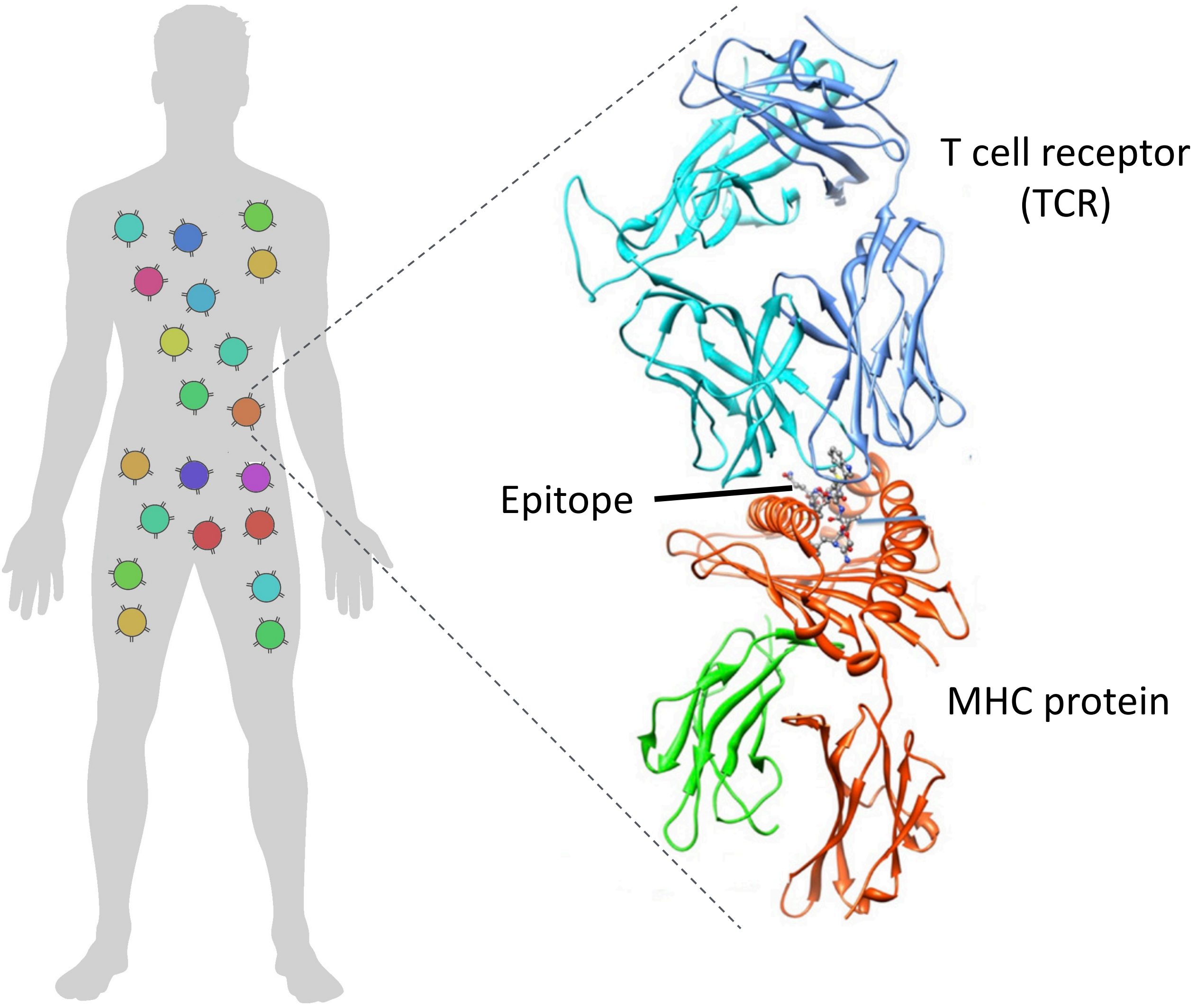
Interpretable prediction of T cell receptor binding

