Mechanism and Potential Diversity of T-Cell Receptor Rearrangement from Sequence Repertoires

Anand Murugan[1], Thierry Mora[2], Aleksandra Walczak[3] and Curtis G. Callan[1]

***Short Abstract* — T-cell receptor (TCR) diversity is generated by VDJ recombination. In this process, germline DNA is rearranged and edited stochastically to produce a random TCR coding sequence. The mechanism and potential diversity of this molecular process is still not fully understood. We apply maximum likelihood methods to deep sequencing data of TCR beta chain CDR3 repertoires to infer a probabilistic generative model of the VDJ rearrangement process by focusing on out-of-frame sequences that have not been subjected to selective forces. We find high consistency of the model between different individuals and we characterize the potential diversity of TCRs. The generative model provides insight into the molecular mechanism of rearrangement and also serves as a baseline for signatures of selection on the repertoire. We also study the relationship between the generation probability of TCRs and the sharing of TCRs between different individuals.**

***Keywords*— Immunology, T-cell Receptors, VDJ rearrangement, Statistical Inference, Expectation-Maximization.**

I. Introduction

THE ability to sequence entire immune cell repertoires [1] provides an opportunity for quantitative insight into the mechanism of diversity generation in VDJ recombination. Each T-cell Receptor (TCR) is created by stochastic editing of germline DNA: V, D and J genes are chosen and joined together with random deletions of nucleotides from each gene end, random insertions of nucleotides at the V-D and the D-J junctions, and creation of short palindromic nucleotides [2],[3],[4]. This results in a large diversity of potential receptor coding sequences. The properties of this molecular mechanism have not yet been quantitatively characterized.

We use repertoires of T-cell receptor beta chain CDR3 sequences of 8 human individuals to infer the statistical properties of these basic biochemical events. Since the repertoire is subject to selection in the thymus and elsewhere, we focus on the non-productive out-of-frame sequences that occur in a fraction of T-cells where the other copy of the chromosome carries a productive TCR. This subset of CDR3s has not been subject to selection.

Since any given CDR3 sequence can be produced in multiple ways, the probability distribution of underlying recombination events cannot be inferred directly from the observed sequences. We present a maximum likelihood method that uses an Expectation-Maximization (EM) algorithm to infer a probabilistic generative model of VDJ rearrangement.

II. Results

The generative model of VDJ rearrangement that we infer is highly consistent between individuals, suggesting its origin in a universal biochemical process. We are able to quantify the potential diversity of the T-cell repertoire and assess the significance of shared sequences between individuals in our data set. The model also serves as a baseline to probe the effect of selection on the repertoire.

The model provides insight into the molecular mechanism of rearrangement. In particular, we find that the distribution of nucleotide deletions is highly dependent on the gene being deleted, reflecting a sequence dependent activity of the nuclease involved, which we characterize. The nucleotide insertions at V-D and D-J junctions are independent and identical. The nucleotide bias in the insertions is well modeled by a di-nucleotide Markov model.

More generally, we argue that the use of maximum likelihood statistical inference methods is essential for quantitative understanding of the generation and evolution of diversity in the adaptive immune system.

Acknowledgements: The work of CGC and AM was supported in part by National Science Foundation grant PHY-0957573. The work of CGC was supported in part by US Department of Energy grant DE-FG02-91ER40671.

[1]Department of Physics, Princeton University, Princeton NJ 08544. E-mail: anandm@princeton.edu

[2]Laboratoire de physique statistique, UMR8550, CNRS, Paris, France. E-mail: tmora@lps.ens.fr

[3]Laboratoire de physique theorique, UMR8549, CNRS, Paris, France. E-mail: awalczak@lpt.ens.fr

References

[1] Robins et al., Comprehensive assessment of T-cell receptor beta-chain diversity in alpha-beta T cells. Blood, 114(19): 4099–4107, November 2009.

[2] David G Schatz and Patrick C Swanson. V(D)J recombination: mechanisms of initiation. Annual review of genetics, 45:167–202, 2011.

[3] Michael R Lieber and Thomas E Wilson. SnapShot: Nonhomologous DNA end joining (NHEJ). Cell, 142(3): 496–496.e1, August 2010.

[4] David G Schatz. V(D)J recombination. Immunological Reviews, 200(1): 5–11, August 2004.