Raditya Utama
Cold Spring Harbor Laboratory, NY, 11724
e-mail: Utama@cshl.edu

Abstract

Recognition of antigens in adaptive immune responses plays a crucial role in fighting cancer progression and has been essential in developing cancer immunotherapy. It has been widely understood that the clonal expansion of antigen-binding T-cells highly depends on the type of disease and its associated pathogens. Here we show the significance of T-cell receptor (TCR) sequences pattern and clonality in determining the tissue type where a specific cancer developed. We collected 8742 cancerous samples from 22 different tissue types in The Cancer Genome Atlas (TCGA) database. Repertoire of TCR sequences and abundance were obtained by aligning the bulk RNA-Seq database onto reference sequences of TCR beta chain. We implemented deep neural networks method for learning informative values within complementarity-determining region 3 (CDR3) that are useful to predict the cancerous tissue of origin. We achieved significant average area under the curve (AUC) of 0.62 and 0.77 (including HLA-type) representing the amounts of informative values within the specific region. High variability of learning accuracies is suggestive of close association among contributing factors including clonality, HLA-type, immune infiltration, and mutational load. Nearly random prediction of skin cancer melanoma (SKCM) supports the role of melanoma-associated antigens (MAGE) expression in modulating T-cell responses, despite the high clonality and mutational load. On the other hand, we suggest that the rate of immune infiltration due to immune-privileged sites and dense fibrotic tissue contributed to the clonality and predictability of tissue types in brain and pancreatic cancer. Together, these results support the significant association between TCR sequence clonality and tissue specificity due to antigen abundance pattern and microenvironmental factors.