

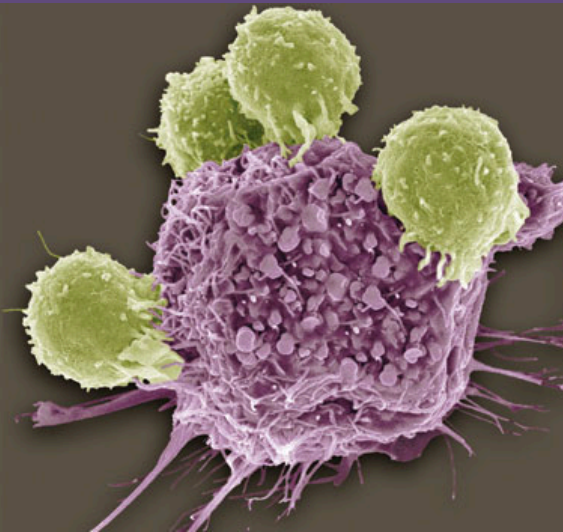
Detect Cancer from the Peripheral Blood TCR Repertoire

Bo Li

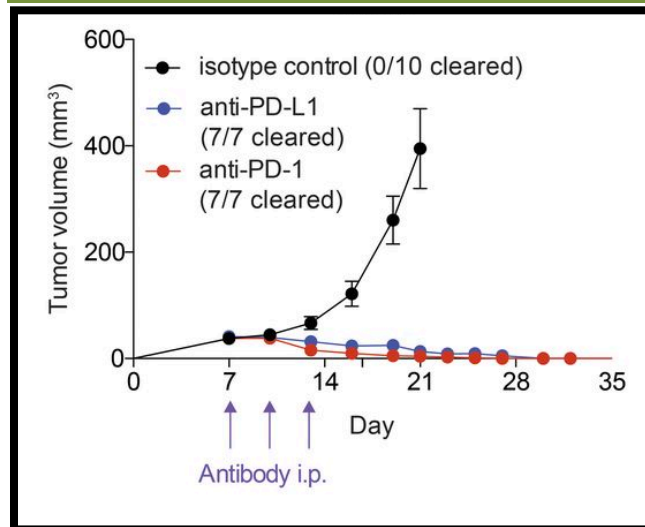
Department of Bioinformatics
UT Southwestern Medical Center

Tumor-reactive T cells are key components in immunotherapies

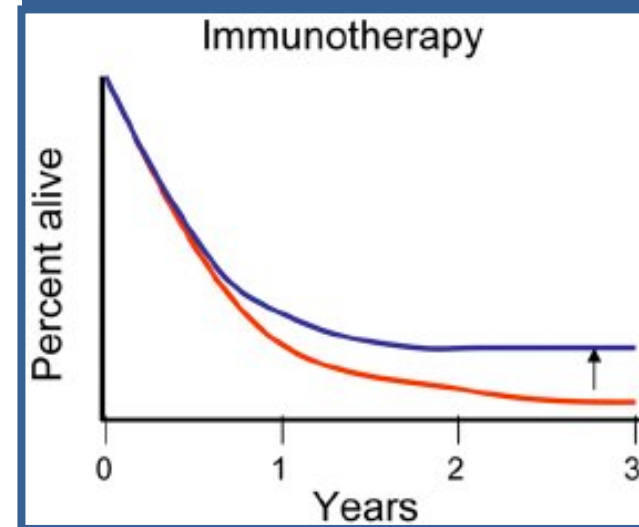
Tumor-specific killing by T cells



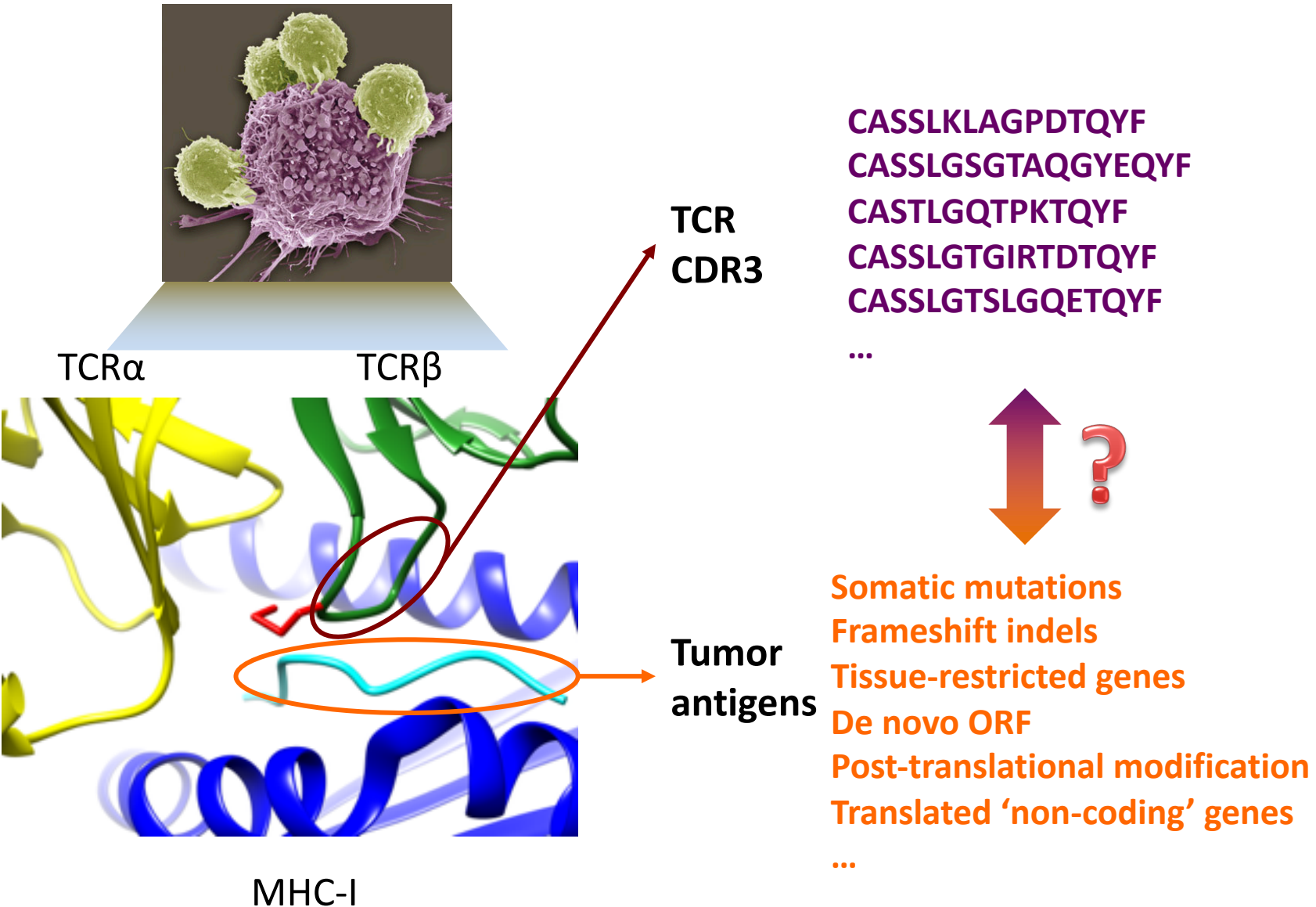
Reduced tumor growth under immunotherapy



Sustained outcome in late-stage cancer patients



Extremely diversified interactions between TCR and MHC-antigen complex



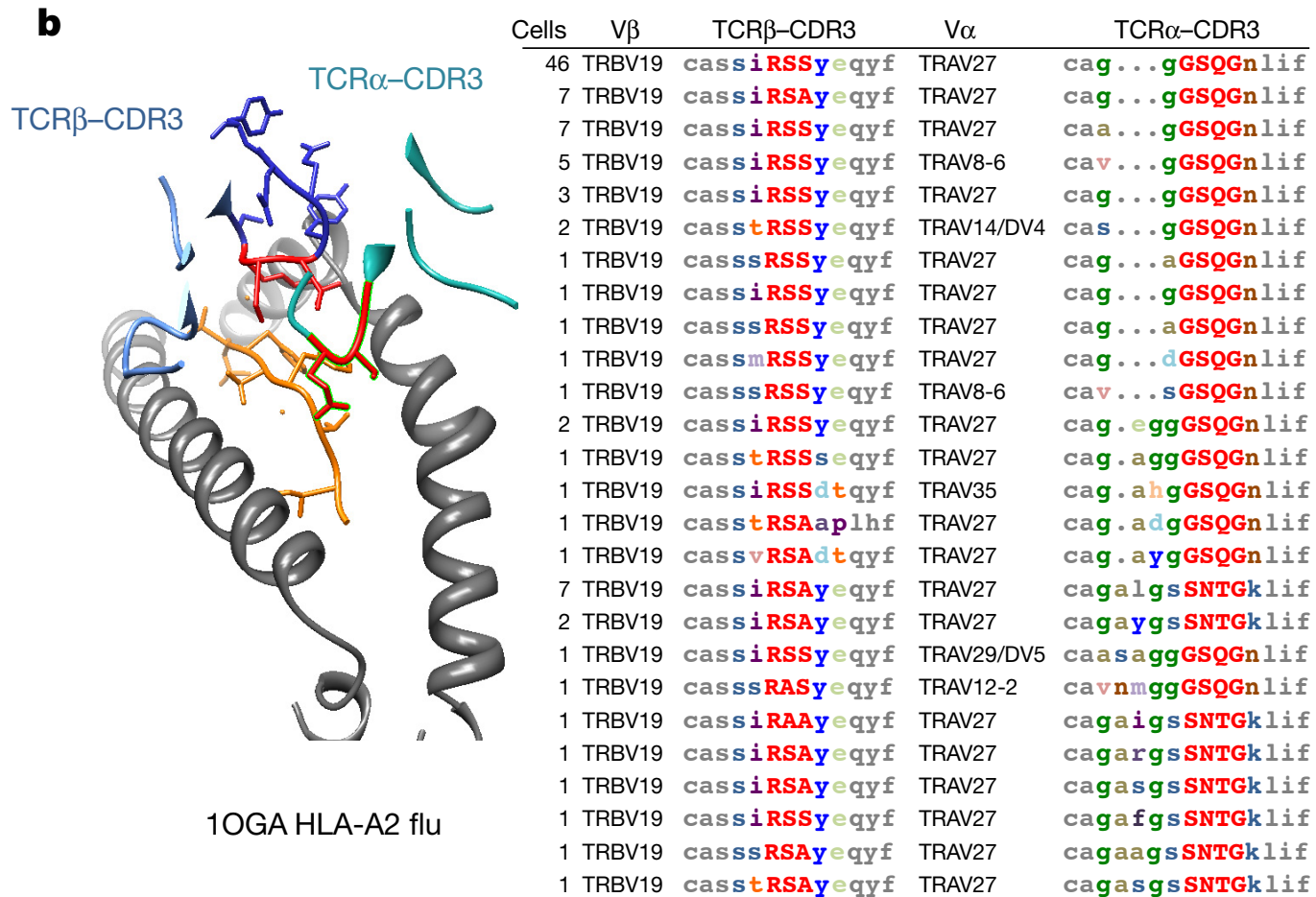
Outline

- Clustering TCRs into antigen-specific groups
- Identification of a novel cancer antigen with TCR clusters
- Novel machine learning method for de novo prediction of cancer-associated T cells

