From Single Cells to Systems:
Network Maps for Cluster Identification in Single-Cell Data

Dan Ryder[1,3], Haider Hassan[1], Jessica Corrado [1], Alissa Cait[1], Erica Scott[1], Jonathan Hsu[2],
[1] Bridge Informatics, Inc., 160 Alewife Brook Pkwy, Cambridge, MA 02138; [2]Marengo Therapeutics, Inc., 840 Memorial Dr 4th Floor, Cambridge, MA 02139; [3]Corresponding author and Chief Executive Officer of Bridge Informatics, Inc.

Mining Public TCRs

- **Harnessing Single-Cell Data**: Advanced computational analysis reveals shared ("public") TCRs across individuals, uncovering universal immune responses
- **Mapping Functional networks**: Network analysis of TCR clusters identifies patterns and relationships crucial for understanding adaptive immunity
- **Accelerating Immunotherapy**: Insights from public TCRs pave the way for developing broad-spectrum cell therapies targeting common immune threats

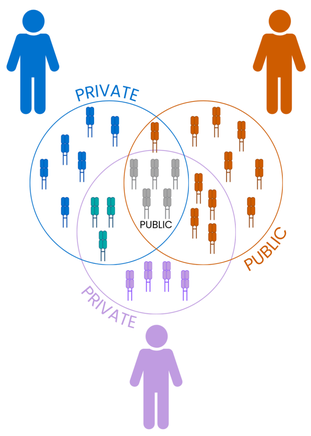


Figure 1. What are Public TCRs? A T-cell receptor (TCR) is like a lock that only fits specific "keys" (antigens) to help the immune system recognize threats. A public TCR is a receptor found in many different people, meaning their immune systems have independently developed the same way to recognize a particular antigen. In contrast, a private TCR is unique to an individual, shaped by their personal immune history and genetics, meaning it's not commonly found in others.

Decoding T-Cell Receptors: Insights from Single Cell Analysis

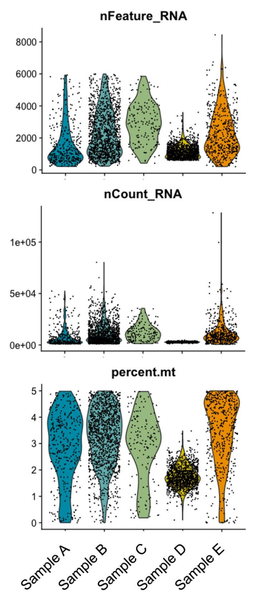
- **Single-Cell Immunology Revolution**: Single-cell technologies provide unprecedented insights into T-cell diversity, function, and immunological outcomes.
- **Public vs. Private TCRs:** While most TCR sequences are unique to individuals, some "public" TCRs recur across multiple people, revealing shared antigen targets for immunotherapy.
- **The Role of CDR3**: The complementarity-determining region 3 (CDR3) is the most variable and functionally critical part of the TCR, playing a key role in antigen recognition and T-cell specificity.

Figure 2. The T cell receptor.
(A) Overview of the T cell receptor (TCR). **(B)** Overview of the complementary determining regions (CDRs) within the antigen binding region of the TCR.

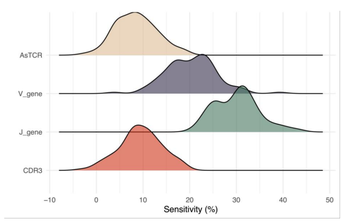
- **Computational Power for Discovery**: Identifying public TCRs requires large-scale data mining, advanced computational pipelines, and robust analytical frameworks.
- **Challenges in Data Analysis**: The complexity of single-cell data demands specialized bioinformatics expertise to integrate, clean, and analyse datasets effectively.

Identifying Public TCRs: Computational Workflow

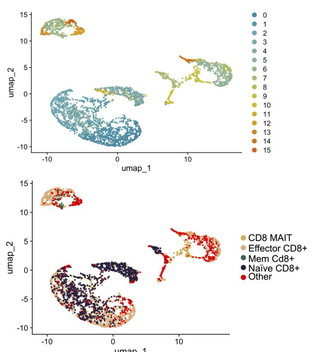
Select and consolidate samples



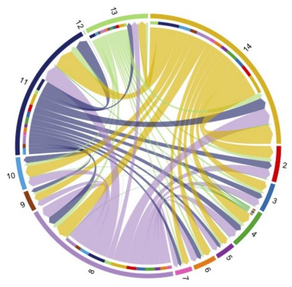
The count matrices are filtered to remove cells, empty barcodes, and cell multiplets. Separate samples are integrated to compensate for technical variability



TCR extraction. Showing the sensitivity of extraction of assembled-TCR (AsTCR), V gene, J gene and CDR3 amino acid sequences from scRNA-seq datasets



