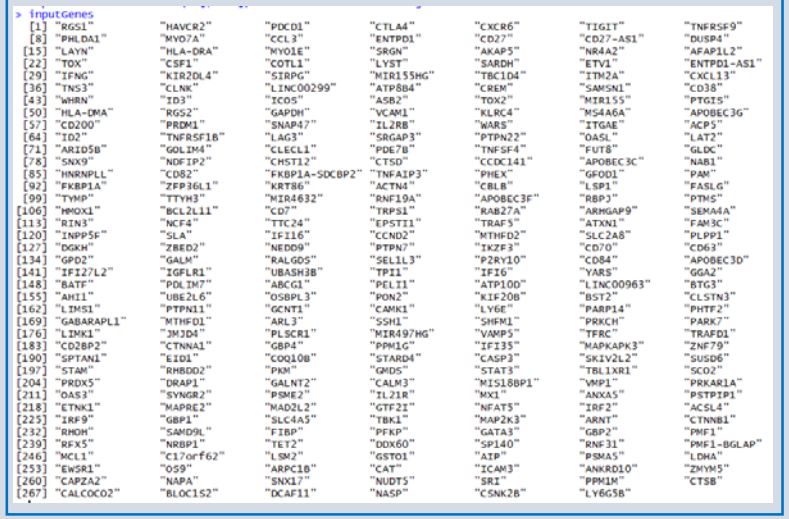
Tissue Enrichment analysis – Scheme (20180901)

Table S2
Zheng et al., 2017

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

Extract GeneID (272 genes)

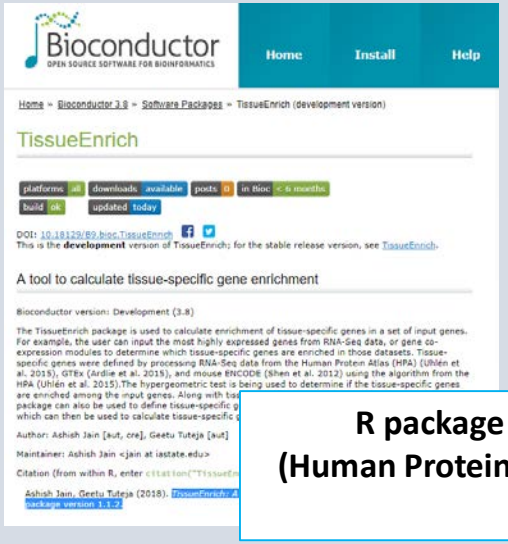
2 entrez ID were not assigned with gene name (“NA”)



```
> inputGenes
  [1] "RGS1"      "HAVCR2"    "PDCD1"     "CTLA4"     "CXCR6"     "TIGIT"     "TNFRSF9"
  [8] "PHLDA1"    "MYO7A"     "CCL3"      "ENTPD1"    "CD27"      "CD27-AS1"  "DUSP4"
 [15] "LAYN"      "HLA-DRA"   "MYO1E"     "SRGN"      "AKAP5"     "NR4A2"     "AFAP1L2"
 [22] "TOX"       "CSF1"      "COTL1"     "LYST"      "SARDH"     "ETV1"      "ENTPD1-AS1"
 [29] "IFNG"      "KIR2DL4"   "SIRPG"     "MIR155HG"  "TBC1D4"    "ITM2A"     "CXCL13"
 [36] "TNS3"      "CLNK"      "LINC00299" "ATP8B4"    "CREM"      "SAMSN1"    "CD38"
 [43] "WHRN"      "ID3"       "ICOS"      "ASB2"      "TOX2"      "MIR155"    "PTGIS"
 [50] "HLA-DMA"   "RGS2"      "GAPDH"     "VCAM1"     "KLRC4"     "MS4A6A"    "APOBEC3G"
 [57] "CD200"     "PRDM1"     "SNAP47"    "IL2RB"     "WARS"      "ITGAE"     "ACP5"
 [64] "ID2"       "TNFRSF1B"  "LAG3"      "SRGAP3"    "PTPN22"    "OASL"      "LAT2"
 [71] "ARID5B"    "GOLIM4"    "CLEC1L"    "PDE7B"     "TNFSF4"    "FUT8"      "GLDC"
 [78] "SNX9"      "NDFIP2"    "CHST12"    "CTSO"      "CCDC141"   "APOBEC3C"  "NAB1"
 [85] "HNRNPLL"   "CD82"      "FKBP1A-SDCBP2" "TNFAIP3"   "PHEX"      "GFOD1"     "PAM"
 [92] "FKBP1A"    "ZFP36L1"   "KRT86"     "ACTN4"     "CBLB"      "LSP1"      "FASLG"
 [99] "TYMP"      "TTYH3"     "MIR4632"   "RNF19A"    "APOBEC3F"  "RBPJ"      "PTMS"
[106] "HMOX1"     "BCL2L11"   "CD7"       "TRPS1"     "RAB27A"    "ARHGAP9"   "SEMA4A"
[113] "RIN3"      "NCF4"      "TTC24"     "EPSTI1"    "TRAF5"     "ATXN1"     "FAM3C"
[120] "INPP5F"    "SLA"       "IFI16"     "CCND2"     "MTHFD2"    "SLC2A8"    "PLPP1"
[127] "DGKH"      "ZBED2"     "NEDD9"     "PTPN7"     "IKZF3"     "CD70"      "CD69"
[134] "GPD2"      "GALM"      "RALGDS"    "SELL3"     "P2RY10"    "CD84"      "APOBEC3D"
[141] "IFI27L2"   "IGFLR1"   "UBASH3B"   "TPI1"      "IFI6"      "YARS"      "GGA2"
[148] "BATF"      "PDLIM7"    "ABCG1"     "PELI1"     "ATP10D"    "LINC00963" "BTG3"
[155] "AHI1"      "UBE2L6"    "OSBPL3"    "PON2"      "KIF20B"    "BST2"      "CLSTN3"
[162] "LIMS1"     "PTPN11"    "GCNT1"     "CAMK1"     "LY6E"      "PARP14"    "PHTF2"
[169] "GABARAPL1" "MTHFD1"    "ARL3"      "SSH1"      "SHFM1"     "PRKCH"     "PARK7"
[176] "LIMK1"     "JMJD4"     "PLSCR1"    "MIR497HG"  "VAMP5"     "TFRC"      "TRAFD1"
[183] "CD28P2"    "CTNNA1"    "GBP4"      "PPM1G"     "IFI35"     "MAPKAPK3"  "ZNF79"
[190] "SPTAN1"    "EID1"      "COQ10B"    "STARD4"    "CASP3"     "SKIV2L2"   "SUSD6"
[197] "STAM"      "RHBDD2"    "PKM"       "GMDS"      "STAT3"     "TBL1XR1"   "SCO2"
[204] "PRDX5"     "DRAP1"     "GALNT2"    "CALM3"     "MIS18BP1"  "VMP1"      "PRKAR1A"
[211] "OAS3"      "SYNGR2"    "PSME2"     "IL21R"     "MX1"       "ANXA5"     "PSTPIP1"
[218] "ETNK1"     "MAPRE2"    "MAD2L2"    "GTF2I"     "NFAT5"     "IRF2"      "AC5L4"
[225] "IRF9"      "GBP1"      "SLC4A5"    "TBK1"      "MAP2K3"    "ARNT"      "CTNNB1"
[232] "RHOH"      "SAMD9L"    "FIBP"      "PFKP"      "GATA3"     "GBP2"      "PMF1"
[239] "RFX5"      "NRBP1"     "TET2"      "DDX60"     "SP140"     "RNF31"     "PMF1-BGLAP"
[246] "MCL1"      "C17orf62"  "LSM2"      "GSTO1"     "AIP"       "PSMA5"     "LDHA"
[253] "EWSR1"     "OS9"       "ARPC1B"    "CAT"       "ICAM3"     "ANKRD10"   "ZMYM5"
[260] "CAPZA2"    "NAPA"      "SNX17"     "NUDT5"     "SRI"       "PPM1M"     "CTSB"
[267] "CALCOCO2"  "BLOC1S2"    "DCAF11"    "NASP"      "C5NK2B"    "LY6G5B"
```

