

Obtaining T cell receptor pairs from
high-throughput sequencing data with the
ALPHABETR package

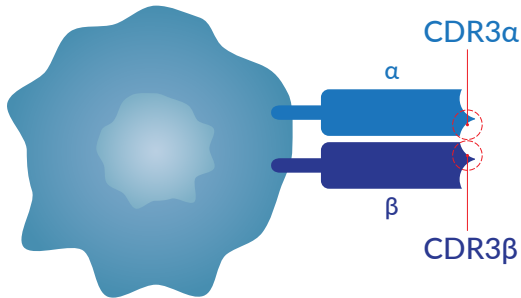
Edward S. Lee

University of Glasgow, Albert Einstein College of Medicine

24 June 2016

Introduction to T cells

- T cells play a central role in the adaptive immune system
- Every $\alpha\beta$ T cell has a T cell receptor (TCR) made up of an α and a β chain
- The immune system utilizes many TCRs in order to recognize virtually any antigen
- Identifying TCR $\alpha\beta$ pairs could have important implications in immunotherapy design



Obtaining TCR pairs is difficult

- High-throughput sequencing cannot be used due to loss of pairing information
- Current experimental protocols use some variation of single-cell sequencing
- Our approach uses **frequency-based pairing**

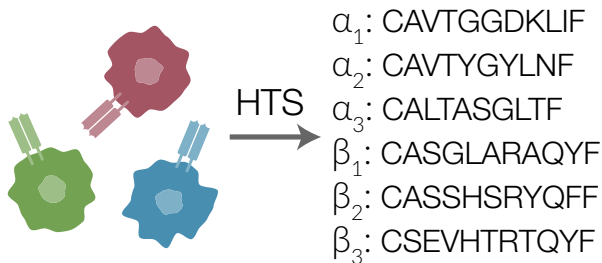
Obtaining TCR pairs is difficult

- High-throughput sequencing cannot be used due to loss of pairing information
- Current experimental protocols use some variation of single-cell sequencing
- Our approach uses **frequency-based pairing**



Obtaining TCR pairs is difficult

- High-throughput sequencing cannot be used due to loss of pairing information
- Current experimental protocols use some variation of single-cell sequencing
- Our approach uses **frequency-based pairing**



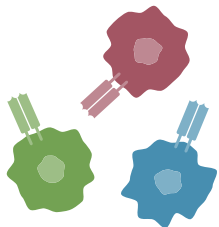
Obtaining TCR pairs is difficult

- High-throughput sequencing cannot be used due to loss of pairing information
- Current experimental protocols use some variation of single-cell sequencing
- Our approach uses **frequency-based pairing**



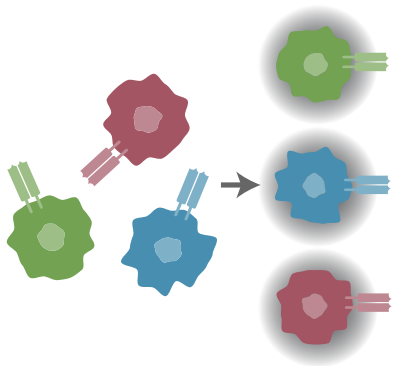
Obtaining TCR pairs is difficult

- High-throughput sequencing cannot be used due to loss of pairing information
- Current experimental protocols use some variation of single-cell sequencing
- Our approach uses **frequency-based pairing**



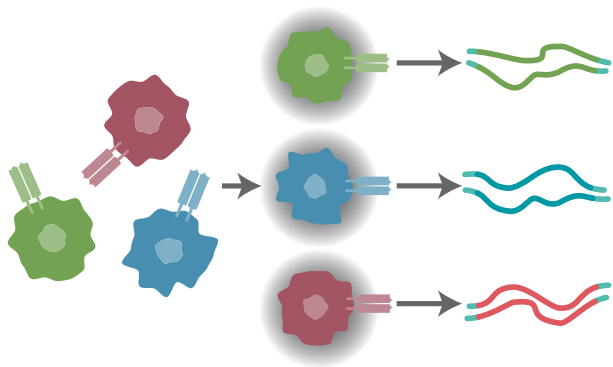
Obtaining TCR pairs is difficult

- High-throughput sequencing cannot be used due to loss of pairing information
- Current experimental protocols use some variation of single-cell sequencing
- Our approach uses **frequency-based pairing**



