

catELMo: Context-Aware Amino Acid Embedding Advances Analysis of TCR-epitope Interactions

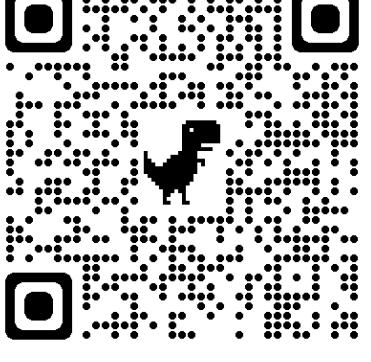


Pengfei Zhang^{1,2}, Seojin Bang², Michael Cai^{1,2}, and Heewook Lee^{1,2,*}

¹School of Computing and Augmented Intelligence, Arizona State University

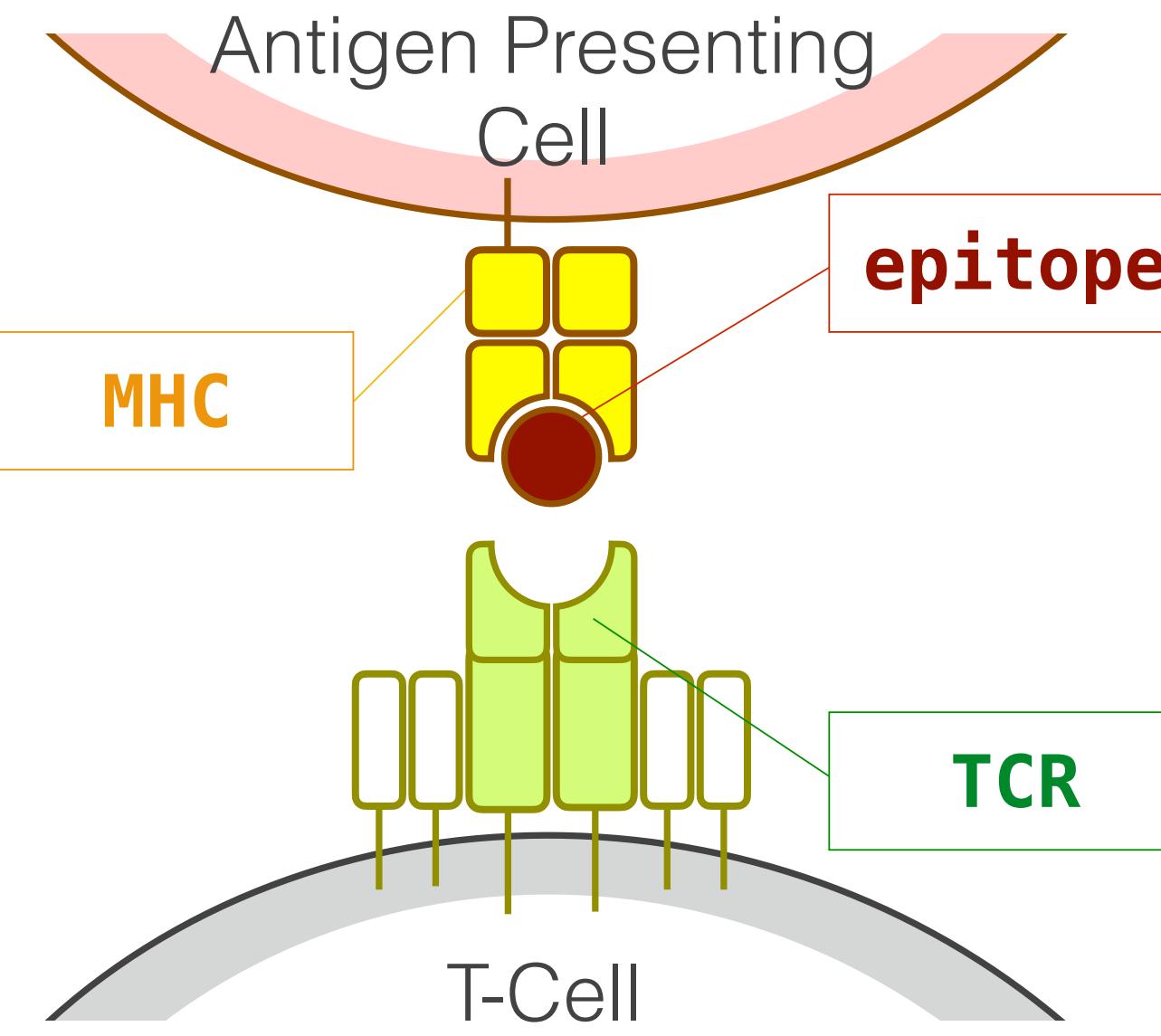
²Biodesign Institute, Arizona State University

View code



Background.