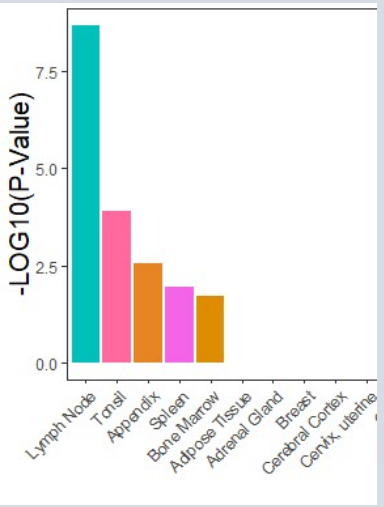
Tissue Enrichment Analysis

Human RNA-seq dataset



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

Visualization

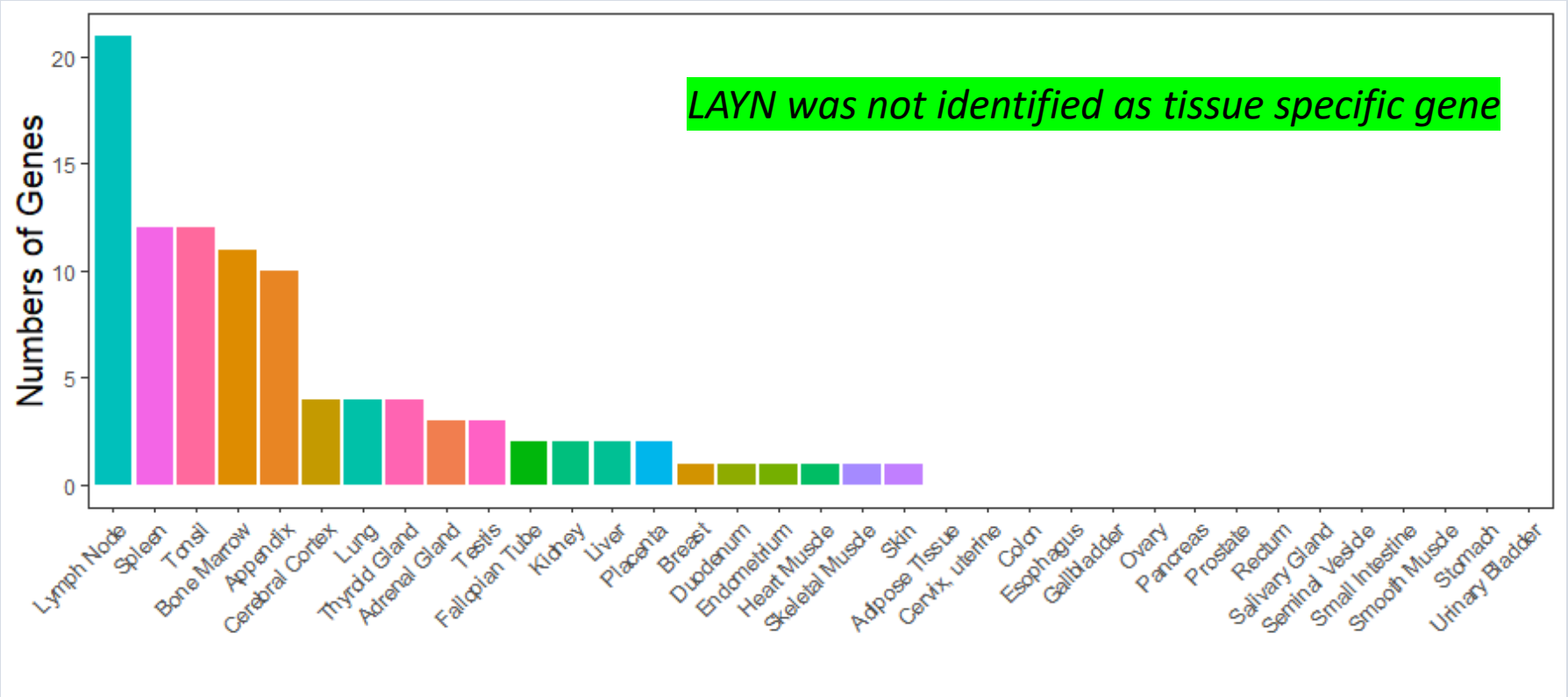


enrichPlot
ggplot2

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

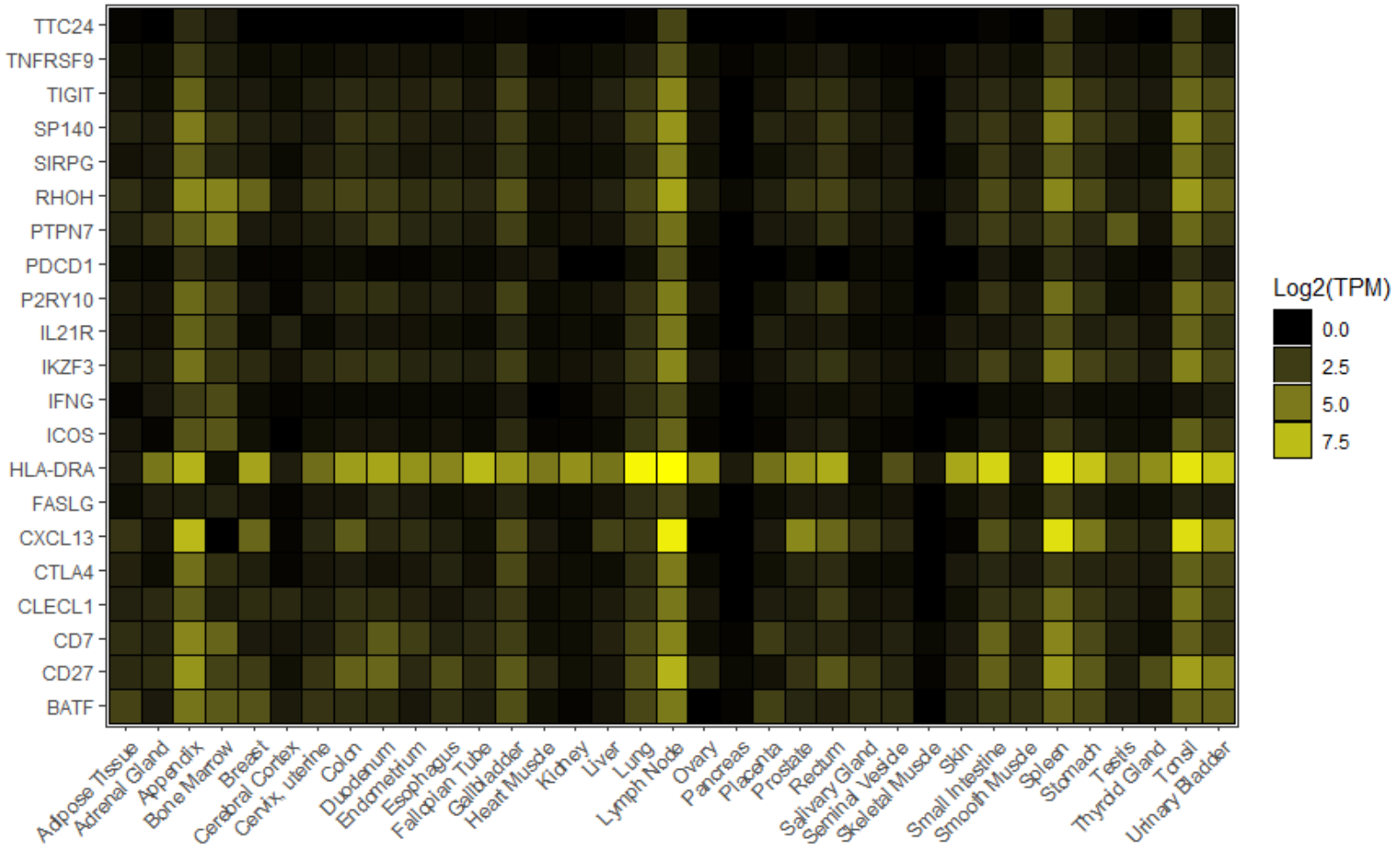


21 Lymph node specific genes

(1) Lymph node, spleen, appendix and bone marrow partially shared gene list

(2) LAYN was not identified as enriched genes from any tissues. Probably because LAYN is activation/exhaustion specific, not because universal expression.