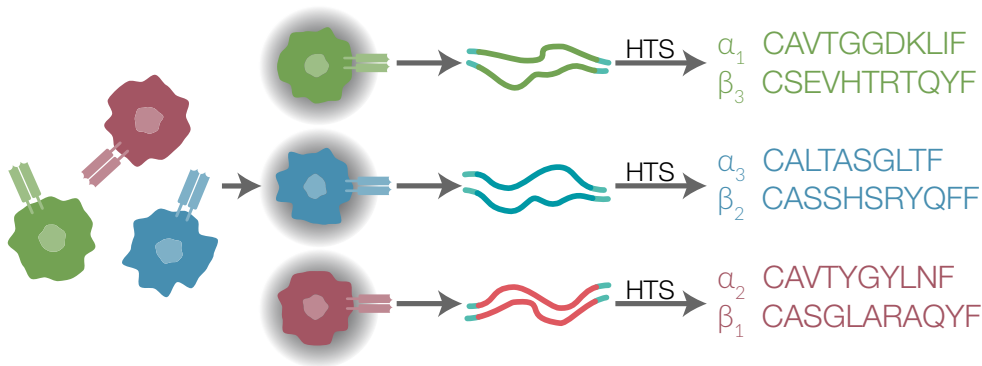
Obtaining TCR pairs is difficult

- High-throughput sequencing cannot be used due to loss of pairing information
- Current experimental protocols use some variation of single-cell sequencing
- Our approach uses **frequency-based pairing**



Obtaining TCR pairs is difficult

- High-throughput sequencing cannot be used due to loss of pairing information
- Current experimental protocols use some variation of single-cell sequencing
- Our approach uses **frequency-based pairing**

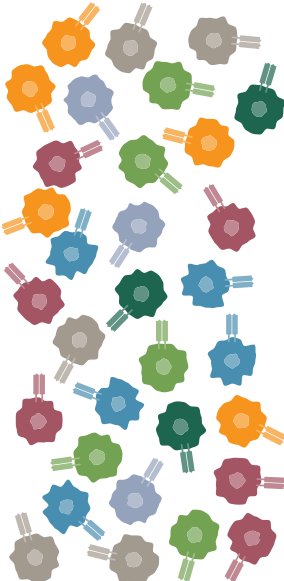


Obtaining TCR pairs is difficult

- High-throughput sequencing cannot be used due to loss of pairing information
- Current experimental protocols use some variation of single-cell sequencing
- Our approach uses **frequency-based pairing**