María Rodríguez Martínez

IBM, Zürich Research Lab

*https://precise-project.eu/

T cells are essential for adaptive immune responses



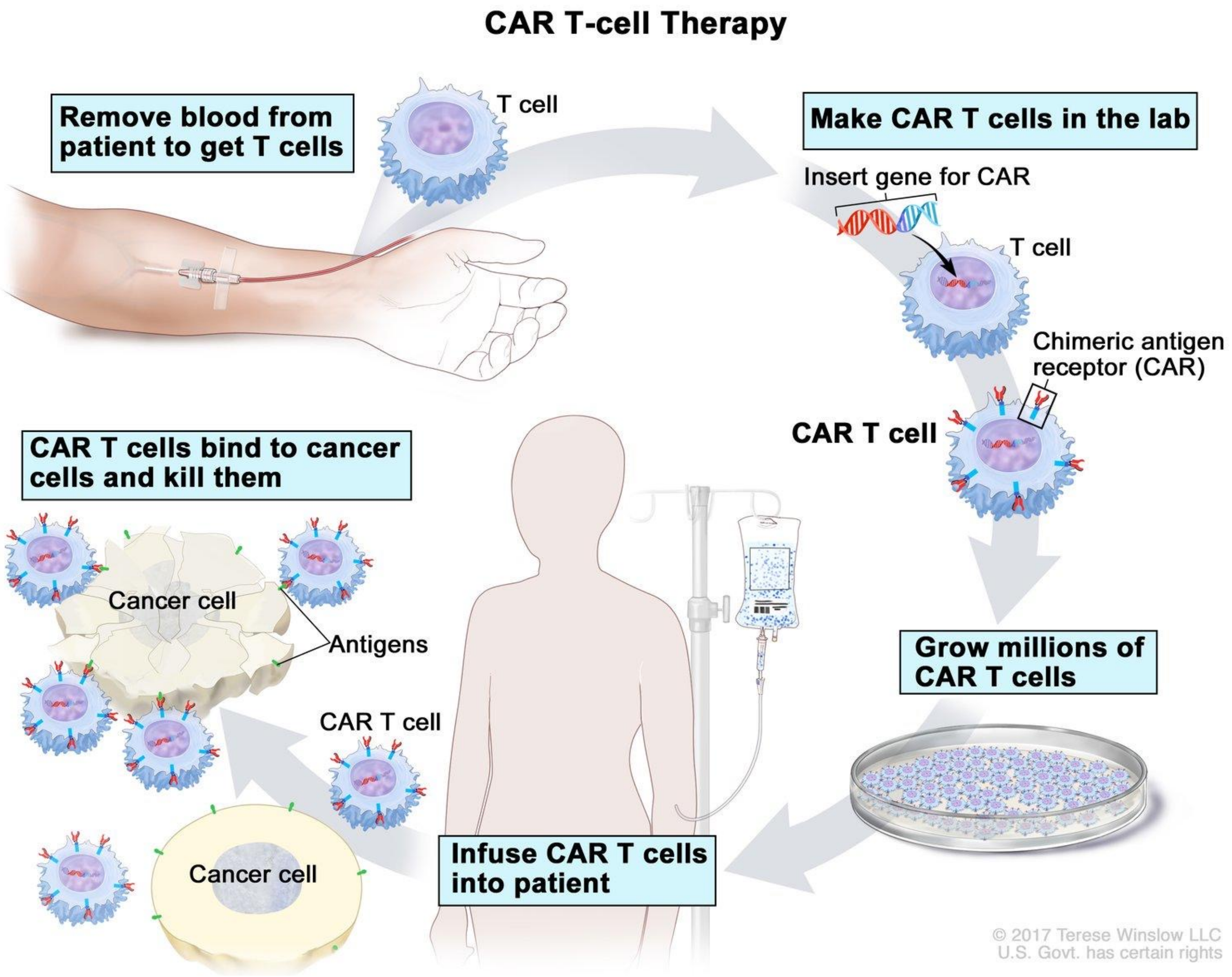
Immense Diversity of T Cells

Highly Specific Binding to Foreign Epitope

- T cells
  - → distinct T cell receptor (TCR)
  - → selectively recognize foreign epitopes
  - → cross-reactive
  - → regulate adaptive immune responses

Modeling adoptive cell transfer therapies

- Chimeric Antigen Receptor (CAR) T cell therapies
- T cell-based therapies



- Extract T cells from a patient
- (Engineer them to express a CAR)
- Expand them in vivo
- Reinfuse them into the patient

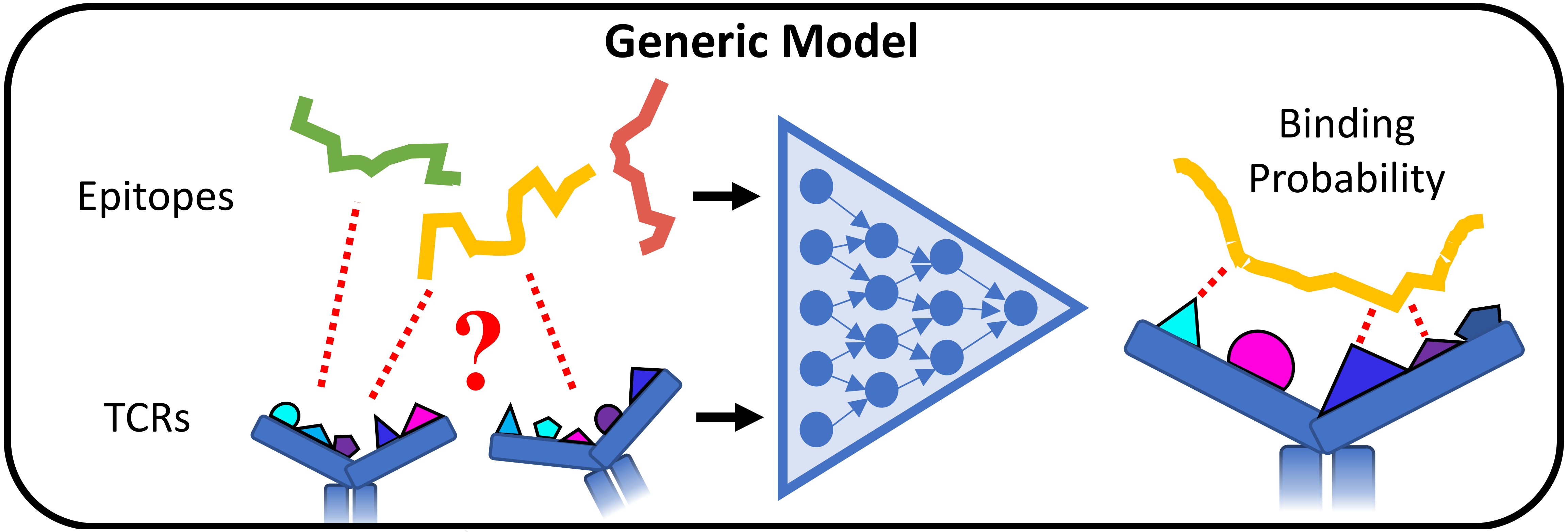
Credit: US National Cancer Institute

TITAN, an interpretable model to predict T cell receptor – antigen binding.

*https://precise-project.eu/

TITAN: AI model to predict T cell receptor – antigen binding

- TITAN: multimodal neural network to predict TCR binding



- 10^{15}-10^{20} theoretical diversity of TCRs
- Binding data of very few epitopes

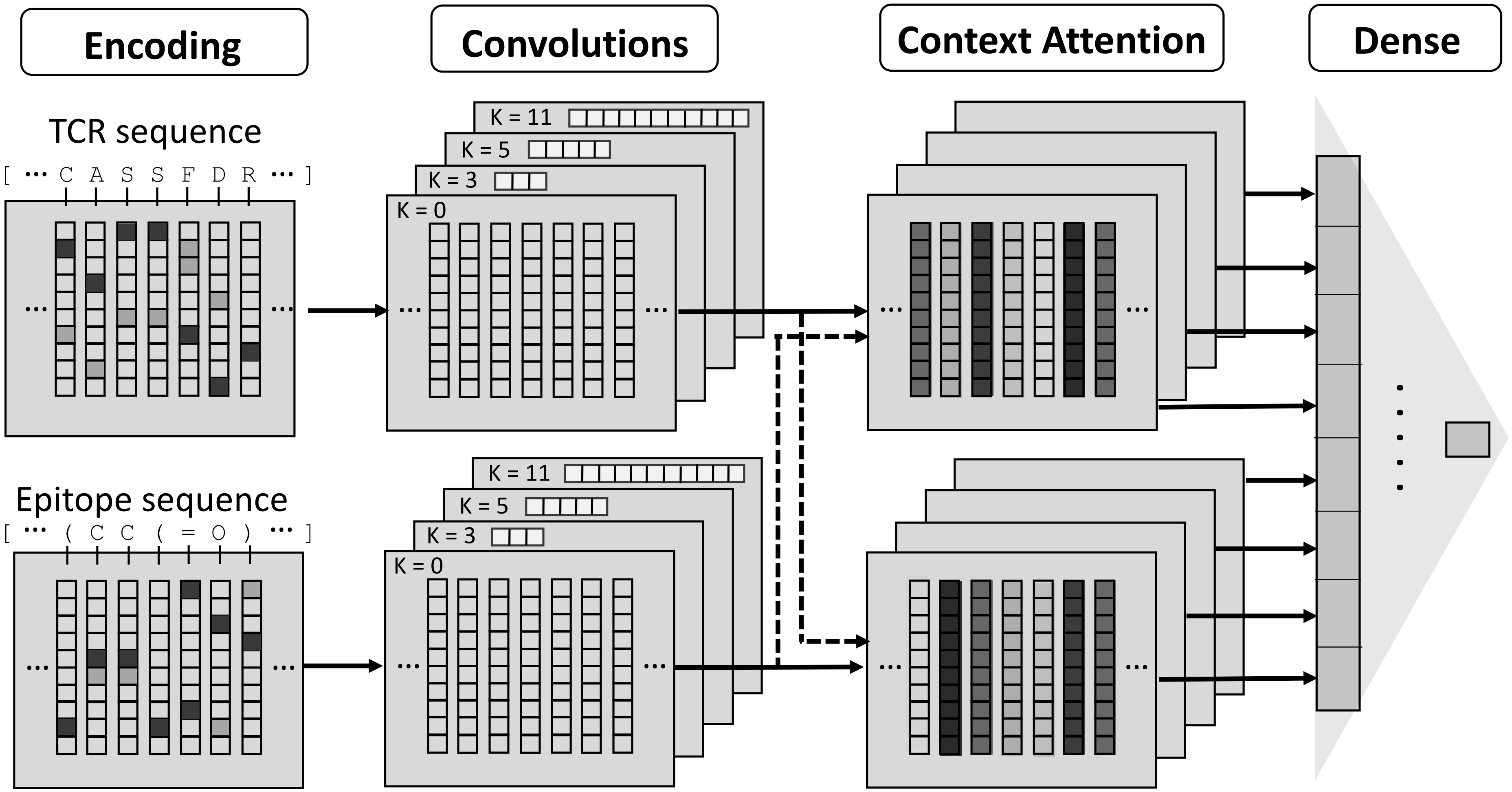
Weber et al, Bioinformatics, 2021.