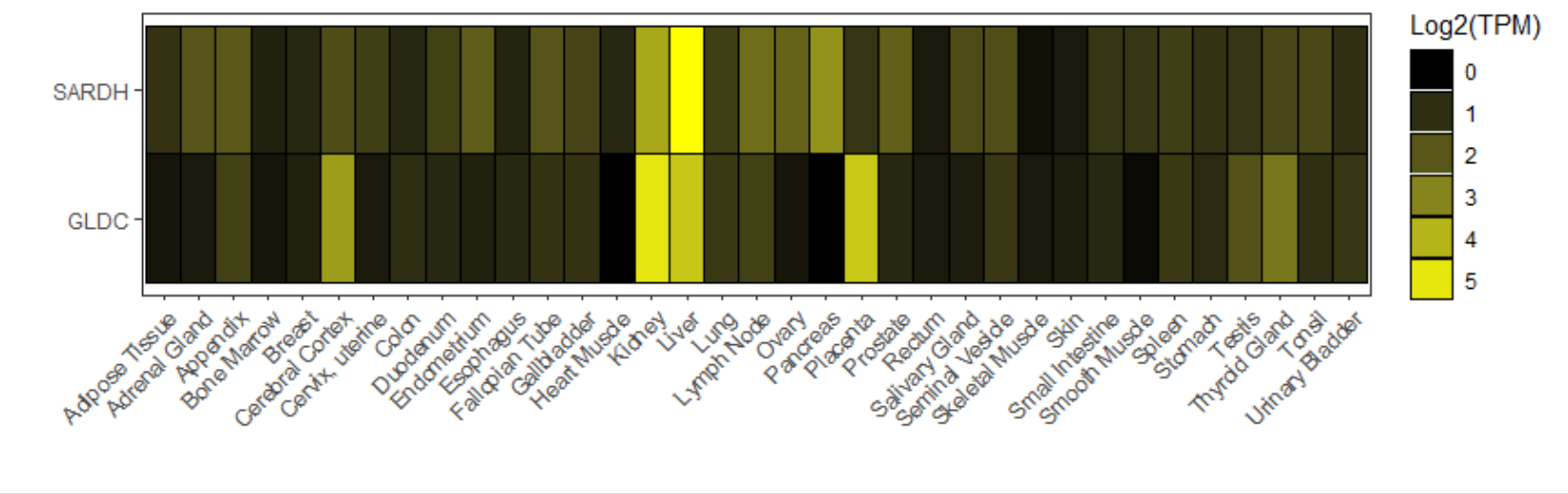
(3) These could be potential genes to follow if trying to target only immune cells.



Two liver specific genes

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

SARDH and GLDC

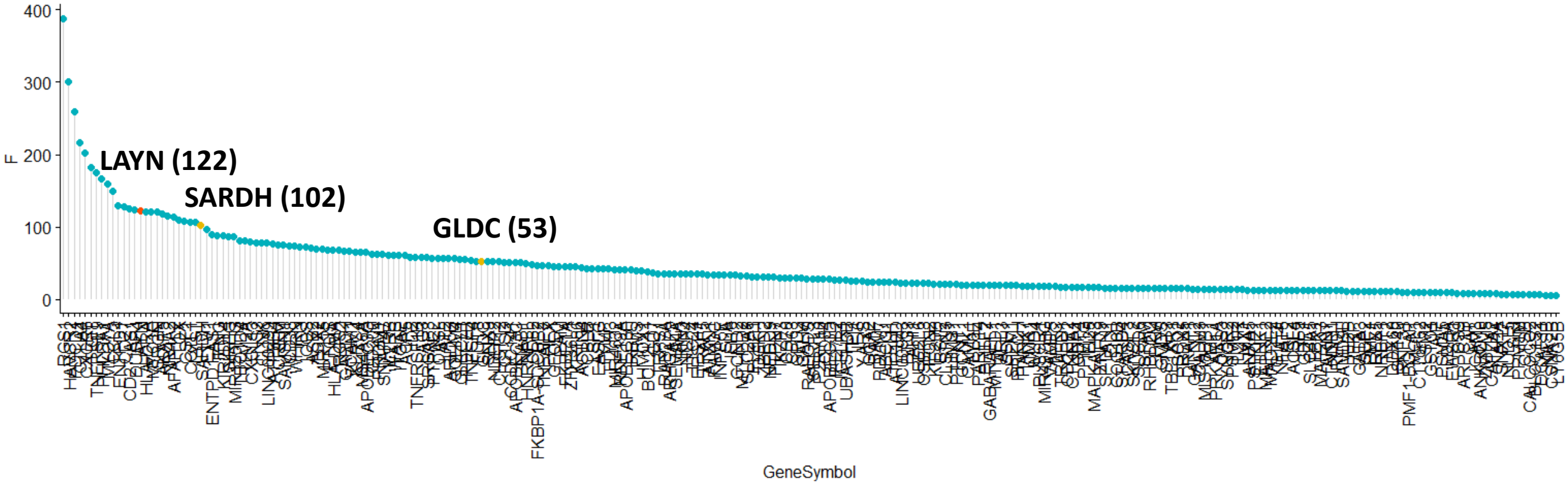


Plan: Validate this results using mouse tissue with qRT-PCR
(Perfusion → liver, lung, heart, lymph node, spleen, bone marrow, Thymus, and Kidney) :
Do I need animal protocol for this?

