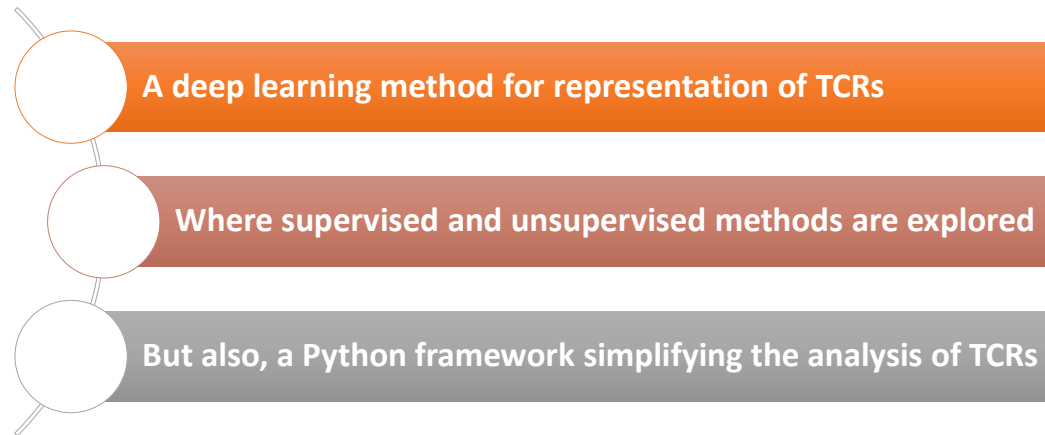
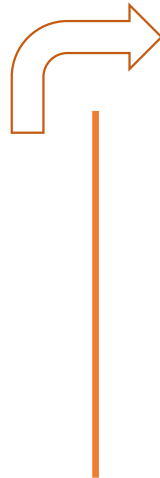


What is DeepTCR ?



MAIN OBJECTIVES

- Use Deep Learning (in particular CNNs) in order to **extract important features** from sequencing data
- Learn sequence motifs and hence patterns useful for both **descriptive and predictive purposes**