Data

- VDJ database contains 40,438 TCR sequences binding 191 epitopes
- COVID dataset with 154,320 TCR sequences binding 269 epitopes
- Generate negative examples by shuffling

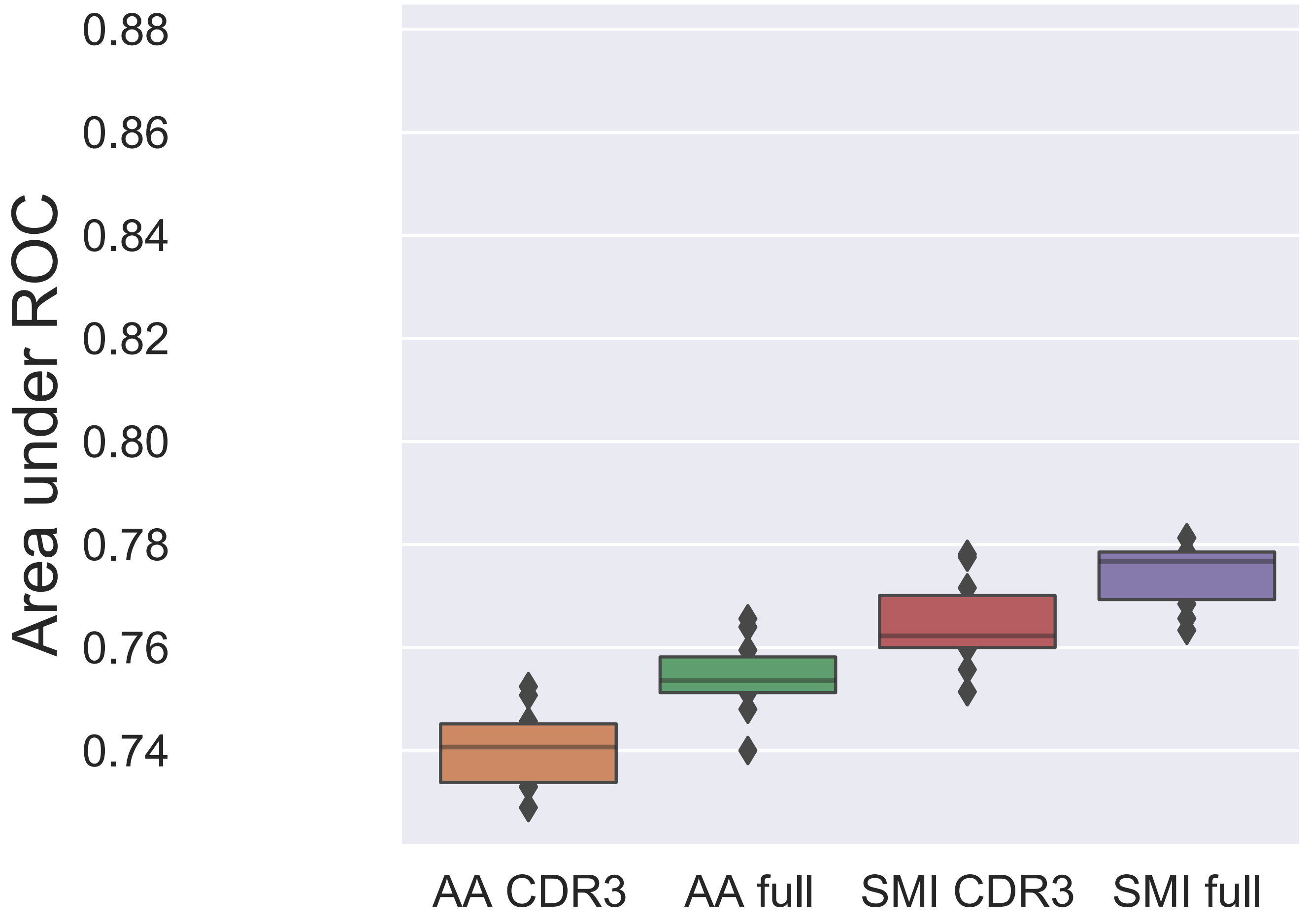
- Final data:
    - **46,290 TCRs binding 192 epitopes (15 – 400 TCRs/epitope)**

