Data-driven Approaches to Identify the Regulators of the Anticancer Immune Response

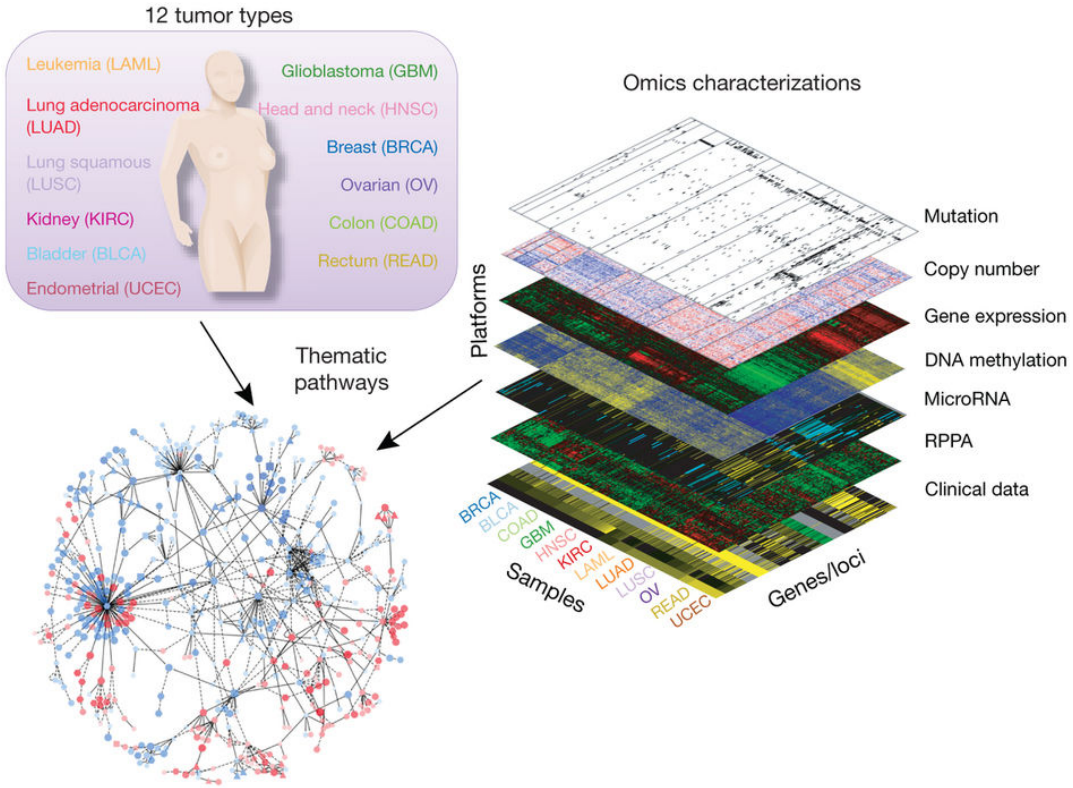
Peng Jiang, Ph.D.

July 9th, 2020

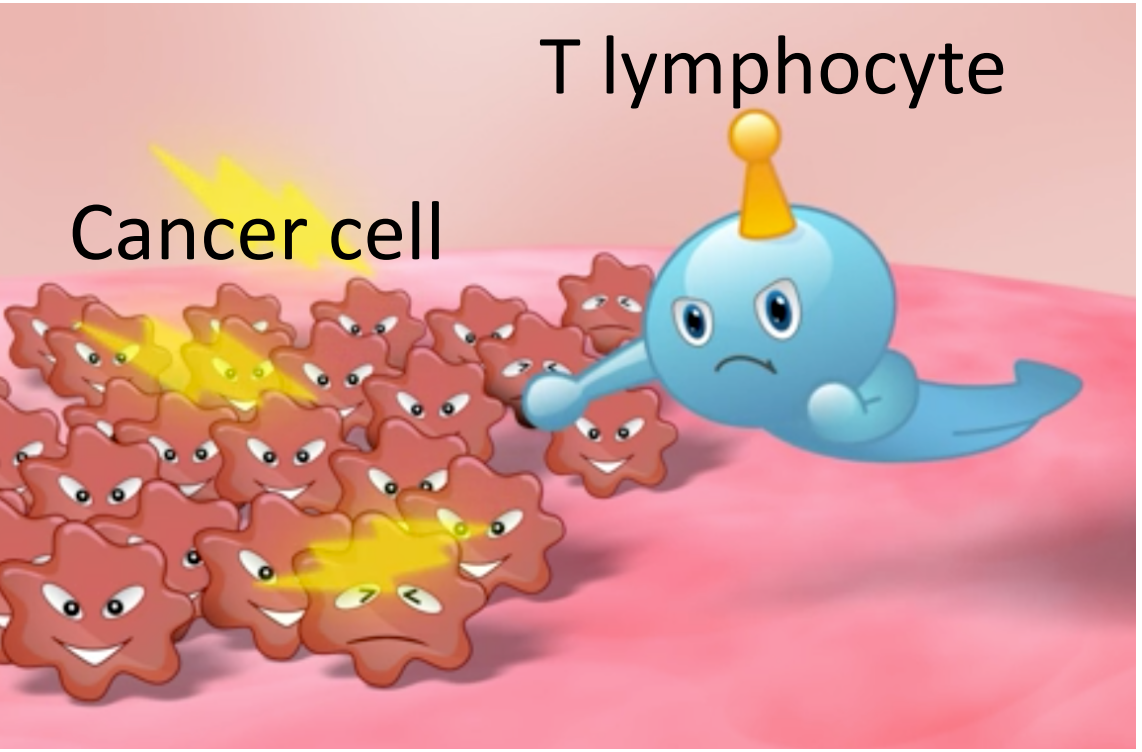
Bioinformatics Training and Education Program

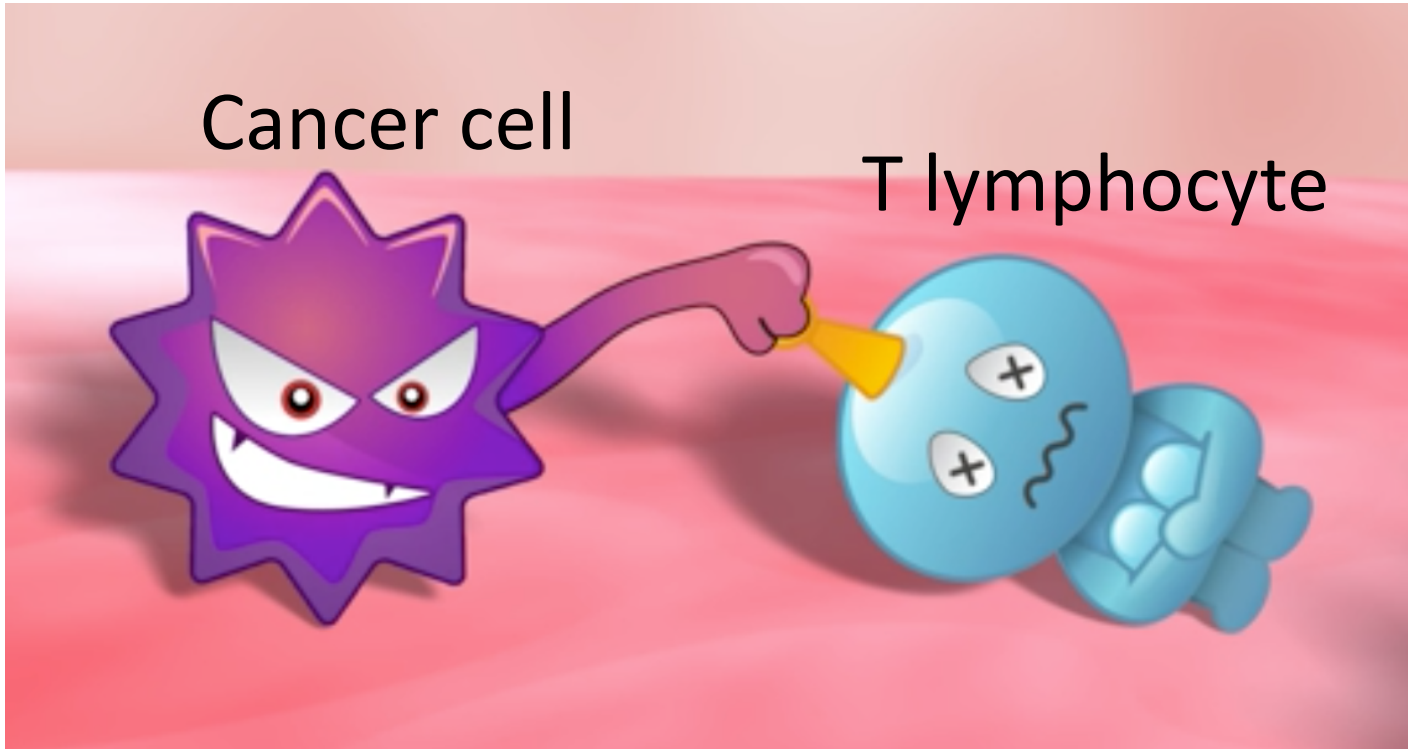
NIH NATIONAL CANCER INSTITUTE
Center for Cancer Research

Is my result supported by other datasets?



Tumor Immune Evasion





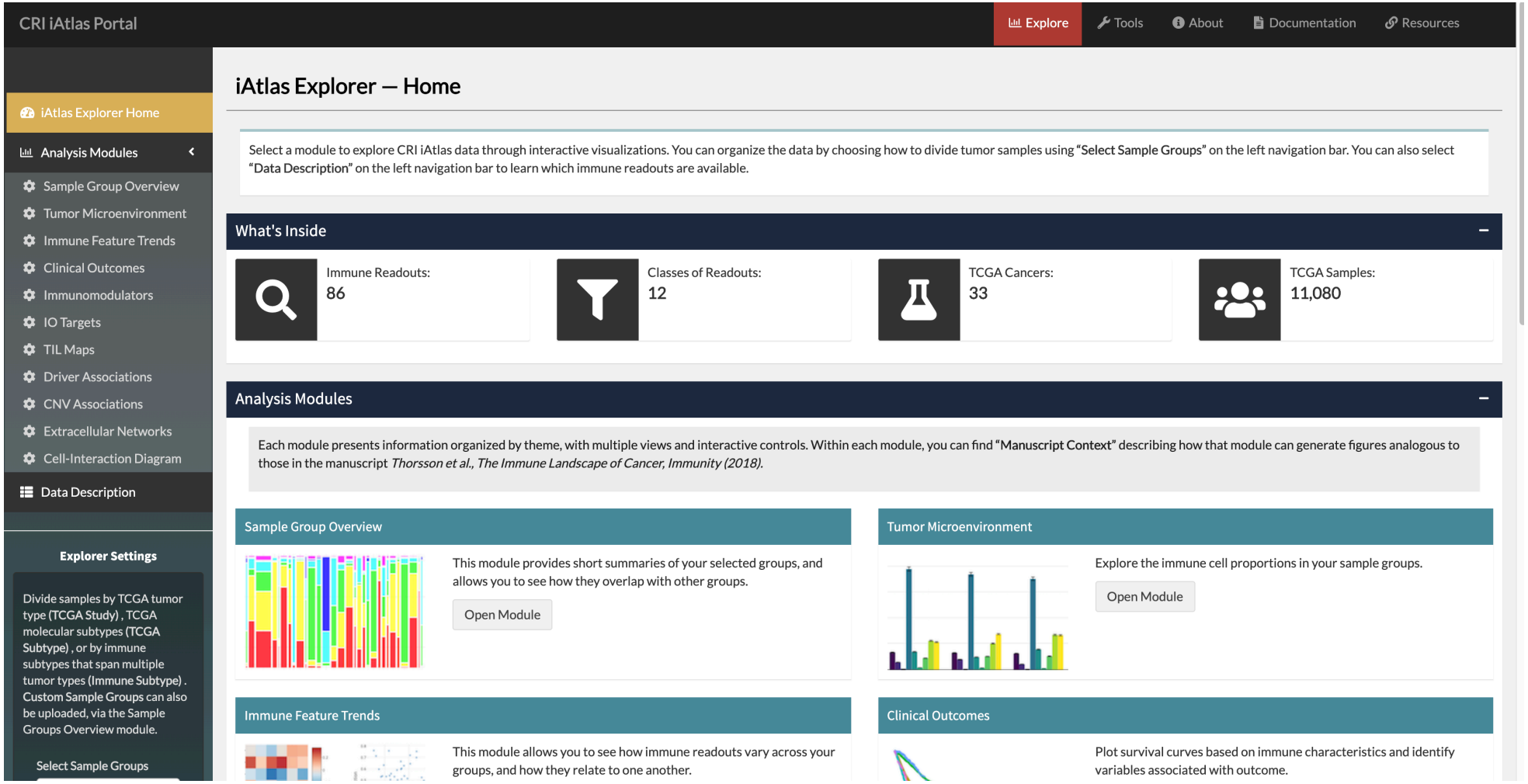
Knowledge database

- Focused Presentation
  - iATLAS (Thorsson et al., Immunity 2018)
  - TCIA (Charoentong et al., Cell Report 2017)

- Customized Analysis
  - TIDE (Jiang et al., Nature Medicine 2018)
  - TISIDB (Ru et al., Bioinformatics 2019)

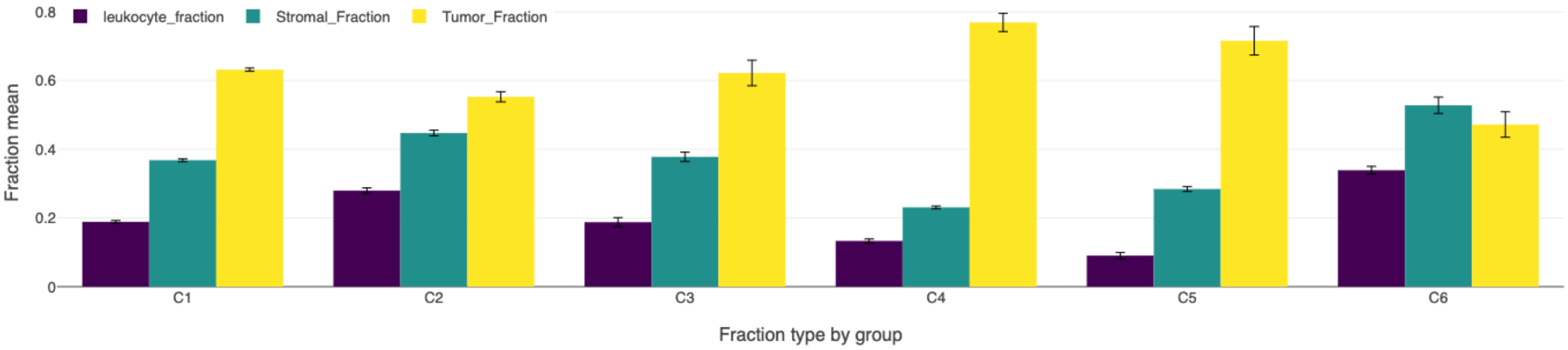
iATLAS: Flagship website of TCGA

Study the immune ecosystem from bulk tumor profiles

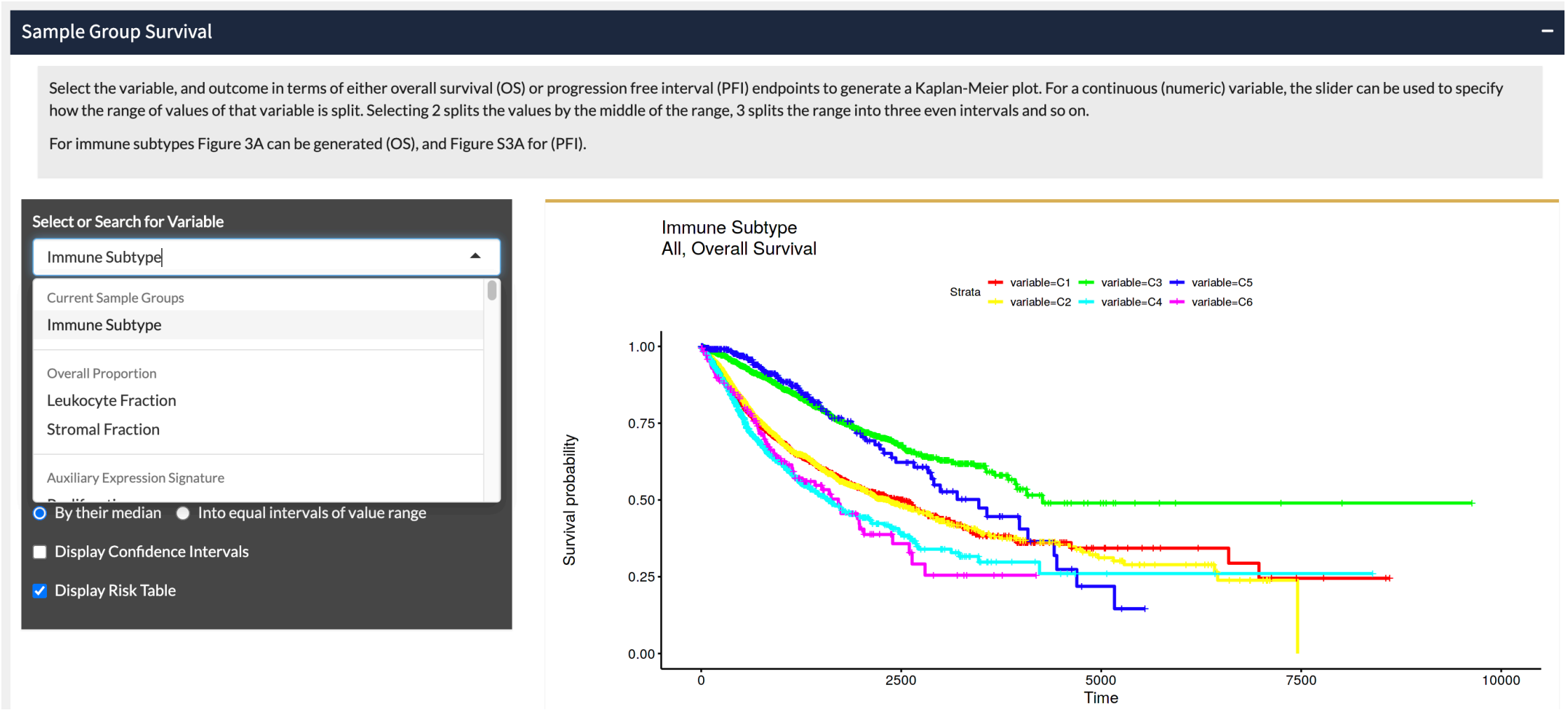


iATLAS: Flagship website of TCGA

Major result to browser, the six clusters they defined



Select subtypes to browse TCGA association



Knowledge database

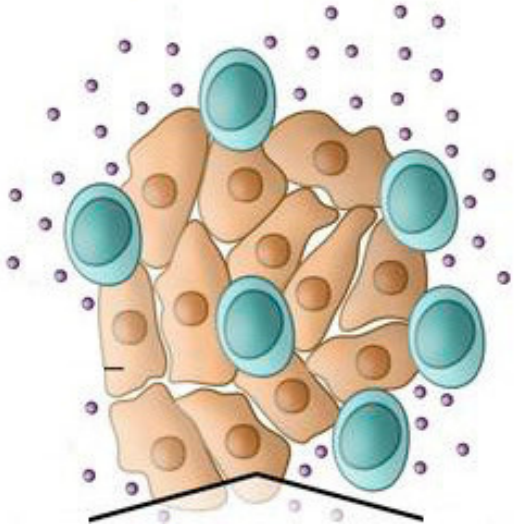
- Focused Presentation
  - iATLAS (Thorsson et al., Immunity 2018)
  - TCIA (Charoentong et al., Cell Report 2017)

- Customized Analysis
  - TIDE (Jiang et al., Nature Medicine 2018)
  - TISIDB (Ru et al., Bioinformatics 2019)

