

Immunological Sequencing (Immuno-Seq)

Mark James Thompson (github = Amuraivel)

2018-10-07

Introduction

T-cells have receptor (TCR) proteins that recognize antigens. These exhibit significant diversity within and across subjects. TCR alpha-chain and beta-chain each have complementarity determining regions (CDRs). These regions are highly variable hence characterizing the T cell activity requires an understanding of the underlying statistical distribution of the T-cell and their CDRs.

T-Cell Characterization Technologies

The diversity of T-cell expressions are measured by the protein surfaces with flow cytometry or at the DNA level with spectratyping.

In terms of DNA techniques “Immuno-Seq” permits the simultaneous sequencing of genomic DNA from the rearranged TCR CD3 regions carried in millions of T cells. “This approach enables direct sequencing of a significant fraction of the uniquely rearranged TCR beta CR3 regions in populations of alpha T cells whose diversity far exceeds the capabilities of conventional capillary-based DNA sequencing instruments, and also permits estimation of the relative frequency of each CDR3 sequence in the population (Harlan S. Robins et al., “Comprehensive Assessment of T-cell Receptor beta-chain Diversity in AB T Cells,” Blood 114, no. 19 (November 5, 2009): 4099–4107, doi:10.1182/blood-2009-04-217604.).”

Other authors use both use flow cytometry and spectratype data, e.g. (Ciupe et al. BMC Immunology 2013, 14:35 <http://www.biomedcentral.com/1471-2172/14/35>).

Relevant Statistics

The main difference of T-cell characterization to other DNA phenotyping is the cells themselves have a distribution within the host.

Since not all the variations are observed in the target population in any given sample, statistical techniques are needed to infer the diversity itself within a given patient, and then variability across patients.

Diversity Metrics

In overall metrics, we can use the Kullback-Leibler divergence to measure the deviation of T-cell receptor diversity from normal.

Unseen species model for estimation of CDR3 sequence diversity

“The total number of unseen species, or CDR3 sequences, then, is the number of new TCR CD3 sequences that would be detected if the experiment were repeated an infinite number of times. The main assumption required is that T cells circulate freely in the blood (Harlan S. Robins et al.).”

This key assumption may not hold where T-cell distributions affect certain tissues. Other statistical models may be needed.

Clustering Techniques

It follows that the T-cell characterizing distribution can then be used to classify autoimmune pathologies without relying on other more general biological markers.

Applications

The application could be extended to researchers where certain types of T-cells are active in a particular active region of the body. For example, it might be used to characterize the T-cells in multiple sclerosis, where “T cells in these (MS) experiments revealed that at least some of the pervasive T-cell clones belonged to the CD8+ compartment, supporting the pathogenic relevance of this T-cell subset (Brain. 2007 Nov;130(Pt 11):2789-99. Epub 2007 Sep 21. Multiple sclerosis: T-cell receptor expression in distinct brain regions. Junker A1, Ivanidze J, Malotka J, Eiglmeier I, Lassmann H, Wekerle H, Meinl E, Hohlfeld R, Dornmair K.)”

The technique can be used to characterize the distributions of cells between healthy and sick patients.