Outline

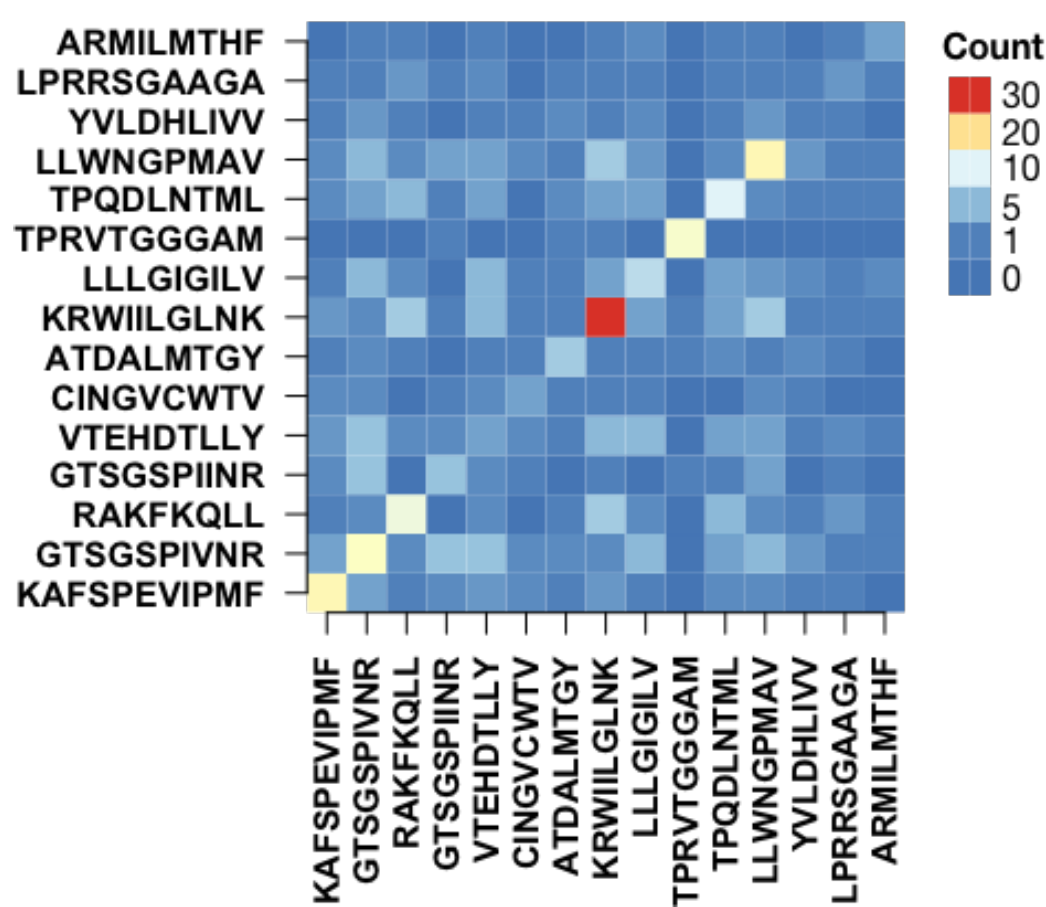
- Clustering TCRs into antigen-specific groups
- Identification of a novel cancer antigen with TCR clusters
- Novel machine learning method for de novo prediction of cancer-associated T cells

GLIPH: Clustering similar TCRs for antigen-specificity



**Shared CDR3 motifs may be surrogates
for shared antigen-specificity**

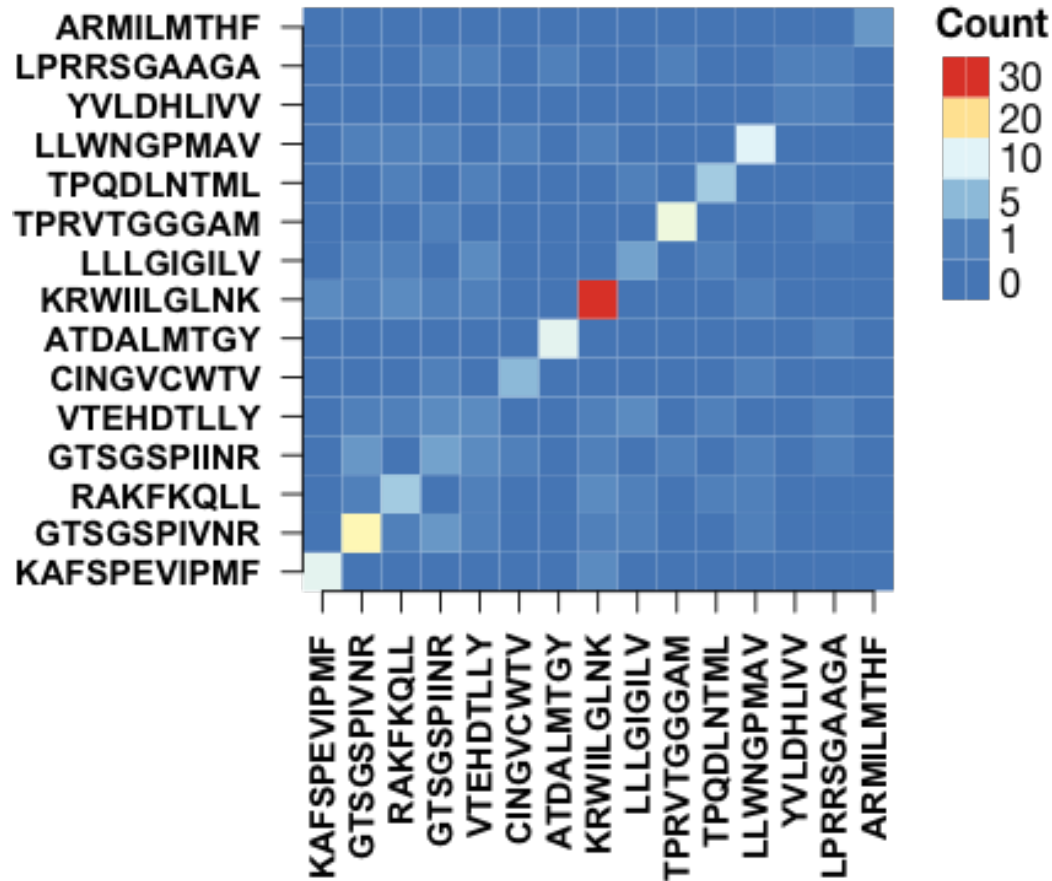
Benchmarking GLIPH antigen-specificity with antigen-specific TCR data



2,000 TCRs from:

Epitope	Epitope Species
KAFSPEVIPMF	HIV-1
GTSGSPIVNR	DENV1
RAKFKQLL	EBV
GTSGSPINR	DENV3/4
VTEHDTLLY	CMV
CINGVCWTV	HCV
ATDALMTGY	HCV
KRWILGLNK	HIV-1
LLLGIGILV	Homo Sapiens
TPRVTGGGAM	CMV
TPQDLNTML	HIV-1
LLWNGPMAV	Yellow Fever Virus
YVLDHLIVV	EBV
LPRRSGAAGA	InfluenzaA
ARMILMTHF	HCV

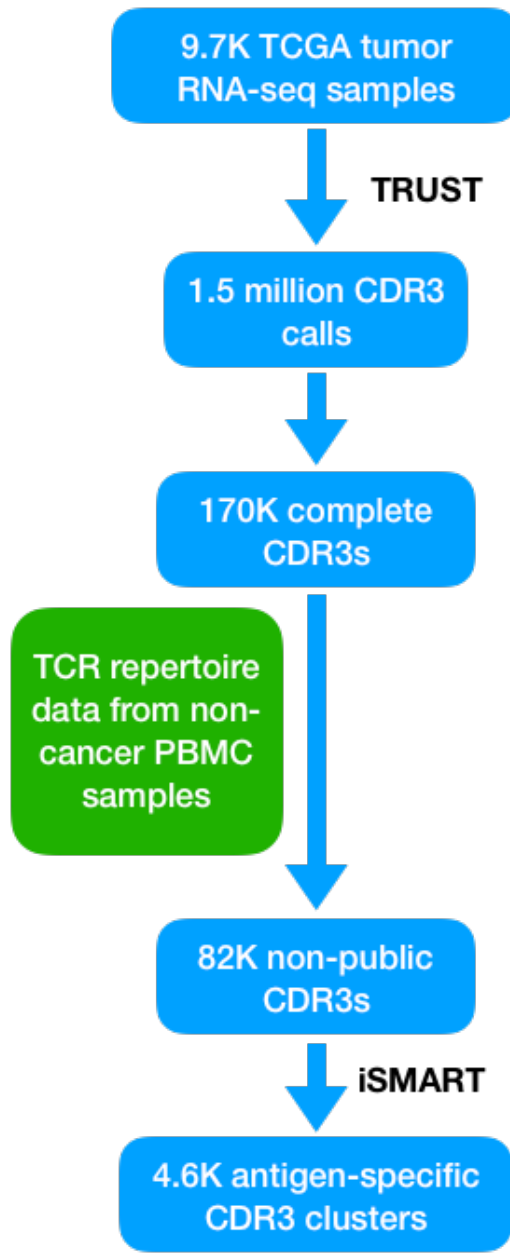
iSMART: immuno-Similarity Measurement via Aligning Receptors of T cells



iSMART workflow:

1. Perform pairwise alignment of CDR3 sequences
2. Build connectivity matrix of CDR3 sequences based on high alignment scores
3. Call CDR3 clusters based on connectivity matrix

Identification of clusters from TIL CDR3 data

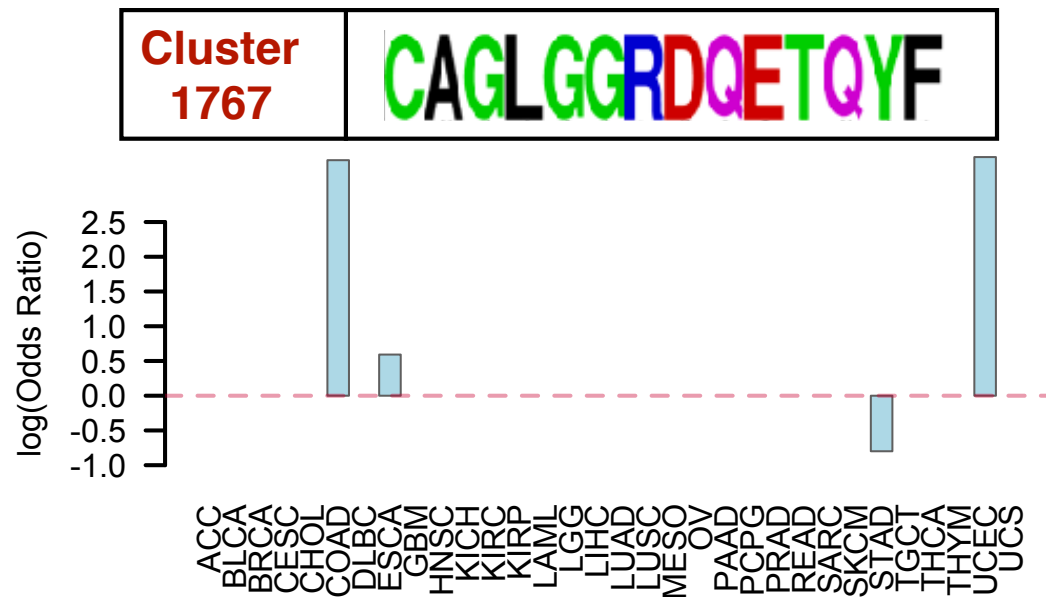
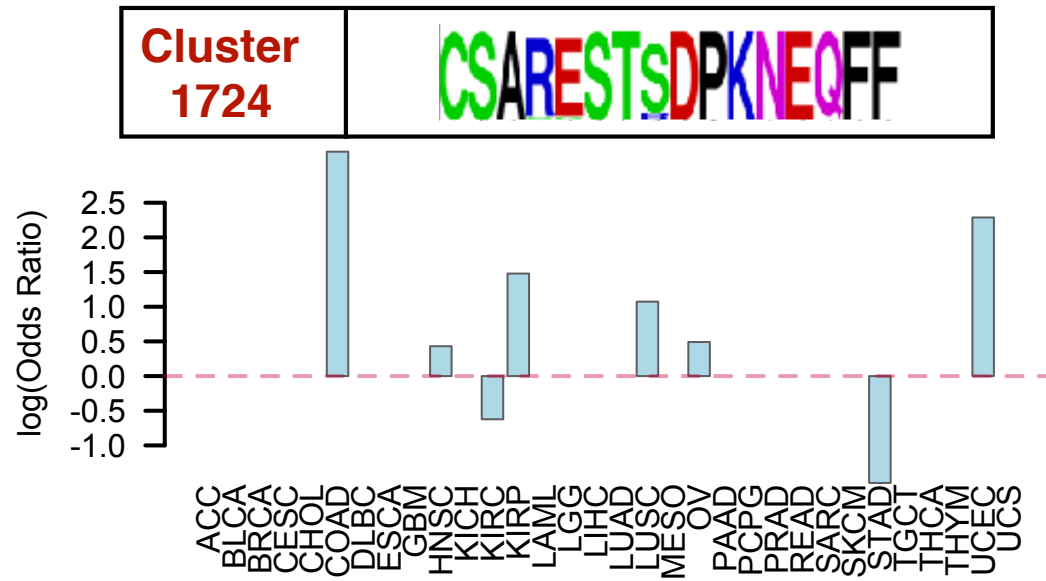


CDR3	Sample ID	Cancer	Group
CAVRDWNAGNMLTF	TCGA-4K-AA1I-01A	TGCT	1720
CAVRDFNAGNMLTF	TCGA-29-1698-01A	OV	1720
CASPKNTGFQKLVF	TCGA-EW-A1IY-01A	BRCA	1721
CAAPRNTGFQKLVF	TCGA-BR-8687-01A	STAD	1721
CASSLHSRAETQYF	TCGA-E2-A1IO-01A	BRCA	1722
CASSAHSRAETQYF	TCGA-TS-A8AV-01A	MESO	1722
CAVSEWSGGSNYKLTF	TCGA-BR-8361-01A	STAD	1723
CAVSEGS GGSNYKLTF	TCGA-L5-A4OS-01A	ESCA	1723

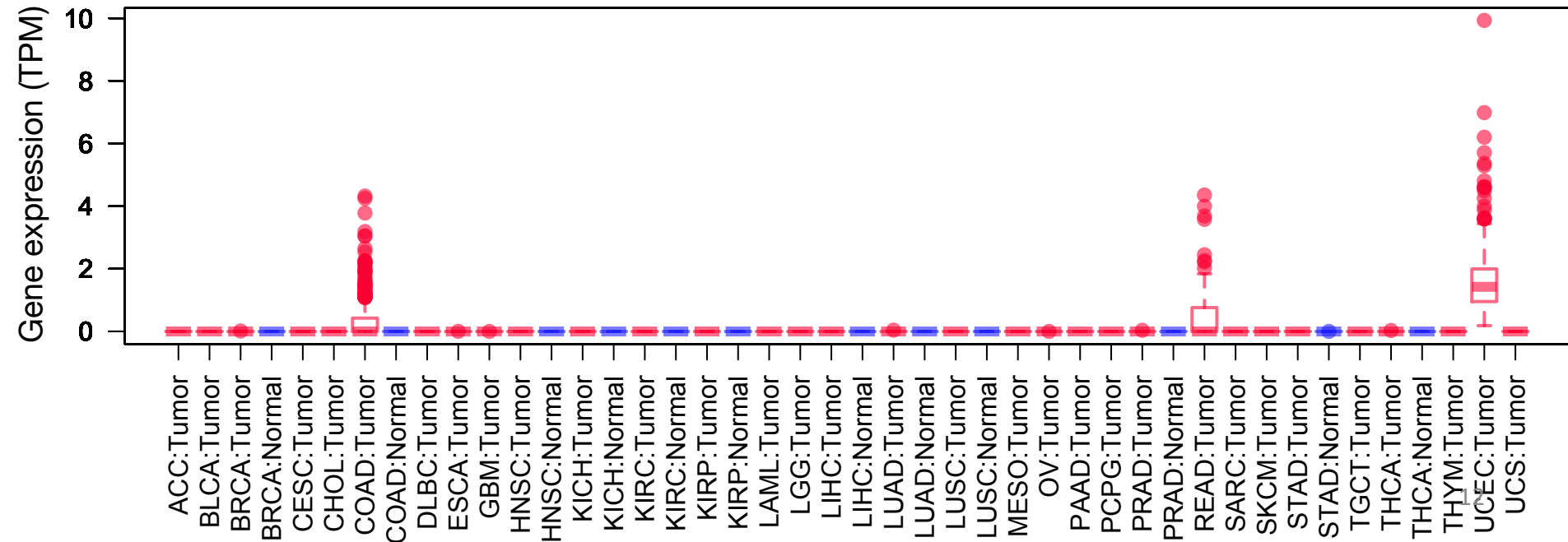
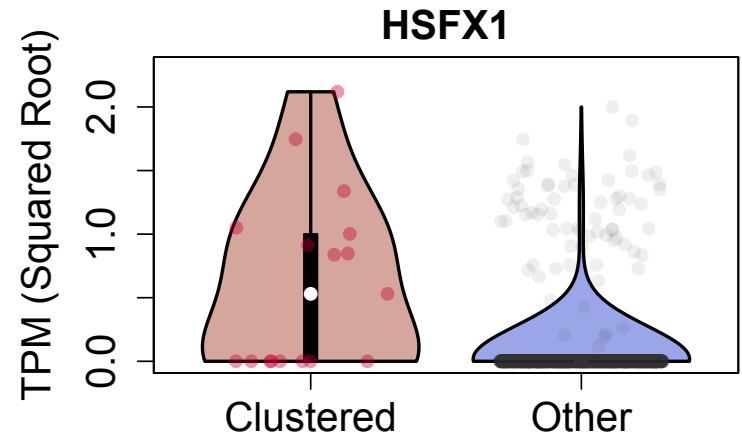
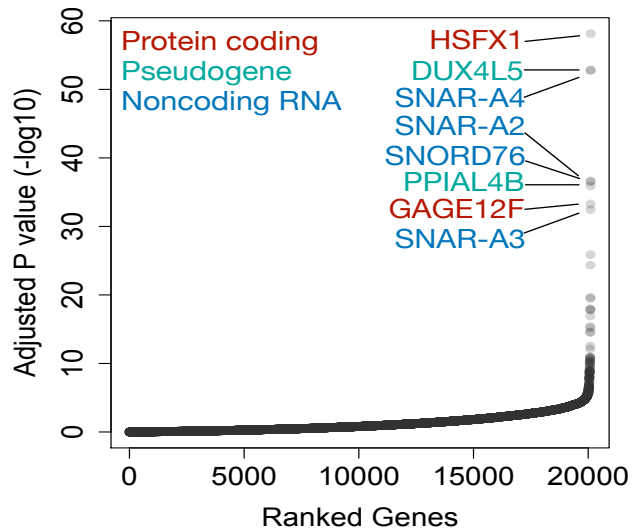
Outline

