Identification of T-cell receptor sequences associated with Type 1 diabetes through computational analysis

Camryn Pettenger-Willey[1,2], Sebastiaan Valkiers[1,3], Victor Greiff[4], Todd Brusko[5], Dr. Philip Bradley[1]

Fred Hutch Cancer Center [1]Fred Hutchinson Cancer Center, Seattle, WA, [2]Whitman College, Walla Walla, WA, [3]University of Antwerp, Antwerp, Belgium, [4]University of Oslo, Oslo, Norway, [5]University of Florida, Gainesville, FL

Whitman College

SUMMARY

Objective: Identify self-reactive T-cell receptors (TCRs) involved in the auto-immune response of Type 1 diabetes (T1D) to be used in the future as potential biomarkers or in therapeutic approaches for the disease.

- Employed large-scale TCR repertoire analysis on a dataset of T1D patients alongside control patients.
- Identified Significantly Neighbor Enriched (SNE) TCR sequences that are associated with T1D based on Fisher's exact test.
- Performed additional analyses with clustering, HLA T1D risk alleles, auto-antigen reactivity, and age.
- These findings contribute to the understanding of islet-reactive T-cells and their potential implications in T1D diagnosis and treatment strategies.
- Further research and the availability of larger datasets comprised of paired-chain disease-specific TCR sequences are vital for developing biomarkers and therapeutics.

BACKGROUND

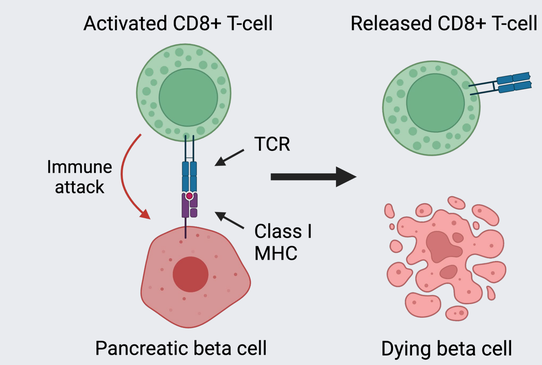

Figure 1: Diagram illustrating the death of pancreatic beta cells in Type 1 diabetes through an activated CD8+ T-cell.

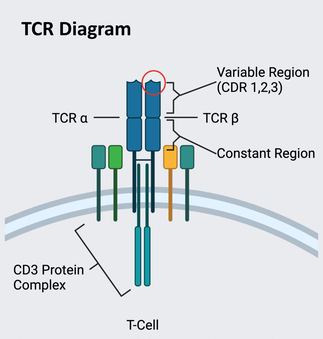

Figure 2: Diagram of the parts of a TCR. The red circle points out the variable binding region of interest on the beta chain.

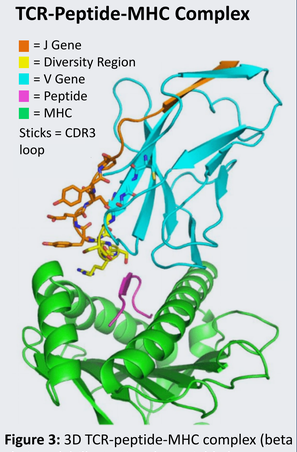

Figure 3: 3D TCR-peptide-MHC complex (beta chain only) illustrating the variable loops essential to binding. Adapted from P. Bradley.