- T cell receptors (TCRs) play a critical role in adaptive immune systems as they enable killer T cells to recognize abnormal cells from healthy cells.
- Computationally predicting their integration results can drastically reduce the cost and the time needed to narrow down a set of candidate TCR targets and accelerating the development of personalized immunotherapy.



Key Ideas.

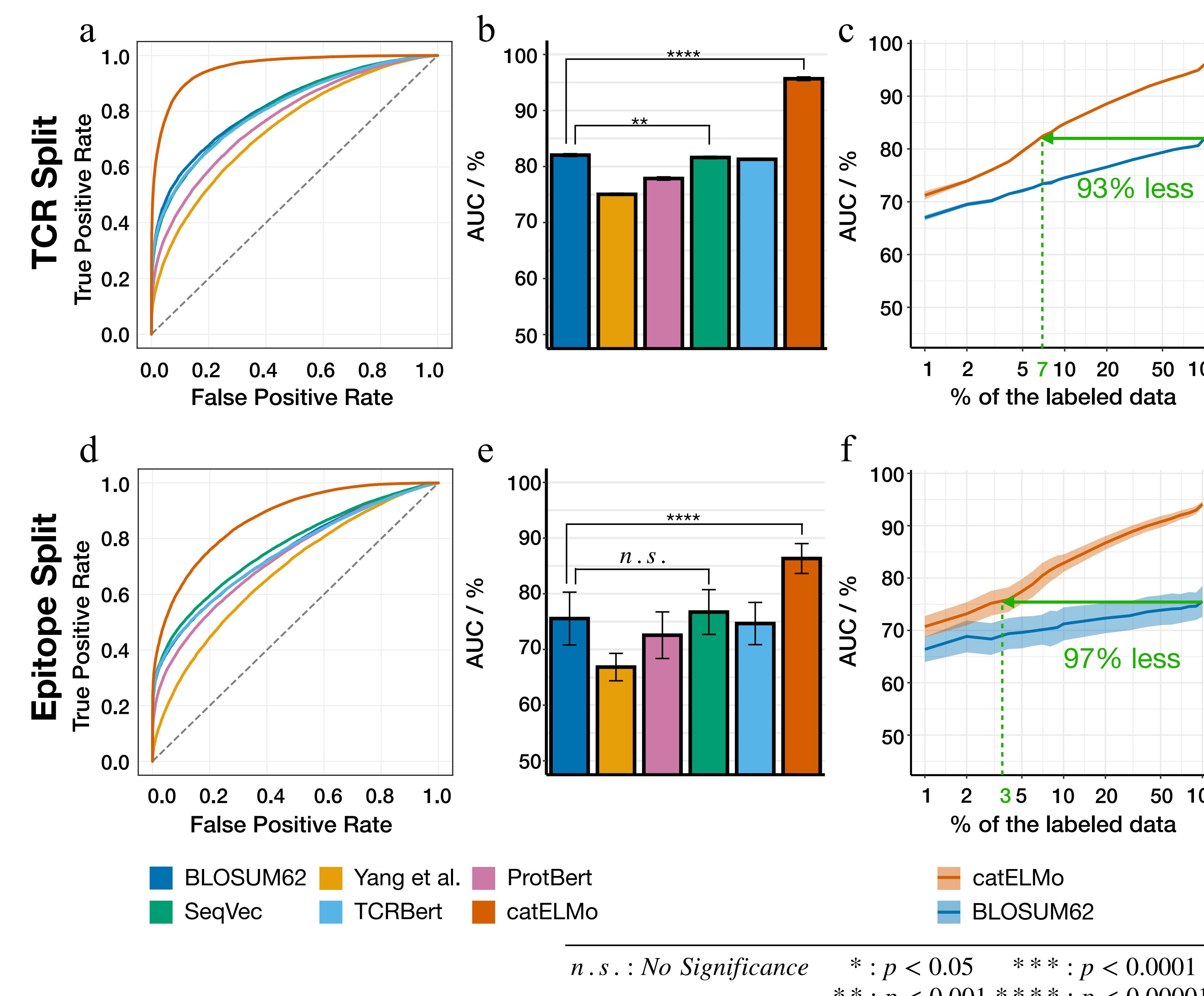
Learn dynamic amino acid embedding models to better represent amino acid residues for TCR analysis.