- Clustering TCRs into antigen-specific groups
- Identification of a novel cancer antigen with TCR clusters
- Novel machine learning method for de novo prediction of cancer-associated T cells

Cancer-type enrichment for two CDR3 clusters



HSFX1 is overexpressed in the two clusters



HSFX1 is a potential cancer-associated antigen

Peptide **HLA allele** **Rank** **Binder**

HSFX1: 384-392

VMFPHLPAL A*02:01 0.02 SB

VMFPHLPAL C*07:01 0.25 SB

VMFPHLPAL C*07:02 0.125 SB

Sample **Disease** **TPM** **HLA-A1** **HLA-A2** **HLA-B1** **HLA-B2** **HLA-C1** **HLA-C2**

TCGA-AA-3502 COAD 0.28 A*02:01 A*25:01 B*44:02 B*51:01 C*07:04 C*16:02

TCGA-AP-A1DM UCEC 1.00 A*03:01 A*01:01 B*08:01 B*35:01 C*07:01 C*04:01

TCGA-AA-A01K COAD 0.83 A*01:01 A*03:01 B*07:02 B*51:01 C*07:02 C*12:02

TCGA-AA-3684 COAD 1.10 A*02:01 A*24:02 B*44:05 B*52:01 C*02:02 C*12:02

TCGA-D1-A17F UCEC 1.79 A*02:01 A*02:01 B*15:01 B*51:01 C*01:02 C*03:04

TCGA-AA-3875 COAD 0.72 A*01:01 A*68:01 B*08:01 B*15:01 C*03:03 C*07:01

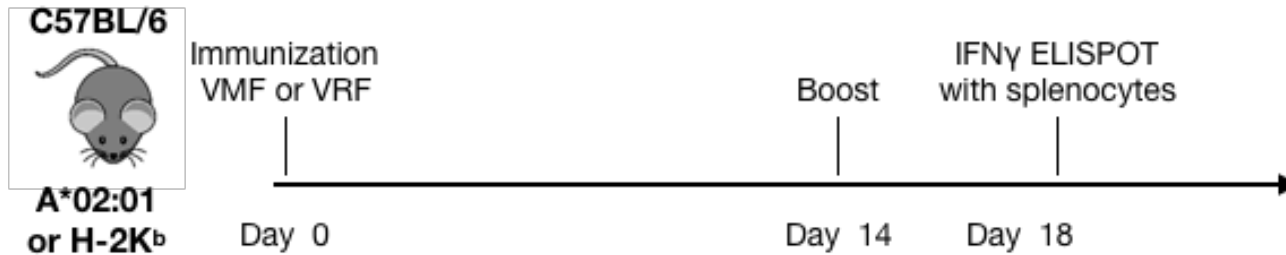
TCGA-AA-3848 COAD 3.05 A*02:01 A*29:02 B*41:02 B*44:03 C*16:01 C*17:01

TCGA-BG-A0MT UCEC 0.70 A*02:01 A*02:01 B*44:02 B*50:01 C*05:01 C*06:02

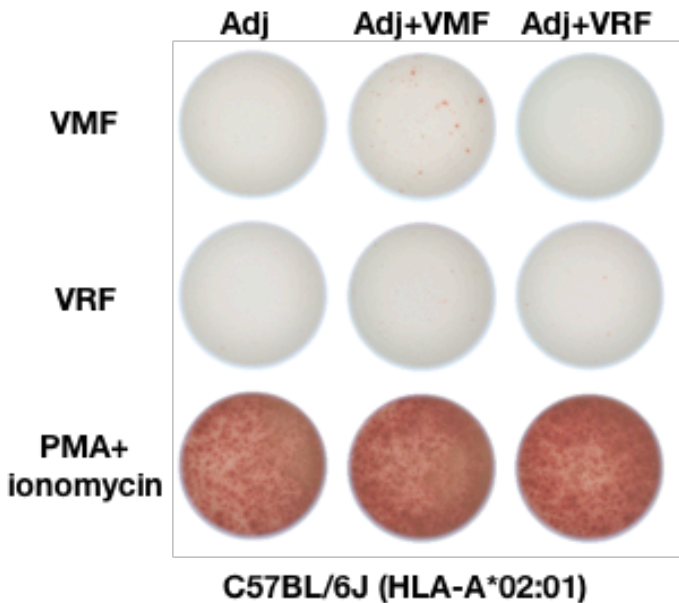
TCGA-BG-A18C UCEC 4.49 A*03:01 A*24:02 B*07:02 B*07:02 C*07:02 C*07:02