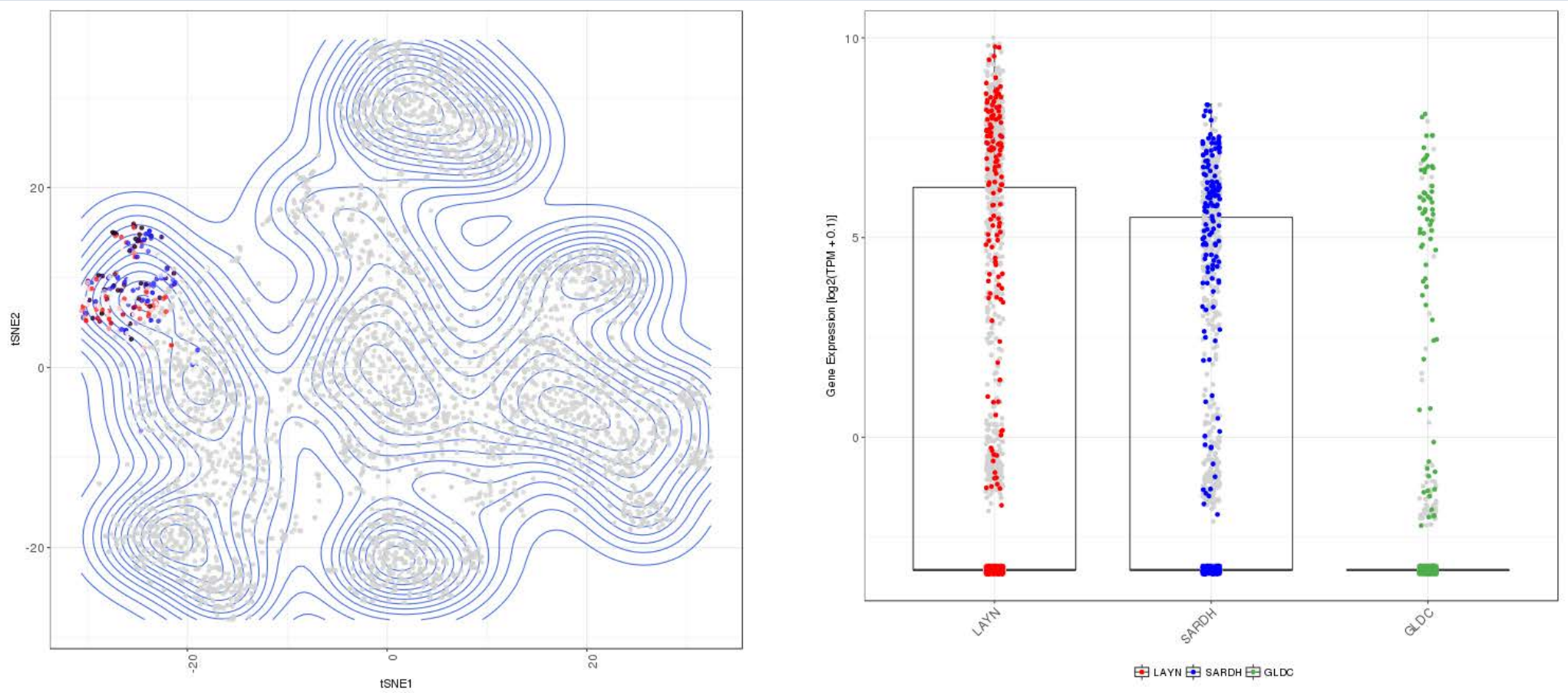
Fold Changes of Targets of interests

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

What is F value?

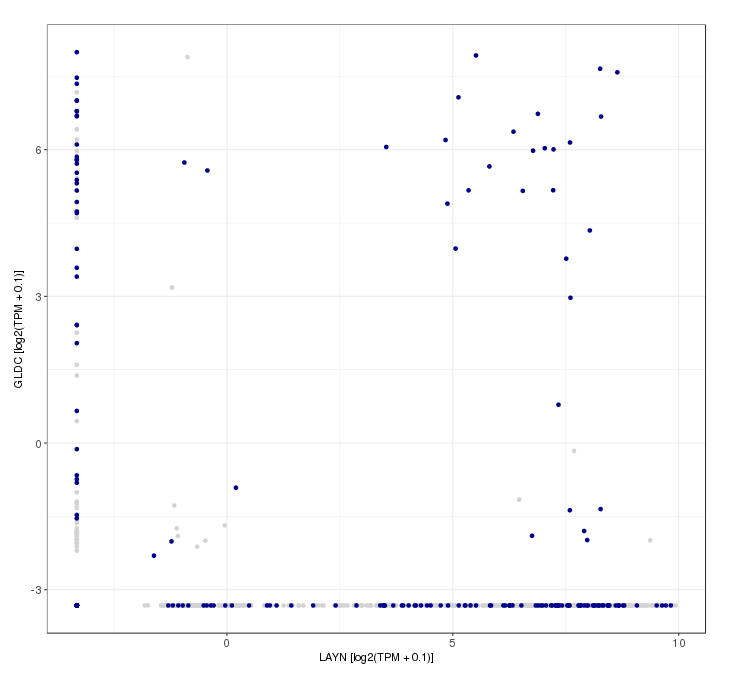


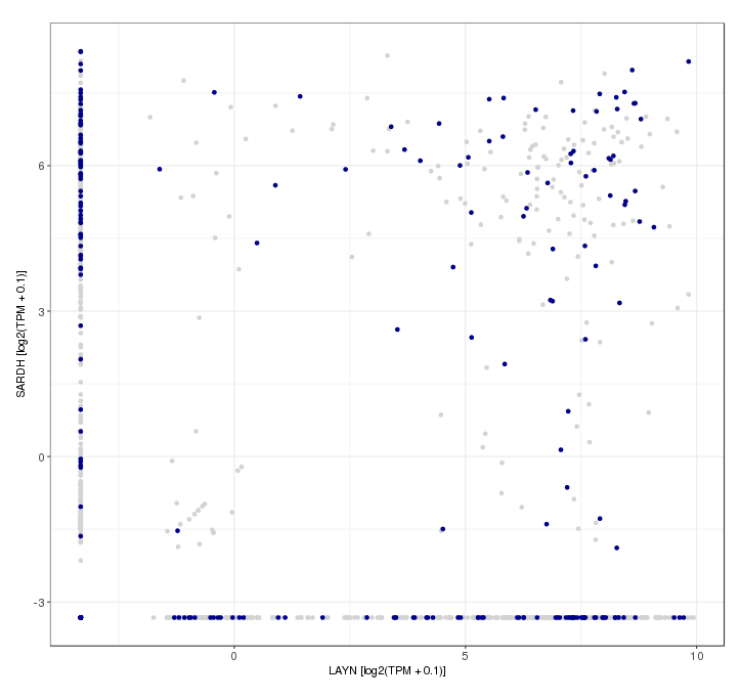
Three gene expression in C4-CD8-LAYN cluster (Liver)