Takeaways of catELMo.

- Pre-trained on ~4M TCRs with **no supervision labels**.
- An **easy-to-use** feature extractor for any TCR analysis.
- An alternative **more effective** embedding solution achieved **significant performance gains**.

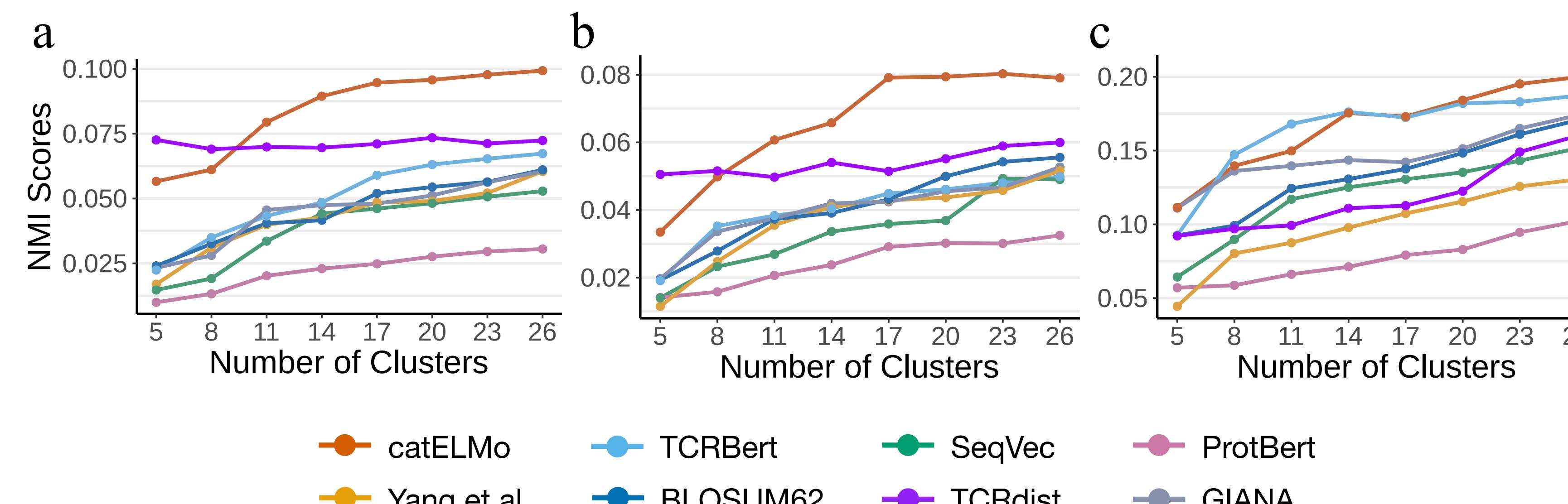
Downstream 1. TCR-epitope binding affinity prediction.

- catELMo significantly outperforms the state-of-the-art method by >14% AUCs.
- catELMo dramatically reduces >93% annotation cost while achieves an equivalent performance to the state-of-the-art method.