In vivo immunogenicity for the 9-mer peptide generated from HSFX1

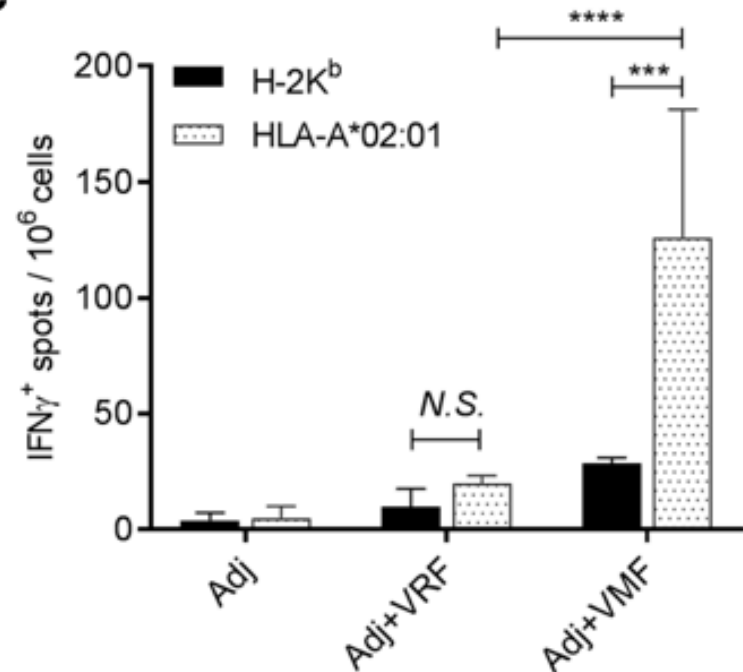
a



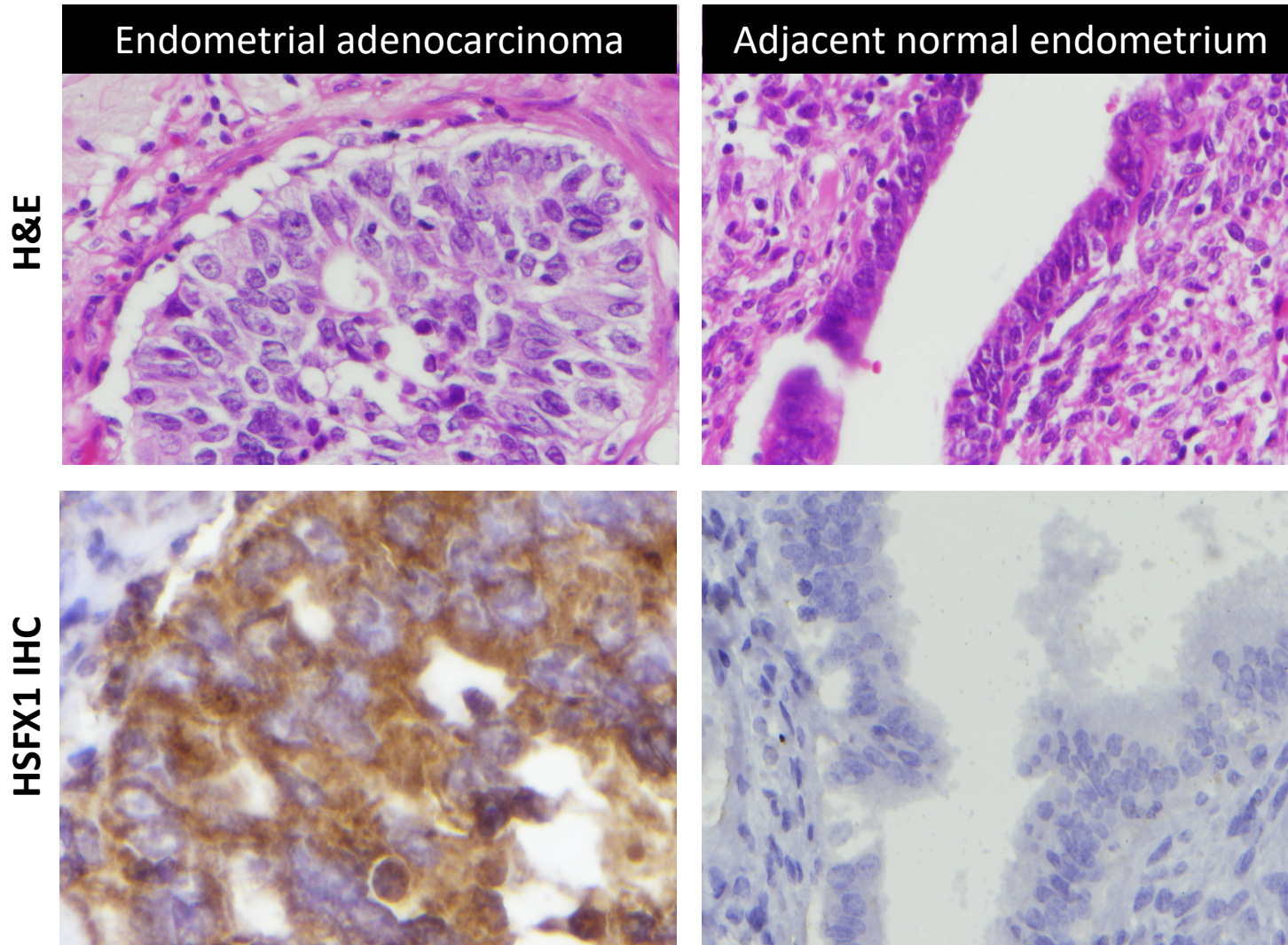
b



c



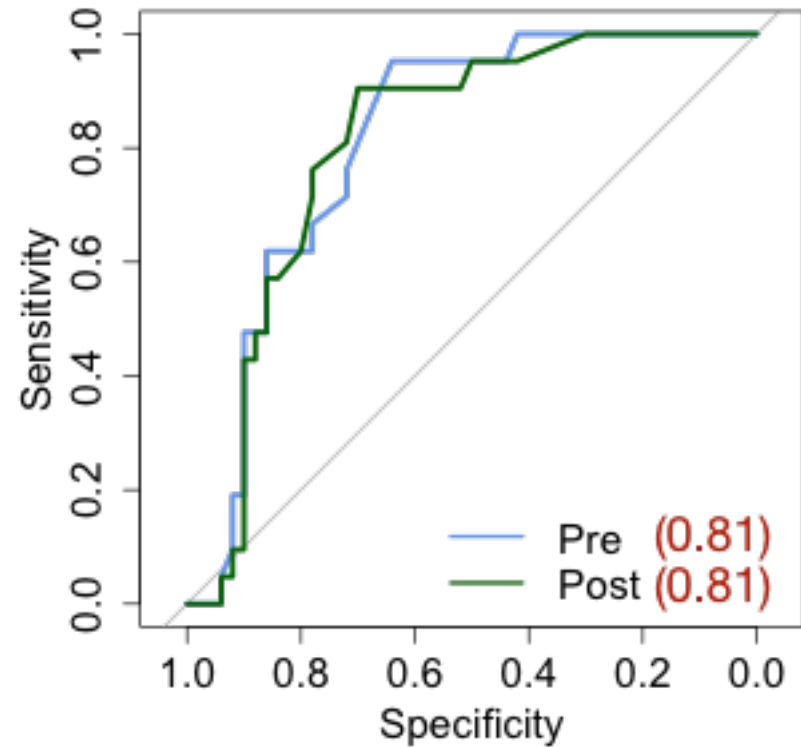
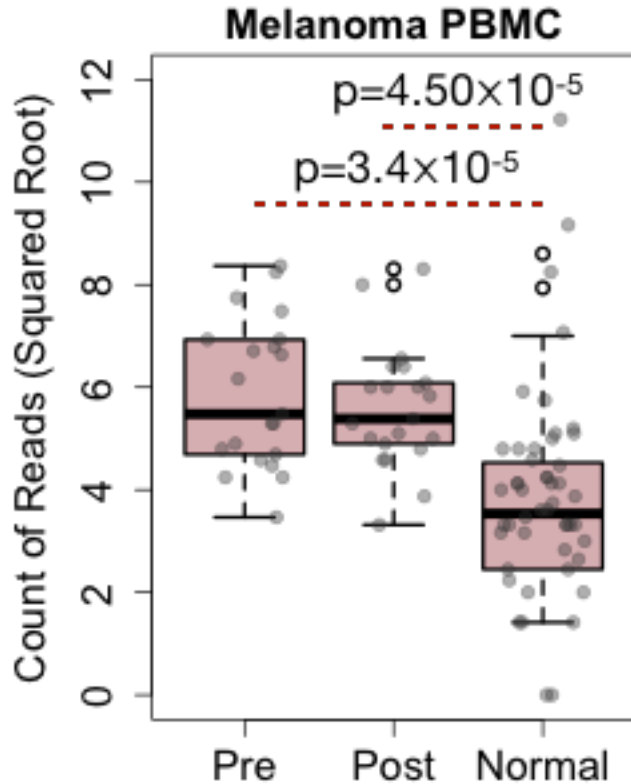
HSFX1 protein is specifically expressed in high-grade endometrial tumors



Outline

- Clustering TCRs into antigen-specific groups
- Identification of a novel cancer antigen with TCR clusters
- **Novel machine learning method for de novo prediction of cancer-associated T cells**

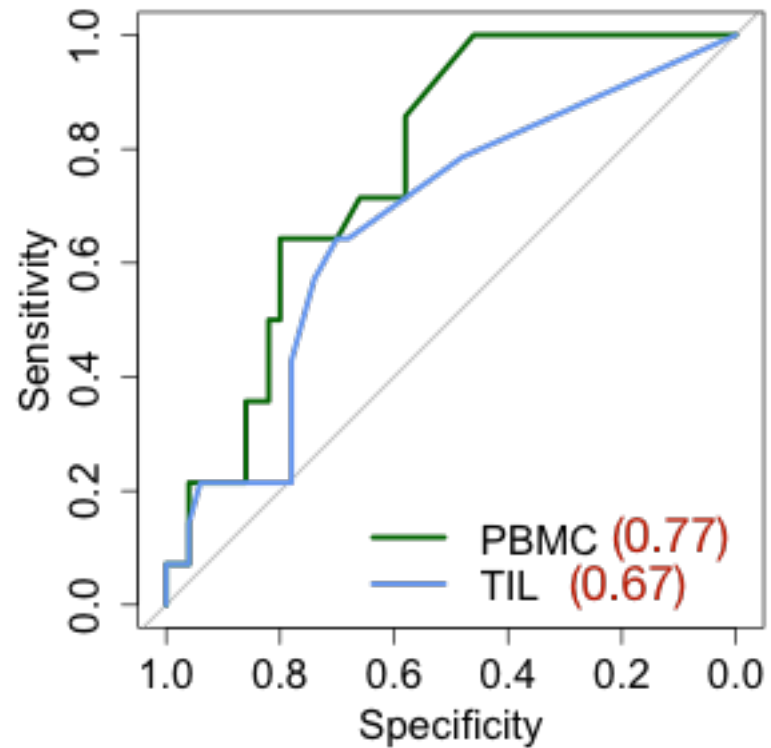
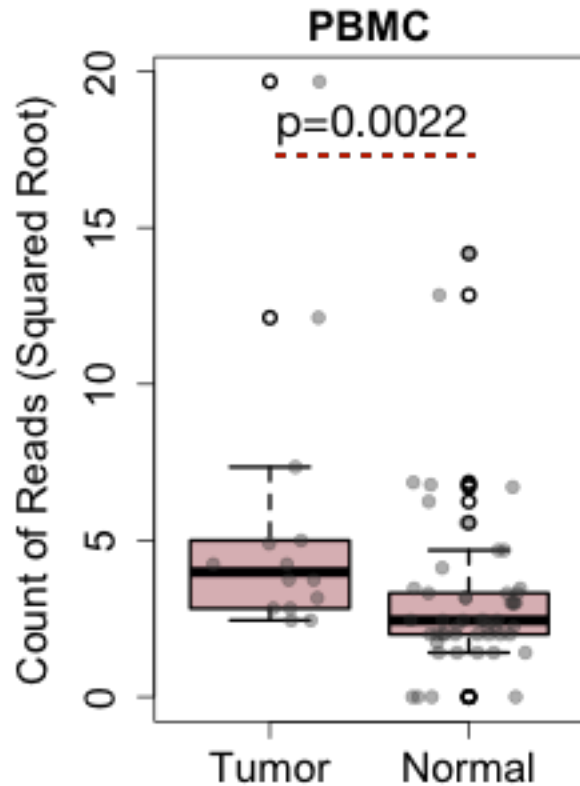
Presence of clustered CDR3s in late stage melanoma patients' blood TCR repertoire



21 TCR-seq samples of melanoma patients coming from:

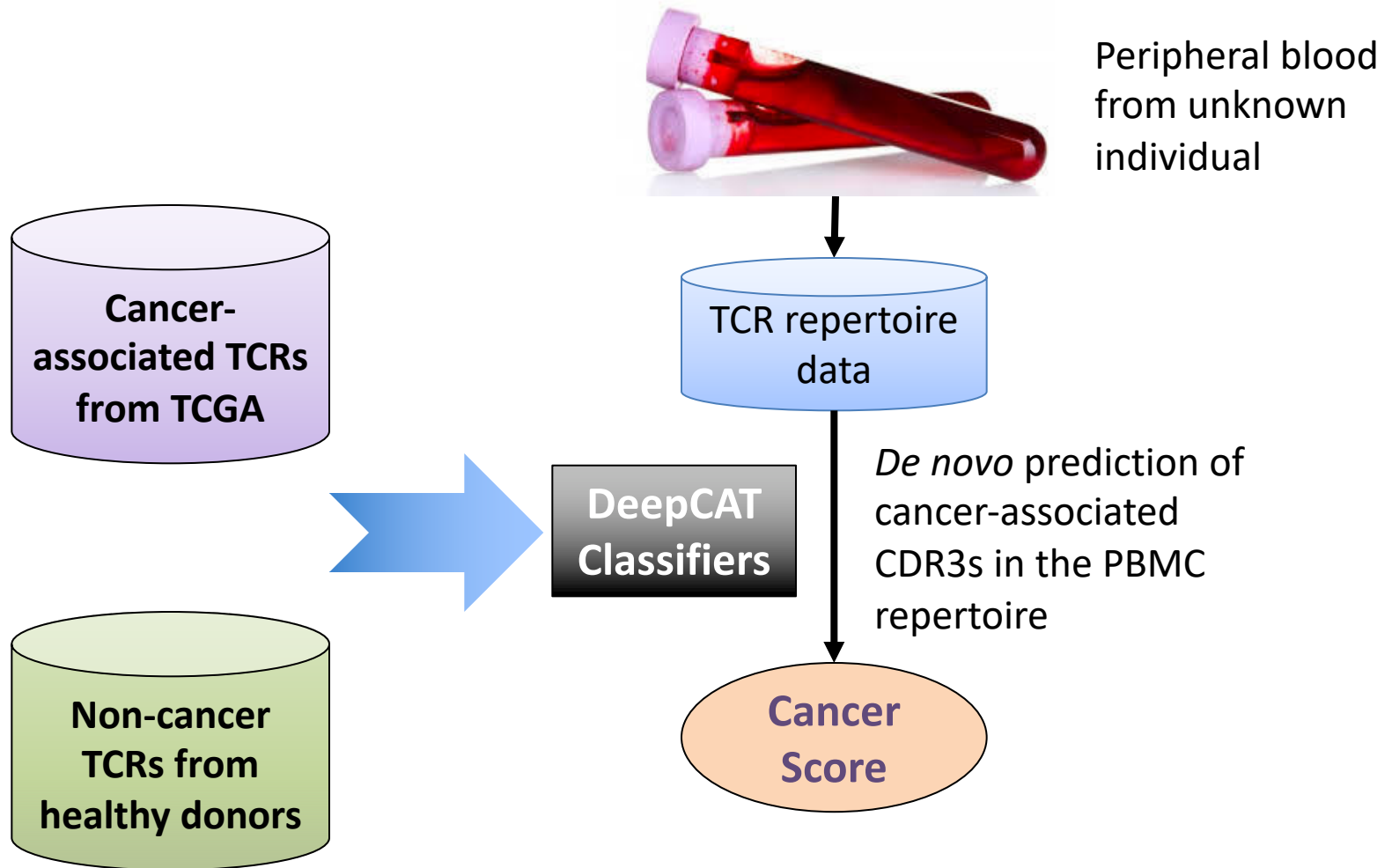
Robert et al., 2014, Clin Cancer Res.