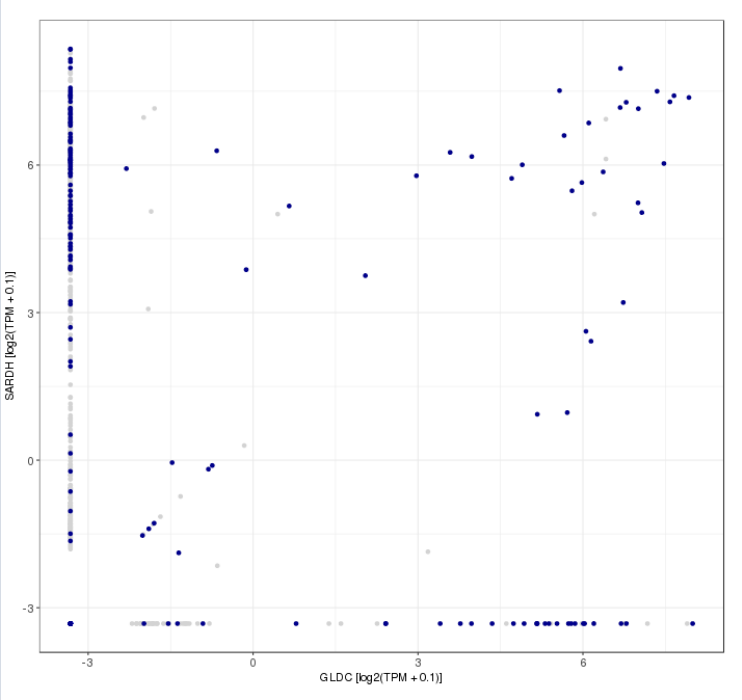


Co-expression profile of LAYN, SARdh and gldc

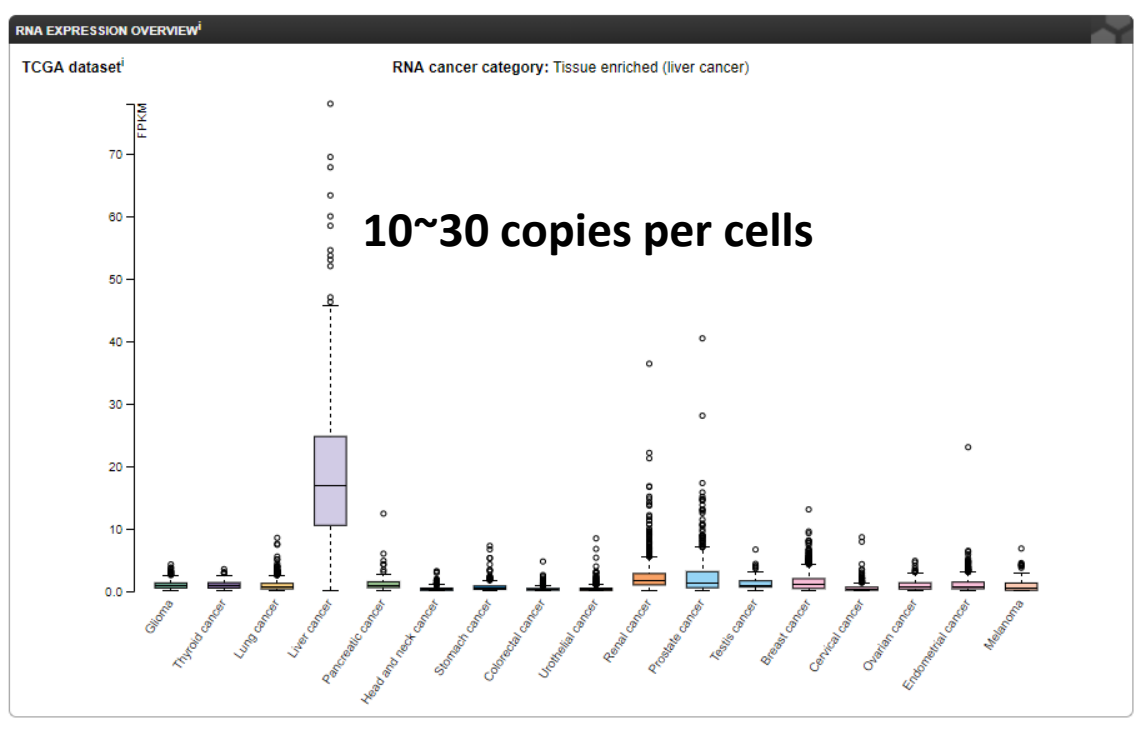
Poor co-expression? Detection issue?

