
Seminar

Daron M. Standley

Dept. of Genome Informatics/Genome Information Research Center, Osaka University

“Integration of protein sequence and structural data for prediction of biological function”

Immune-based medicine, including checkpoint therapy, chimeric antigen receptors, and bispecific antibodies, is one of the best approaches to treat a wide range of cancers. Emerging B cell receptor (BCR) and T cell receptor (TCR) sequencing data provides an unprecedented view of how our immune system interacts with tumor cells and other diseased tissues (infection, autoimmunity, graft versus host disease, etc.). In order to make practical use of such data, there is a need to integrate protein sequence and structural information. To address this need, we have developed a new version of the multiple sequence alignment software MAFFT¹ that can efficiently interact with our newly released database of aligned structural homologs, DASH (<https://sysimm.org/dash/>)². First, we show that MAFFT-DASH alignments are more accurate, overall, than any tested software, without introducing significant burdens with regard to usage or computational resources². Next, we show that by using MAFFT-DASH alignments we were able develop a new tool, Repertoire Builder (https://sysimm.org/rep_builder/), that is capable of generating BCR or TCR structural models in a high-throughput manner that are more accurate than any other tested software³. We have further extended this technology by accurately clustering Repertoire Builder BCR and TCR models according to their antigen and epitope specificity using the new tool, InterClone⁴. In another direction, we have extended Repertoire Builder TCR models in order to generate BCR-antigen and TCR-epitope-MHC structural models, the latter of which has been released as ImmuneScape (<https://sysimm.org/immune-scape/>)⁵. This interoperable collection of tools will enable interrogation of the immune microenvironment of disease tissues at the molecular level and thus can assist in the development of novel immune-based biomarkers and therapeutics.

Friday, March 8, 2019

2:30 PM

***Institute for Advanced Computational
Science seminar room***

Host: Dima Kozakov

***Refreshments following the lecture
Laufer Hub 110***

References

1. Katoh, K. and Standley, D.M. MAFFT multiple sequence alignment software version 7: improvements in performance and usability *Mol Biol Evol* (2013)
2. Rozewicki, J. et al. Integrated sequence and structural alignment using MAFFT and DASH (in prep)
3. Schmitt, D. et al. Repertoire Builder: High-throughput structural modeling of B and T cell receptors (in prep)
4. Xu, Z. and Li, S. et al. Functional clustering of B cell receptors using sequence and structural features (in prep)
5. Li, S. et al. Structural modeling of lymphocyte receptors and their antigens (in press) *Methods Mol Biol* (2019)



Stony Brook University