TITAN: predicting TCR- epitope binding affinity



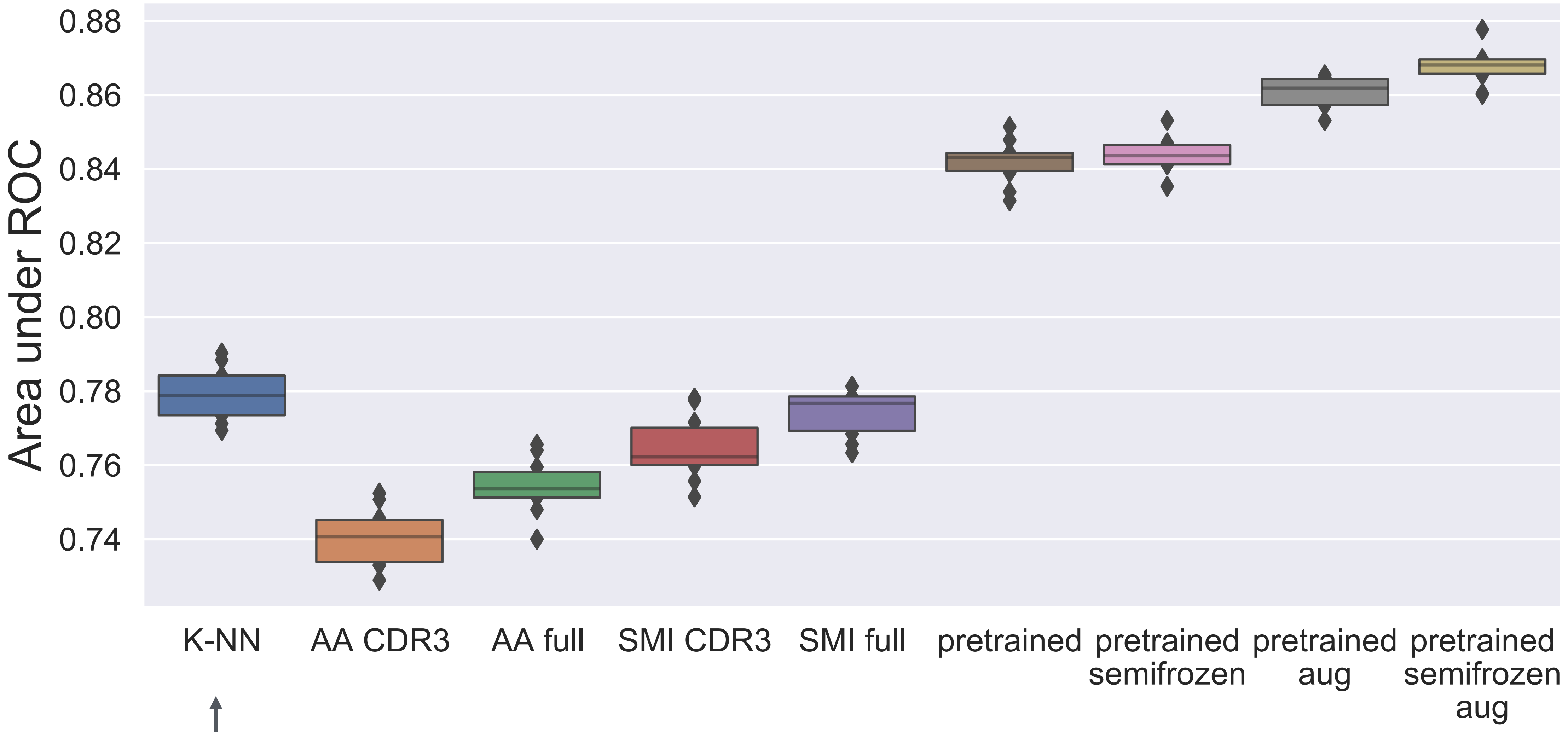
Weber et al, Bioinformatics, 2021.

TITAN performance



Weber et al, Bioinformatics (2021)

TITAN performance: good TCR generalization

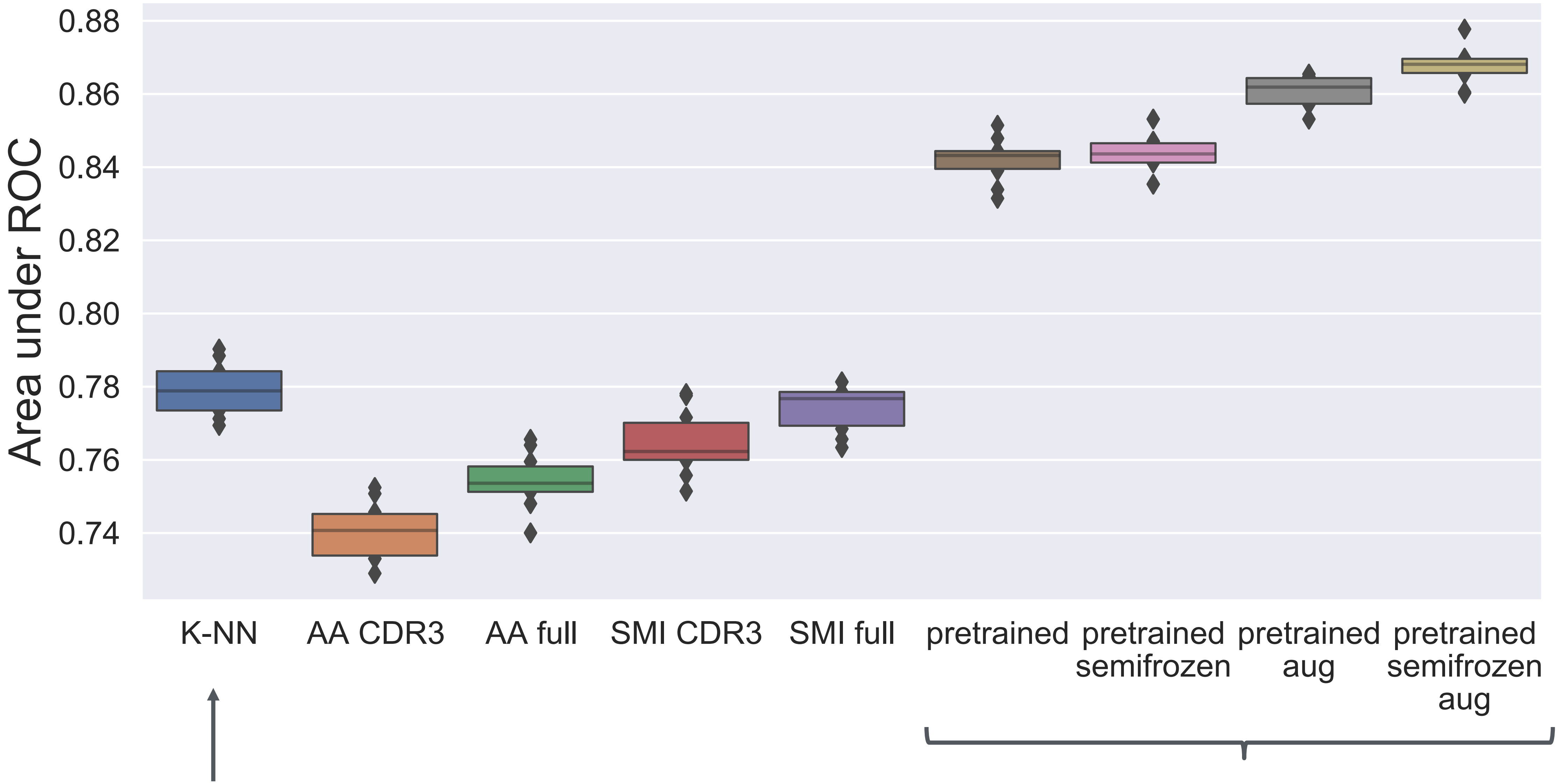


Baseline

TITAN models pretrained on BindingDB

Weber et al, Bioinformatics (2021)