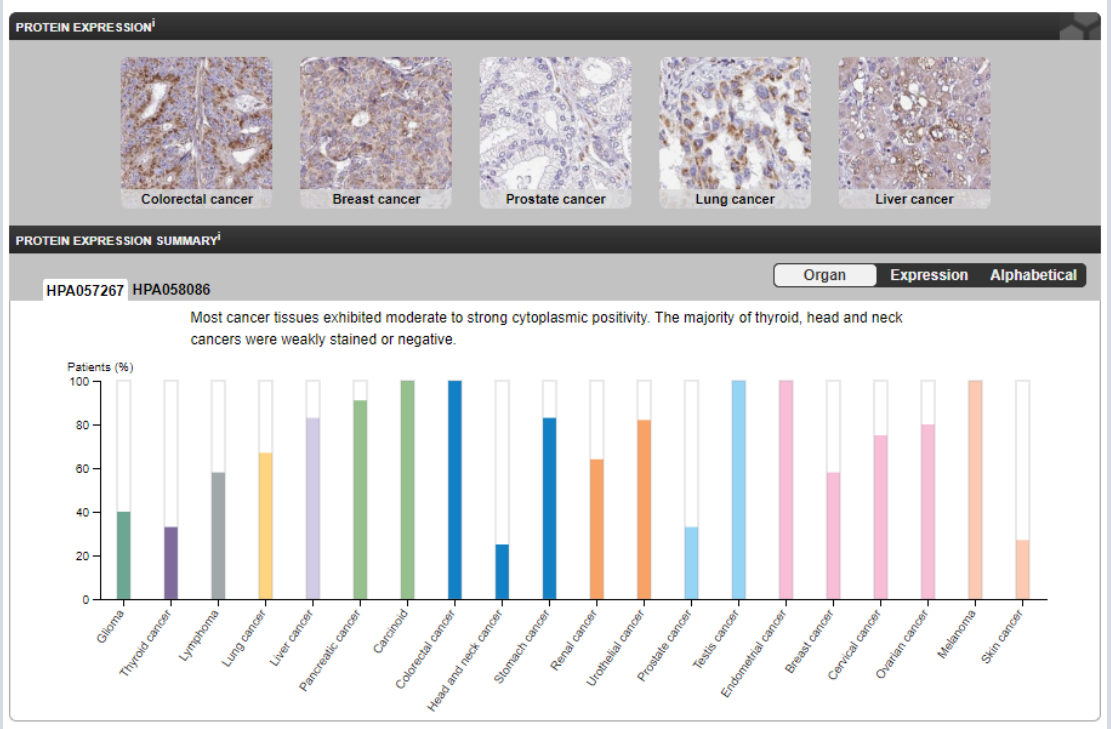




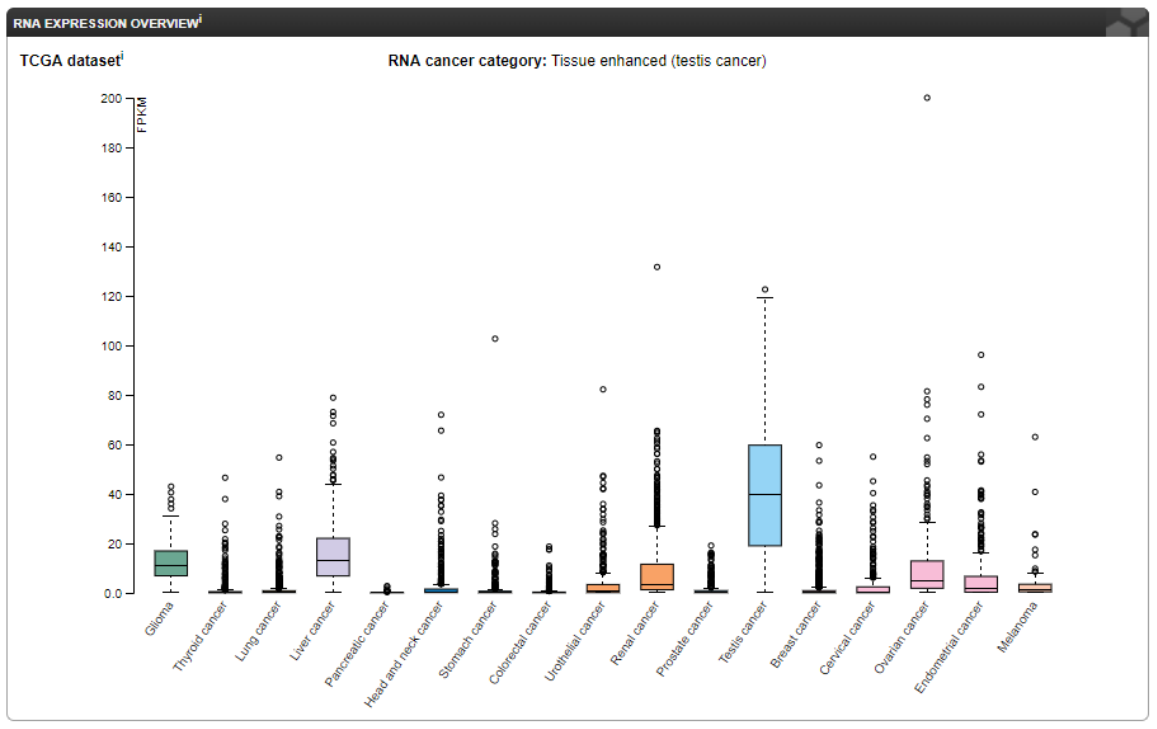


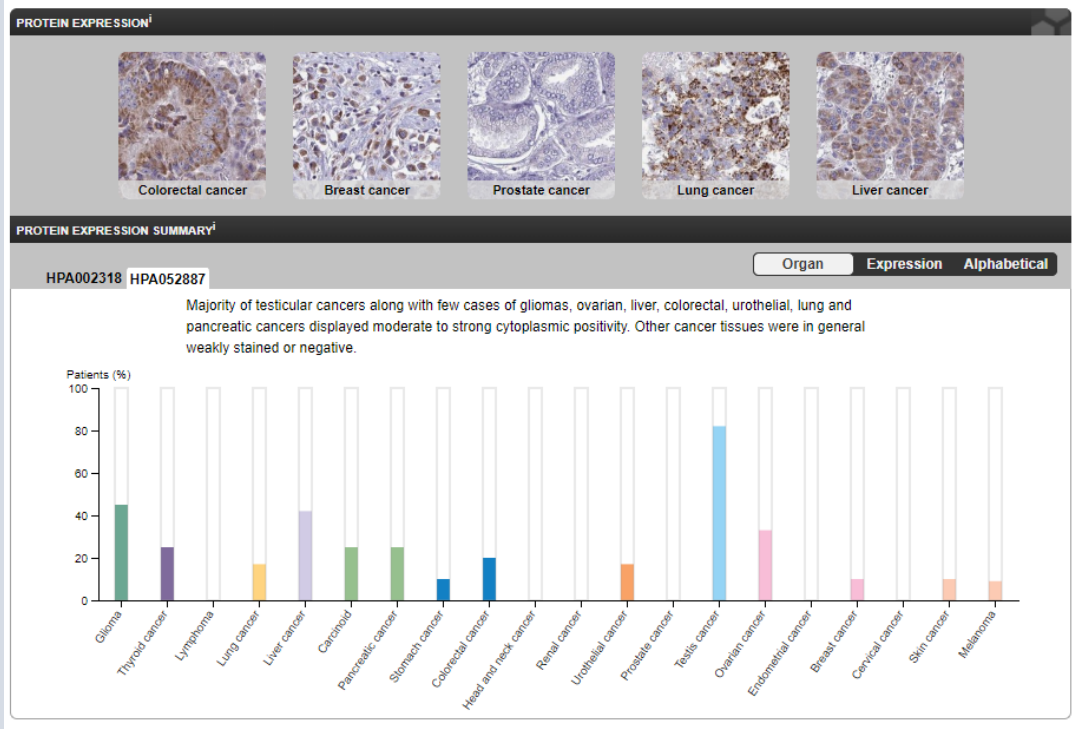
SARDH expression in Cancer (The Human Protein Atlas)





GLDC expression in Cancer (The Human Protein Atlas)





Without comparison with adjacent normal tissue, this inter-tissue comparison does not mean mu