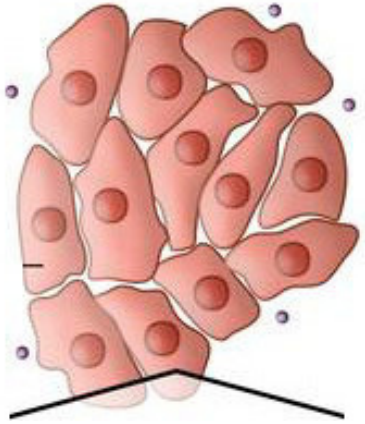
TIDE: Tumor Immune Dysfunction and Exclusion

http://tide.dfci.harvard.edu

Cyotoxic T Lymphocyte



T cell Dysfunction
CTL high tumor

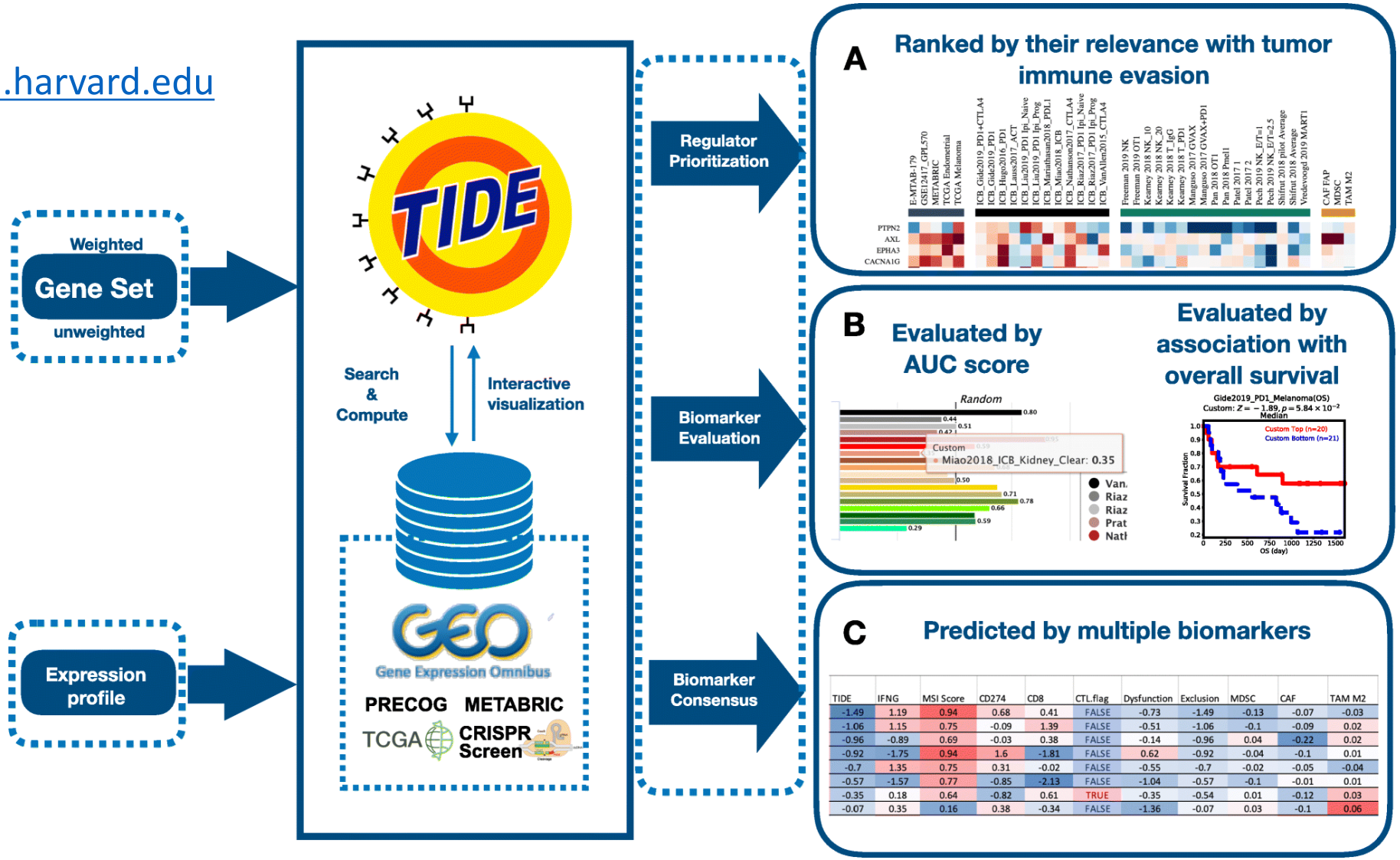


T cell Exclusion
CTL low tumor

(Gajewski, Schreiber, Fu et al., 2013)

TIDE: Tumor Immune Dysfunction and Exclusion

http://tide.dfci.harvard.edu



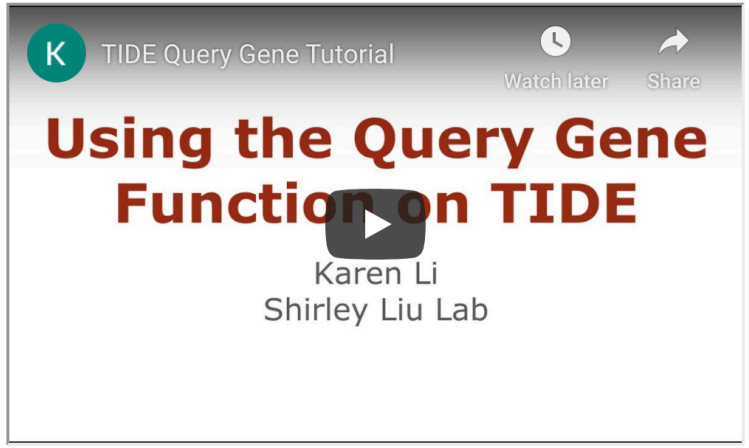
TIDE	IFNG	MSI Score	CD274	CD8	CTL.flag	Dysfunction	Exclusion	MDSC	CAF	TAM M2
-1.49	1.19	0.94	0.68	0.41	FALSE	-0.73	-1.49	-0.13	-0.07	-0.03
-1.06	1.15	0.75	-0.09	1.39	FALSE	-0.51	-1.06	-0.1	-0.09	0.02
-0.96	-0.89	0.69	-0.03	0.38	FALSE	-0.14	-0.96	0.04	-0.22	0.02
-0.92	-1.75	0.94	1.6	-1.81	FALSE	0.62	-0.92	-0.04	-0.1	0.01
-0.7	1.35	0.75	0.31	-0.02	FALSE	-0.55	-0.7	-0.02	-0.05	-0.04
-0.57	-1.57	0.77	-0.85	-2.13	FALSE	-1.04	-0.57	-0.1	-0.01	0.01
-0.35	0.18	0.64	-0.82	0.61	TRUE	-0.35	-0.54	0.01	-0.12	0.03
-0.07	0.35	0.16	0.38	-0.34	FALSE	-1.36	-0.07	0.03	-0.1	0.06

Gene Query

TIDE: Tumor Immune Dysfunction and Exclusion

TIDE also provides data-driven gene signatures of T-cell dysfunction and immunotherapy resistance, computed from a large amount of cancer clinical datasets.

Tutorial video of gene query



Download the video

Type a gene name. (e.g., "PDL1" or "SERPINB9").

Search

Gene Query: Example TGFB1

Result for query: TGFB1 (7040)

We listed the high confidence results (Core) and full range results (Expression, Copy_Number, etc.) from TIDE analysis. We also included several reference results, including Ref_Therapy and Ref_Screen, to show the gene level scores in immunotherapy clinical response and CRISPR screen of anticancer immunity. Please click each tab for individual section.

Core | Expression | Copy_Number | Methylation | Exclusion | Ref_Therapy | Ref_Screen

Most confident results using gene expression data

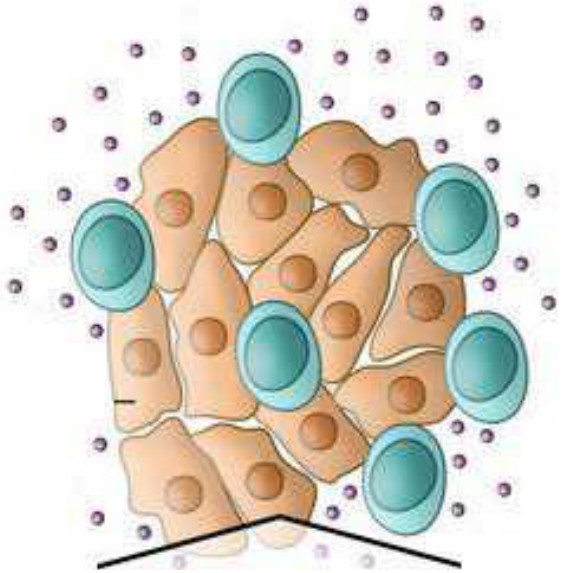
Click table header for sorting. Hover your mouse on a cell for description or full name of abbreviation.

Cohort	Cancer	Subtype	CTL Cor	T Dysfunction	Risk	Risk.adj	Count
TCGA	Melanoma	Metastatic	0.385	3.471	-1.424	-0.519	317
TCGA	Endometrial		0.296	3.02	-0.167	0.015	541
METABRIC	Breast	TN	0.079	1.268	0.766	0.891	233
GSE12417_GPL570	Leukemia	AML	0.013	0.265	0.877	0.511	79
E-MTAB-179	Brain	Neuroblastoma	0.49	0.158	-0.37	0.774	389

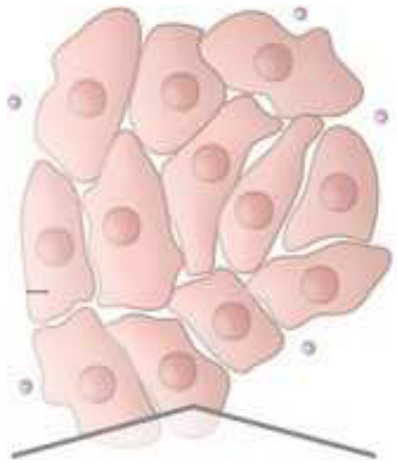
Export to CSV

T-cell Dysfunction in CTL-high Tumors

Cyotoxic T Lymphocyte



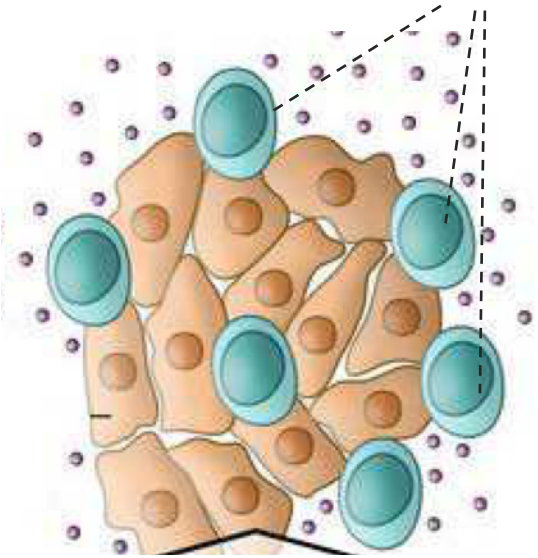
T cell Dysfunction
CTL high tumor



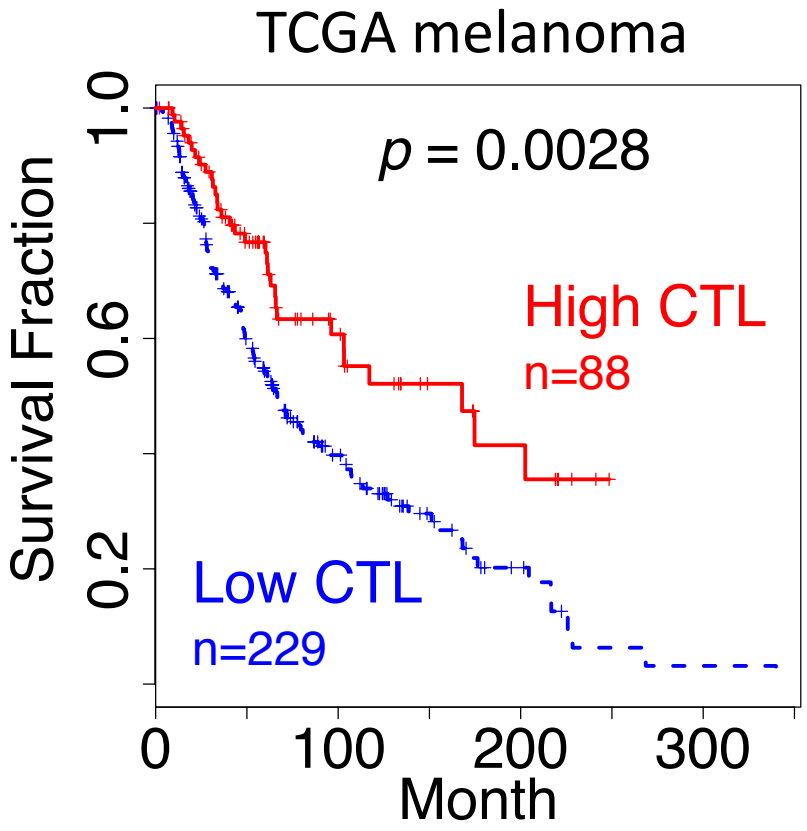
T cell Exclusion
CTL low tumor

High level of cytotoxic T lymphocyte indicates better survival

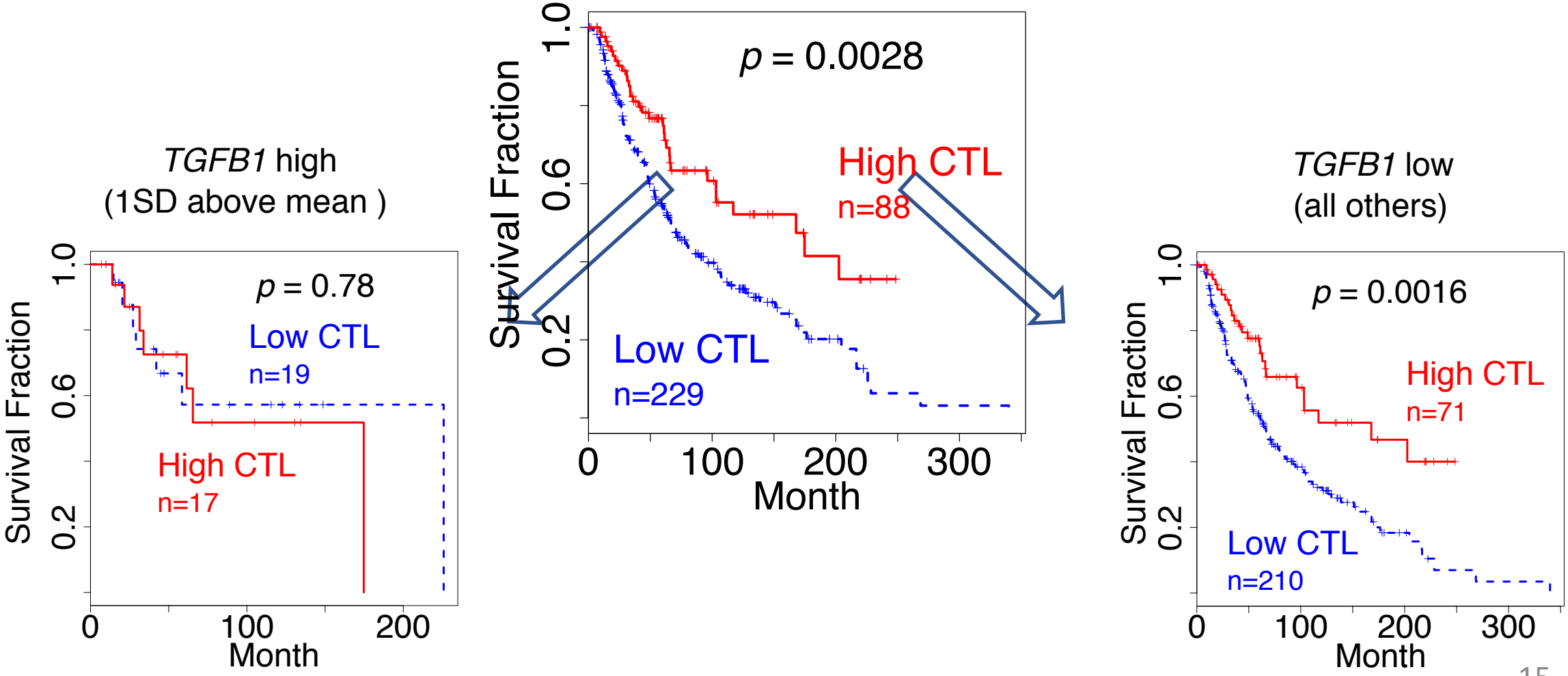
Cytotoxic T Lymphocyte (CTL)



Mark genes:
CD8A, CD8B, GZMA, GZMB, PRF1



Association between CTL and survival depends on *TGFB1* level



T-cell dysfunction test in a linear model

Death Hazard

↑ *TGFB1* expression

CTL (Cytotoxic T Lymphocyte)

$$\text{Hazard} = a * \text{CTL} + b * P + \underline{d * \text{CTL} * P}$$

$$\text{T cell dysfunction score} = d / \text{StdErr}(d)$$

Antagonistic Interaction between CTL and *TGFB1*

	Coef	Stderr	Z	Pr(>|z|)
Age	0.02	0.01	3.55	3.78E-04
Gender	0.02	0.17	0.12	9.07E-01
Stage	0.29	0.09	3.31	9.34E-04
CTL	-0.50	0.15	-3.32	9.05E-04
TGFB1	-0.10	0.10	-1.04	3.00E-01
CTL * *TGFB1*	0.11	0.03	3.47	5.18E-04

Hazard = $\underbrace{\text{c1*Age + c2*Gender + c3*Stage}}_{\text{Background}}$ + $\underbrace{\text{a*CTL + b*}\textit{TGFB1}\text{ + d * CTL * }\textit{TGFB1}}_{\text{Interaction}}$

Interaction between variables matters!