TITAN performance: good TCR generalization

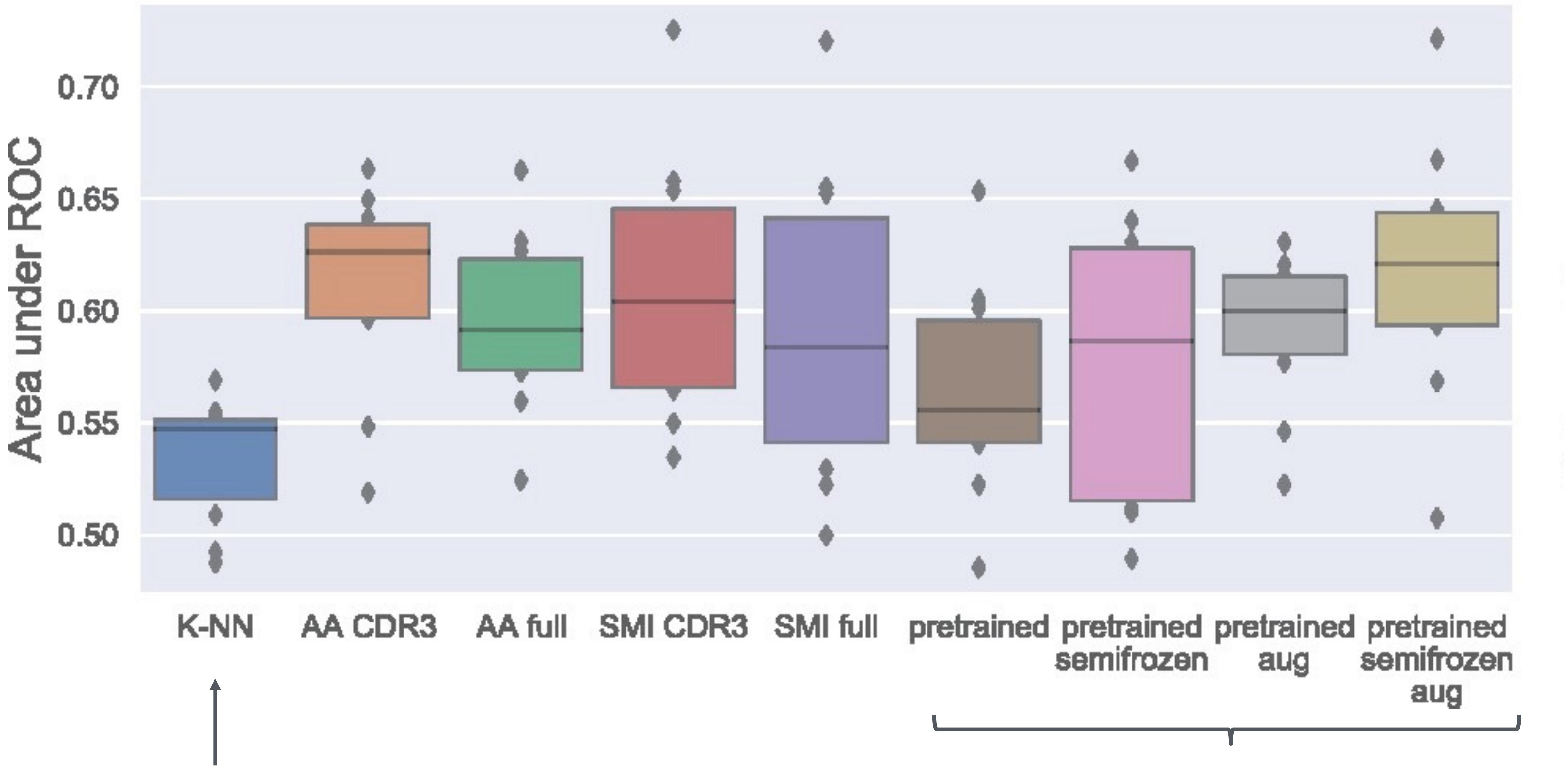


Baseline

TITAN models pretrained on BindingDB

Weber et al, Bioinformatics (2021)