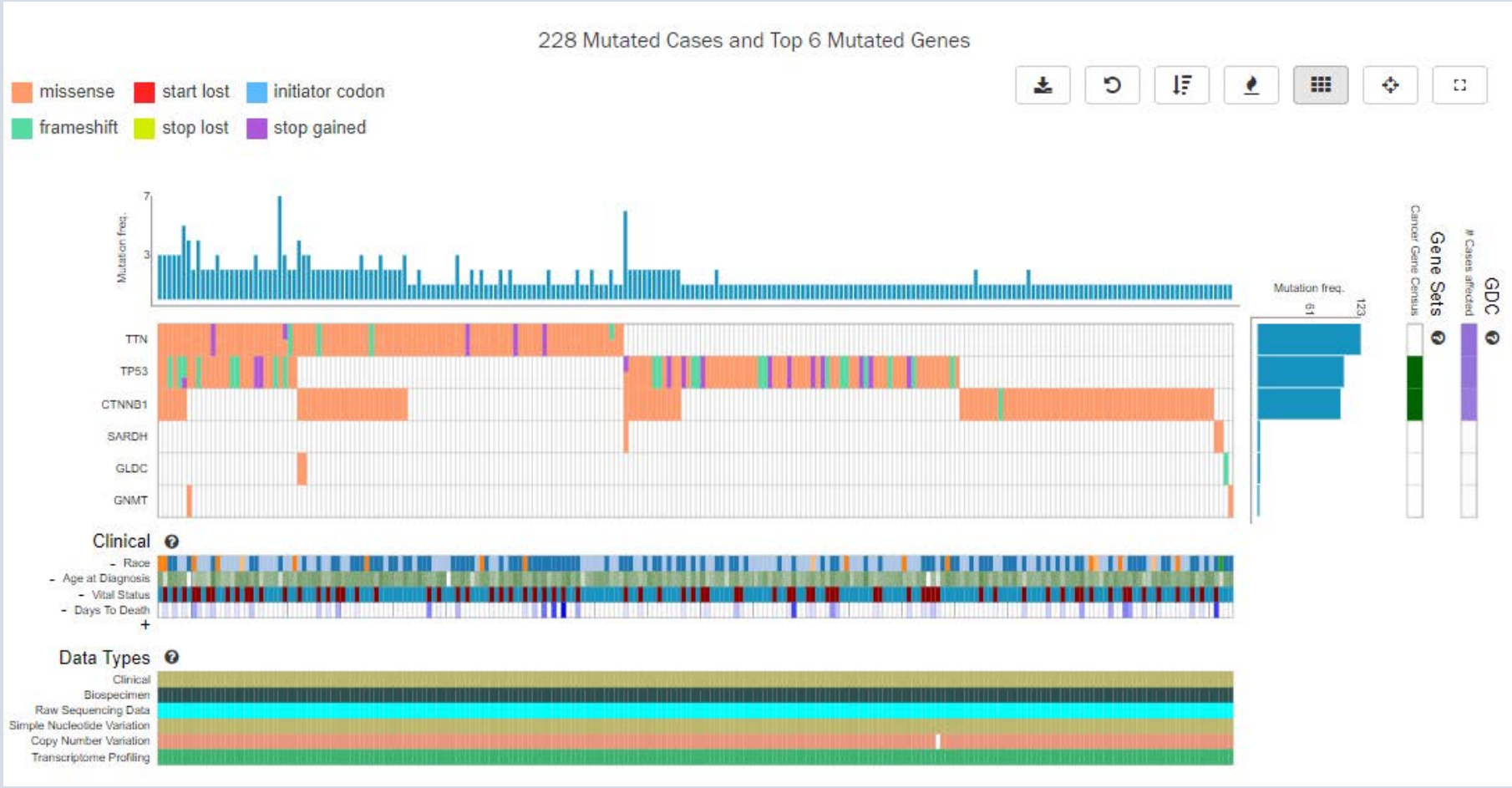
We can just naively say these two genes are at least expressed in liver cancer, and relatively liver

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), Firehorse (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

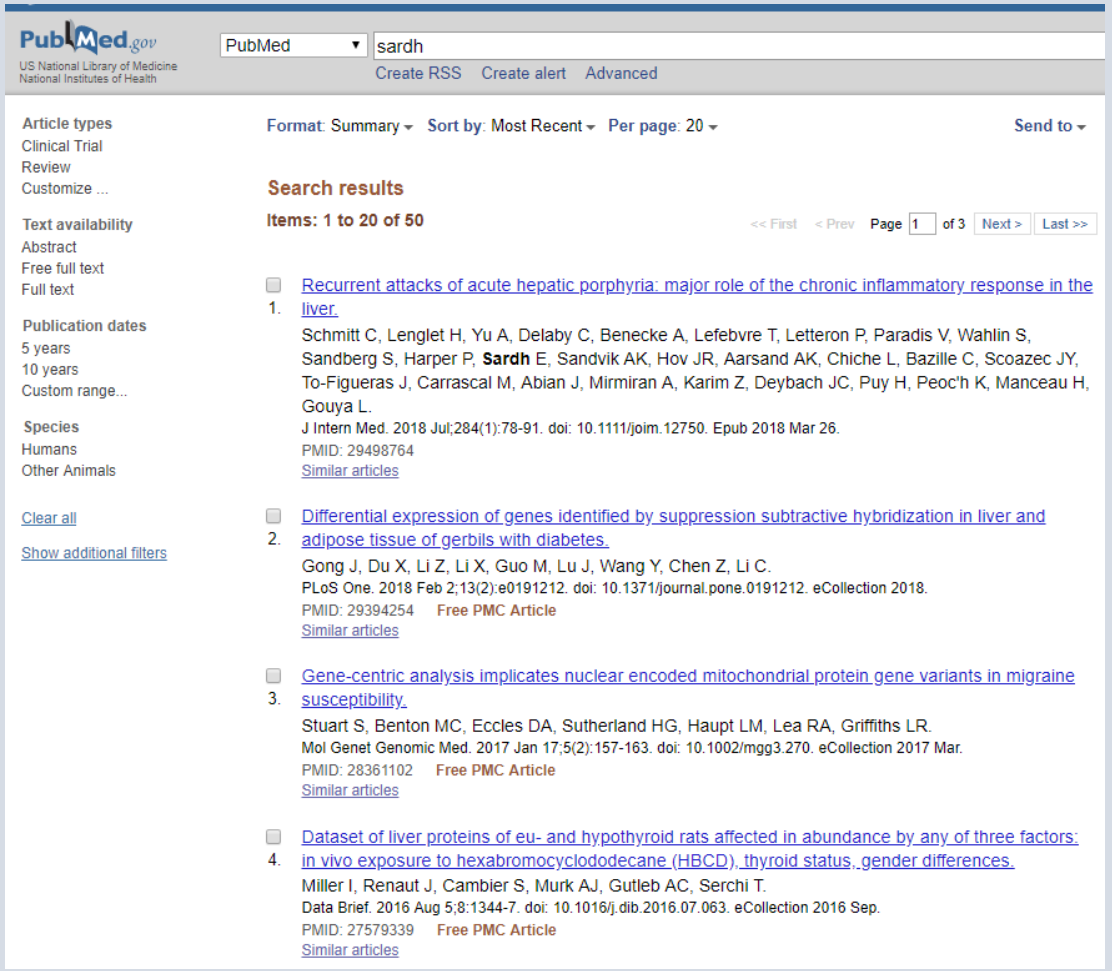
- Download: Liver cancer RNA-seq dataset → I need to normalize by myself
- Download: Methylation dataset

- 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.



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 · Haike Reznik-Wolf · Dina Marek-Yagel · Almogit Abu-Horvitz · Maya Dushnitzky · Nurit Goldstein · Shlomit Rienstein · Michal Dekel · Ben Pode-Shakked · Joseph Zlotnik · Anelia Benarrosh · Philippe Gillery · Niklaus Hofliger · 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.