TGFB1 encodes an immuno-suppressive cytokine (Sharma et al., 2017)

	Coef	StdErr	Z-score	Pr(>|z|)
Age	0.02	0.01	3.95	7.85E-05
Gender	0.00	0.17	0.00	9.97E-01
Stage	0.30	0.09	3.36	7.74E-04
TGFB1	-0.13	0.09	-1.42	1.55E-01

TGFB1 alone associate with lower death risk

Correlation(*TGFB1*, CTL) = 0.39

	Coef	StdErr	Z-score	Pr(>|z|)
Age	0.02	0.01	3.55	3.79E-04
Gender	0.02	0.17	0.12	9.08E-01
Stage	0.29	0.09	3.31	9.34E-04
CTL	-0.54	0.17	-3.29	9.95E-04
TGFB1	-0.10	0.10	-0.97	3.31E-01
CTL**TGFB1*	0.13	0.04	3.47	5.28E-04

TGFB1 antagonizes the effect of cytotoxic T lymphocyte

Synergistic Interaction between CTL and *SOX10*

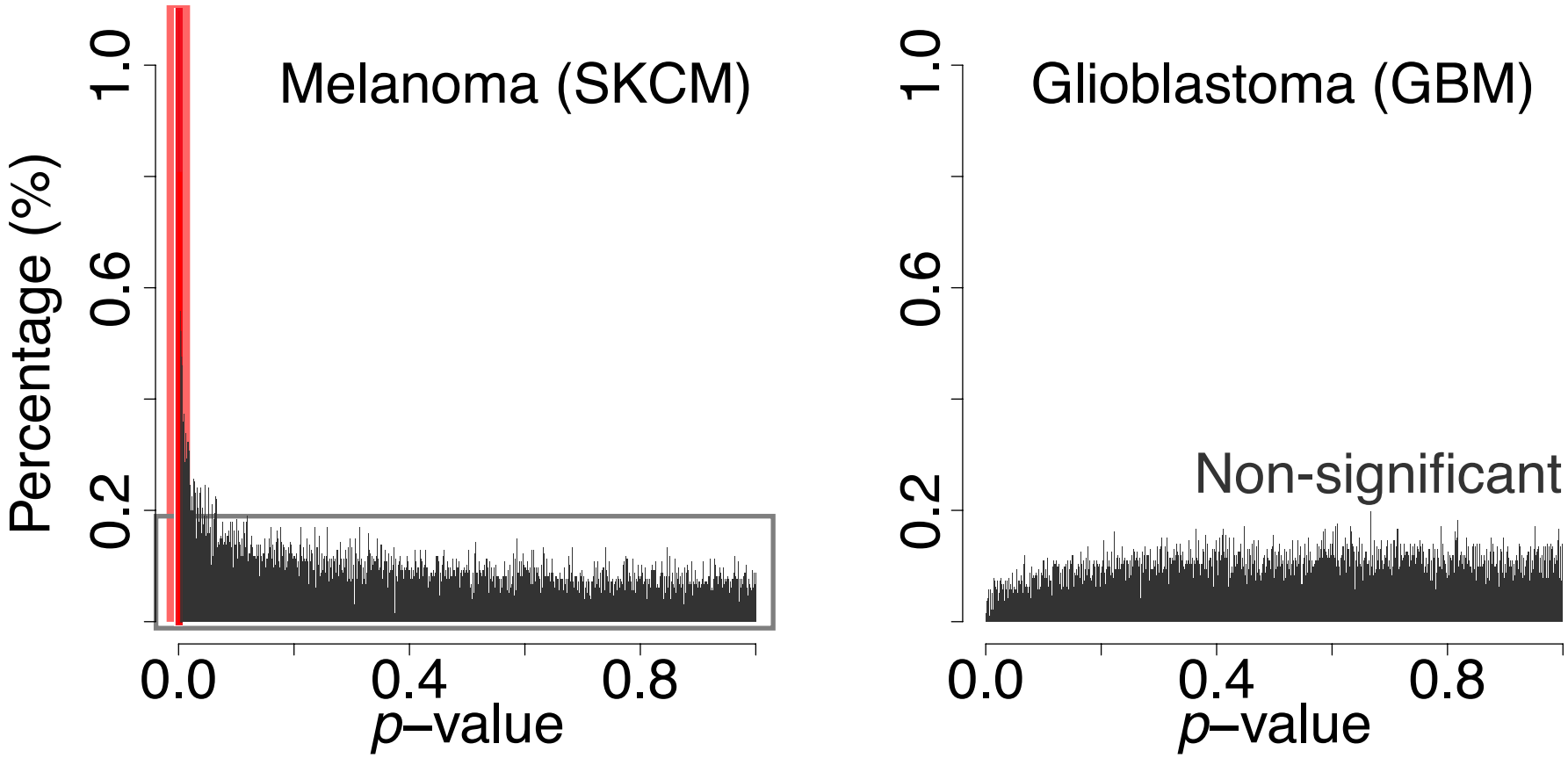
	Coef	Stderr	Z	Pr(>|z|)
Age	0.02	0.01	3.26	1.11E-03
Gender	0.03	0.17	0.15	8.80E-01
Stage	0.29	0.09	3.33	8.63E-04
CTL	-0.79	0.21	-3.79	1.51E-04
SOX10	-0.01	0.10	-0.11	9.10E-01
CTL * *SOX10*	-0.59	0.16	-3.69	2.23E-04

Hazard = c1*Age + c2*Gender + c3*Stage + a*CTL + b*_SOX10_ + d * CTL * _SOX10_

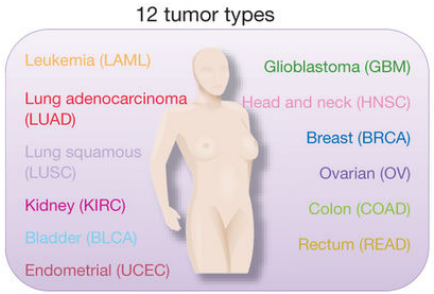
Background Interaction

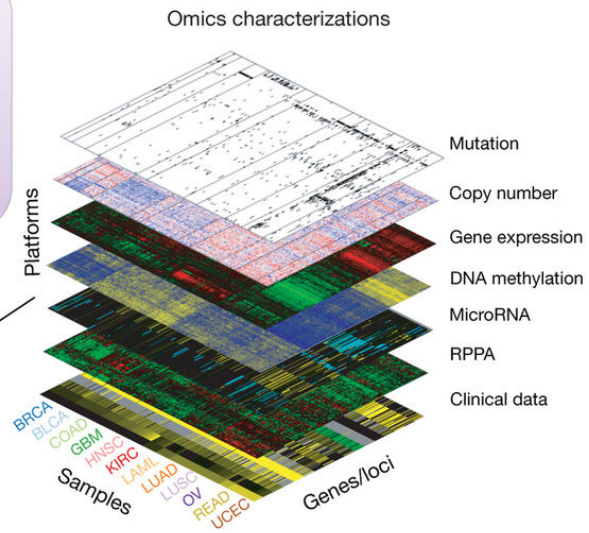
SOX10 level in cancer cells can promote the T-cell mediated tumor killing (Patel et al., 2017; Khong et al., 2002).

The *p*-value histogram should have a significant peak

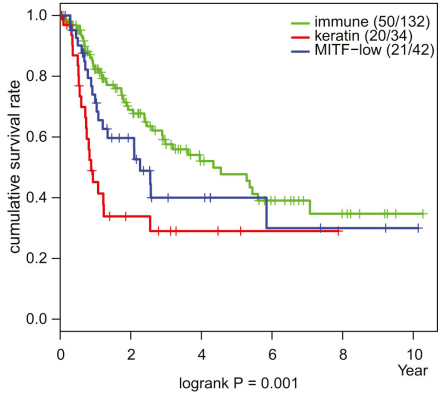


Large-scale Cancer Genomics Data Cohorts



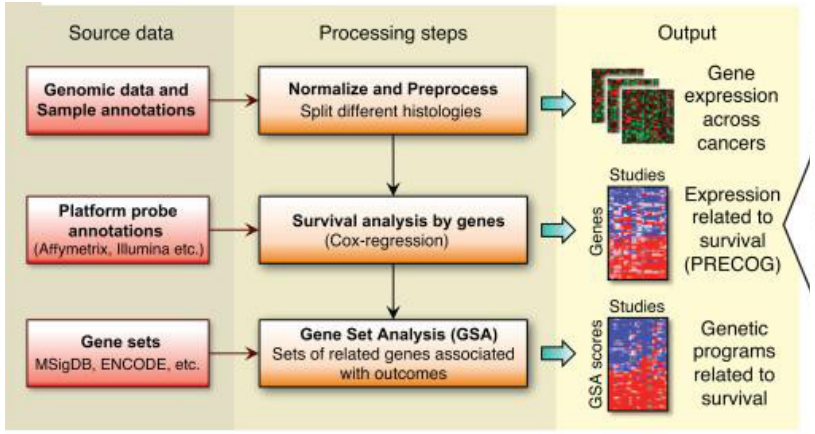


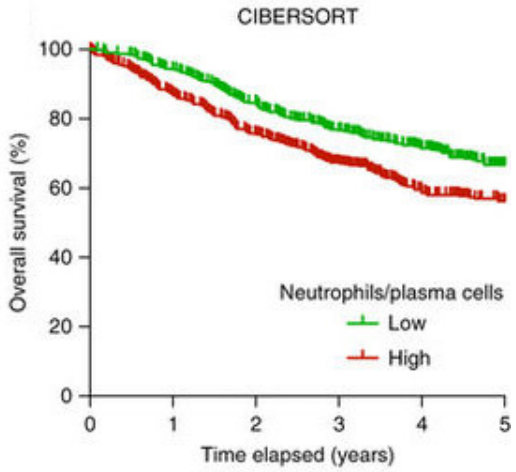
Thematic pathways



TCGA

49 cohorts, 13K samples





PRECOG 135 cohorts 18K samples

METABRIC

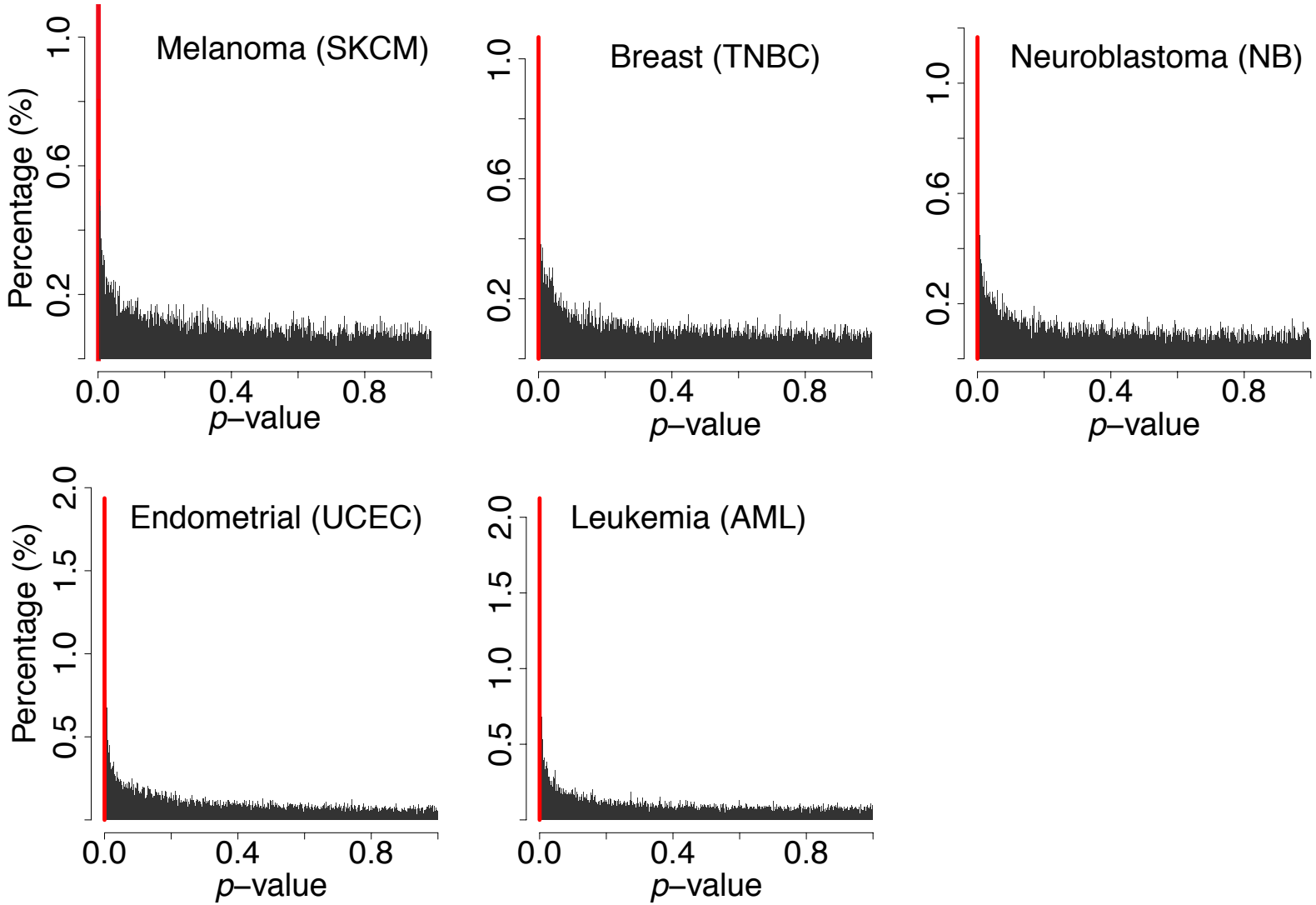
NATURE | ARTICLE

日本語要約

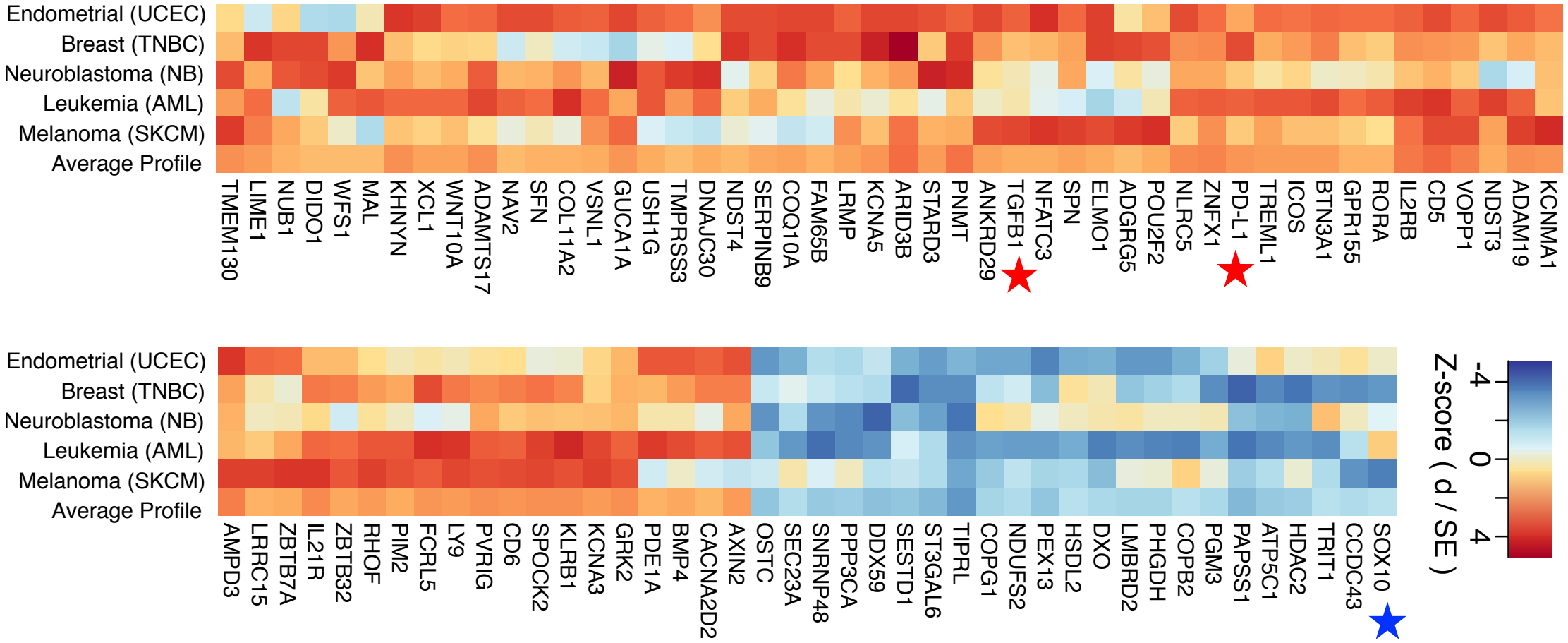
The genomic and transcriptomic architecture of 2,000 breast tumours reveals novel subgroups

Christina Curtis, Sohrab P. Shah, Suet-Feung Chin, Gulisa Turashvili, Oscar M. Rueda, Mark J. Dunning, Doug Speed, Andy G. Lynch, Shamith Samarajiwa, Yinyin Yuan, Stefan Gräf, Gavin Ha, Gholamreza Haffari, Ali Bashashati, Roslin Russell, Steven McKinney, METABRIC Group, Anita Langerød, Andrew Green, Elena Provenzano, Gordon Wishart, Sarah Pinder, Peter Watson, Florian Markowetz, Leigh Murphy + *et al.*

Five datasets with significant *p*-value peaks



Genes with Significant T Dysfunction Scores



Correlation with Cytotoxic T Lymphocyte Level

Result for query: TGFB1 (7040)

We listed the high confidence results (Core) and full range results (Expression, Copy_Number, etc.) from T'
scores in immunotherapy clinical response and CRISPR screen of anticancer immunity. Please click each

_Screen, to show the gene level

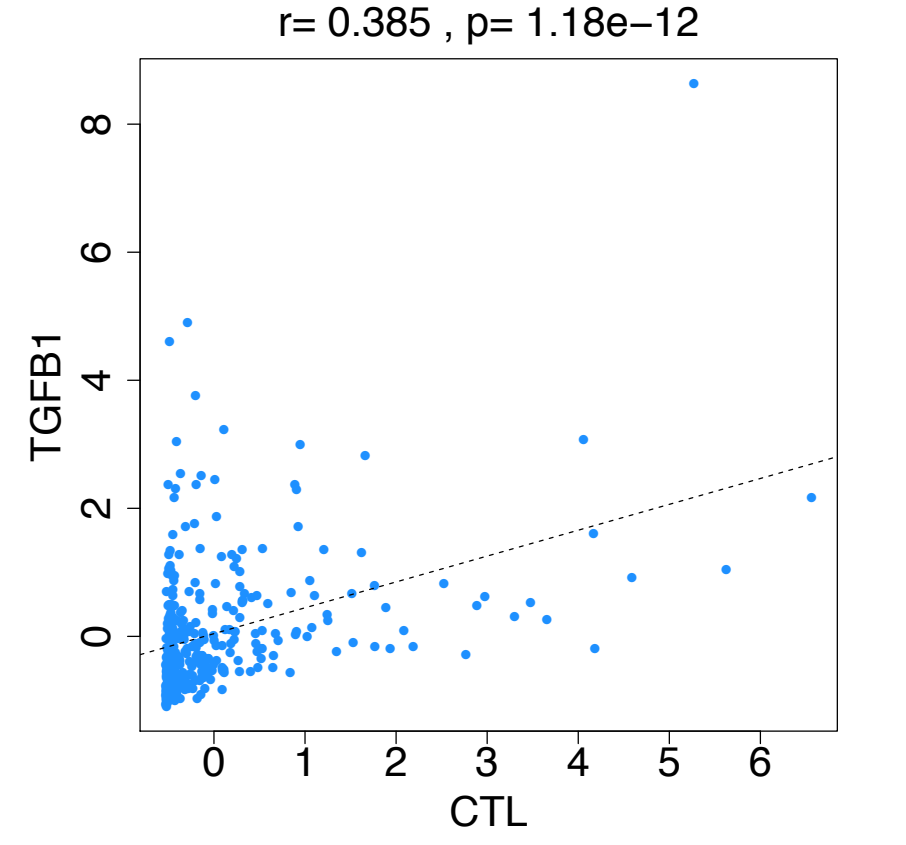
Core | Expression | Copy_Number | Methylation | Exclusion | Ref_

Most confident results using gene expression data

Click table header for sorting. Hover your mouse on a cell for description or

Cohort	Cancer	Subtype	CTL Cor	
TCGA	Melanoma	Metastatic	0.385	
TCGA	Endometrial		0.296	3.02
METABRIC	Breast	TN	0.079	1.268
GSE12417_GPL570	Leukemia	AML	0.013	0.265
E-MTAB-179	Brain	Neuroblastoma	0.49	0.158

Export to CSV



Correlation with Death Risk

Result for query: TGFB1 (7040)

We listed the high confidence results (Core) and full range results (Expression, Copy_Number, etc.) from TIDE analysis. We also included se scores in immunotherapy clinical response and CRISPR screen of anticancer immunity. Please click each tab for individual section.

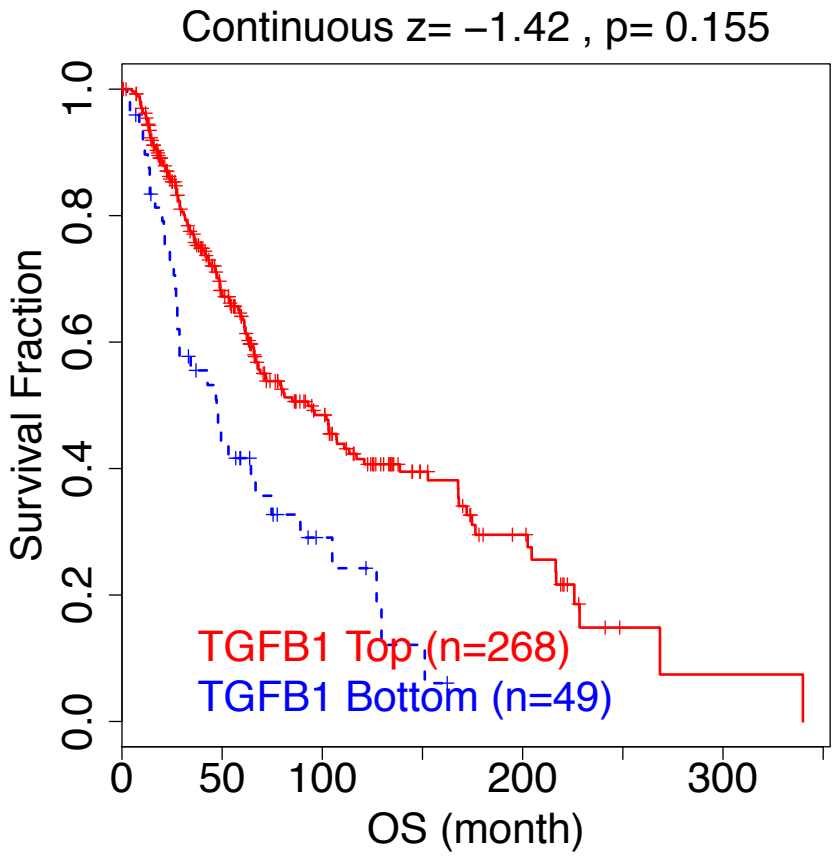
Core | Expression | Copy_Number | Methylation | Exclusion | Ref_Therapy | Ref_Screen

Most confident results using gene expression data

Click table header for sorting. Hover your mouse on a cell for description or full name of abbreviation.