METHODOLOGY

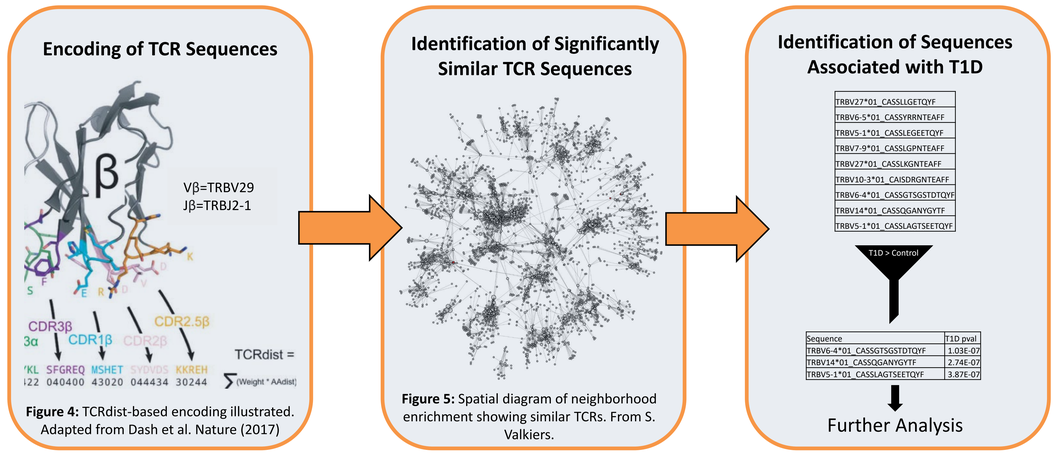

Figure 4: TCRdist-based encoding illustrated. Adapted from Dash et al. Nature (2017)

Figure 5: Spatial diagram of neighborhood enrichment showing similar TCRs. From S. Valkiers.

IDENTIFYING T1D TCRS

86 Million Sequences → 30,000 SNE Sequences→ 130 Significant Sequences

Sequence Association with T1D:

- Fisher exact test based on sequence group assignment to determine association
- In order to account for multiple testing (30,000 sequences) a false-discovery rate was identified by permutation testing. A p-value threshold of 0.0004 was used to minimize the inclusion of false positives.

Example Sequence:
TRBV6-4*01_CASSGTSGSTDTQYF

	T1D Count	Control Count
Yes Seq	260	453
No Seq	169	543
Proportion	0.61	0.45

T1D > Control
p-value = 1.03*E-07

Further Analysis of Sequence Results:

- Clustering (hierarchical clustering with the Leiden algorithm)
- HLA allele association (DR3/DR4 haplotype as risk alleles)
- Evaluate T1D risk alleles as a confounding variable
- Known auto-antigen response association
- Age association

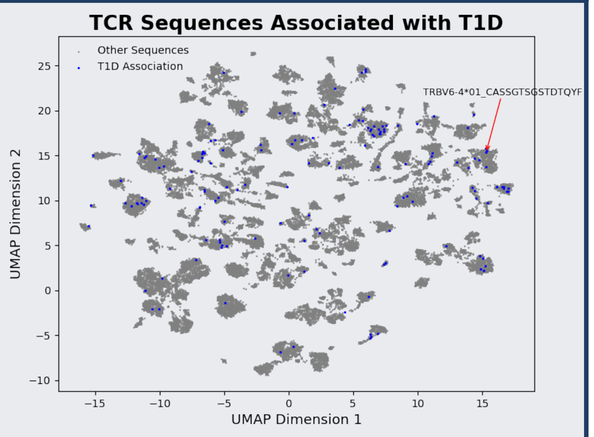

Figure 6: UMAP plot of SNE TCR sequences showing specific sequences associated with T1D.

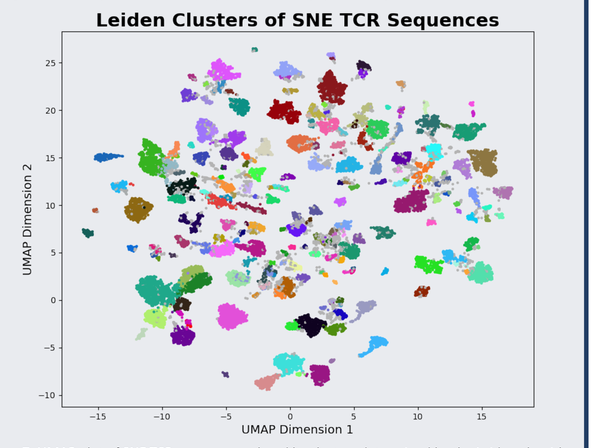

Figure 7: UMAP plot of SNE TCR sequences colored by clusters determined by the Leiden algorithm.