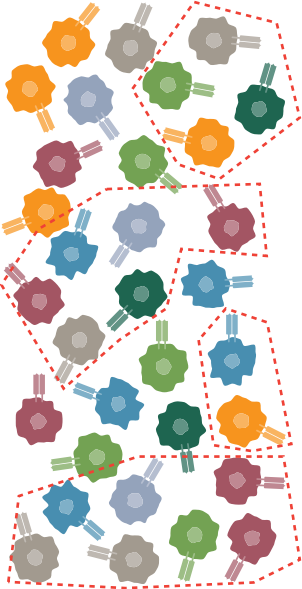
Frequency-based pairing

Sequence multiple samples from the T cell population of interest



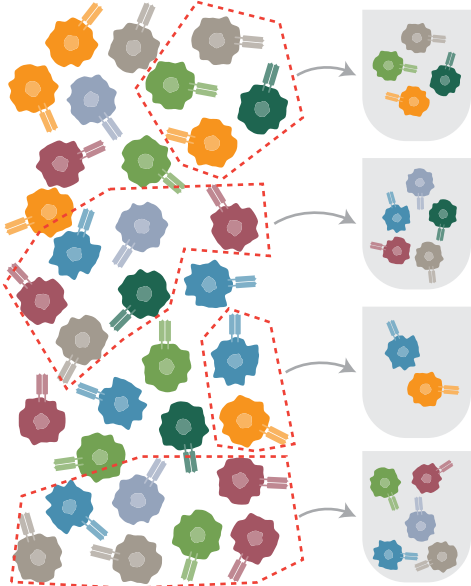
Frequency-based pairing

Sequence multiple samples from the T cell population of interest



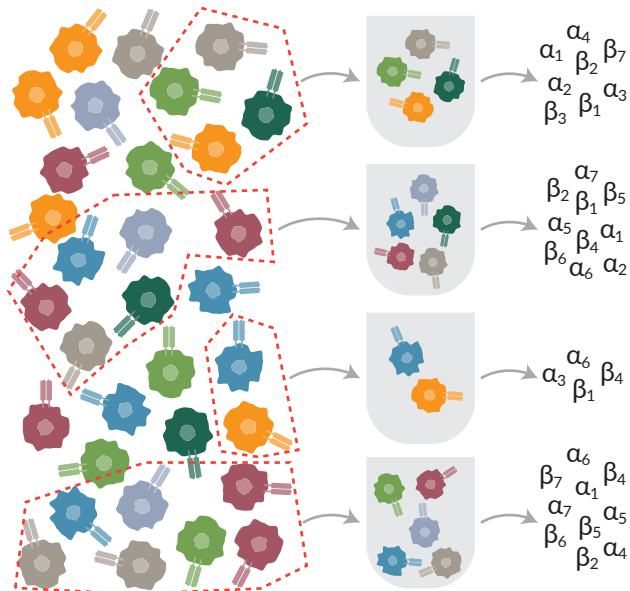
Frequency-based pairing

Sequence multiple samples from the T cell population of interest



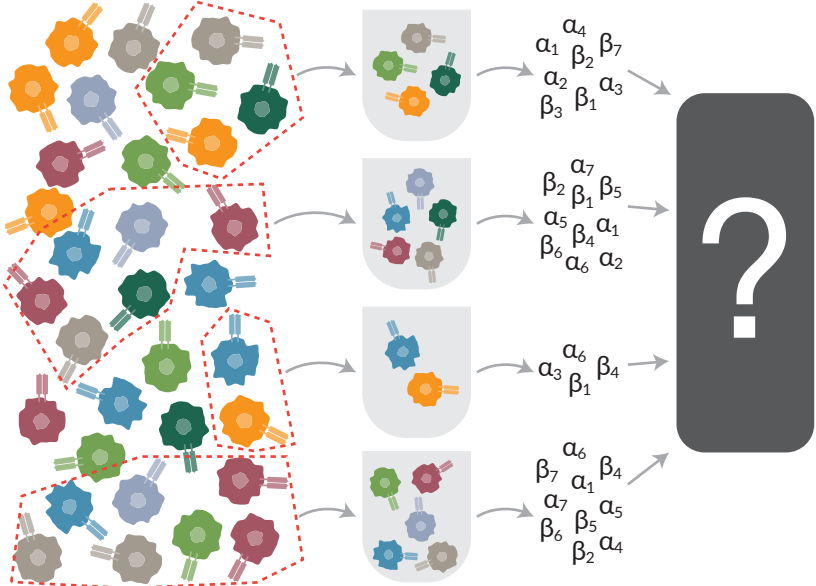
Frequency-based pairing

Sequence multiple samples from the T cell population of interest



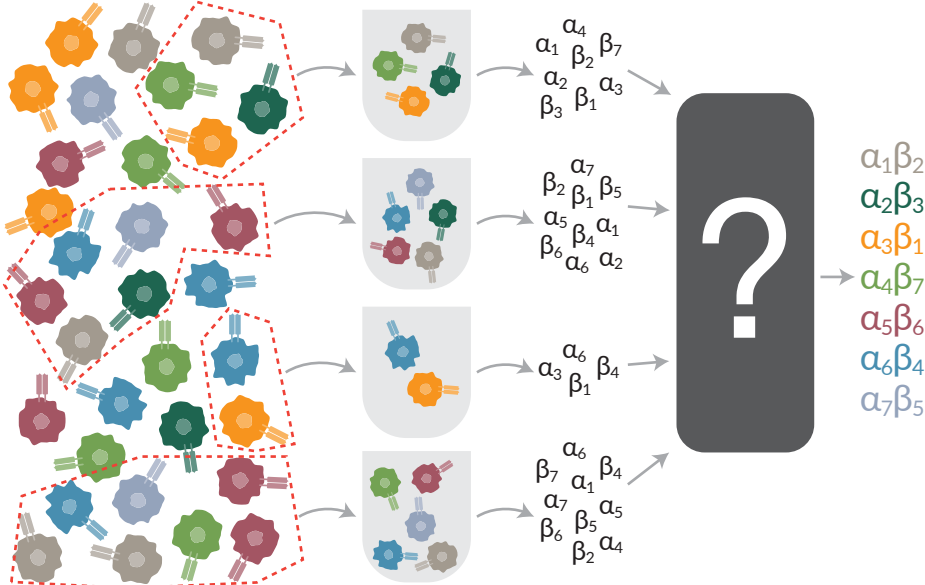
Frequency-based pairing

Sequence multiple samples from the T cell population of interest



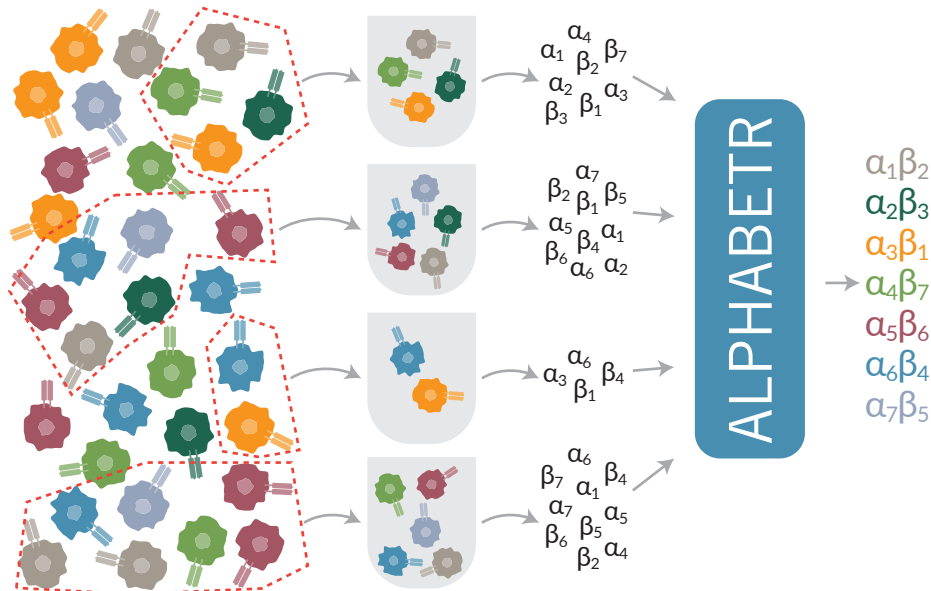
Frequency-based pairing

Sequence multiple samples from the T cell population of interest



Frequency-based pairing

Sequence multiple samples from the T cell population of interest

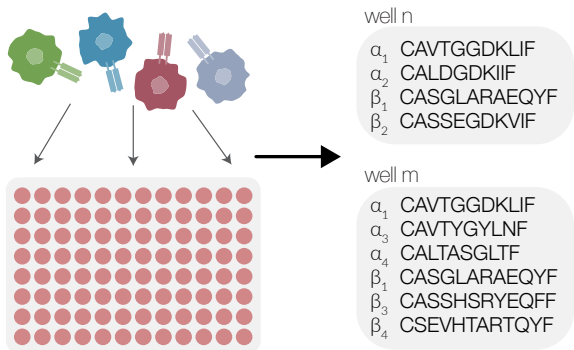


The ALPHABETR package