Cohort	Cancer	Subtype	CTL Cor	T Dysfunction	Risk	
TCGA	Melanoma	Metastatic	0.385	3.471	-1.424	
TCGA	Endometrial		0.296	3.02	-0.167	
METABRIC	Breast	TN	0.079	1.268	0.766	0.
GSE12417_GPL570	Leukemia	AML	0.013	0.265	0.877	0.
E-MTAB-179	Brain	Neuroblastoma	0.49	0.158	-0.37	0.

Export to CSV



Association with T cell Dysfunction : Core

Result for query: TGFB1 (7040)

We listed the high confidence results (Core) and full range results (Expression, Copy_Number, etc.) from TIDE anal
scores in immunotherapy clinical response and CRISPR screen of anticancer immunity. Please click each tab for in

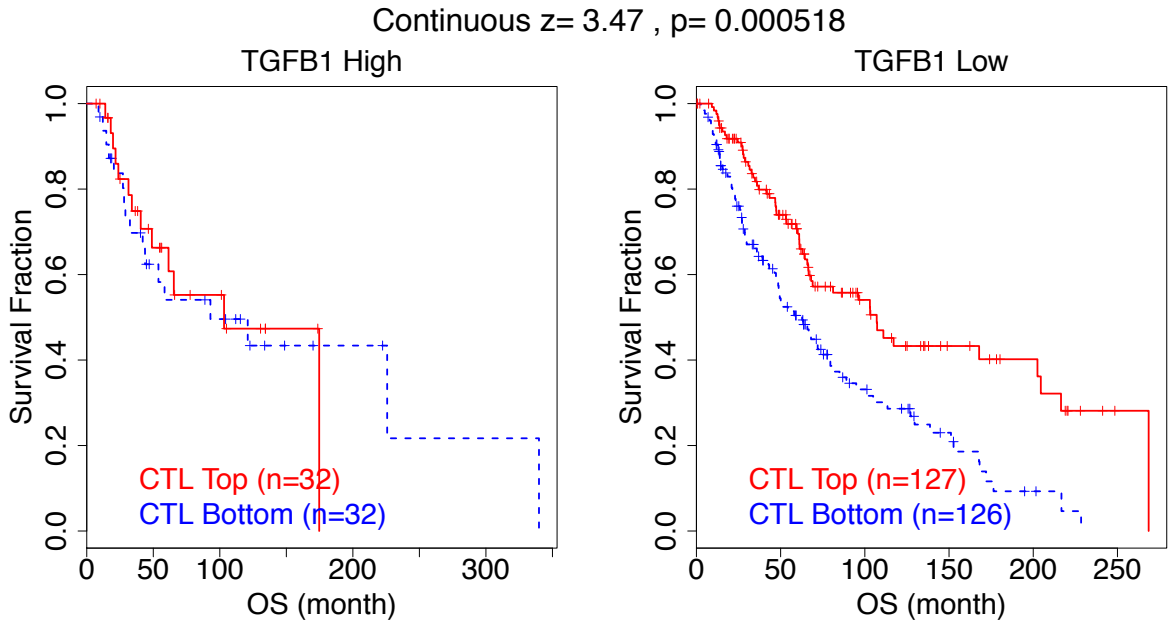
Core | Expression | Copy_Number | Methylation | Exclusion | Ref_Therapy

Most confident results using gene expression data

Click table header for sorting. Hover your mouse on a cell for description or full name of ab

Cohort	Cancer	Subtype	CTL Cor	T Dysfunct
TCGA	Melanoma	Metastatic	0.385	3.471
TCGA	Endometrial		0.296	3.02
METABRIC	Breast	TN	0.079	1.268
GSE12417_GPL570	Leukemia	AML	0.013	0.265
E-MTAB-179	Brain	Neuroblastoma	0.49	0.158

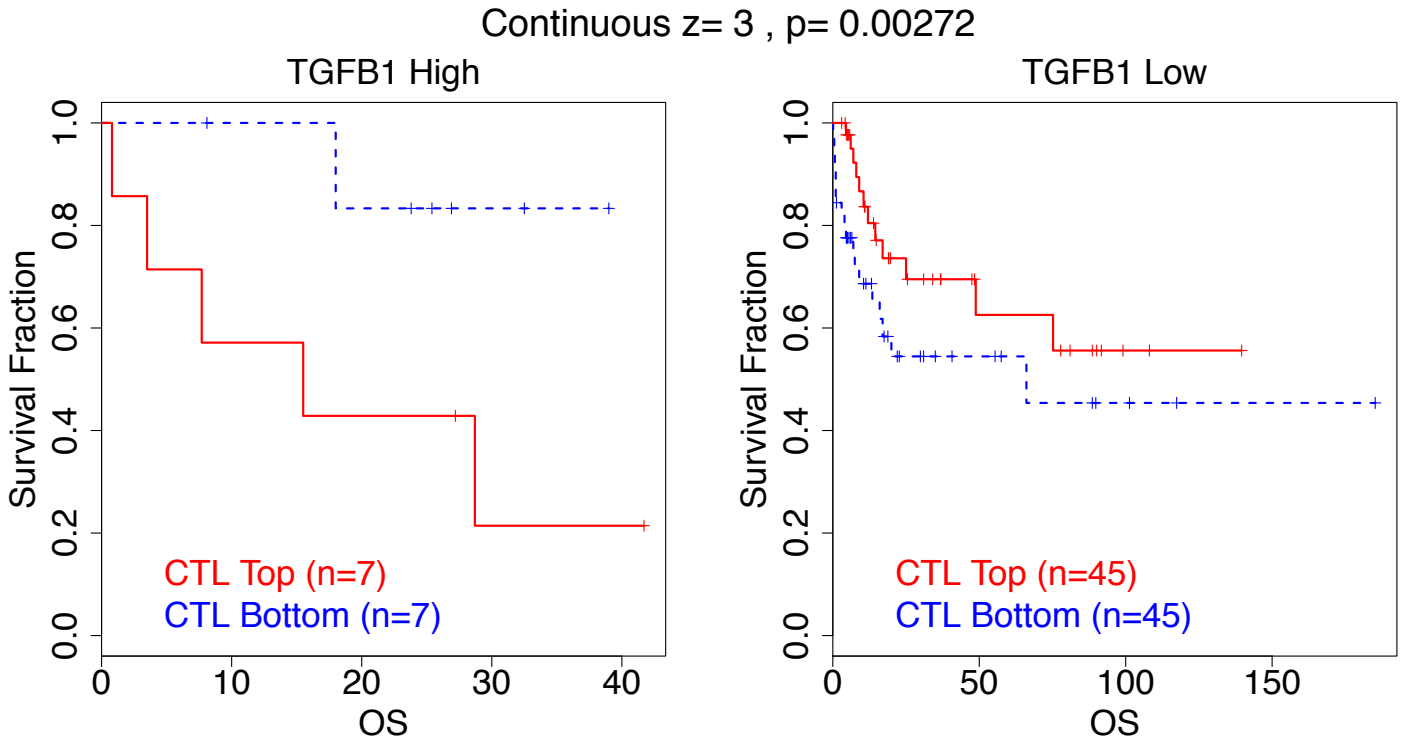
Export to CSV



Association with T cell Dysfunction : More

Result for query: TGFB1 (7040)

We listed the high confidence results (Core) and full range results (Expression, Copy_Number, etc.) from TIDE analysis. We also included several reference results, including Ref_Therapy and Ref_Screen, to show the gene level scores in immunotherapy clinical response and CRISPR screen of anticancer immunity. Please click each tab for individual section.



GSE30219	Lung		OS	0.353	2.062	0.529	0.991	263
GSE21501@PRECOG	Pancreatic		OS	0.28	2.041	1.394	1.406	102

sfunction	Risk	Risk.adj	Count
71	-1.424	-0.519	317
8	-1.4	-0.363	413
	-0.167	0.015	541
	0.032	0.029	104
2.868	-0.262	-0.395	243
2.509	1.224	2.025	562
2.359	-0.222	-0.384	92
2.347	-2.494	-2.365	58
2.307	0.052	-0.127	74
2.212	1.218	1.24	484
2.075	2.438	2.521	553

Exclusion : Example PTGS2

Result for query: PTGS2 (5743)

We listed the high confidence results (Core) and full range results (Expression, Copy_Number, etc.) from TIDE analysis. We also included several reference results, including Ref_Therapy and Ref_Screen, to show the gene level scores in immunotherapy clinical response and CRISPR screen of anticancer immunity. Please click each tab for individual section.

Core | Expression | Copy_Number | Methylation | Exclusion | Ref_Therapy | Ref_Screen

The gene expression value in T cell exclusion signatures

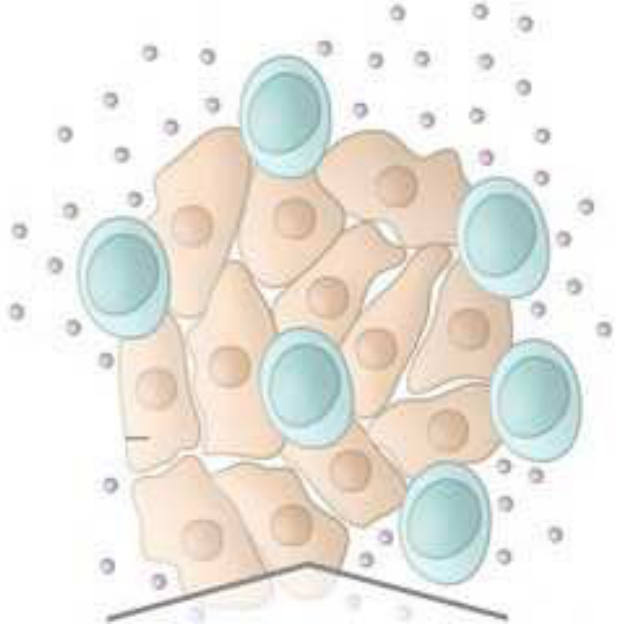
Click table header for sorting. Hover your mouse on a cell for description or full name of abbreviation.

Condition	Z-score
MDSC	3.424
CAF FAP	1.526
TAM M2	-5.725

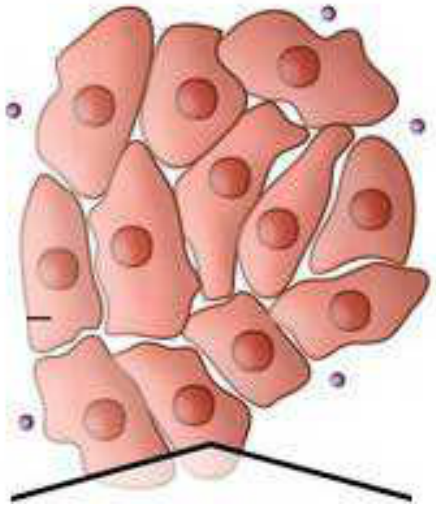
Export to CSV

Tumor Immune Evasion through T-cell Exclusion

Cyotoxic T Lymphocyte



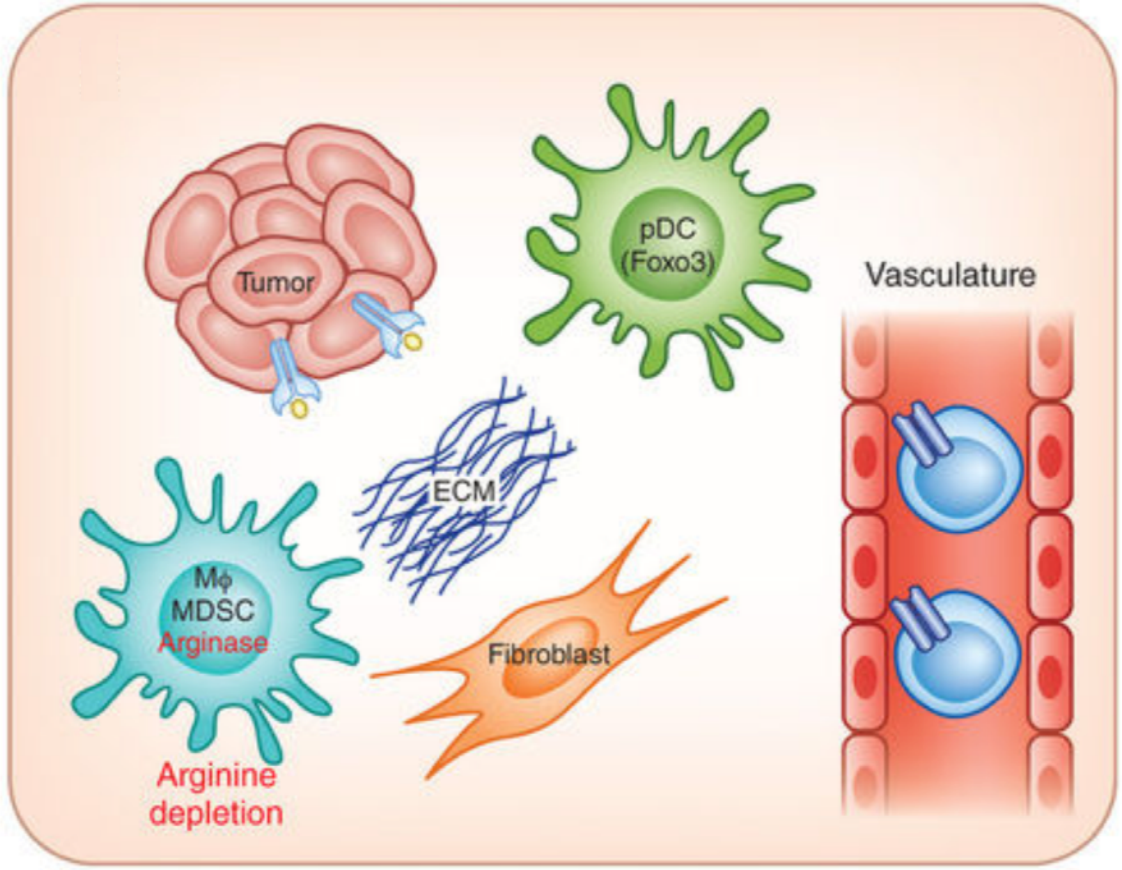
T cell Dysfunction
CTL high tumor

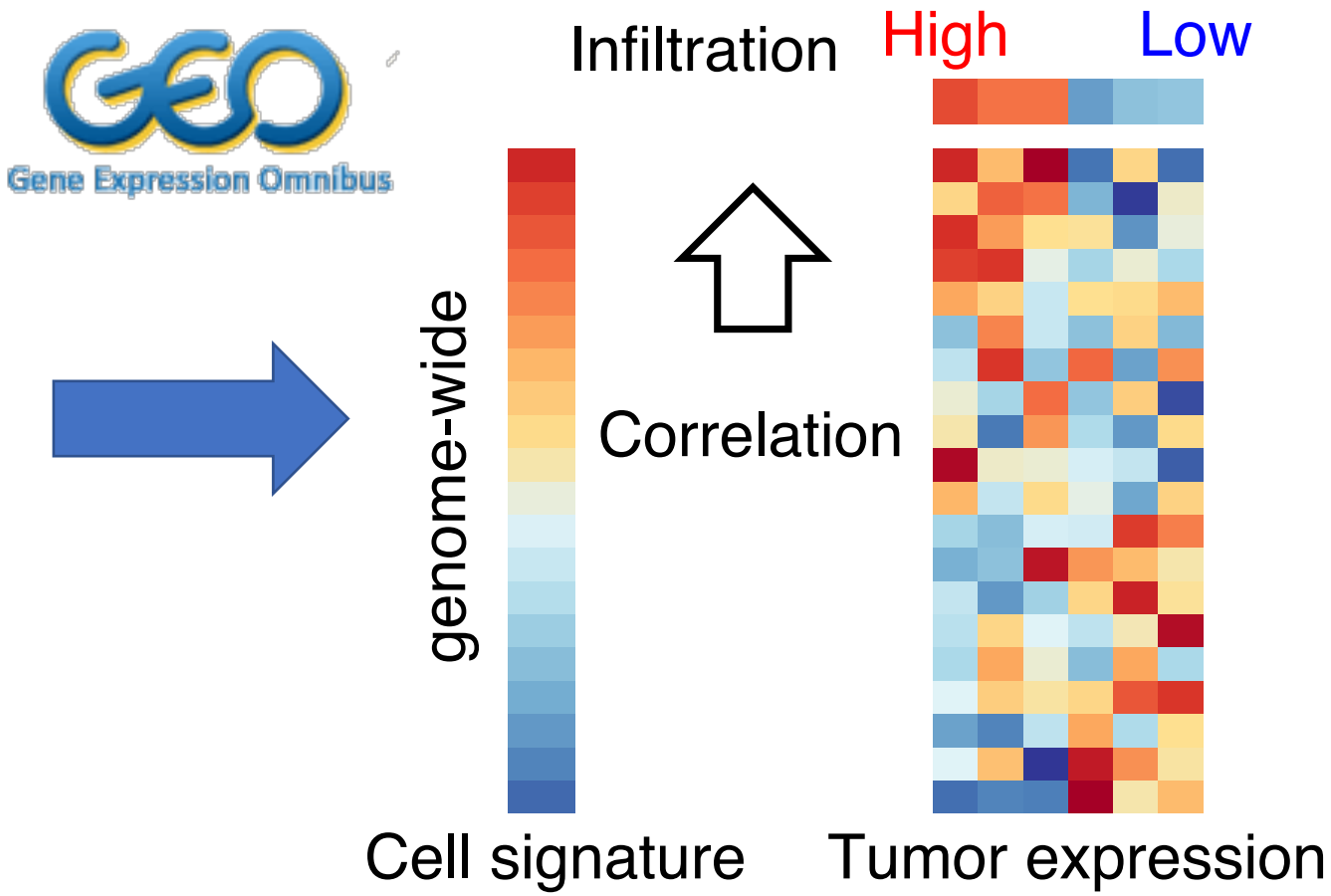


T cell Exclusion
CTL low tumor

Immunosuppressive Cell Types in T-cell Exclusion

T cell exclusion factors

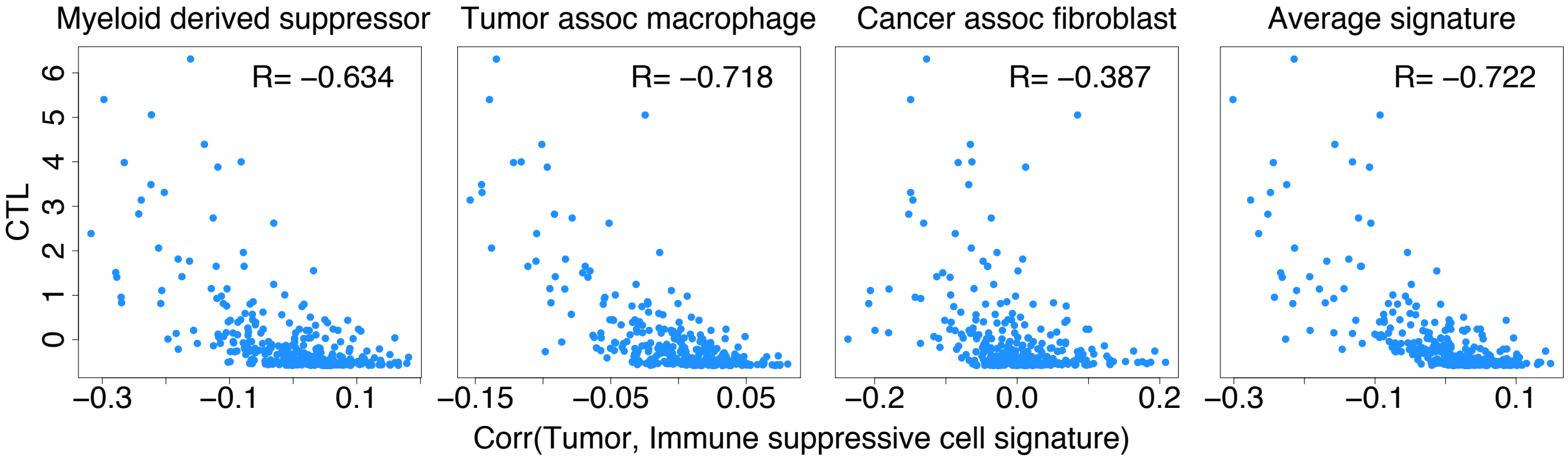




(Gajewski, Schreiber, Fu et al., 2013)