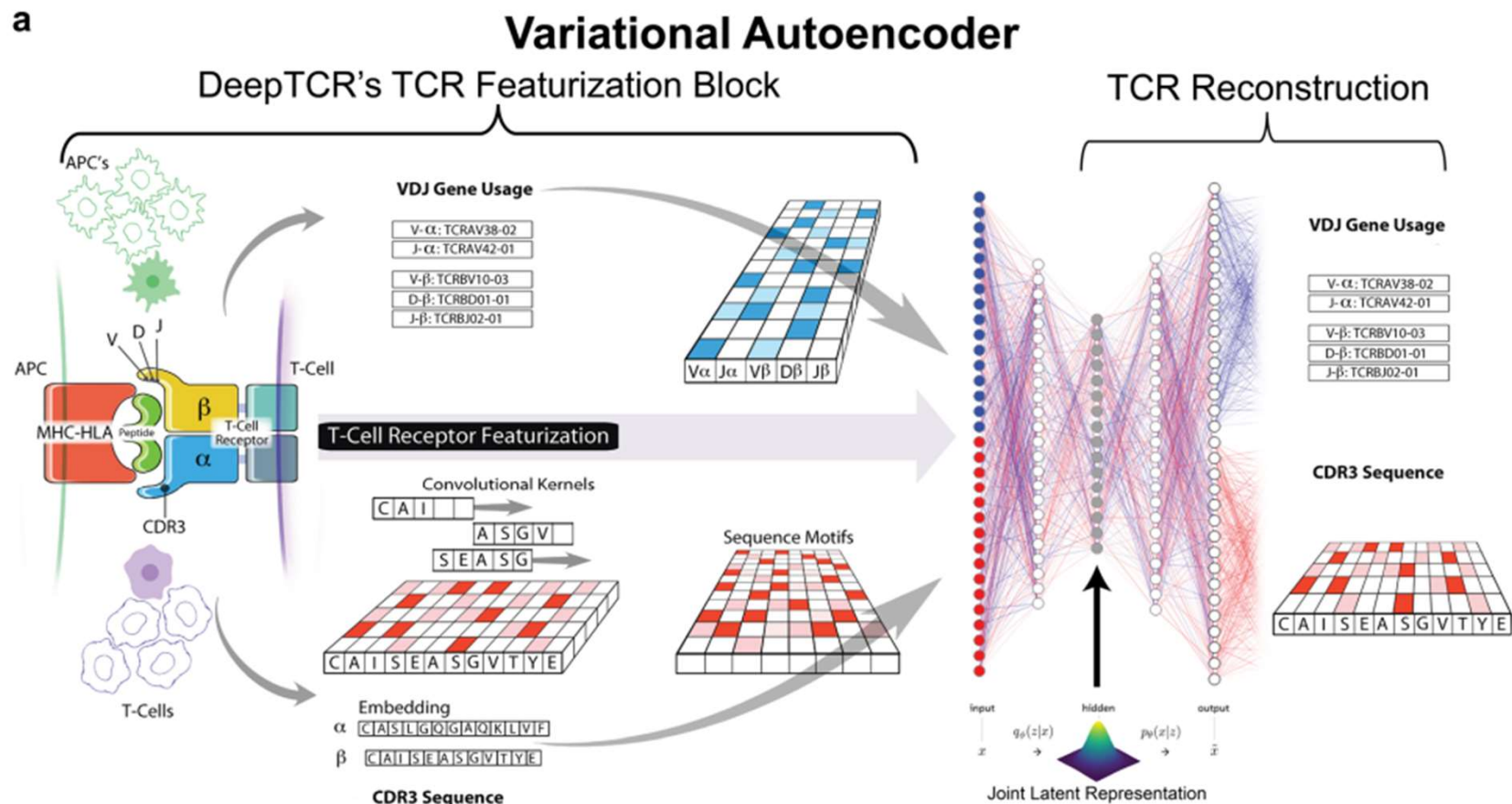
METHODS USED

- One common method used for TCR embedding and featurization
- Variational autoencoder for unsupervised learning
- Perturbation analysis explored for the supervised learning case → sensitivity and importance analysis
- “Weakly supervised” method explored

TCR featurization:

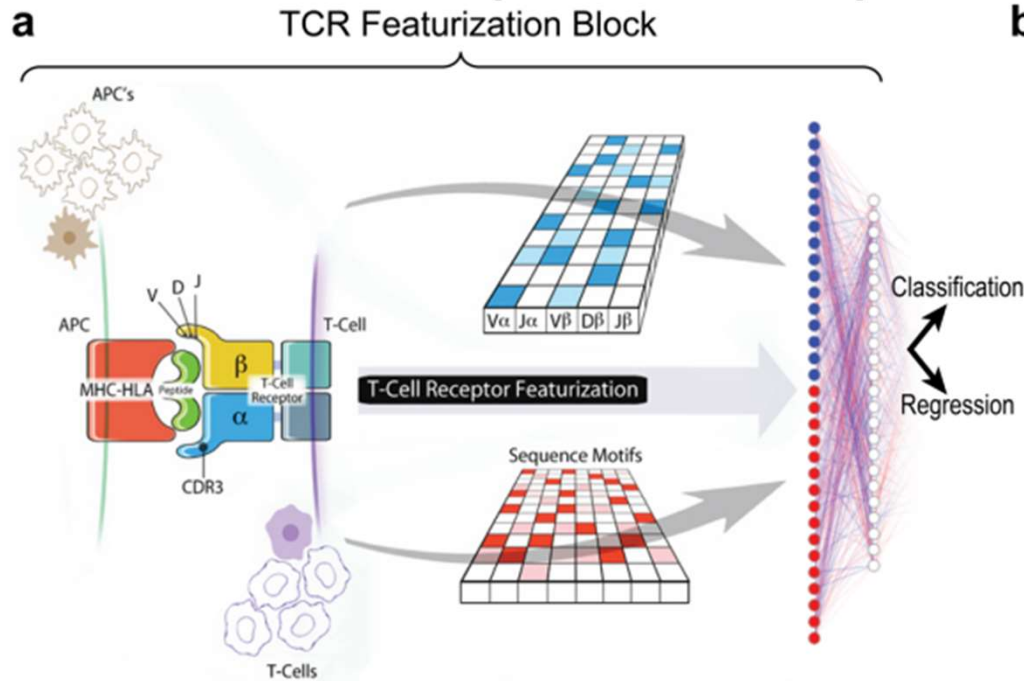
- **One common method** used for multiple architectures:
 1. Embedding of the sequences with a trainable (or not) embedding layer
 2. Three-layer CNN used to extract the most important features from the chains
 3. Use of the V/D/J gene sequences with “one-hot” representation through an additional embedding layer
 4. Combination of both CDR3 embeddings and V/D/J features to obtain an accurate and complete joint representation

- Use of Variational Autoencoder in order to learn the underlying distribution of the sequence data to cluster TCR sequences (based on Hamming distances)