Presence of clustered CDR3s in early breast cancer patients' blood TCR repertoire

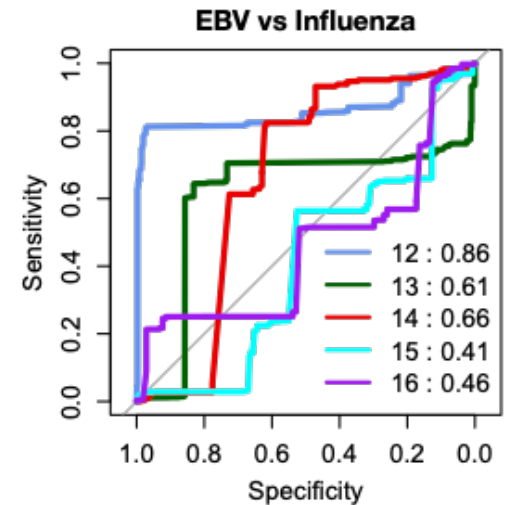
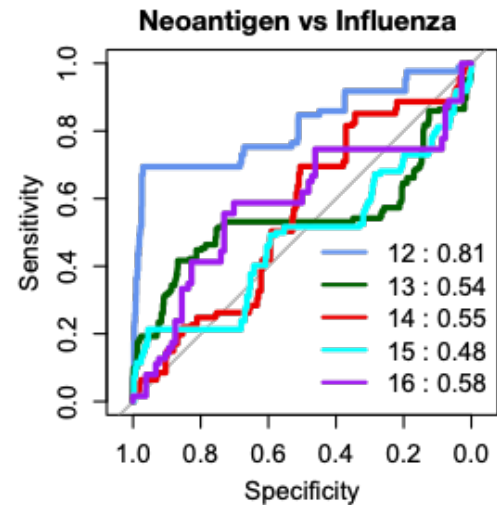
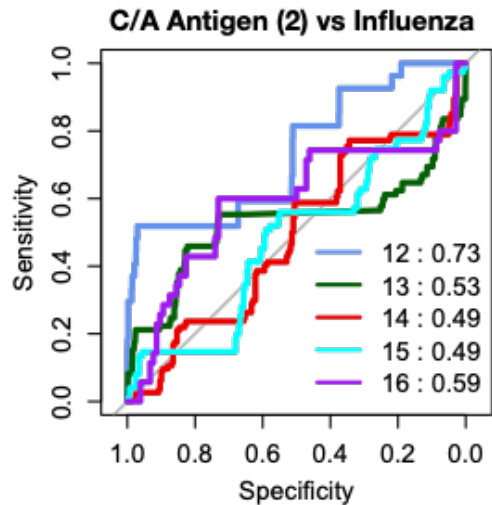
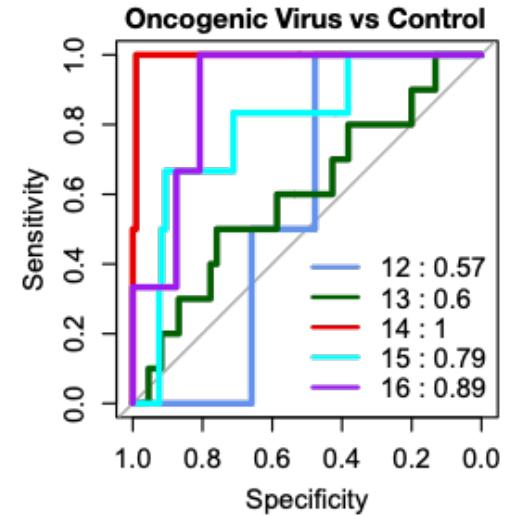
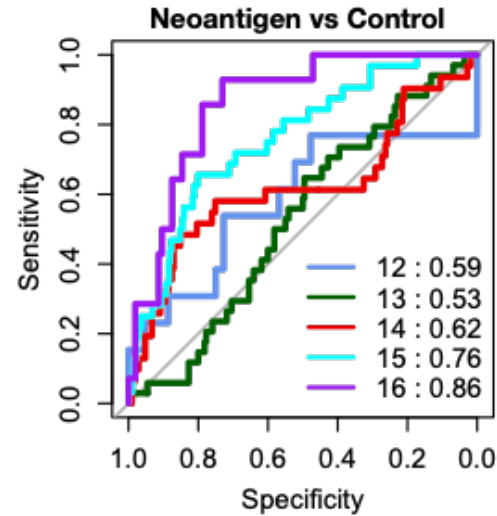
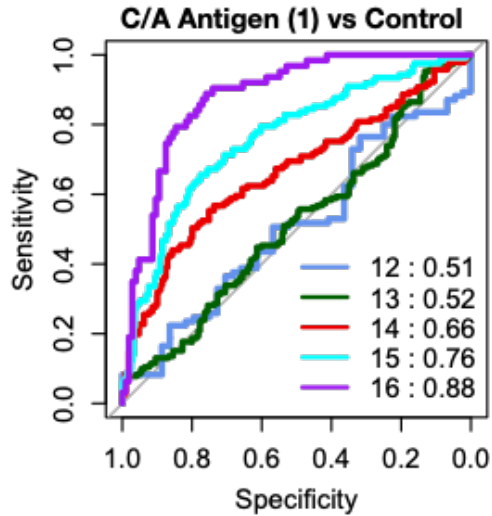


16 early breast cancer PBMC samples from:
Beausang et al., 2017, PNAS

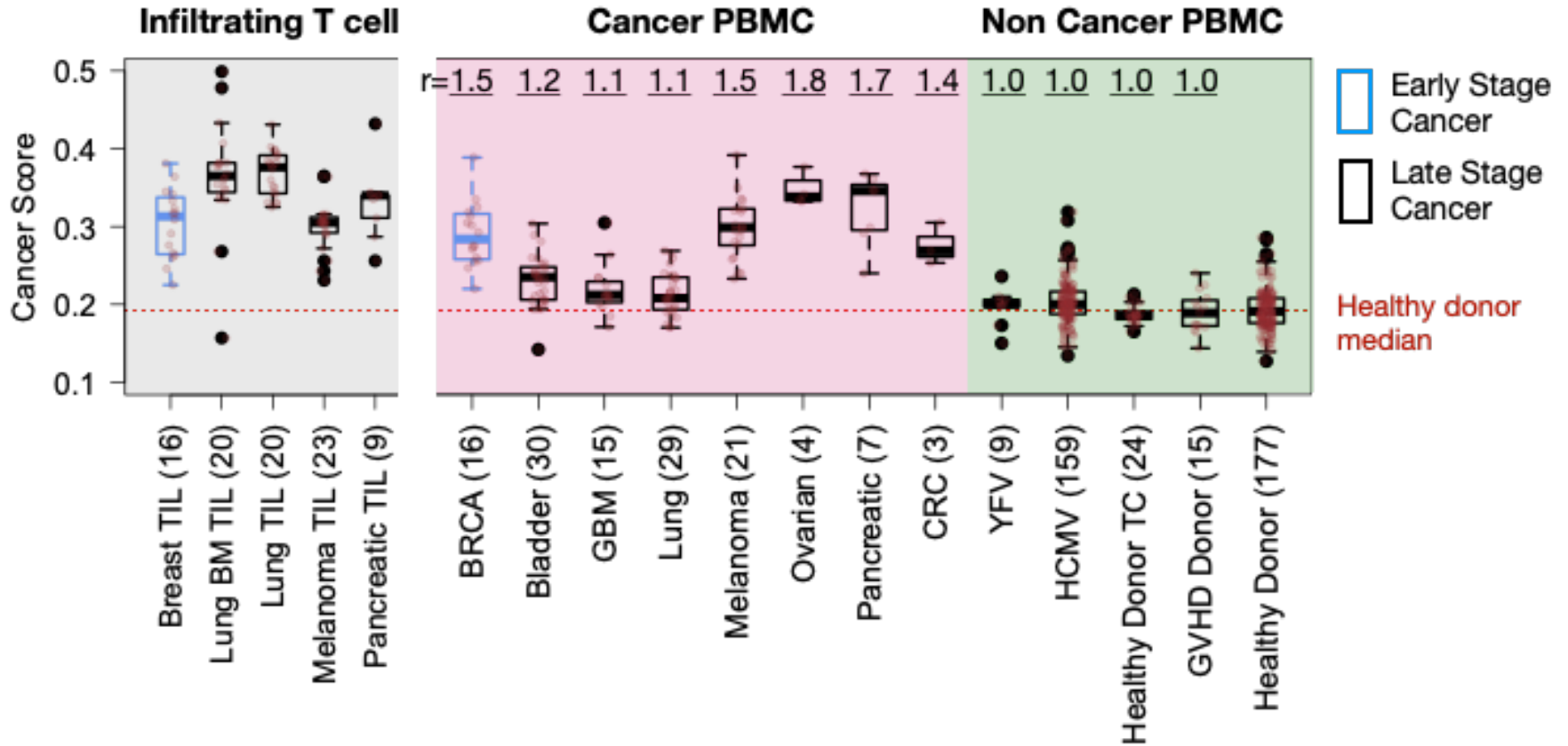
DeepCAT: machine learning method for *de novo* prediction of Cancer-Associated TCRs