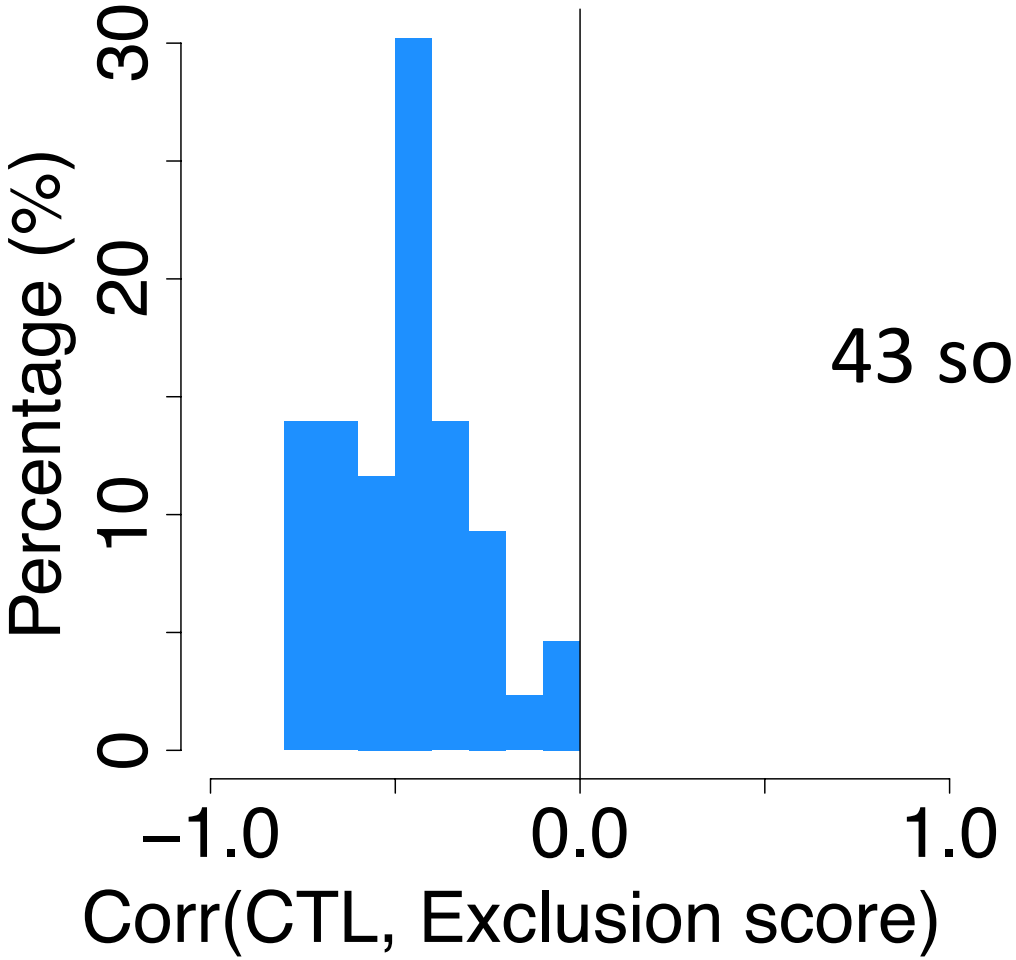
(Dougan et al., 2017)

Immunosuppressive cell signatures predict T-cell exclusion in tumors



CTL : expression average of *CD8A, CD8B, GZMA, GZMB, PRF1*

Immunosuppressive cell signatures predict T-cell exclusion in all solid tumors



43 solid tumor types

Exclusion : Example PTGS2

Result for query: PTGS2 (5743)

We listed the high confidence results (Core) and full range results (Expression, Copy_Number, etc.) from TIDE analysis. We also included several reference results, including Ref_Therapy and Ref_Screen, to show the gene level scores in immunotherapy clinical response and CRISPR screen of anticancer immunity. Please click each tab for individual section.

Core | Expression | Copy_Number | Methylation | Exclusion | Ref_Therapy | Ref_Screen

The gene expression value in T cell exclusion signatures

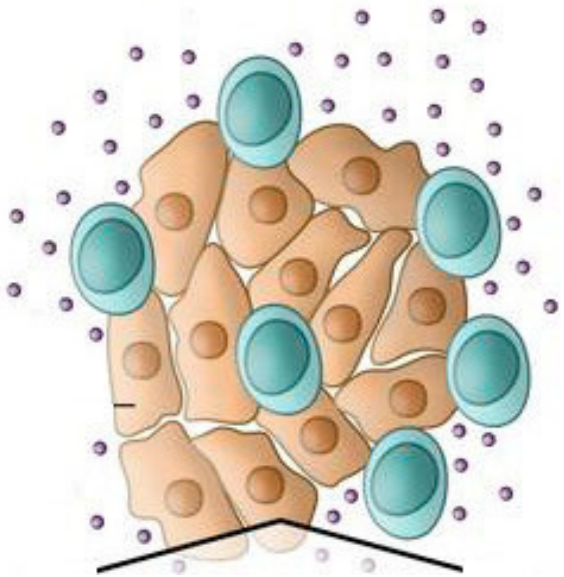
Click table header for sorting. Hover your mouse on a cell for description or full name of abbreviation.

Condition	Z-score
MDSC	3.424
CAF FAP	1.526
TAM M2	-5.725

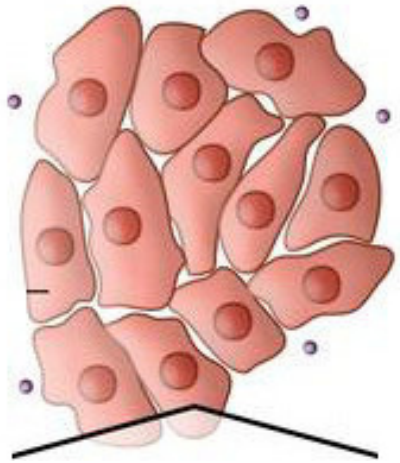
Export to CSV

Two Categories of Tumor Immune Evasion

Cyotoxic T Lymphocyte



T cell Dysfunction
CTL high tumor



T cell Exclusion
CTL low tumor

(Gajewski, Schreiber, Fu et al., 2013)