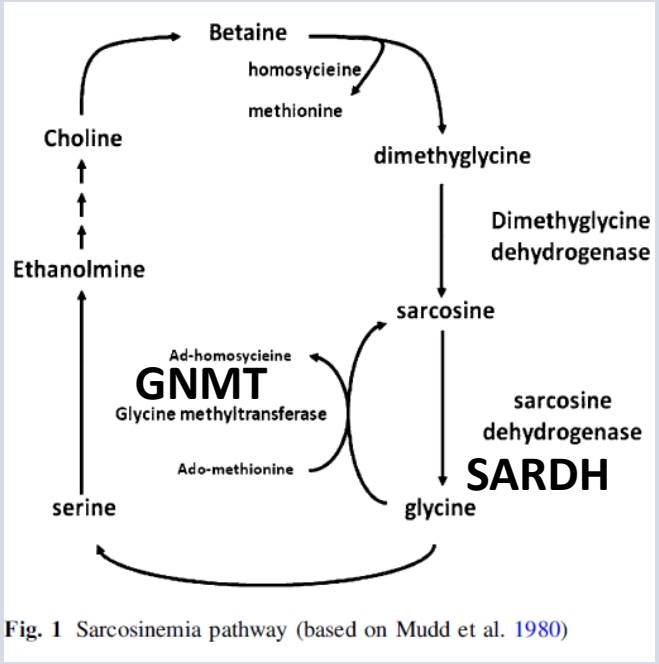
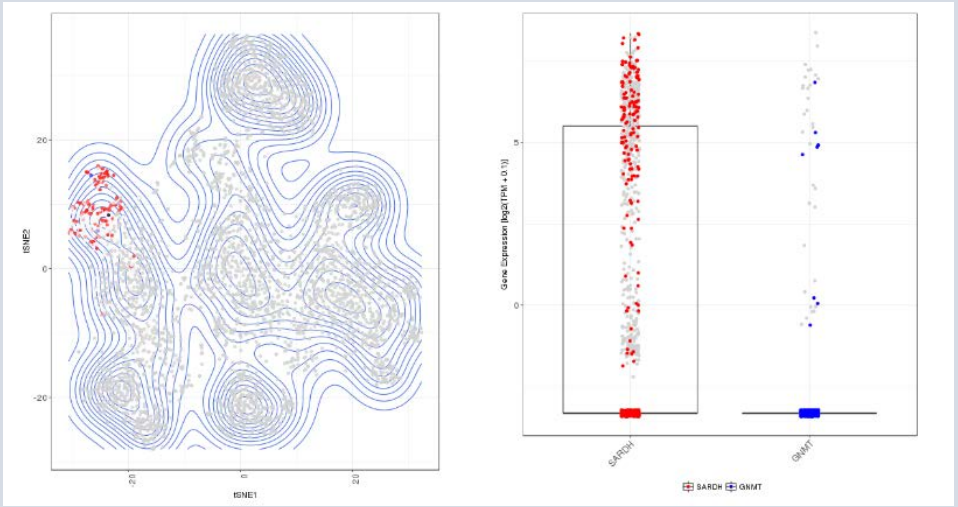


Fig. 1 Sarcosinemia pathway (based on Mudd et al. 1980)

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[1], Shih-Ping Liu[2,3], Cheng-Ming Lee[2], Chia-Hung Yen[1], Pei-Chun Chuang[1], Chia-Yen Chen[2], Ting-Fen Tsai[4], Shiu-Feng Huang[5], Yan-Hwa Wu Lee[6] and Yi-Ming Arthur Chen[1,2,7]*

[1]*Molecular Medicine Program, Institute of Public Health, School of Medicine, National Yang-Ming University, Taipei, Taiwan*
[2]*AIDS Prevention and Research Center, National Yang-Ming University, Taipei, Taiwan*
[3]*Center for Neuropsychiatry, China Medical University and Hospital, Taichung, Taiwan*
[4]*Faculty of Life Sciences and Institute of Genome Sciences, National Yang-Ming University, Taipei, Taiwan*
[5]*Division of Molecular and Genomic Medicine, National Health Research Institute, Miaoli, Taiwan*
[6]*Institute of Biochemistry and Molecular Biology, School of Life Sciences, National Yang-Ming University, Taipei, Taiwan*
[7]*Department of Microbiology, School of Medicine, National Yang-Ming University, Taipei, Taiwan*

GNMT -/- presumably phenocopies SARDH overexpression

GNMT -/- caused spontaneous HCC.

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

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

[1]Institute of Microbiology and Immunology, National Yang-Ming University, Taipei, Taiwan; [2]Center for Infectious Disease and Cancer Research (CICAR), Kaohsiung Medical University, Kaohsiung, Taiwan; [3]Department and Graduate Institute of Microbiology and Immunology, National Defense Medical Center, Taipei, Taiwan; [4]Institute of Biomedical Sciences, Academia Sinica, Taipei, Taiwan; [5]Graduate Institute of Natural Products, College of Pharmacy, Kaohsiung Medical University, Kaohsiung, Taiwan; [6]Department of Pathology, Kaohsiung Medical University Hospital, Kaohsiung Medical University, Kaohsiung, Taiwan; [7]Department of Biotechnology and Laboratory Science in Medicine, National Yang-Ming University, Taipei, Taiwan; [8]National Laboratory Animal Center, National Applied Research Laboratories; [9]Institute of Biomedical Sciences, National Sun Yat-sen University, Kaohsiung, Taiwan; and [10]Department of Microbiology and Immunology, Institute of Medical Research and Institute of Clinical Medicine, College of Medicine, Kaohsiung Medical University, Kaohsiung, Taiwan

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

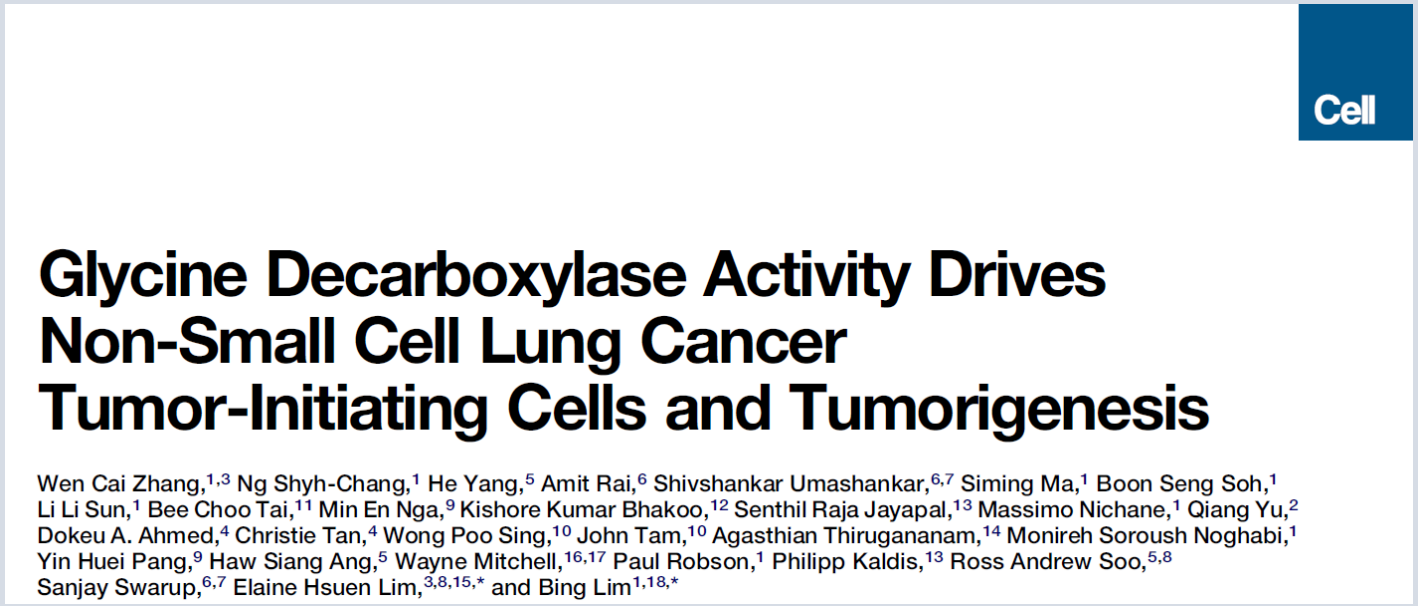
Proposed mechanism: GNMT→mTORC1→T cell activation

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

Future Plan

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