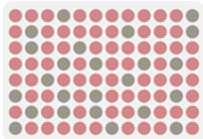
Algorithm for Pairing alpha BETA T cell Receptors

- Identifies $TCR\alpha$ and $TCR\beta$ chains of T cells found in antigen-specific populations
- Calculates the frequencies of the identified T cells
- Handles features unique to antigen-specific populations

Walkthrough of the package



Walkthrough of the package

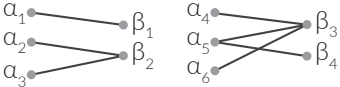


$$S(\alpha_i, \beta_j) = \sum_{\text{well } k} \frac{\delta_{ij}}{N_{\alpha}^{(k)}}$$
$$S(\beta_i, \alpha_j) = \sum_{\text{well } k} \frac{\delta_{ij}}{N_{\beta}^{(k)}}$$

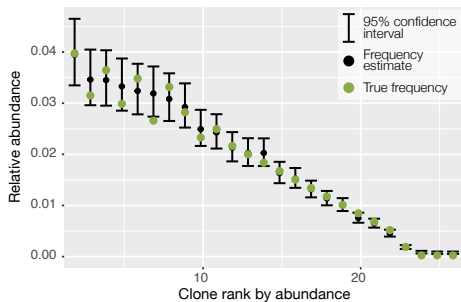
chain_scores()

	α_1	α_2	α_3	α_4	α_5
β_1	24.0	0.6	1.2	4.3	8.2
β_2	0.4	0.2	60.2	0.7	2.2
β_3	1.0	0.2	1.2	9.0	3.0
β_4	3.2	30.1	0.1	0.4	2.1

bagpipe()

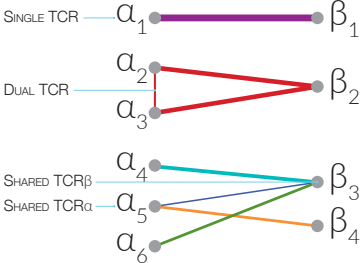


Walkthrough of the package



`freq_estimate()`

Walkthrough of the package



`dual_top()`
`dual_tail()`

Thanks for listening

- `github.com/EdwardSLee/alphabetr`
- `e.lee.2@research.gla.ac.uk`