Cell clustering, and annotation



Advanced analysis, such as TCR clonal sharing between UMAP clusters

STEP ONE:
DATA ACQUISITION AND INTEGRATION

Sample Selection
- Single-cell datasets from multiple studies are identified through public repositories (e.g., GEO, SRA) and/or internal data sources that meet biological and technical selection criteria

Data Consolidation
- Raw and/or processed single-cell datasets are collated
- Study-specific metadata is extracted and harmonized

STEP TWO:
QUALITY CONTROL AND PREPROCESSING

FastQC and Samtools for quality control

STEP THREE:
TCR EXTRACTION AND CLONOTYPING

TCR extraction (TraCeror MiXCR)

STEP FOUR:
IDENTIFICATION OF CONVERGENT ("PUBLIC") TCRS
- Characterization of motif sequences in TCR repertoires for each individual dataset.
- Characterization of shared and unique CDR3s
- Cross-Donor Comparison
- TCR sequences are compared across all donor datasets to identify recurring motifs. TCRs observed in multiple unrelated individuals are tagged as "public"

STEP FIVE:
CLUSTERING AND NETWORK ANALYSIS

Sequence Similarity Metrics
- Use a sequence similarity threshold (Hamming distance) to group TCRs into clusters

Network Construction
- Nodes represented TCR clonotypes and edges indicate similarity or shared antigen specificity
- Visualization: (igraph, NetworkX, Cytoscape)

STEP SIX:
FUNCTIONAL ANNOTATION AND INTEGRATION WITH TRANSCRIPTOMICS

Gene Expression Profiling
- Per-cell transcriptomic profiles can be overlaid on clonotype networks for data with corresponding cell barcodes for each CDR3
- Can use enrichment analyses to identify upregulated pathways in cells harboring shared TCRs to unmask potential relationships between clonotypic expansion and specific biological functions

Results of Network-Based Identification of Public TCR Clones

- Our **network analysis** allowed us to identify the most frequent **CDR3 amino acid sequences and** visualize relationships
- By integrating **TCR sequence data with transcriptomic profiles**, we mapped CDR3 amino acid sequences to distinct T cell subsets, revealing that public TCRs are enriched in activated cytotoxic responses.

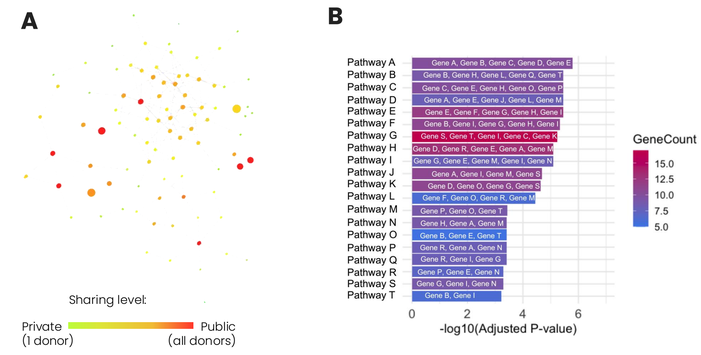


Figure 3: Network represents the most frequent CDR3 amino acid sequences found in the dataset. (**A**) Nodes (CDR3 sequences) are connected by edges defined by a Hamming distance <2. Node size represents log frequency of the sequences. Nodes are colored according to their sharing levels in the dataset: red found in all donors (public TCRs). (**B**) Pathway analysis identifying pathways and genes associated with most frequent CDR3 amino acid sequences.

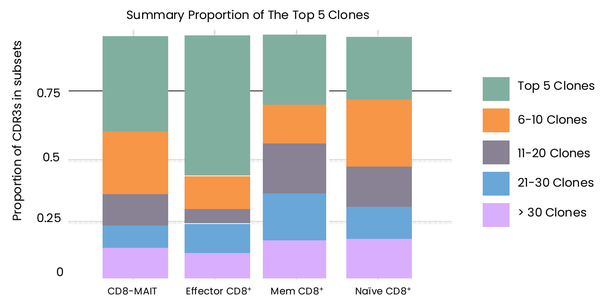


Figure 4: Analysing the distribution of the public TCR clones by CD8⁺ T cell types. This analysis reveals that the top 5 most abundant public TCR clones are found in effector CD8⁺ T cells, suggesting they are primarily involved in active cytotoxic immune responses.

Summary

- **Identifying Public TCRs**: We leveraged single-cell datasets to uncover public TCRs, advancing universal immunotherapy strategies.
- **Sophisticated Bioinformatics Methods**: Data integration, preprocessing, TCR clustering, and network analysis were key to extracting meaningful insights.
- **Accelerating Immunotherapy Discovery**: By combining Marengo's immunotherapy expertise with BI's bioinformatics and computational biology skills, we streamlined workflows and facilitated breakthroughs in T-cell biology!

Let's connect!

linkedin.com/company/bridge-informatics

dan.ryder@bridgeinformatics.com

Acknowledgments

We want to acknowledge and thank Marengo Therapeutics for their discussions and collaboration; and allowing us to include their exploratory analysis in this poster.