Independent validation of DeepCAT using tetramer sorted antigen-specific TCRs

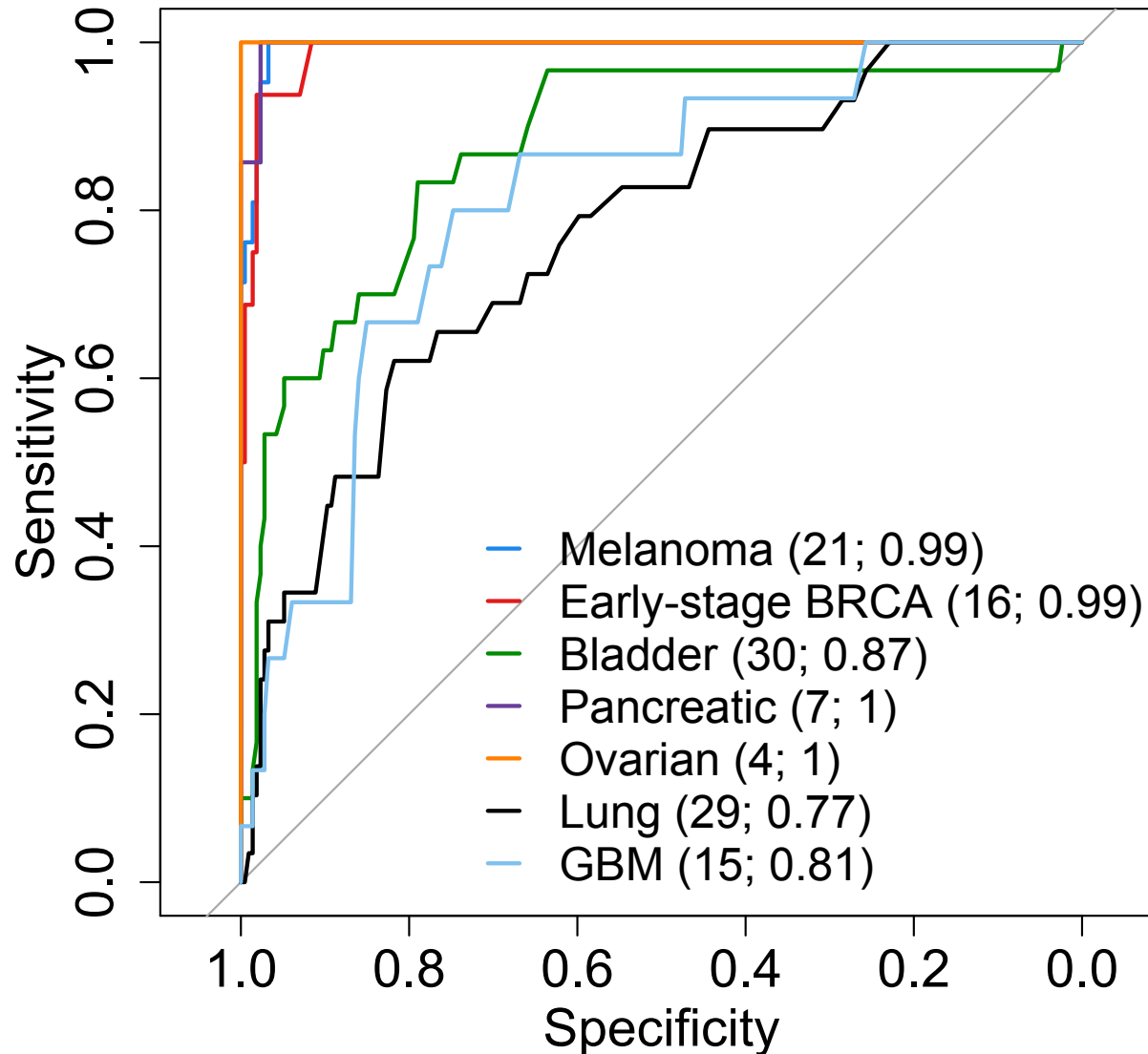


Predicting cancer scores across multiple disease cohorts

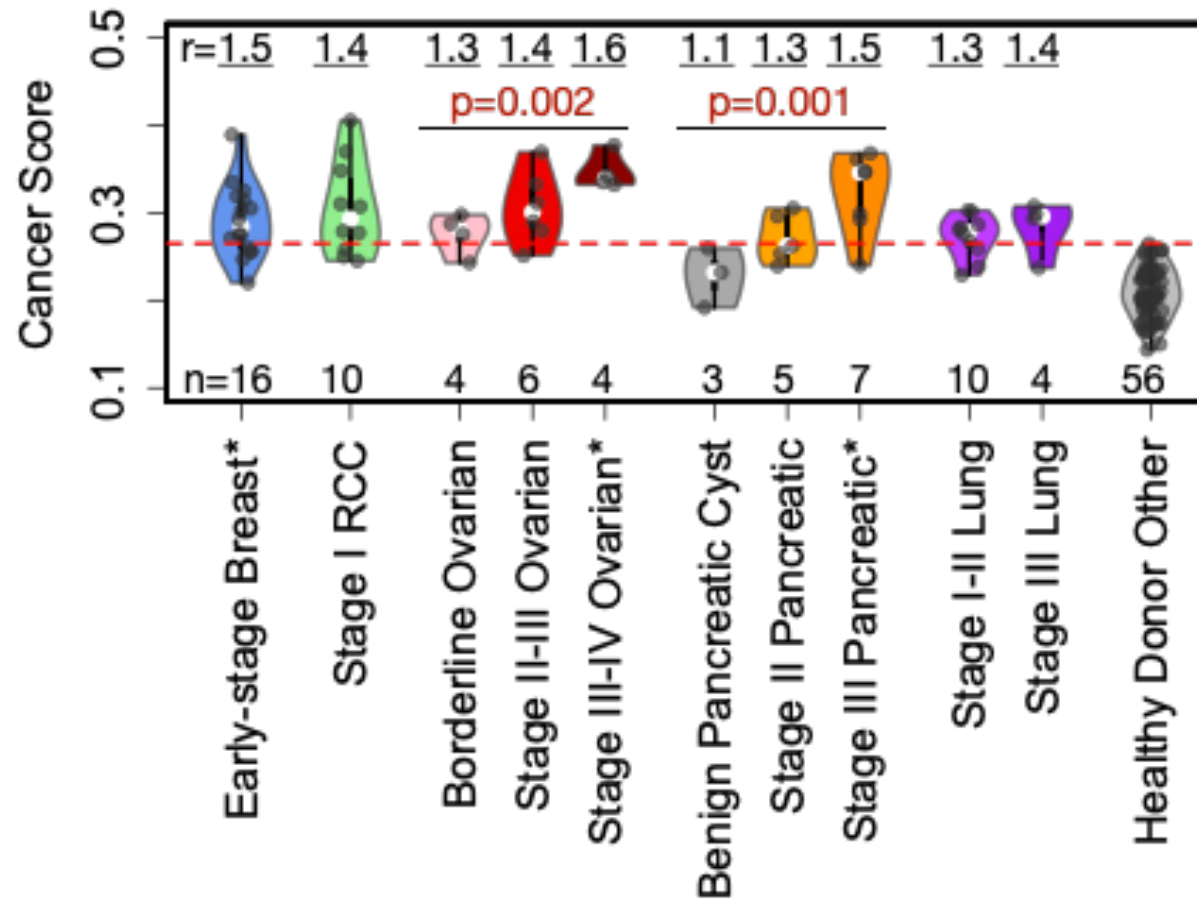


Limitation: Cancer score cannot distinguish different cancer types

High prediction power of cancer score suggests potential non-invasive (early) detection



Cancer score may serve as an additional diagnostic modality for multiple cancers



Summary

- iSMART can group TCRs into antigen-specific clusters
- Predicted cancer antigen HSFX1 induces T cell response *in vivo*
- Development of a *de novo* prediction method (DeepCAT) for cancer-associated TCR repertoire
- DeepCAT predictions are robust against random fluctuations of TCR repertoire over time
- Using PBMC samples, DeepCAT separates cancer patients from healthy individuals with high accuracy

Acknowledgements

Li lab members:

Hongyi Zhang, PhD

Daria Beshnova, PhD

Jiahui Chen, PhD

Jianfeng Ye, PhD

Fu lab members:

Longchao Liu, PhD

Zhenhua Ren, PhD

Jian Qiao, PhD

Zhida Liu, PhD

Clinical collaborators at UTSW:

James Brugarolas, MD

Jayanthi Lea, MD



**Yang-xin Fu,
MD, PhD**



**Wenxin
Zheng, MD**



**Carlos Arteaga,
MD**

Harvard Medical School:



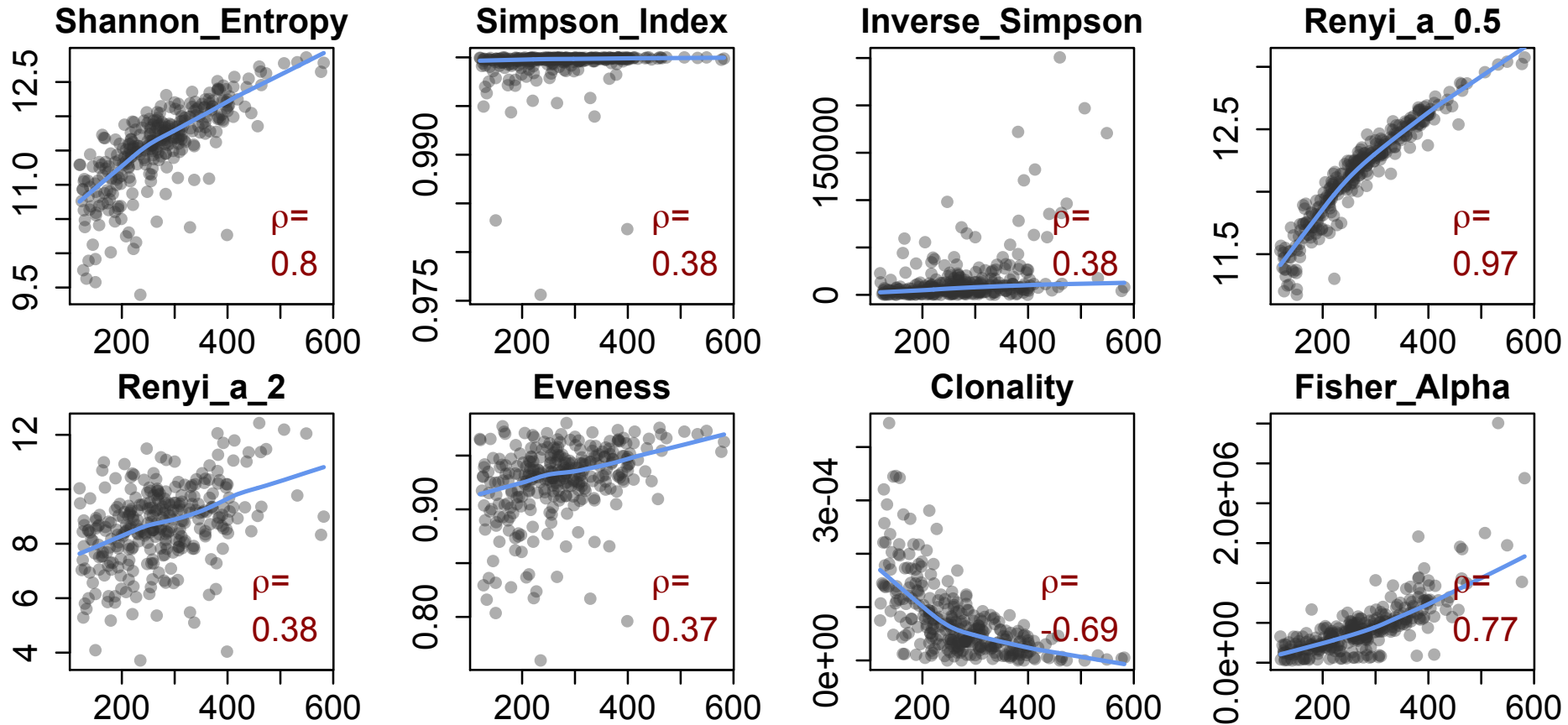
**Catherine
Wu, MD**



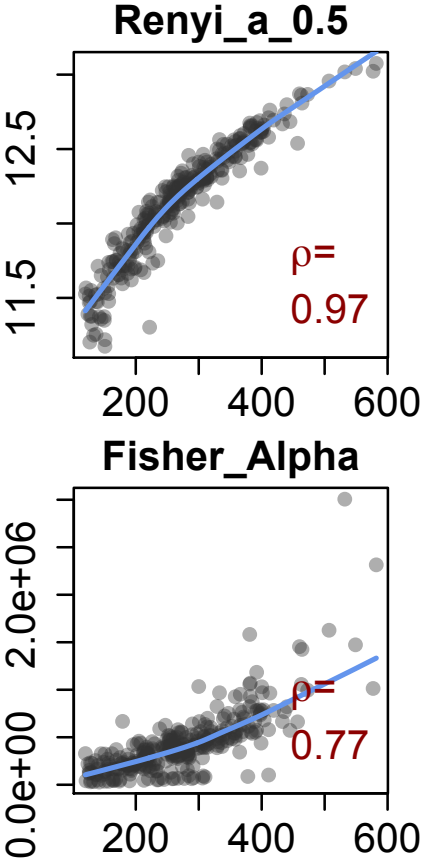
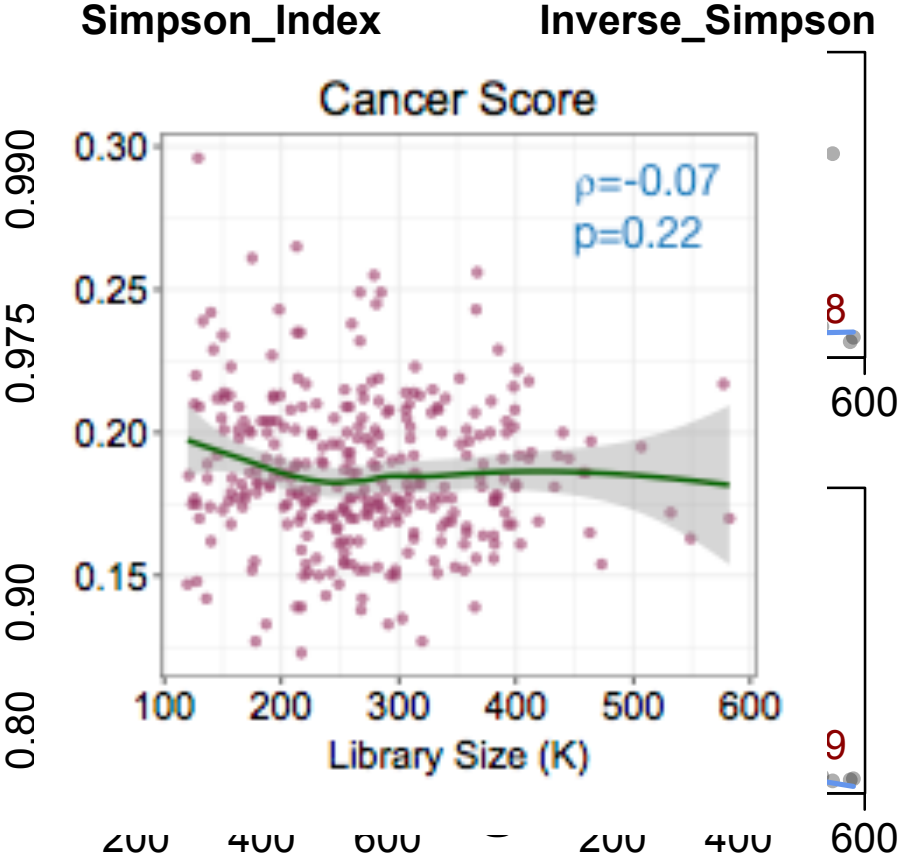
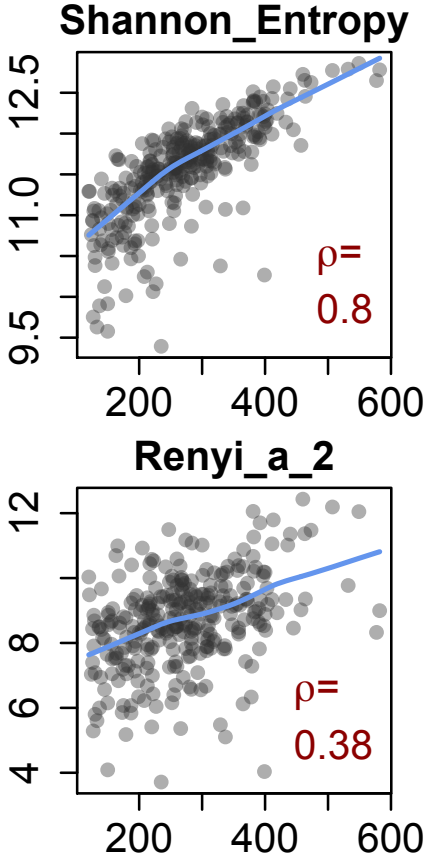
**Nir Hacohen,
PhD**



TCR repertoire diversity indices are usually affected by arbitrary library size

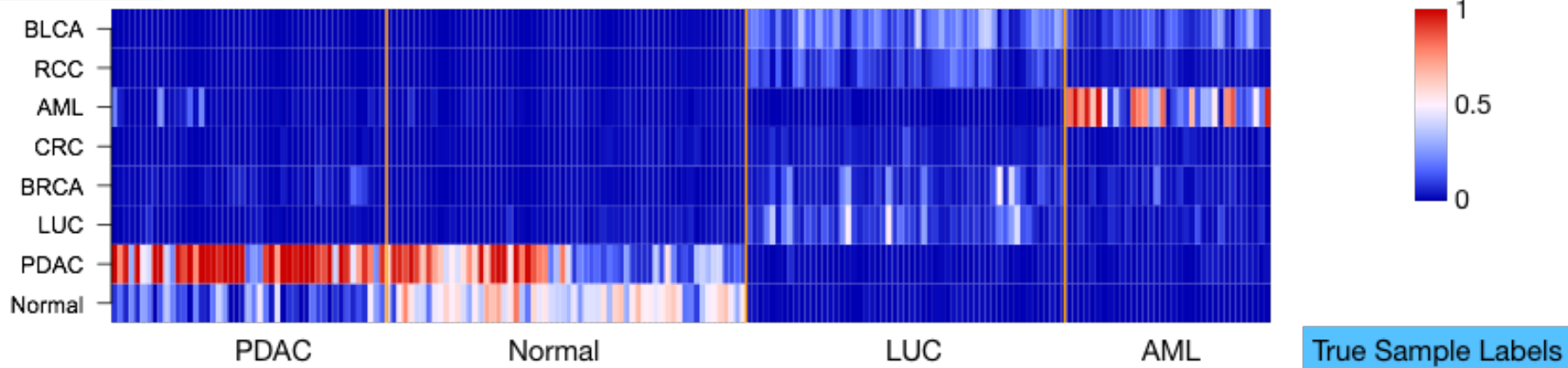


Cancer score is invariant to library size



Re-evaluation of DNA-methylation based blood test for cancer detection


Models Predictions



Confusion matrix for Shen et al.'s method when using all 8 prediction models

	LUC	Normal	PDAC	AML	Specificity
AML	0	0	0	26	1
BLCA	39	0	0	8	-
BRCA	2	0	0	0	-
CRC	1	0	0	0	-
LUC	12	0	0	1	0.92
Normal	0	34	3	0	0.92
PDAC	0	28	44	0	0.61
RCC	1	0	0	0	-
Sensitivity	0.22	0.55	0.94	0.74	0.58

Overall Accuracy



Prediction power for early stage pancreatic cancer using normal tissue as control