TIDE: Tumor Immune Dysfunction and Exclusion

Prediction with a Two-Part Procedure

CD8A, CD8B, GZMA, GZMB, PRF1

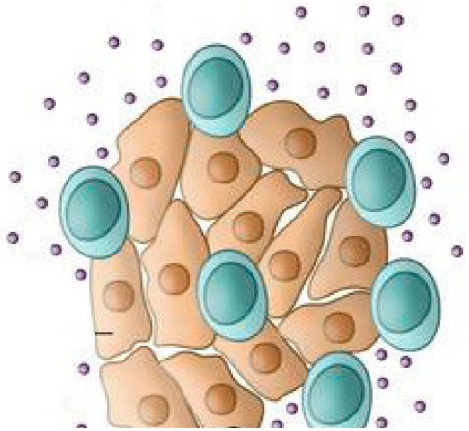
CTL

All positive

Others

CTL high

CTL low

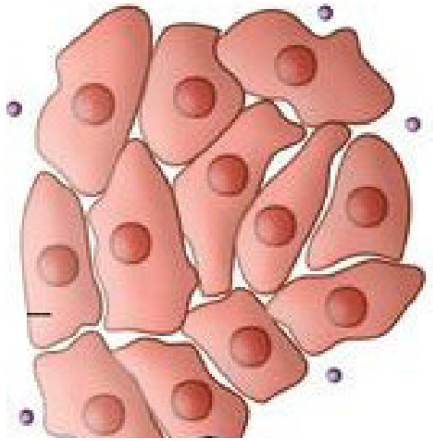


Corr

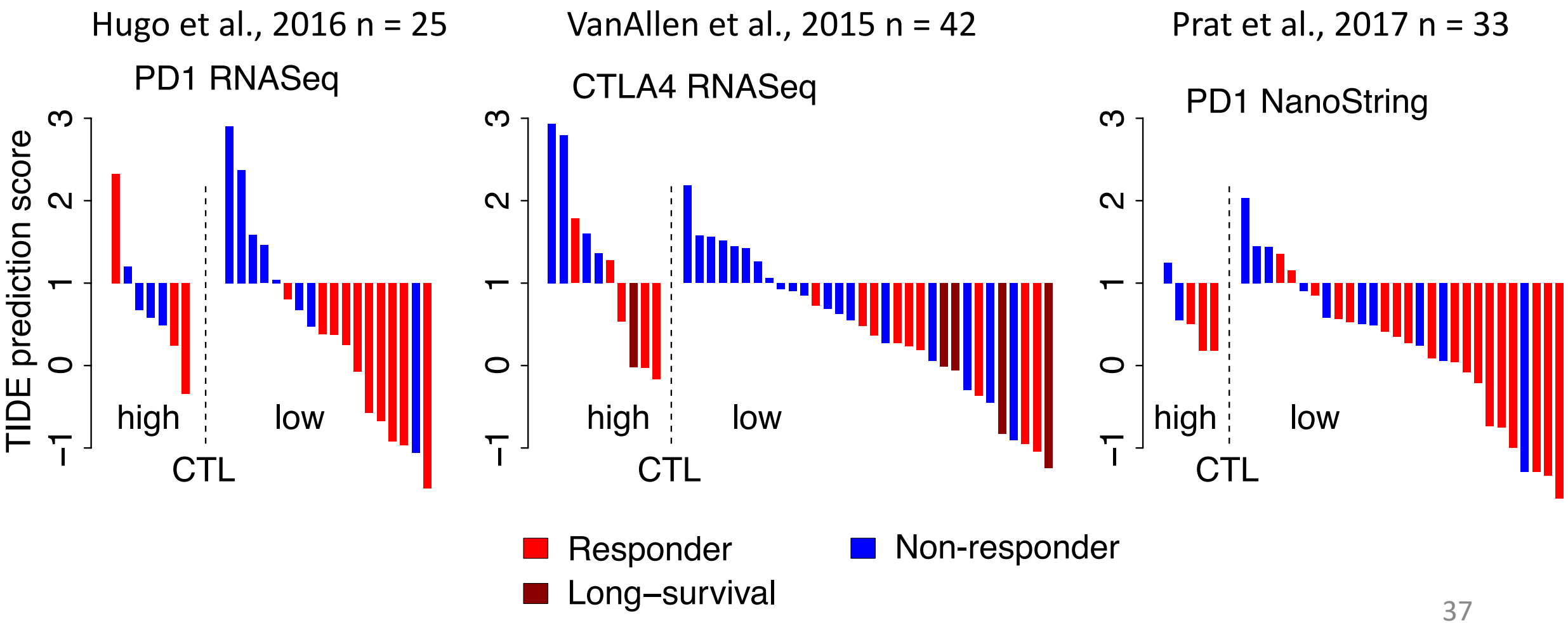
Corr

Dysfunction signature

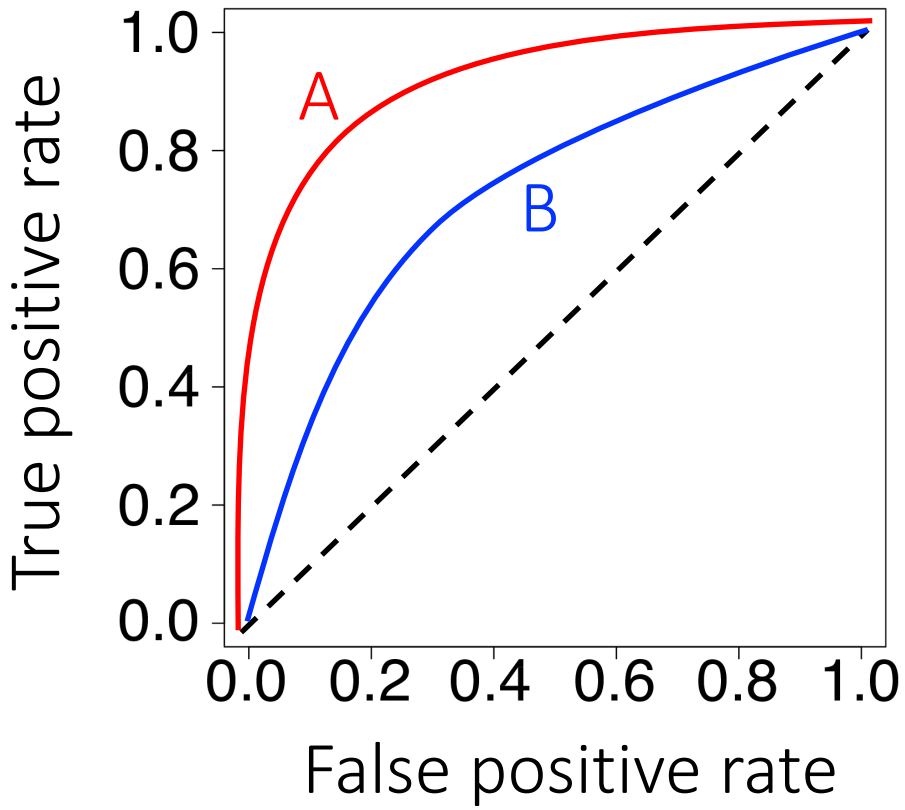
Exclusion signature

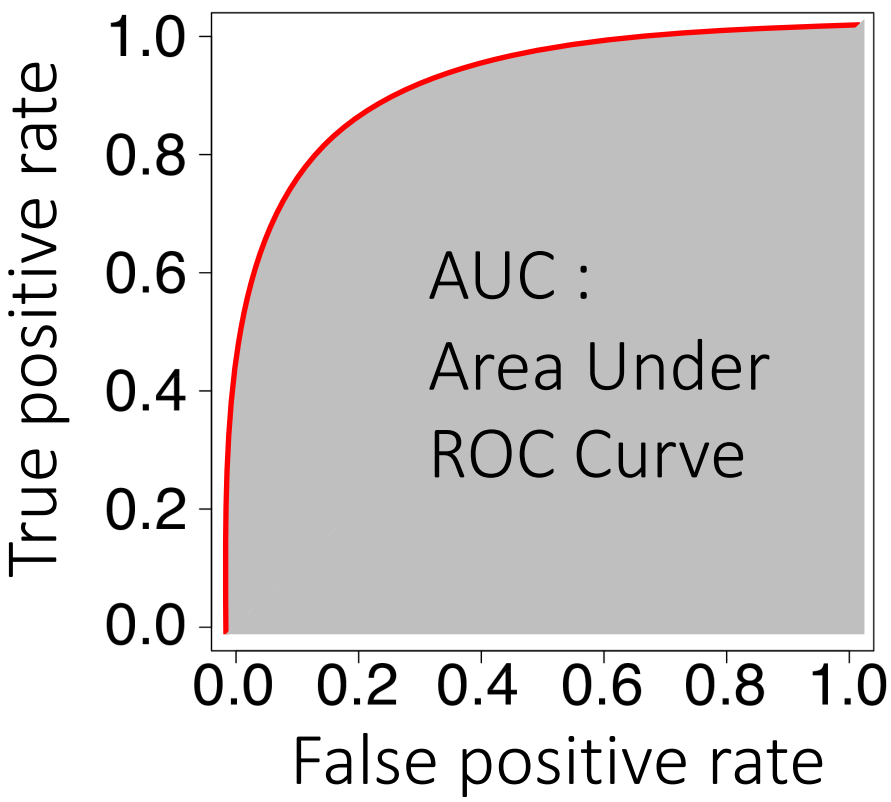


TIDE Scores of Pre-Treatment Tumors



Receiver Operating Characteristic (ROC) Curve

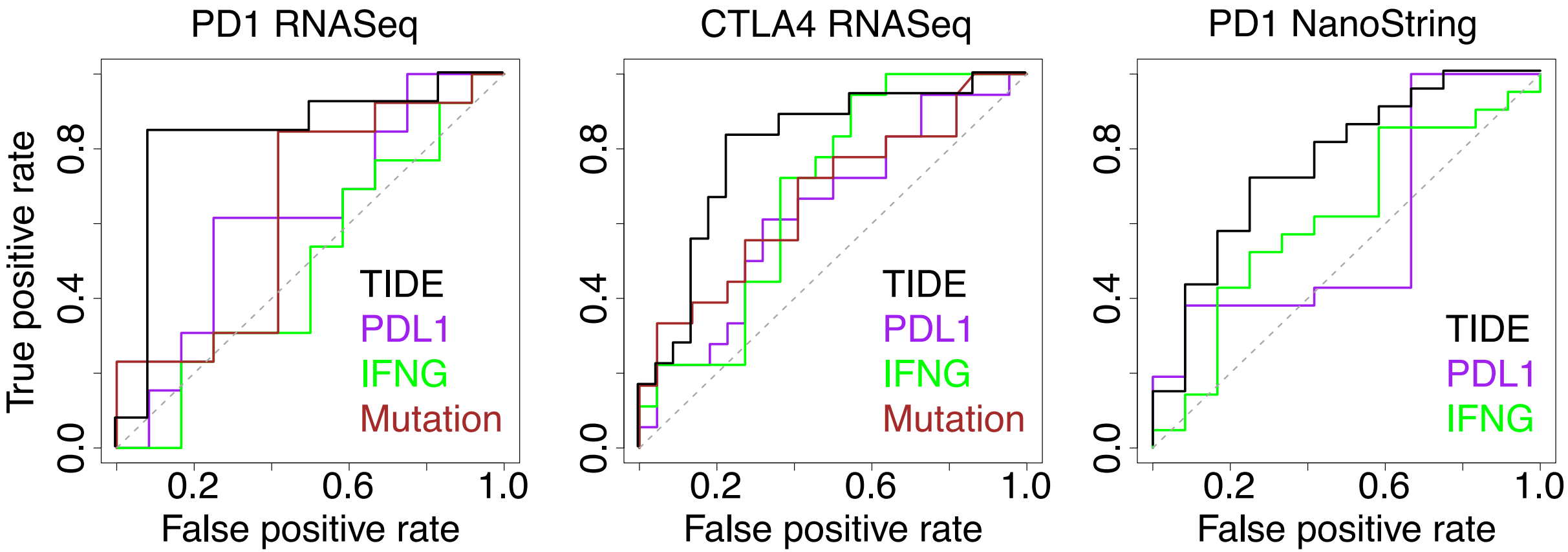




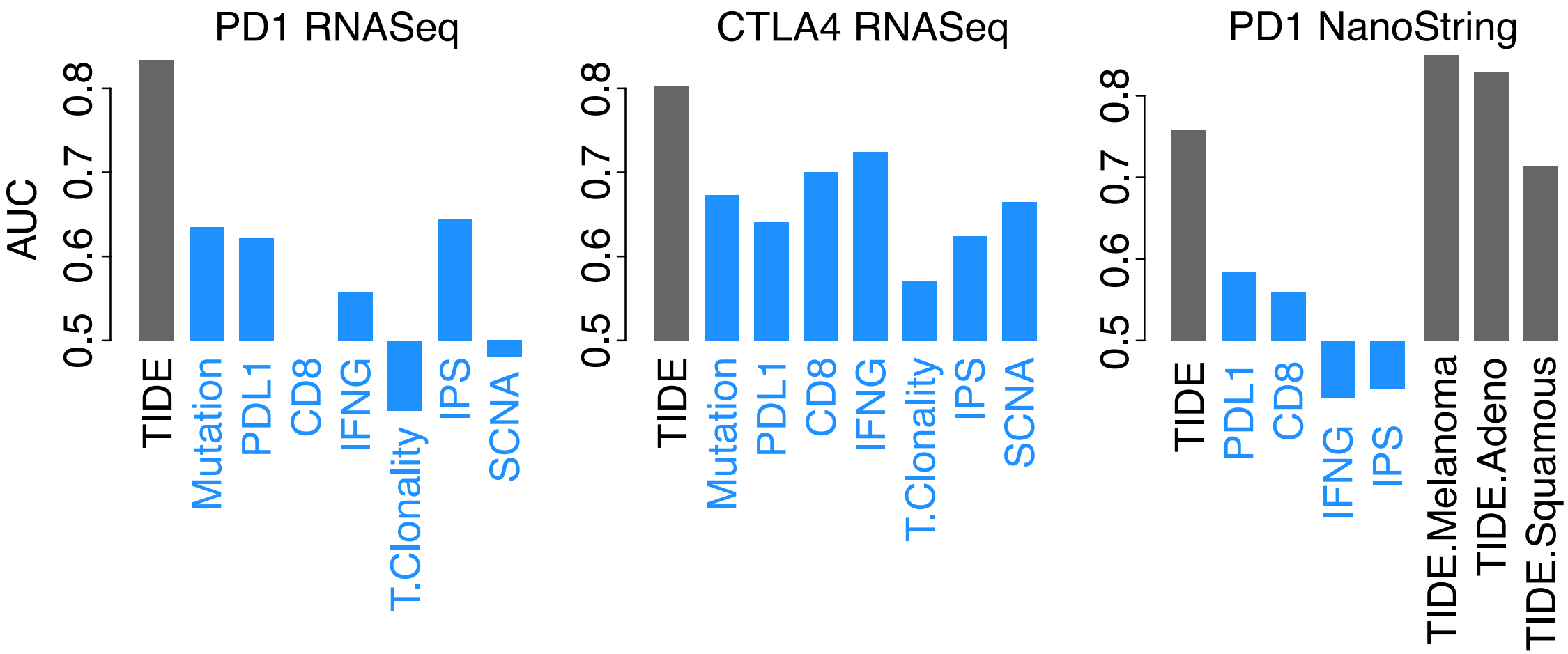
Biomarker A is better than Biomarker B

From 0 to 1 with random performance as 0.5

TIDE outperforms other biomarkers

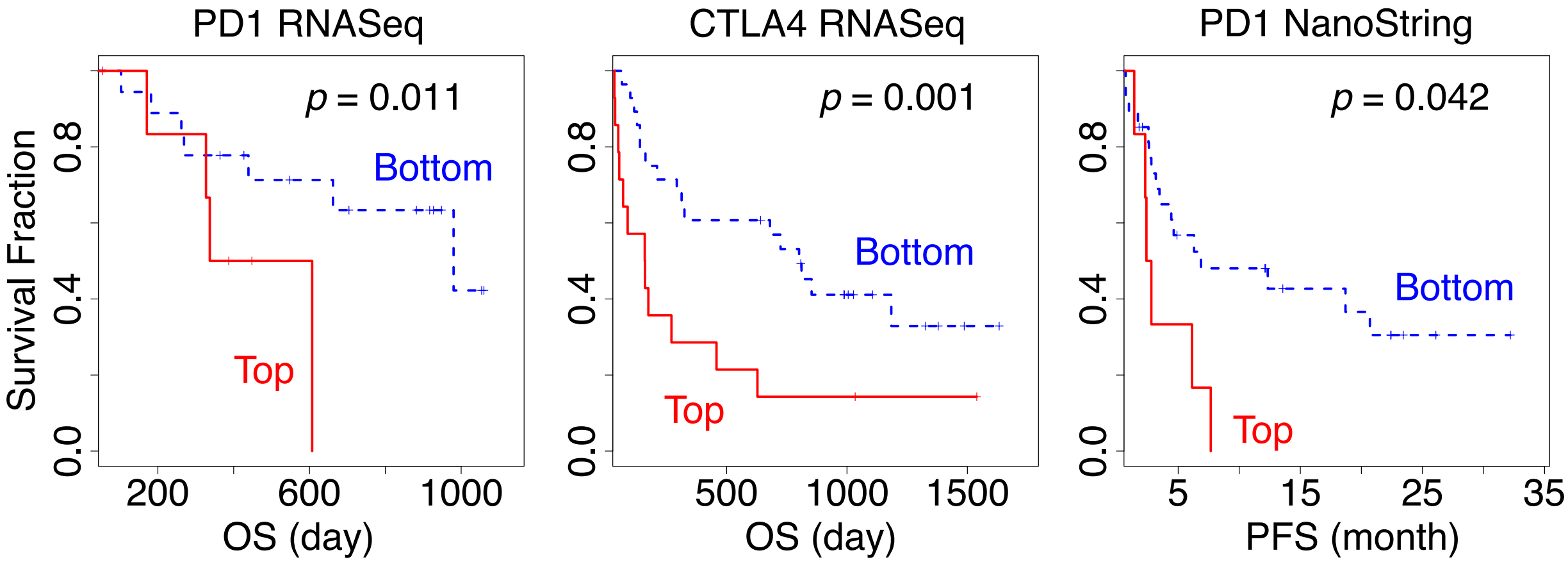


TIDE outperforms other biomarkers

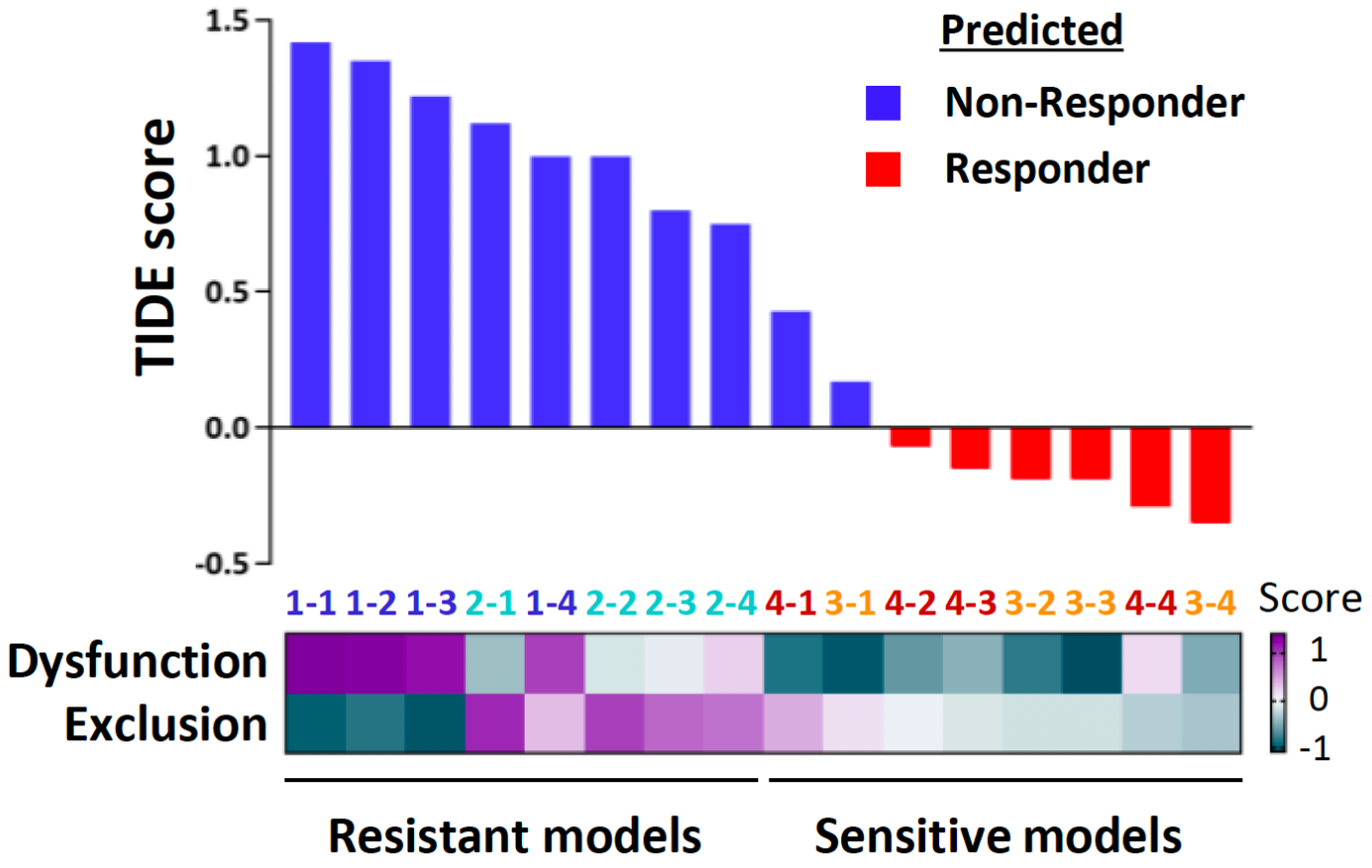


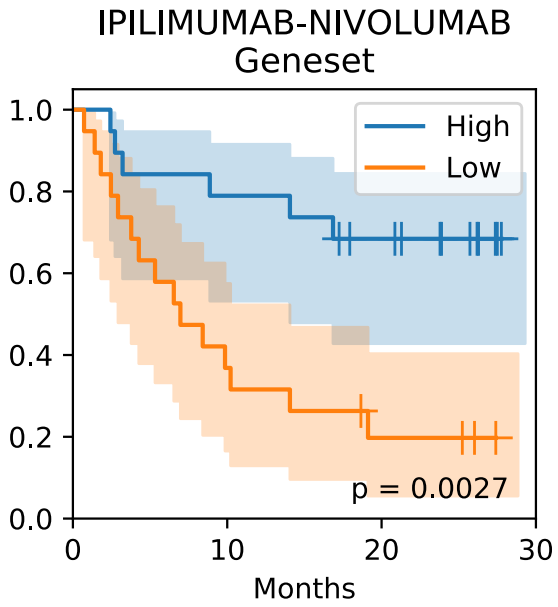
Mutation: nonsynonymous mutation load; **PDL1**: PD-L1 expression; **CD8**: CD8A + CD8B expression;
IFNG: interferon gamma response expression signature (Ayers et al. 2017); **T.Clonality**: Clonality of T cell receptor;
IPS: ImmunePhenoScore (Charoentong et al. 2017); **SCNA**: variation of copy number alterations (Davoli et al. 2017).

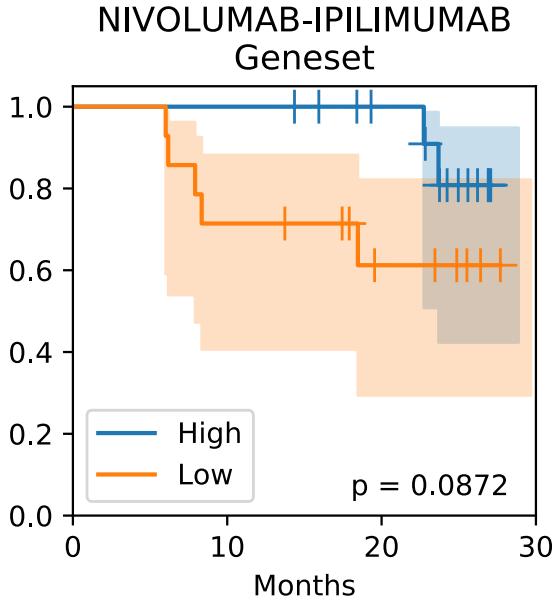
TIDE can predict patient overall survival



Independent validation of TIDE







Eva Pérez-Guijarro et al., Nature Medicine 2020 @ Glenn Merlino Lab

Double-blinded test @ BMS064

Immunotherapy Clinical Cohorts Example : STAT1

Result for query: STAT1 (6772)

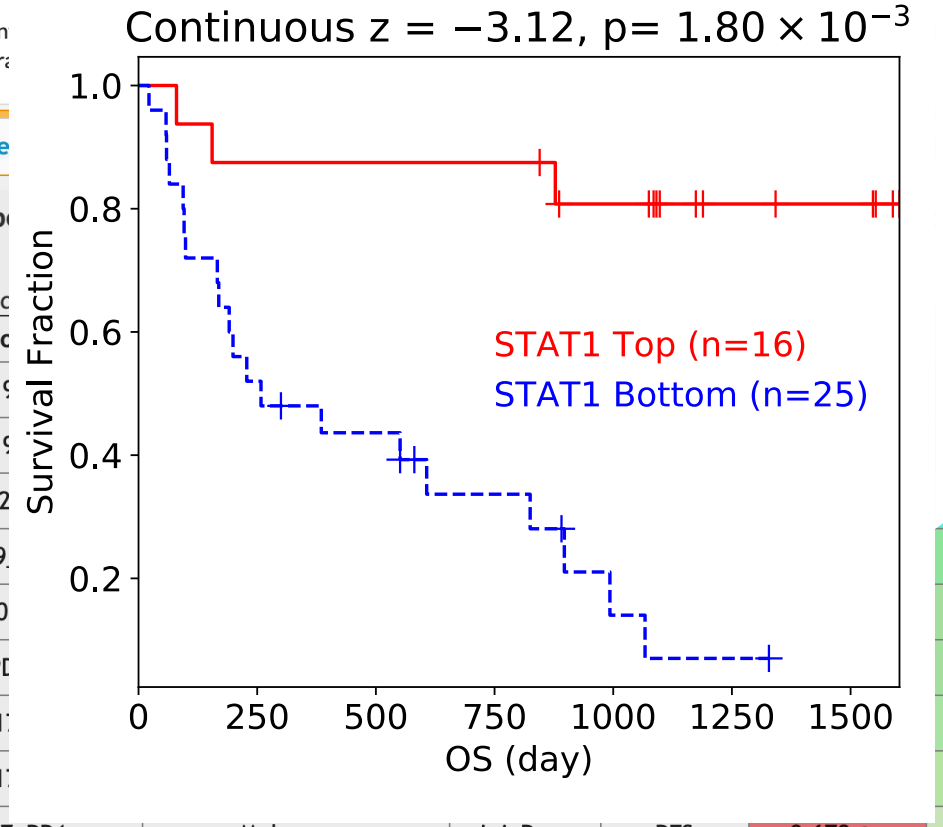
We listed the high con
scores in immunothera

lso included several reference results, including Re
ction.

Core | Expre | en

Association b | point blockade

Click table head



Coho					sk	Risk.adj	Count
Gide2019						-0.851	41
Gide2019						-2.084	41
Mariathasan2					2.474	-1.268	348
Liu2019					-1.909	-0.247	47
Nathanson20					-1.678	0.657	15
Gide2019_PD					-1.673	-0.887	32
Lauss201					-1.608	-2.219	25
Lauss201					-1.543	-1.634	25
Riaz2017_PD1	Melanoma	Ipi_Prog	PFS	0.678	-1.377	-0.367	26
VanAllen2015_CTLA4	Melanoma		OS	0.734	-1.371	0.391	42
VanAllen2015_CTLA4	Melanoma		PFS	0.734	-1.313	-0.124	42
Liu2019_PD1	Melanoma	Ipi_Prog	OS	0.739	-1.089	0.062	47
Gide2019_PD1+CTLA4	Melanoma		OS	0.879	-0.956	-0.237	32
Nathanson2017_CTLA4	Melanoma	Pre	OS	0.584	-0.803	-1.576	9
Riaz2017_PD1	Melanoma	Ipi_Naive	OS	0.757	-0.752	-0.753	25

CRISPR Screens Example: STAT1

Core | Expression | Copy_Number | Methylation | Exclusion | Ref_Therapy | Ref_Screen

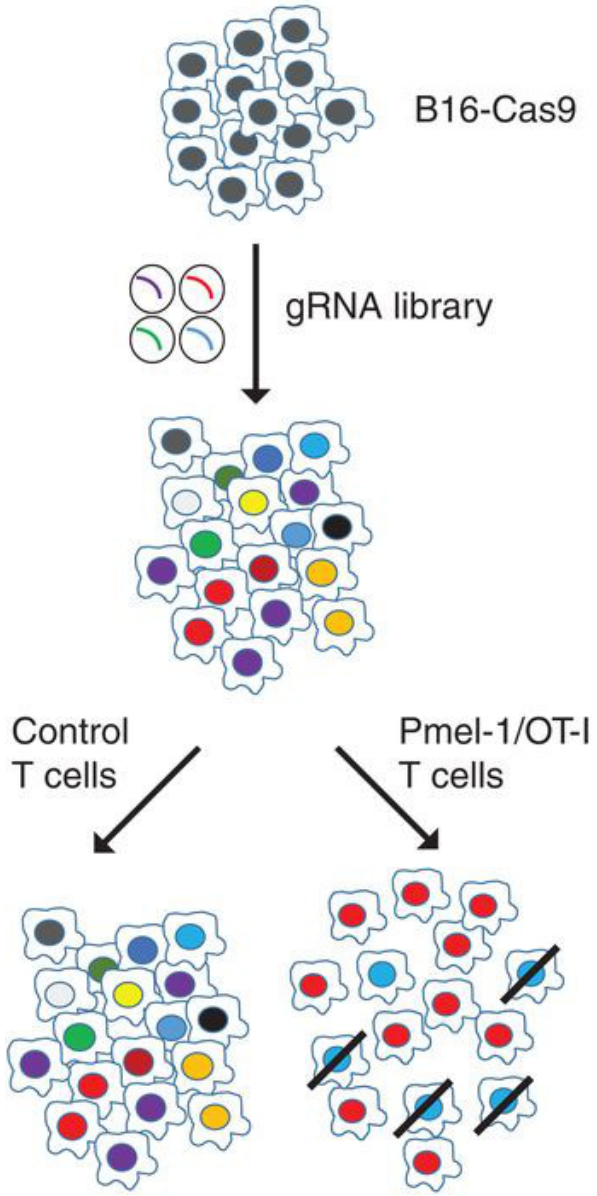
Gene knockout phenotype from genetic screens profiling regulators of T-cell mediated tumor killin

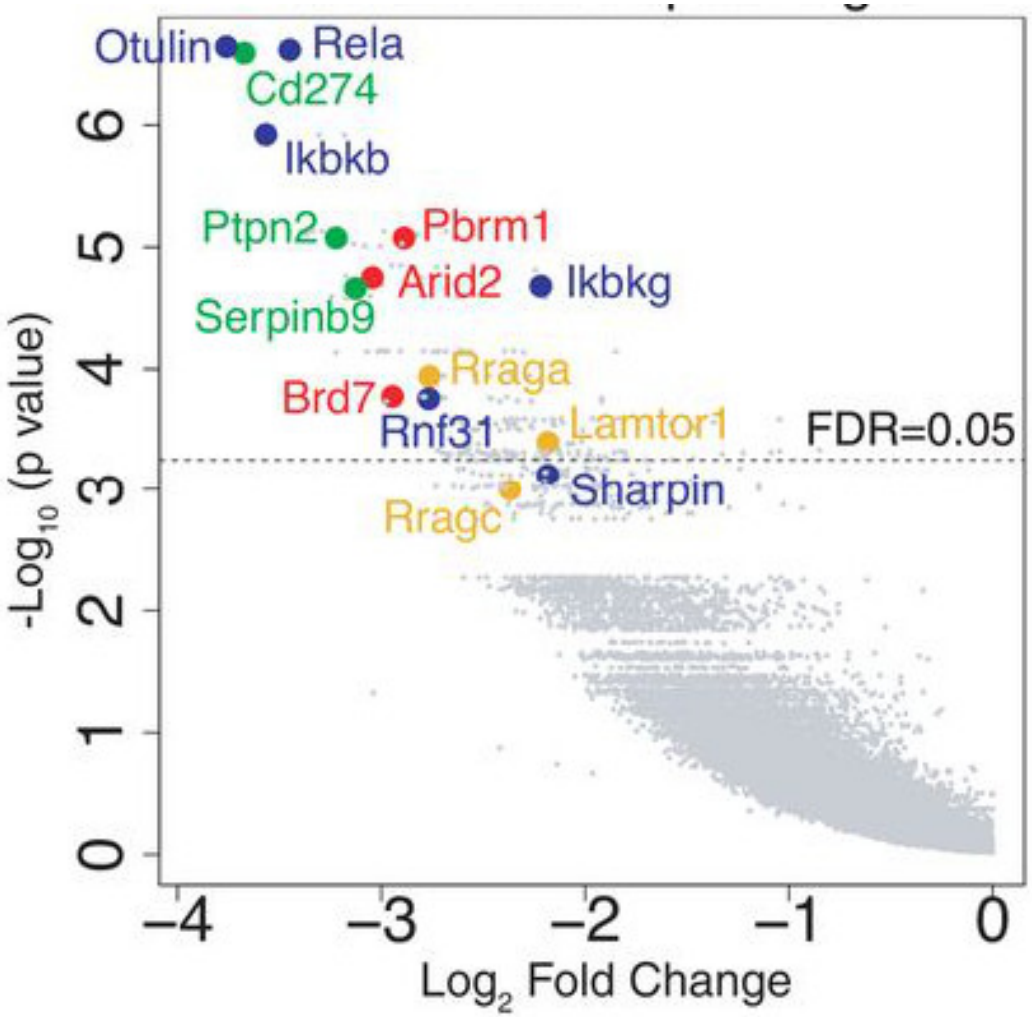
Click table header for sorting. Hover your mouse on a cell for description or full name of abbreviation.

