Measuring immune repertoire rarefaction richness and diversity by genetic sequence fragment quantification

Gregory W. Schwartz[1] and Uri Hershberg[1]

***Short Abstract* — Adaptive immunity of vertebrates is known to be dependent on the diversity of B-cell and T-cell receptor repertoires. However, the exact level of diversity needed is not clear as well as how the level of diversity is functionally characterized in different parts of the genes encoding the B-cell and T-cell receptors. We present a novel fragment method to quantify the richness and diversity of V genes at different levels of precision. We found several common themes but also differences in richness of the V genes of the T-cell and B-cell receptor repertoires and between mice and humans.**

***Keywords* — adaptive immune system, lymphocytes, repertoire, diversity, richness, rarefaction, Shannon index, germline.**

I. Introduction

THE adaptive immune system is the body's means of defending against recurring pathogens and the underlying system through which vaccines function. Adaptive immunity is dependent on the diversity of its B-cell and T-cell receptor repertoire that enable identification and action against novel and recurring pathogens. This diversity is generated through two molecular means and in two distinct developmental stages.

The receptors on the T-cells and B-cells are composed of two heavy chains and two light chains [1]. These sequences are in turn considerably diversified due to somatic recombination [1]. Every individual has multiple genes encoding for potential variable regions (V), diverse regions (D), and joining regions (J). However, every immune cell selects from these germline genes and recombines only a single one of these into a single combined V(D)J gene. In addition, B-cells undergo an additional somatic hypermutation step [2].

The degree of variation for these receptor repertoires can be measured where appropriate for richness (the number of different alleles) and diversity (the number and evenness of different alleles) [3,4]. By splitting the sequences into fragments with position and window length properties, we can track the variation of the repertoires with different degrees of precision using germline sequences available on the ImMunoGeneTics database [5].

[1]School of Biomedical Engineering, Science and Health Systems Drexel University, Philadelphia, PA. E-mail: gs394@drexel.edu

II. Results

In general, we have discovered that receptor repertoire richness and diversity can be represented over all positions using levels of fragments. Additionally, CDR and Framework regions can be explicitly shown using this technique and variation due to CDR is insufficient to explain variation across the entire V region.

More specifically, light chain repertoires have more richness than heavy chain repertoires and T-cell receptor repertoires have more richness than Immunoglobulin repertoires, with the exception of Alpha in mice. The CDR and FW regions in human Alpha repertoires are relatively the same in richness and diversity

Human and Mouse diversities vary widely, whereas mouse diversities are the same with the exception of Beta while human diversities follow richness with the exception of Heavy Chain repertoires.

III. Conclusion

Lymphocyte receptor repertoire richness and diversity are not evenly distributed across all V gene positions for all levels of precision, implying biases for binding regions outside of the predefined CDRs which can be located with a fragment richness and diversity analysis. In addition, variations are not consistent between species which suggests an evolutionary factor in repertoire diversity.

References

[1] Schatz D, Swanson P (2011) V(D)J Recombination: Mechanisms of Initiation. *Annu. Rev. Genet.* **45**, 167-202.

[2] Peled J, et al. (2008) The Biochemistry of Somatic Hypermutation. *Annu. Rev. Immunol.* **26**, 481-511.

[3] Heck K, Belle G, Simberloff D (1975) Explicit Calculation of the Rarefaction Diversity Measurement and the Determination of Sufficient Sample Size. *Ecology* **56**, 1459-1461.

[4] Soetaert K, Heip C (1990) Sample-size dependence of diversity indices and the determination of sufficient sample size in a high-diversity deep-sea environment. *Mar. Ecol. Prog. Ser.* **59**, 305-307.

[5] Giudicelli V et al. (2006) IMGT/LIGM-DB, the IMGT® comprehensive database of immunoglobulin and T cell receptor nucleotide sequences. *Nucl. Acids Res.* **34**, D781-D784.