SIGNIFICANT TCRS

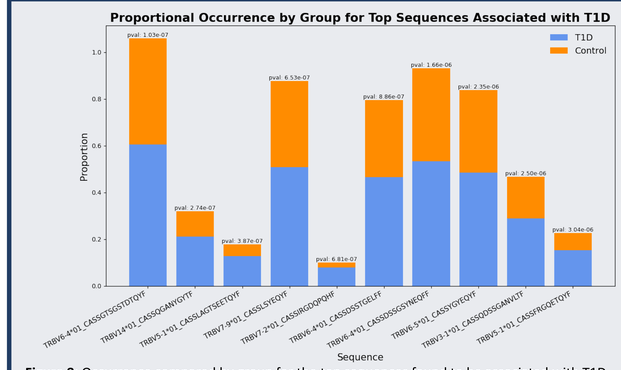

Figure 8: Occurrence compared by group for the top sequences found to be associated with T1D.

Sequence	T1D Proportion	Control Proportion	T1D Association	Top Allele Association	DR3* Association	T1D Association (DR3+ donors)	DR4* Association	T1D Association (DR4+ donors)	Antigen Association
TRBV6-4*01_ CASSGTSGSTDTQYF	0.61	0.45	1.03E-07	DRB1_401*			0.0136	0.0089	0.0013
TRBV14*01_ CASSQGANYGYTF	0.21	0.11	2.74E-07	DRB1_301*	1.75E-41	0.0083	8.51E-05	0.0009	
TRBV5-1*01_ CASSLAGTSEETQYF	0.13	0.05	3.87E-07	DRB1_301*	5.82E-32	0.0012			
TRBV7-9*01_ CASSLSYEQYF	0.51	0.37	6.53E-07	DRB1_401*					0.0023
TRBV7-2*01_ CASSIRGDQPQHF	0.08	0.02	6.81E-07	DPB1_401					
TRBV6-4*01_ CASSDSSTGELFF	0.47	0.33	8.86E-07	A_2902					0.0004
TRBV6-4*01_ CASSDSSGSYNEQFF	0.53	0.40	1.66E-06	DRB4_103			0.0096	0.0004	0.0160
TRBV6-5*01_ CASSYGYEQYF	0.48	0.35	2.35E-06	DRB1_1001					0.0248
TRBV3-1*01_ CASSQDSSGANVLTF	0.29	0.18	2.50E-06	DQB1_501					
TRBV5-1*01_ CASSFRGQETQYF	0.15	0.07	3.04E-06	DRB1_301*	0.0001	0.0004			

Table 1: A subset of the results comparing the sequences associated with T1D. Insignificant results not displayed. * indicates T1D risk alleles.

DISCUSSION

Utilization of the TCR sequences of interest:

- They can be used to build a diagnostic/predictive tool (in addition to antibodies) for T1D.
- Can experiment with regulatory T-cells to shut down these specific TCRs.

Next steps:

- Analyze paired chain data to find alpha chain matches.
- Experimentally confirm reactivity of these TCRs.
- Analyze TCR sequences throughout the progression of T1D in conjunction with epitope spreading.

Limitations:

- Paired-chain data is needed to fully understand these TCRS.
- These sequences need to be experimentally validated.

ACKNOWLEDGEMENTS

Thank you to Sebastiaan Valkiers and Dr. Philip Bradley for their help. Unpublished TCR sequence repertoires provided by Victor Greiff (U. of Oslo) and Todd Brusko (U. of Florida). This work is funded by the National Institutes of Health (Grant # P01 AI42288) and Helmsly Charitable Trust to T.M.B.. The Summer Undergraduate Research Program is supported in parts by Whitman College, the Fred Hutch Internship Program, and individual labs/research groups.