Cohort	Cancer	Model	Condition	logFC	Z-score
Pan 2018	Melanoma	B16	Pmel1	8.801	13.217
Freeman 2019	Melanoma	B16	OT1	12.215	8.91
Manguso 2017	Melanoma	B16	GVAX+PD1	2.393	5.924
Pech 2019	Leukemia	K562	NK_E:T=2.5	1.521	5.782
Pan 2018	Melanoma	B16	OT1	2.616	5.046
Pech 2019	Leukemia	K562	NK_E:T=1	0.987	4.122
Manguso 2017	Melanoma	B16	GVAX	1.199	3.615
Patel 2017	Melanoma	Mel624	1	0.513	3.492
Kearney 2018	Colon	MC38	T_IgG	4.193	2.903
Kearney 2018	Colon	MC38	T_PD1	4.329	2.795
Patel 2017	Melanoma	Mel624	2	0.255	1.462
Shifrut 2018	Donor	T cell	Average	0.214	1.346
Vredevoogd 2019	Melanoma	D10 IFNGR1-	MART1	0.231	1.344
Meyers 2017	Mix	Cell panel	Essentiality	0.126	0.459
Kearney 2018	Colon	MC38	NK_20	0.963	0.376
Kearney 2018	Colon	MC38	NK_10	0.322	0.238
Freeman 2019	Melanoma	B16	NK	-1.425	-1.055

Export to CSV

CRISPR screen for T-cell mediated tumor killing





Pan*, Kobayashi*, Jiang* et al., *Science*, 2018

Gene Prioritization

Putting everything together

Response Prediction | Query Gene | Regulator Prioritization | Biomarker Evaluation | Download | FAQ | Contact | News | Peng J

README

Input

- Input a gene set with whole Entrez ID or whole official gene symbol. **(separated by ",")**
- **Example:**
- Manualy Input:

PTPN2, AXL, EPHA3, CACNA1G, DNMT3A, EPHA2, PDE1B, ROR2, HTRA1, GZMM, KCNK2, KCNF1, SV2A, CLCNKB, SCN9A, SCN2A, HDAC3, PADI4, GRM1, HRH1

- Gene set file:

Download

RUN

Manualy input Genes

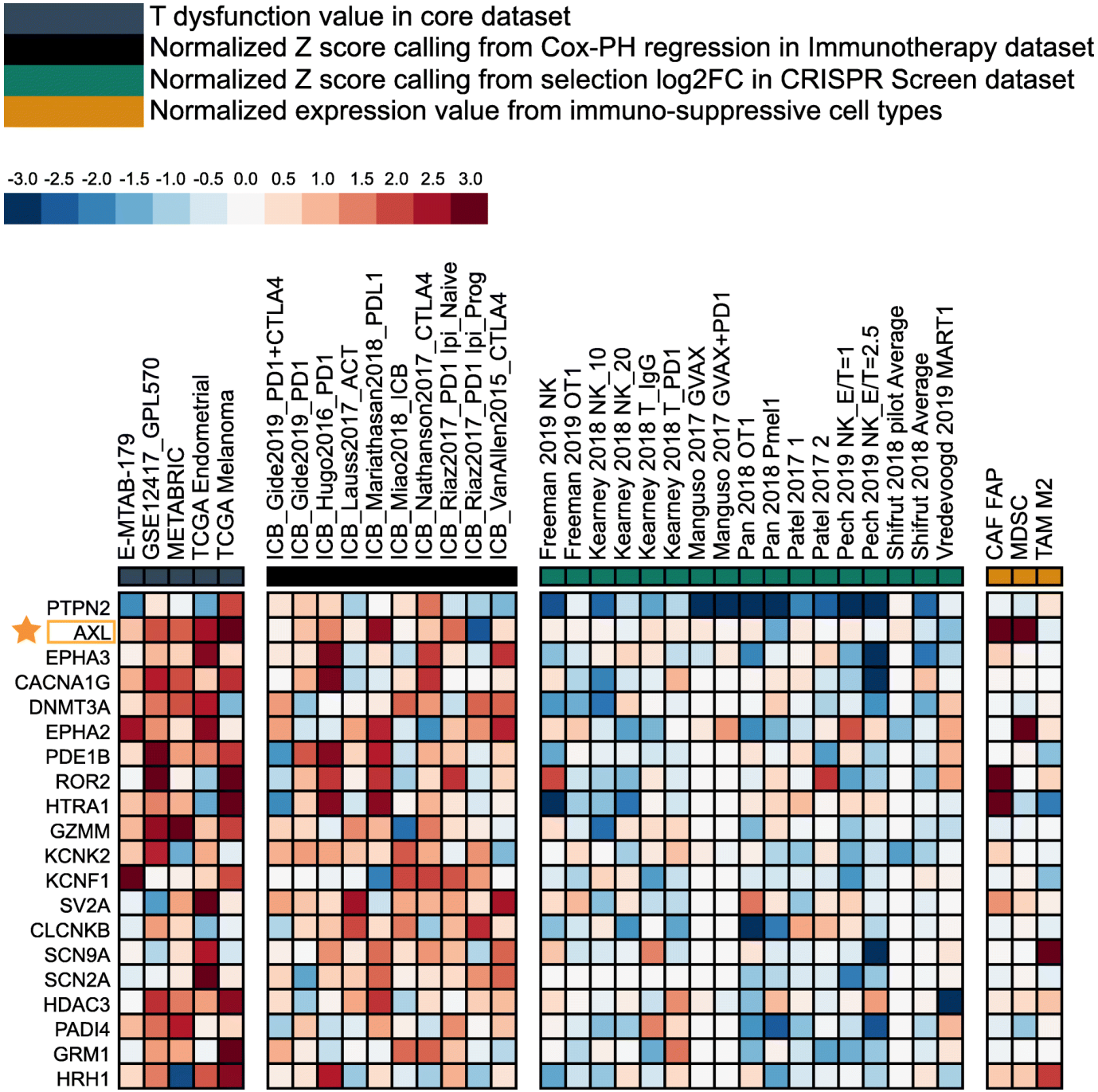
Genes

OR

Upload Geneset File

Submit

Gene Prioritization



Biomarker Evaluation

Response Prediction | Query Gene | Regulator Prioritization | Biomarker Evaluation | Download | FAQ | Contact | News | Peng Jiang

Download the video

TGFB1,-1

PDL1,1

FAP,-1

VEGFA,-1

IFNG,1

ANGPT2,-1

Dowload

README

1 Input genes that you are interested.(Using comma to seprate genes. eg. PD1, CTLA4) or upload a gene file.
2 This function will use the average/weighted(if a gene file was uploaded) expression of the custom geneset as a new biomarker.
3 Comparing the new biomarker with existed biomarkers, showing bar plots.
4 You can download interested plots by clicking the ≡ button.
5 **By clicking each point in the survival plot, you can view the corresponding Kaplan Meier plot on the top.**

Manualy input Genes

Genes

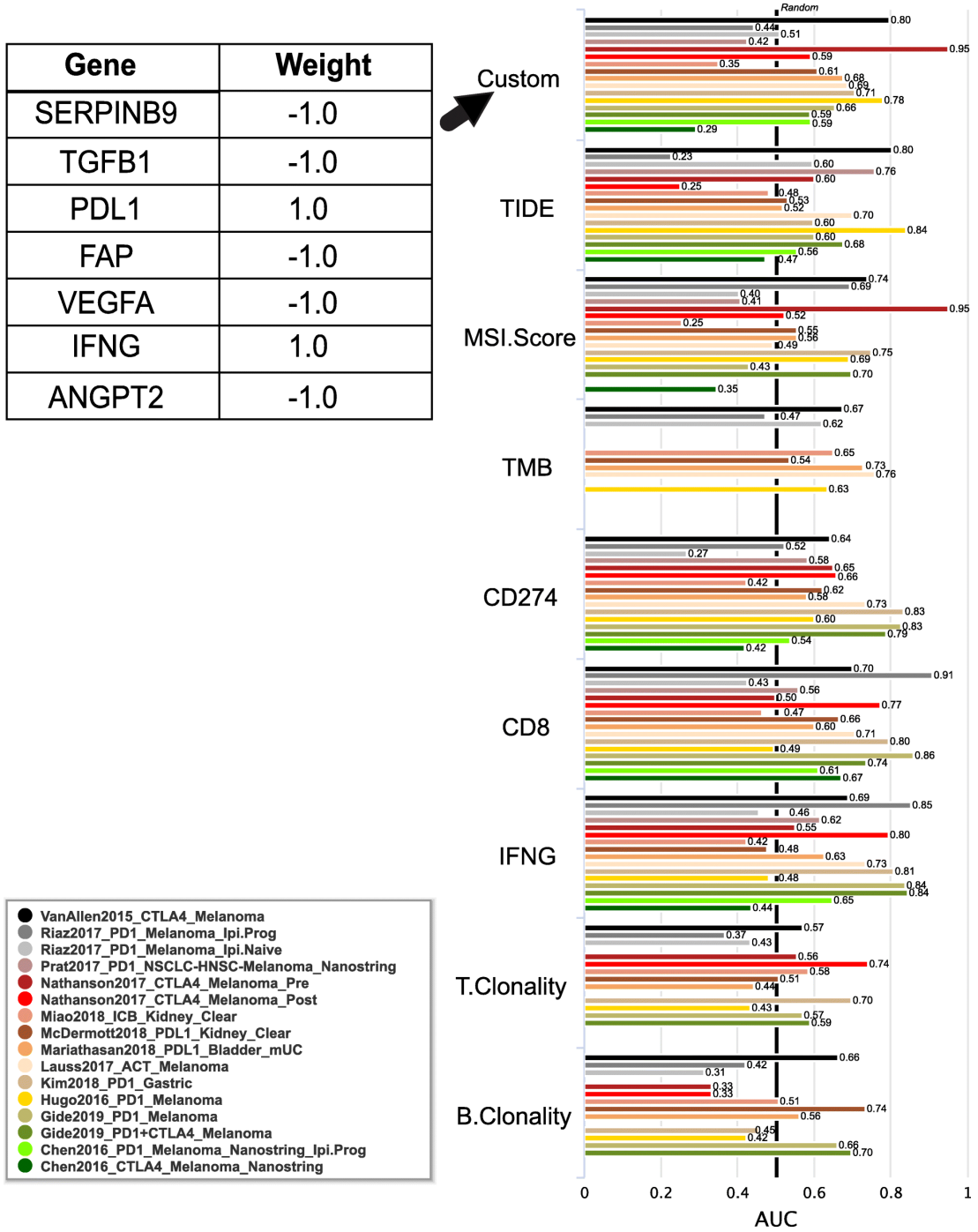
OR

Upload Geneset File

Submit

Biomarker Comparison

Gene	Weight
SERPINB9	-1.0
TGFB1	-1.0
PDL1	1.0
FAP	-1.0
VEGFA	-1.0
IFNG	1.0
ANGPT2	-1.0



Biomarker Comparison : Survival Plot

Running successfully:

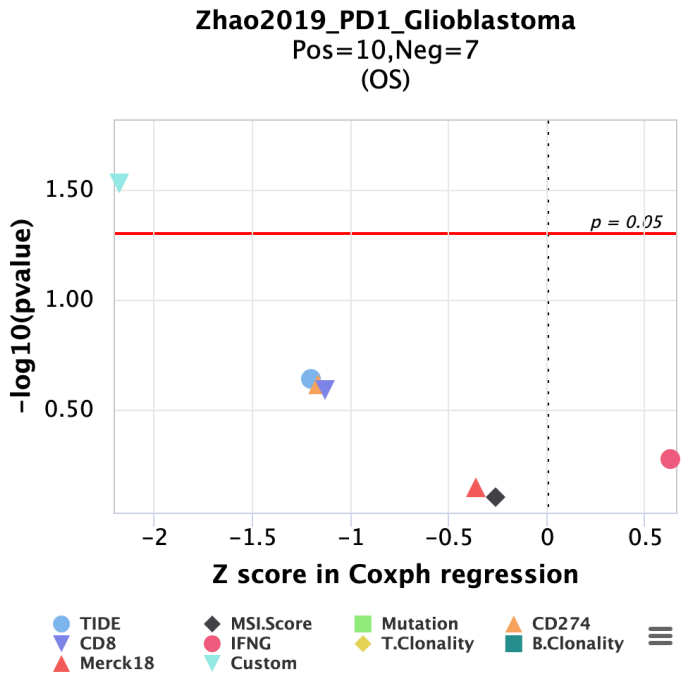
- Input: gene_weight.txt

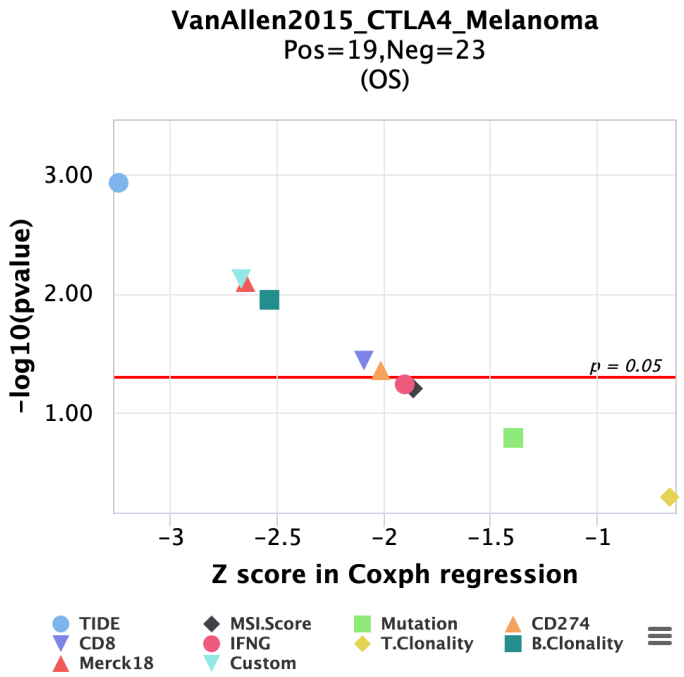
Warning

- 85.7% genes were mapped in Chen2016_PD1_Melanoma_Nanostring_Ipi.Prog
Pos=6,Neg=9
- 85.7% genes were mapped in Chen2016_CTLA4_Melanoma_Nanostring
Pos=5,Neg=11
- 57.1% genes were mapped in Prat2017_PD1_NSCLC-HNSC-Melanoma_Nanostring
Pos=21,Neg=12

- Click ✖ to start new running.

Response | Overall Survival (OS)





Some online demo

2: Cell Signaling Research Platform

https://cellsig-dev.ccr.cancer.gov (temporary domain, only within NIH firewall)

NIH NATIONAL CANCER INSTITUTE

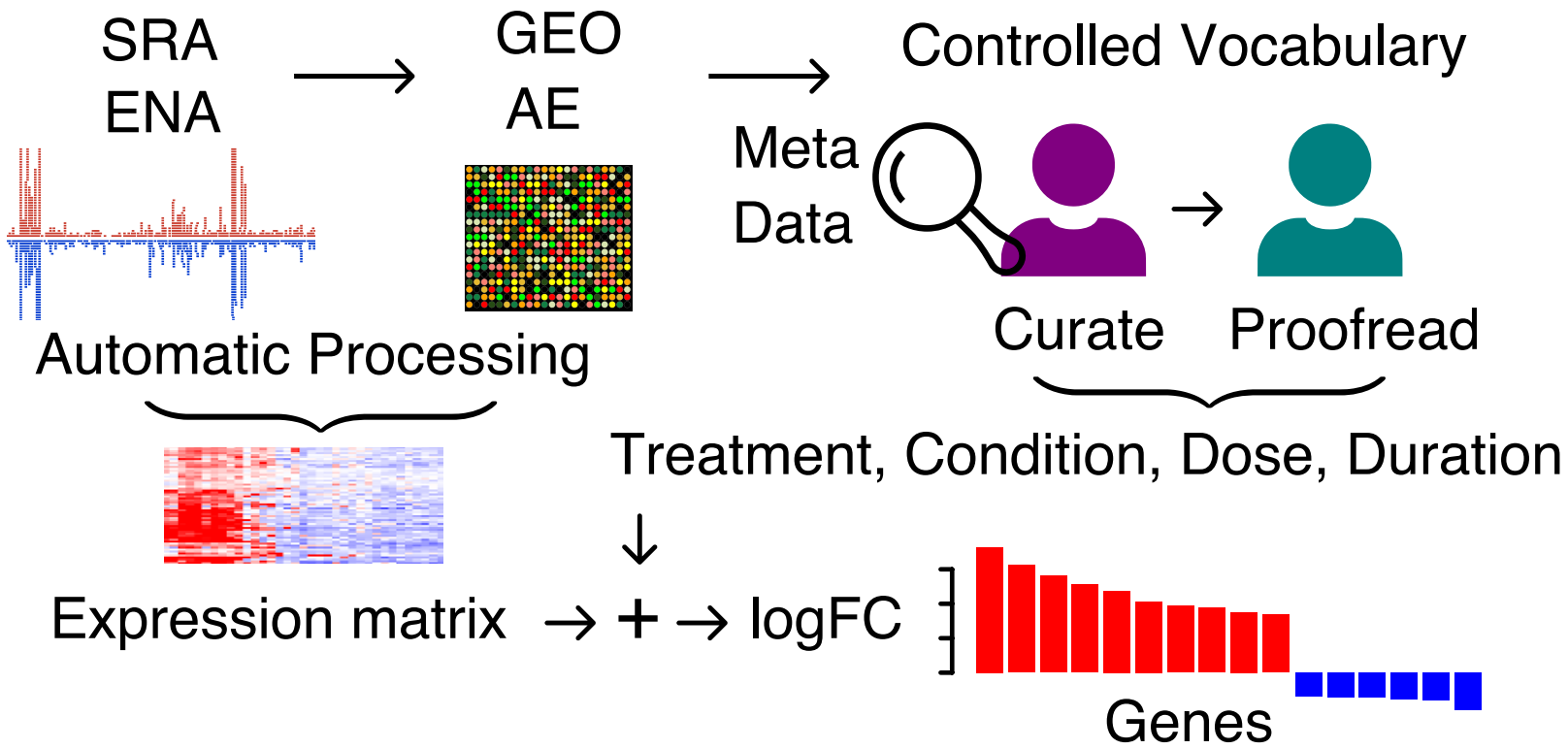
*Cell Sig*naling Database

OVERVIEW | SEARCH | RUN | DOWNLOAD | CONTACT | PENG_JIANG | LOGOUT

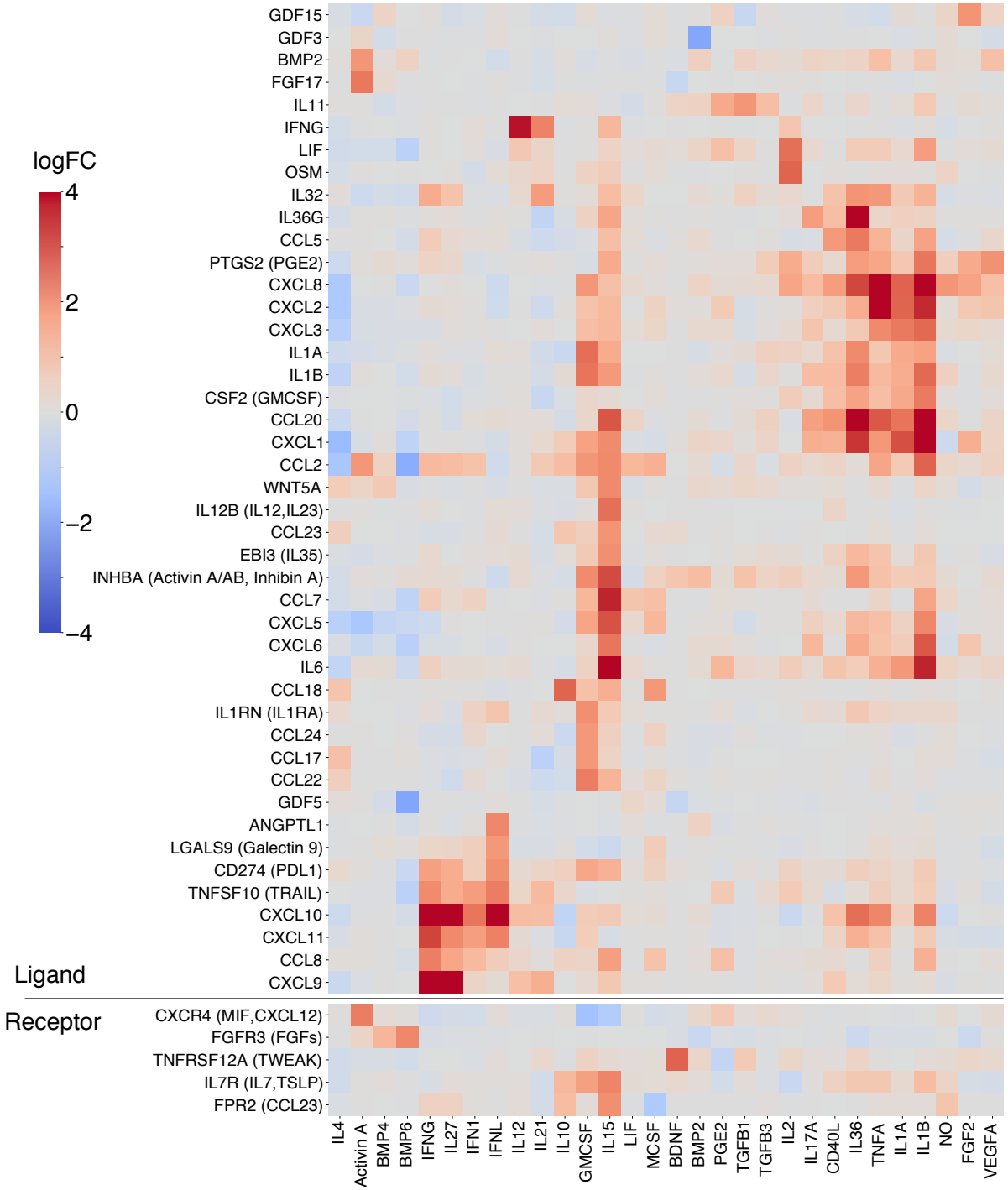
The CellSig (**Cell Sig**naling) platform aims to help biologists to study the cellular response to signaling molecules (e.g., cytokines, chemokines, and growth factors), leveraging the public expression data from treatment experiments deposited in the NCBI GEO and ArrayExpress databases. You can query cell signals and analyze genes induced or repressed (SEARCH module). You can also input a gene expression profile, and analyze the enriched signals, leveraging the treatment response profiles collected (RUN module).