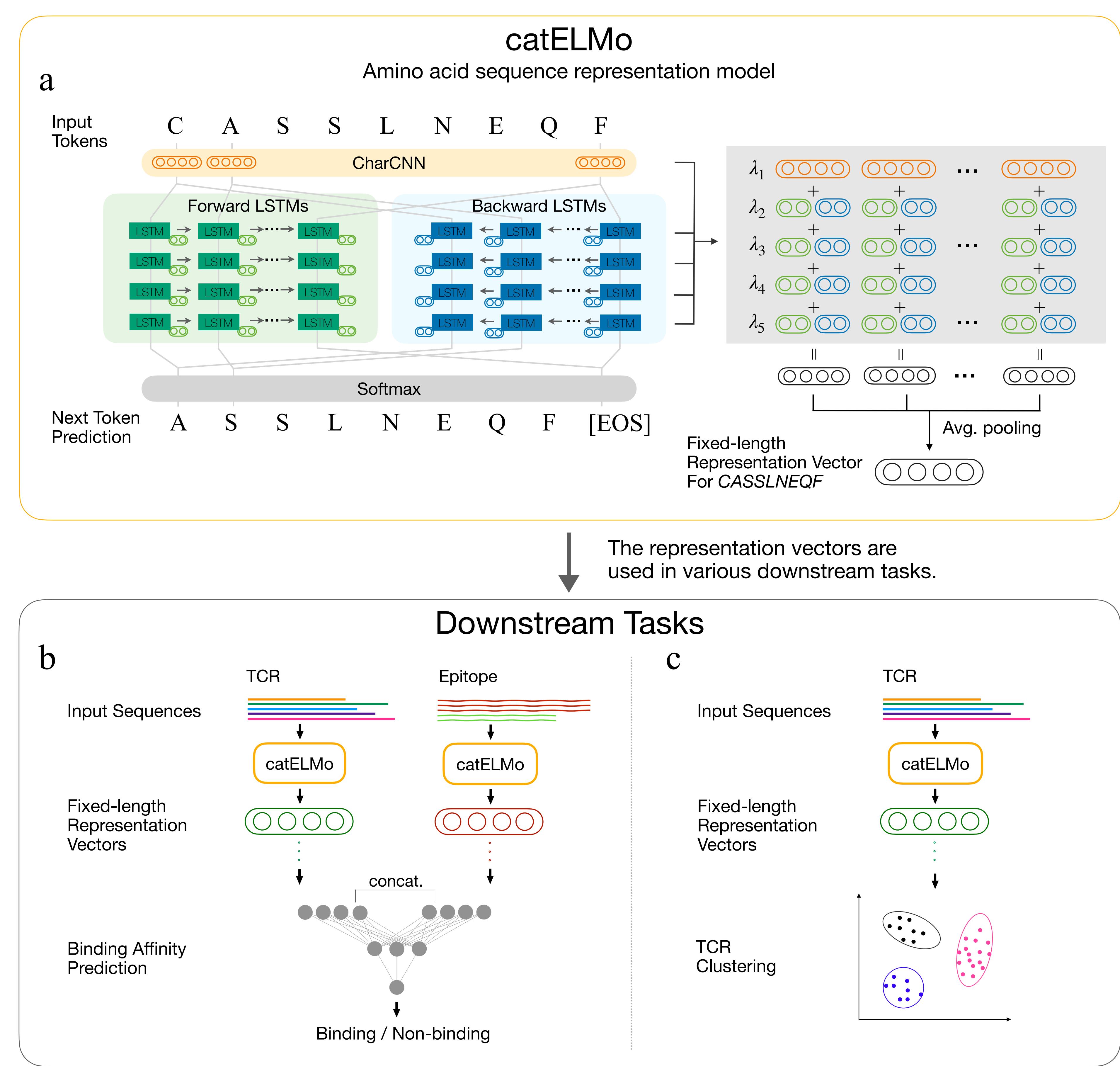
Downstream 2. Epitope-specific TCR clustering.

Hierarchical clustering based on Euclidean distance between TCR representations. Clustering epitopes are from a) both human and mice species, b) human, and c) mouse.



catELMo framework.

- It is a bi-directional amino acid embedding model that learns contextualized amino acid representations.
- It learns patterns of amino acid sequences by predicting the next amino acid token given its previous tokens.
- It yields a real-valued representation vector for amino acid sequences, and can be used as input features of various downstream tasks.



tSNE of latent space from last hidden layer of binding affinity prediction models.

Clearer separation between similar color represents better decision boundary.