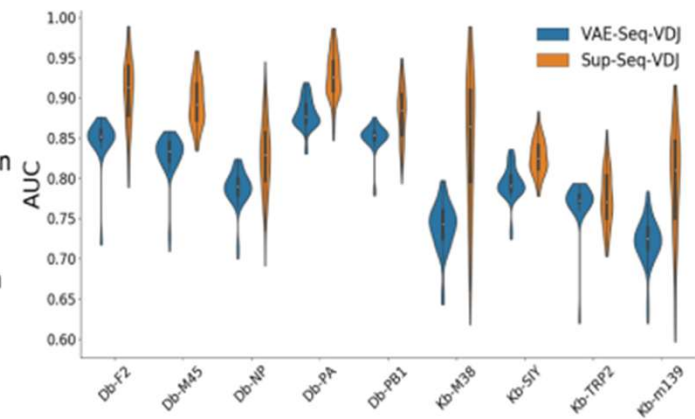
Supervised Classification:

Supervised TCR Sequence Classification



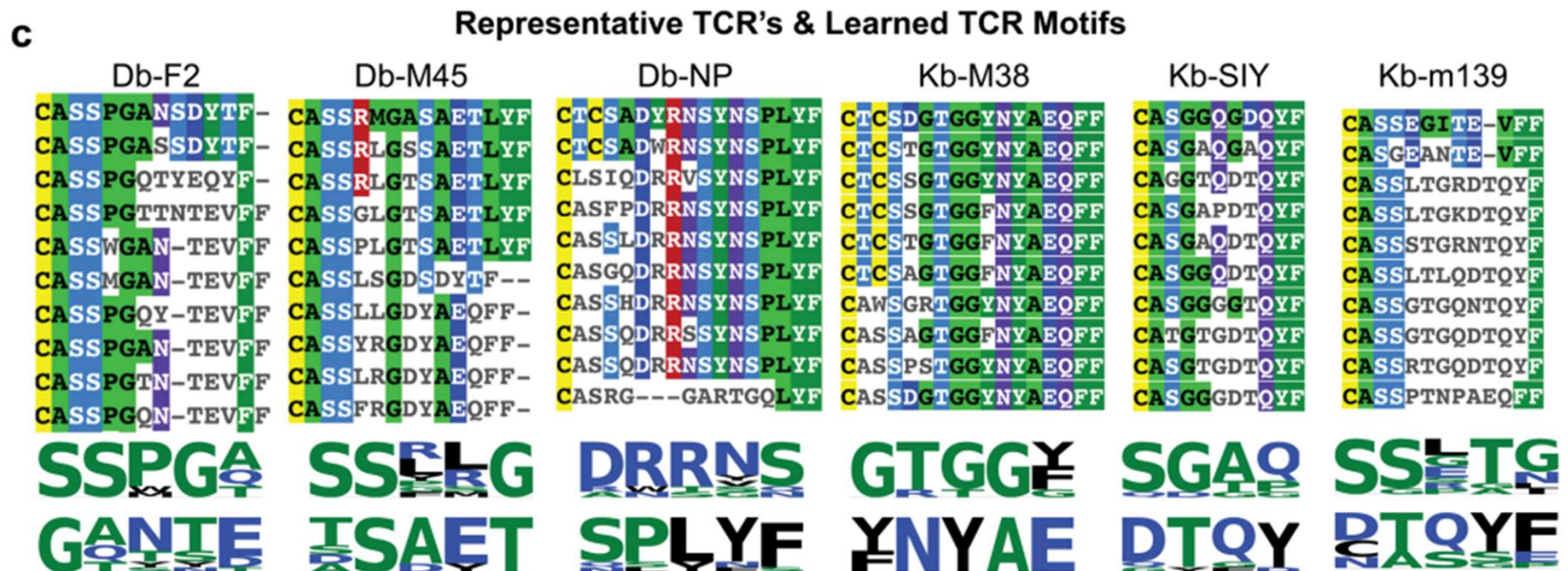
b

Antigen Classification Performance



Perturbation analysis:

- Ability to perturb the inputs and measure the change in the output of the supervised model → Hence, determining the relative importance of antigen-specific binding



DeepTCR Python package:

- ☐ All the code is available freely on GitHub and it can be easily re-used if necessary
- ☐ Clear and understandable code used throughout the package
- ☐ Availability of easy tutorials to reproduce easy tasks with Jupyter notebooks

[LINK TO REPO](#)

Paths not explored by DeepTCR:

- Transfer learning pipeline for more performant model results → use today state-of-the-art NLP-like methodology
 - Determine the interpretability of the weights for the unsupervised case through a perturbation analysis
 - Try different distances metrics to evaluate our model (TCRdist3?)
 - Not clear which regularization is used
 - Generate “artificial TCR sequences” that would have a similar binding specificity as other TCRs
- Clustering methods in general are not really explored → use and evaluation of the embeddings obtained with VAE in clustering