U.S. Department of Health and Human Services | National Institutes of Health | National Cancer Institute | USA.gov

Data Collection



Response within signaling molecules



Gene Query

NIH NATIONAL CANCER INSTITUTE *Cell Sig*naling Database

OVERVIEW | SEARCH | RUN | DOWNLOAD | CONTACT | PENG_JIANG | LOGOUT

In the search module, you can query either a cell **signal molecule** (e.g., *TGFB*) or a **target gene** (e.g., *SOX4*). With a cell signal molecule as the input, the platform will report the log2 fold change (logFC) of target genes upon treatments from that signal in all collected studies. Please use the auto-complete function in the input box for available signal names. With a target gene as the input, the platform will report the logFC upon the treatment of all cell signals.

something like TGFB or SOX4

Search

Signal Prediction

NIH NATIONAL CANCER INSTITUTE

*Cell Sig*naling Database

OVERVIEW | SEARCH | RUN | DOWNLOAD | CONTACT | PENG_JIANG | LOGOUT

Please upload a matrix file with columns as samples and rows as genes. The value separator should be the tab, and the gene names should follow the standard gene symbols. The platform accepts Gzip files as input. Please see a sample file here. The file size should be smaller than 10MB, and the number of samples should be less than 20. If you need to analyze more samples or complicated functions (e.g., background correction), please download our software and run locally.

Note: We recommend input differential profiles between the two conditions. If your data is from a collection of samples without pairs, please mean-centralize the value of each gene across all samples or apply background correction available in the download version (Refer to the help document for details).

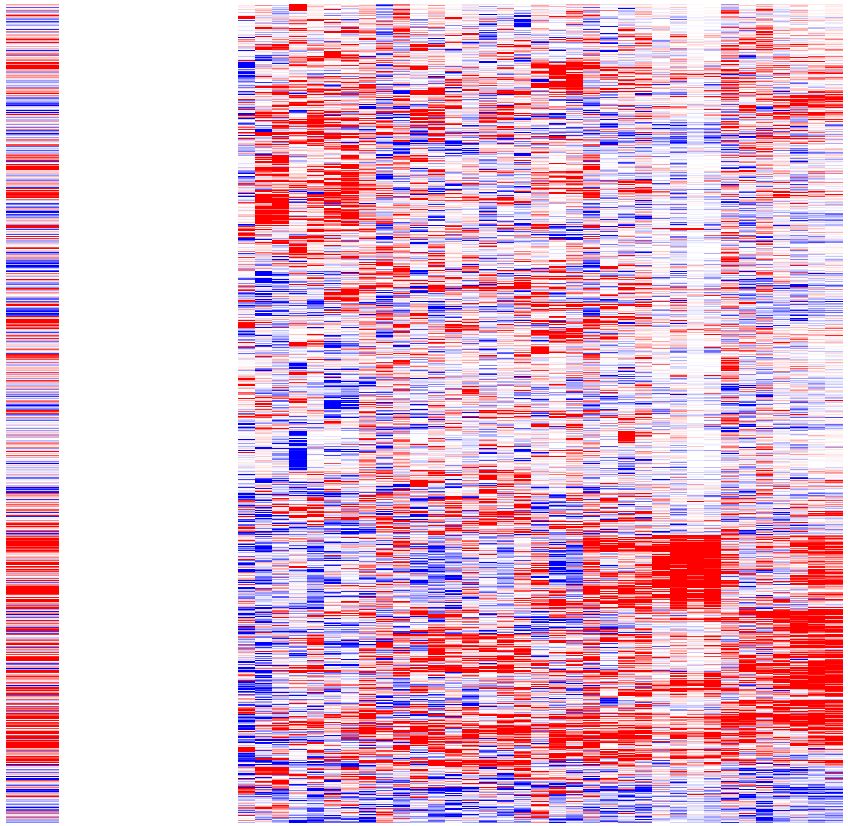
Task_file:

Choose File No file chosen

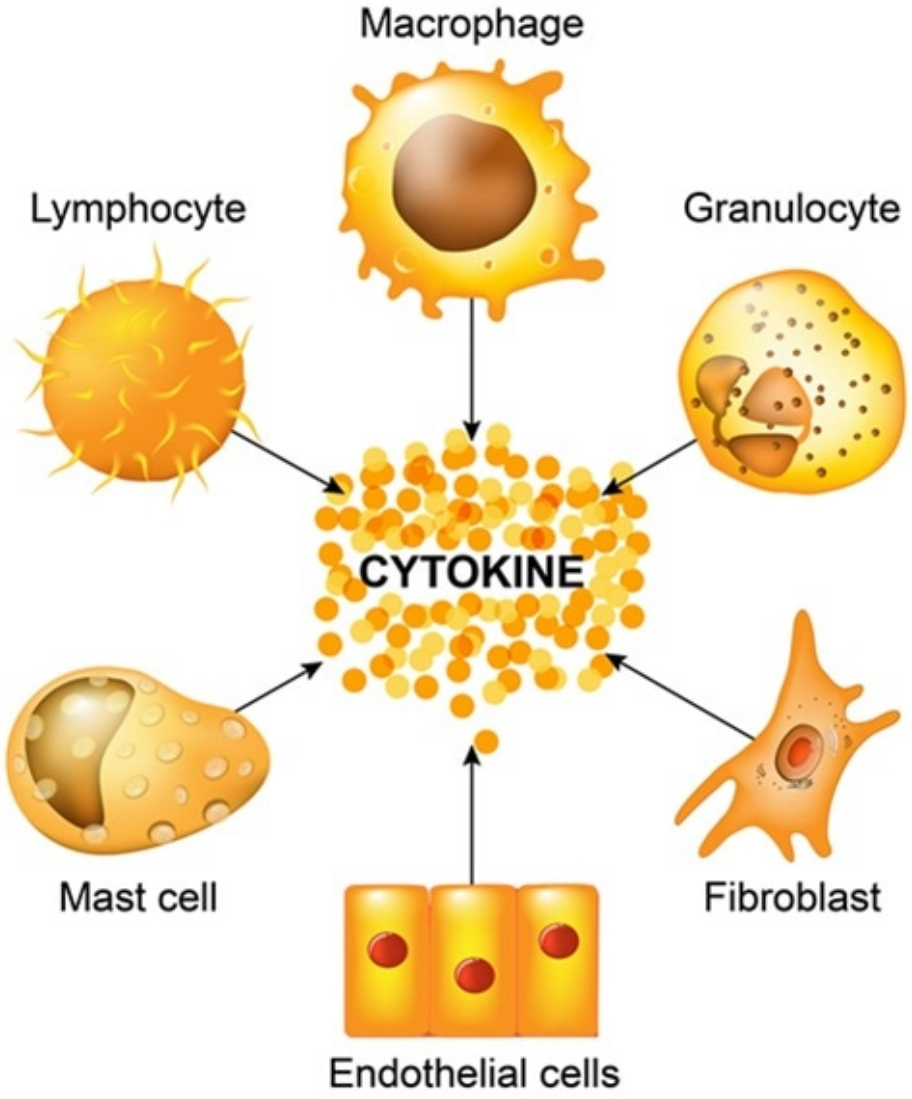
Submit

Signal Prediction

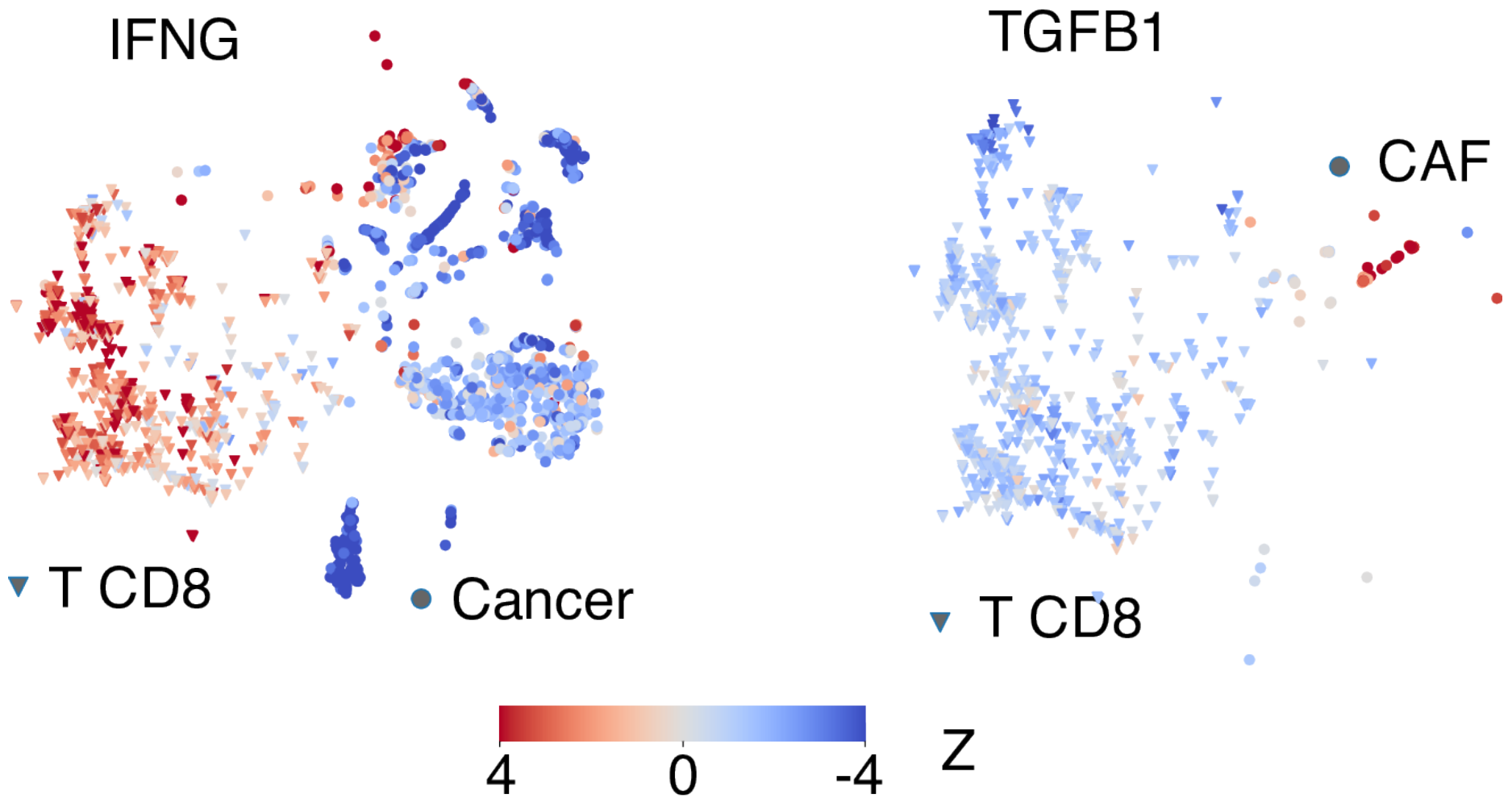
input = $a_1 s_1 + \ldots + a_n s_n$



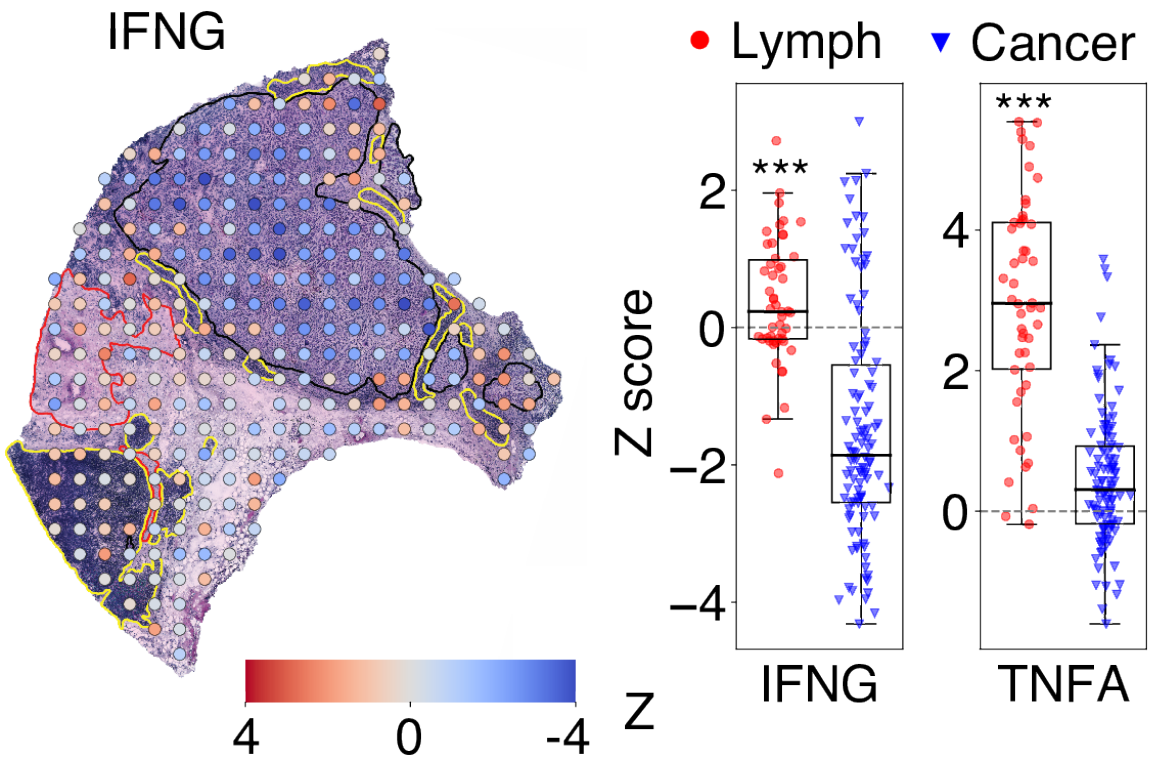
signaling molecules

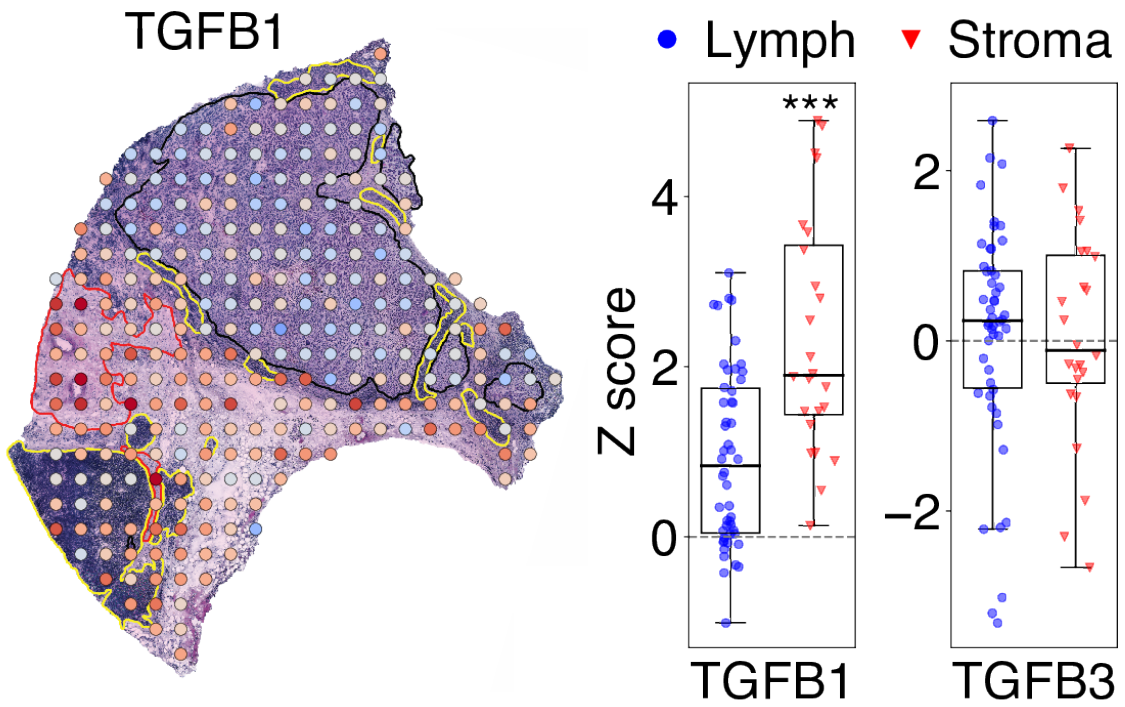


Application on Single Cell RNA-Seq



Application on Spatial Transcriptomics





Some online demo

Postdoc mentors

X. Shirley Liu

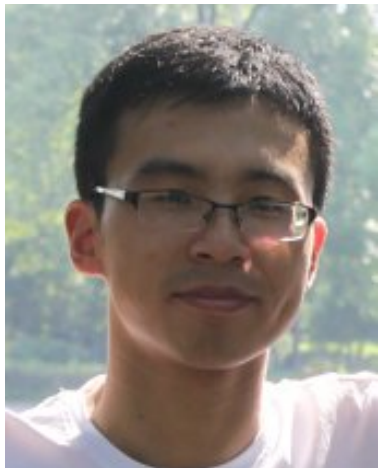
- Shengqing Gu
- Jingxin Fu

Kai Wucherpfennig

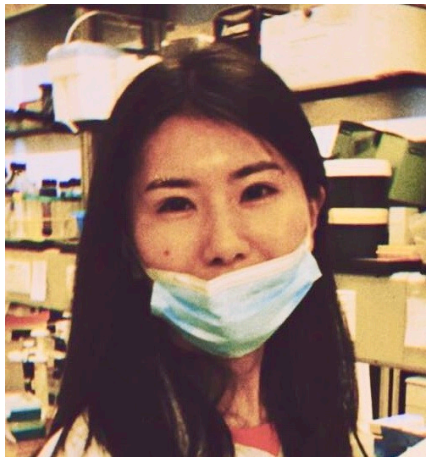
- Deng Pan
- Aya Kobayashi

Current Lab Members

Beibei Ru



Yu Zhang



Jie Yuan (expected on June, 2020)

CDSL PIs

Eytan Ruppin
Branch Chief

S. Cenk Sahinalp

Sridhar Hannenhalli