TITAN performance: poor epitope generalization



Baseline

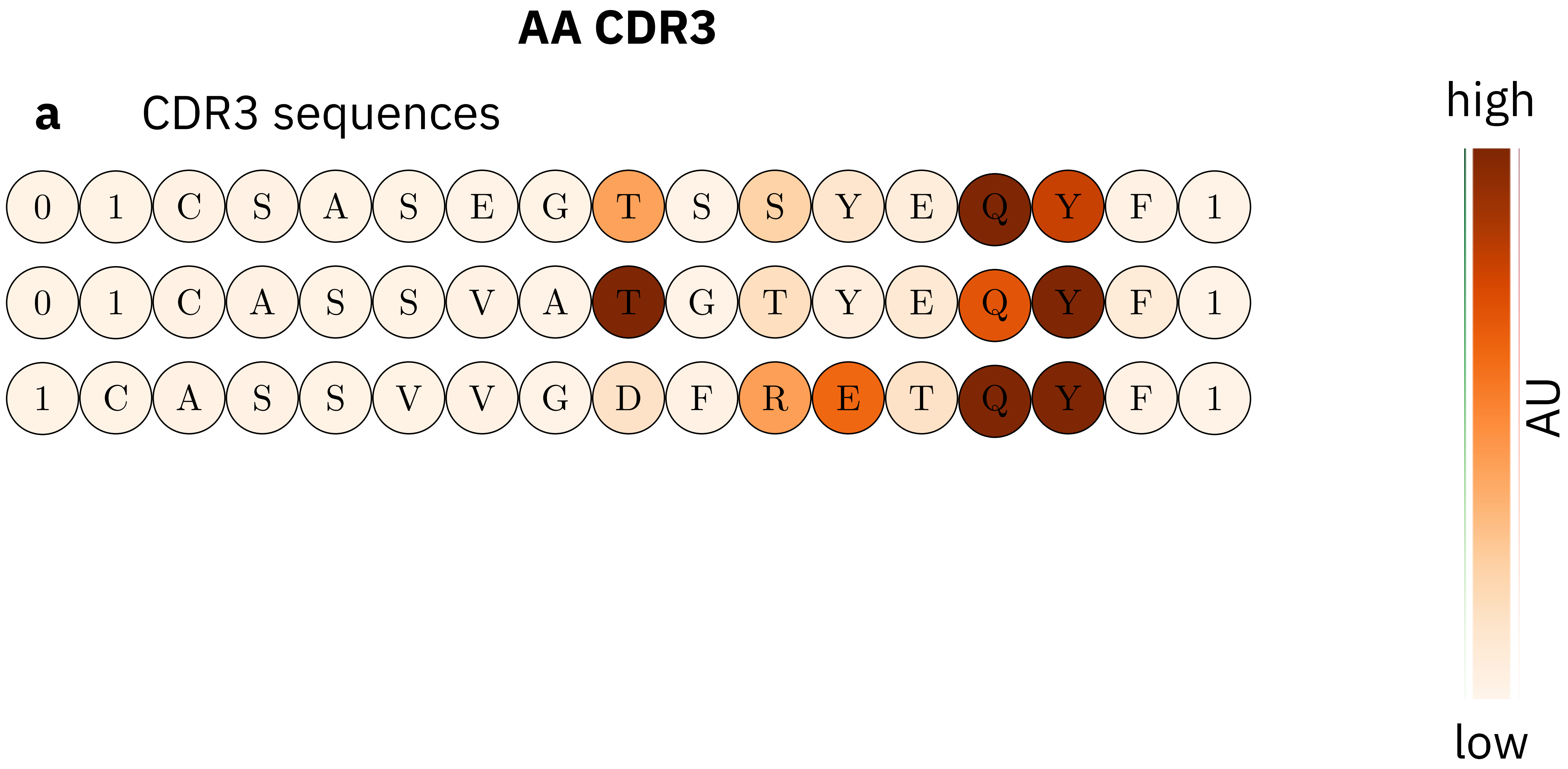
TITAN models pretrained on BindingDB

Weber et al, Bioinformatics (2021)

Extracting insights with attention mechanisms

Attention mechanisms:

→ Identify amino acids most important to predict the binding

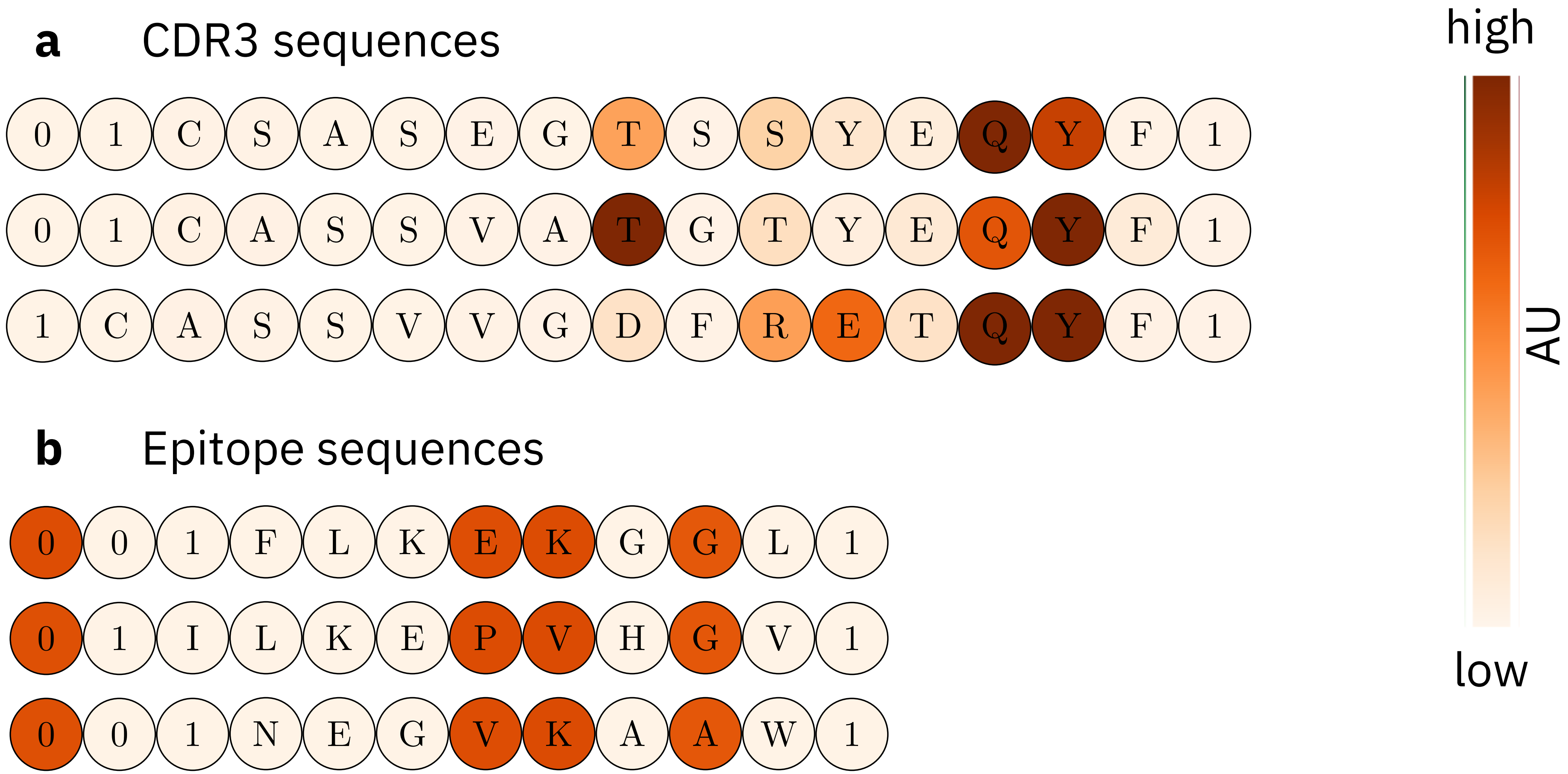


Extracting insights with attention mechanisms

Attention mechanisms:

→ Identify amino acids most important to predict the binding

AA CDR3



a CDR3 sequences

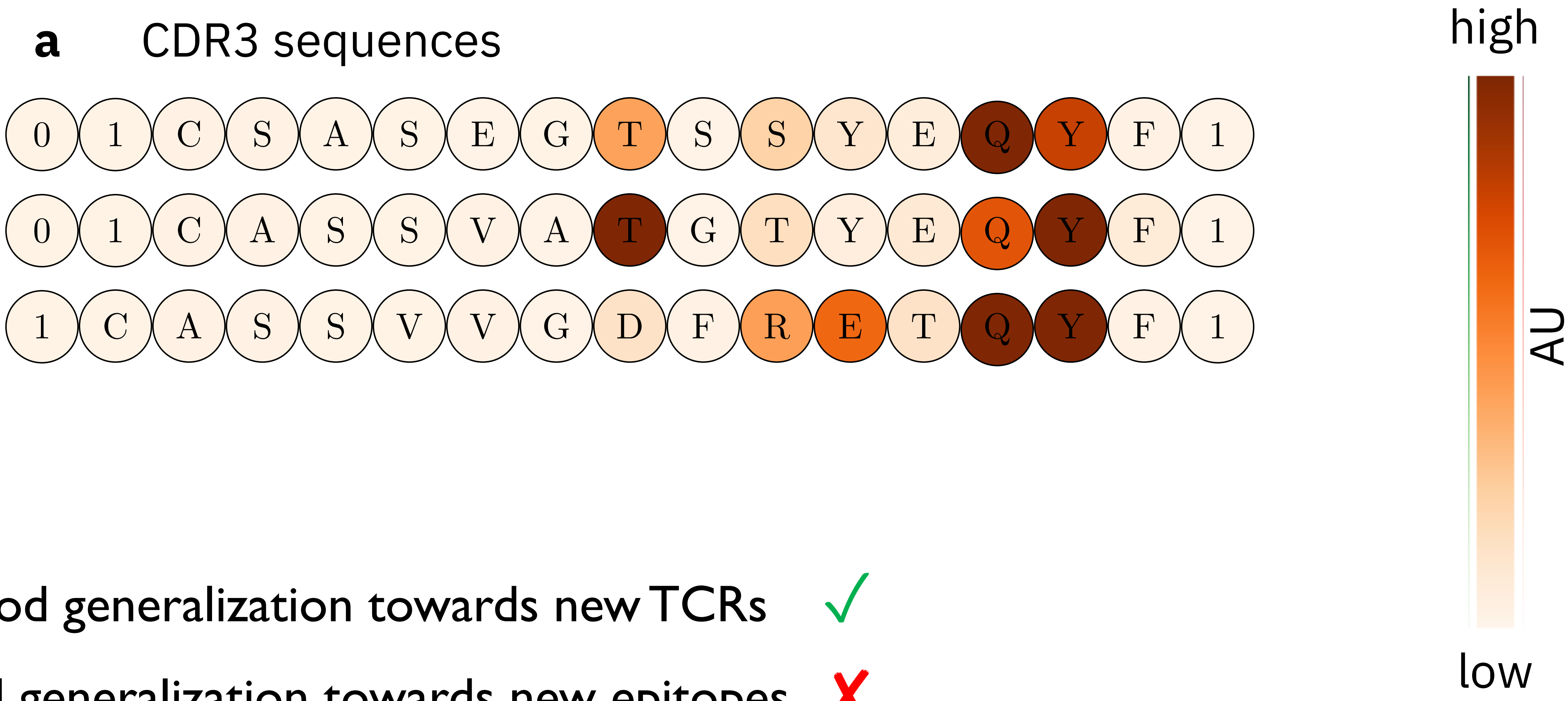
b Epitope sequences

Extracting insights with attention mechanisms

Attention mechanisms:

→ Identify amino acids most important to predict the binding

AA CDR3



- Good generalization towards new TCRs ✓
- Bad generalization towards new epitopes ✗

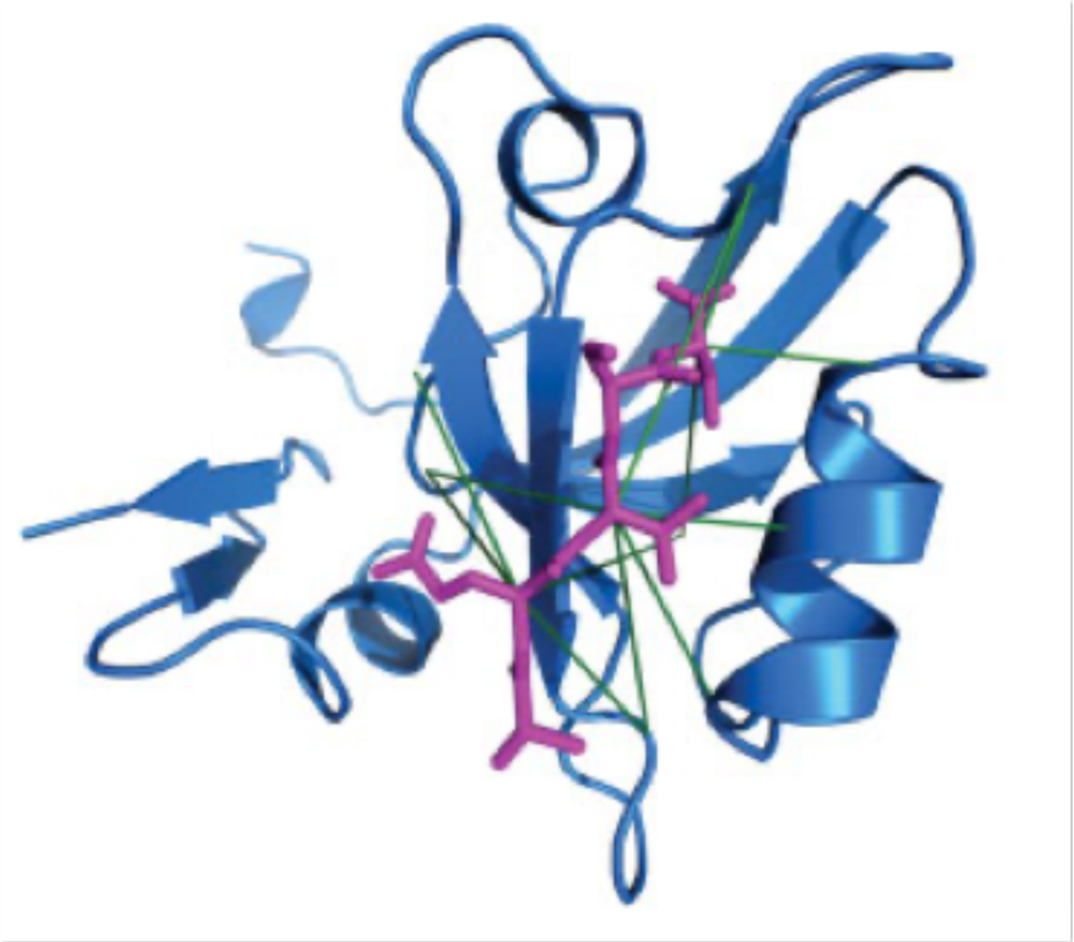
Unravelling TCR-antigen binding rules

Predict cross-reactivity:

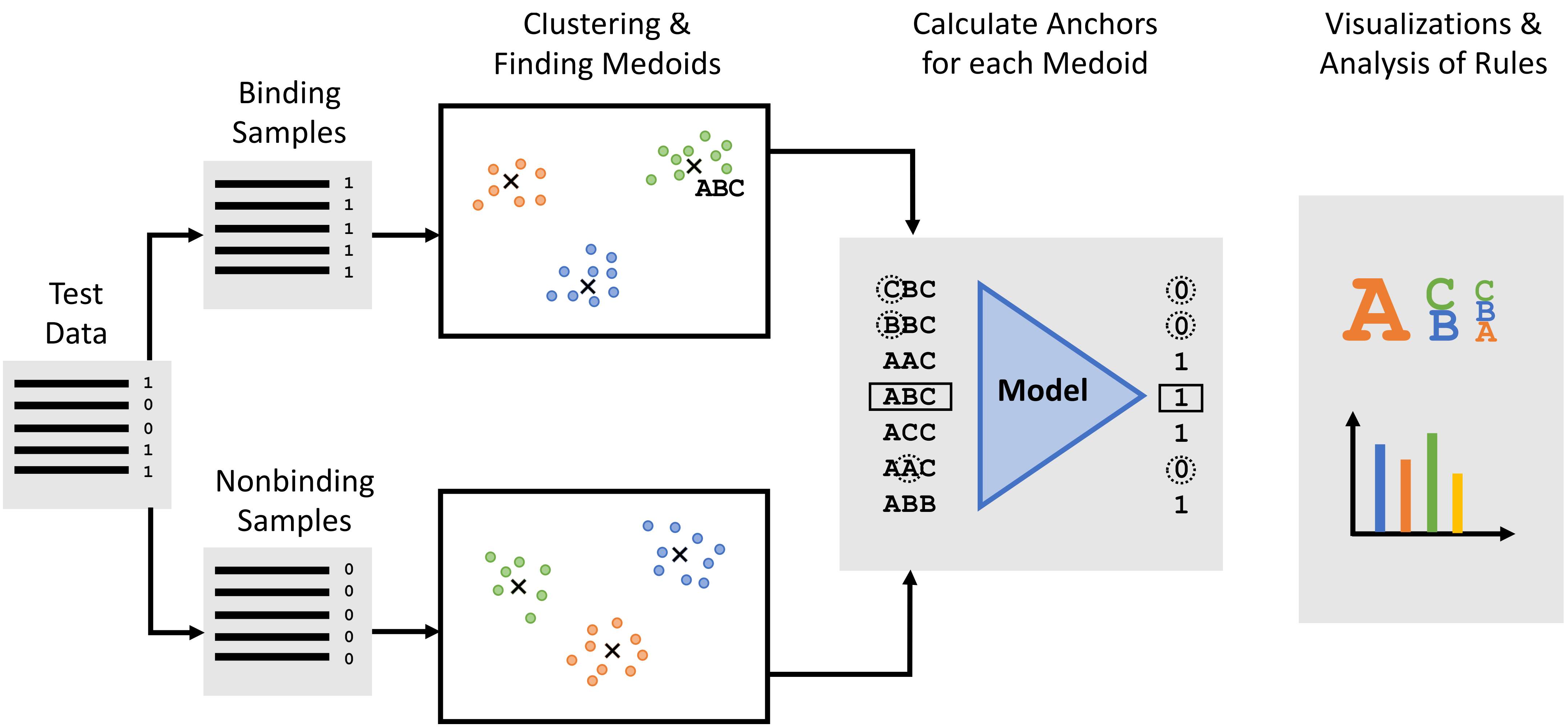
- A single TCR can recognize a million peptides
    - Off-target toxicities
    - Autoimmune diseases

Engineer TCRs with improved affinity

Mechanistic insights



DECODE: Interpretable pipeline for TCR binding



Test Data

Binding Samples

Nonbinding Samples

Clustering & Finding Medoids

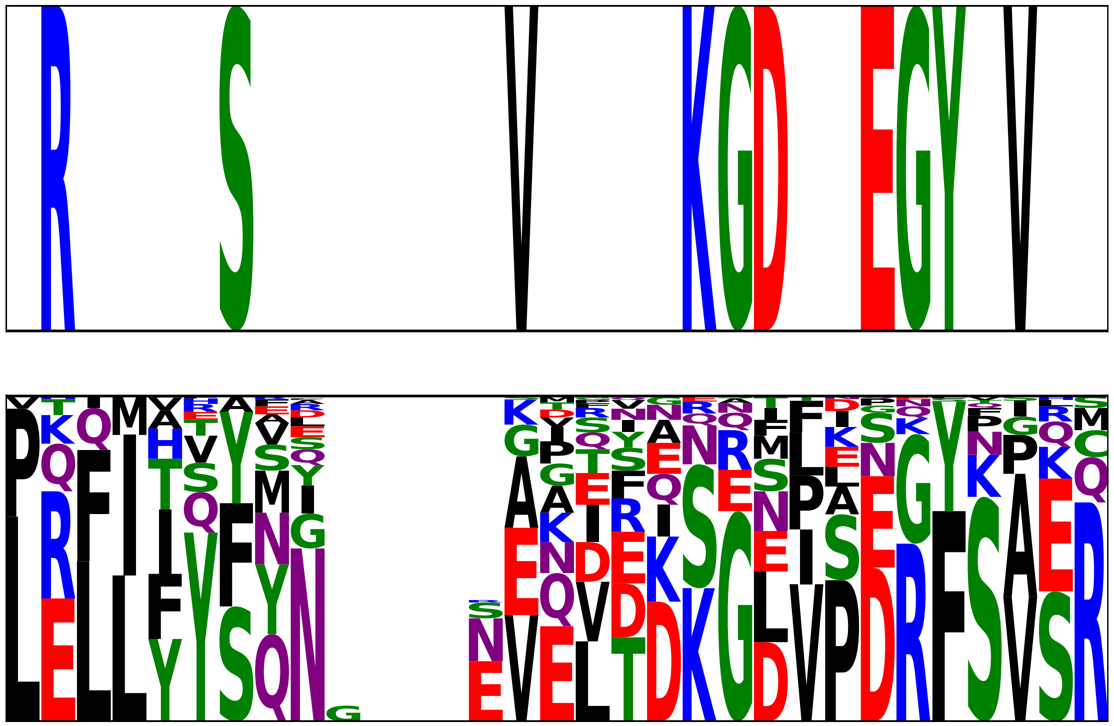
Calculate Anchors for each Medoid

Visualizations & Analysis of Rules

- Understand TCR binding rules
- Generate *if-then* rules

DECODE: Interpretable pipeline for TCR binding

Explanations can be visualized and evaluated



TITAN Anchor CDR2 Region

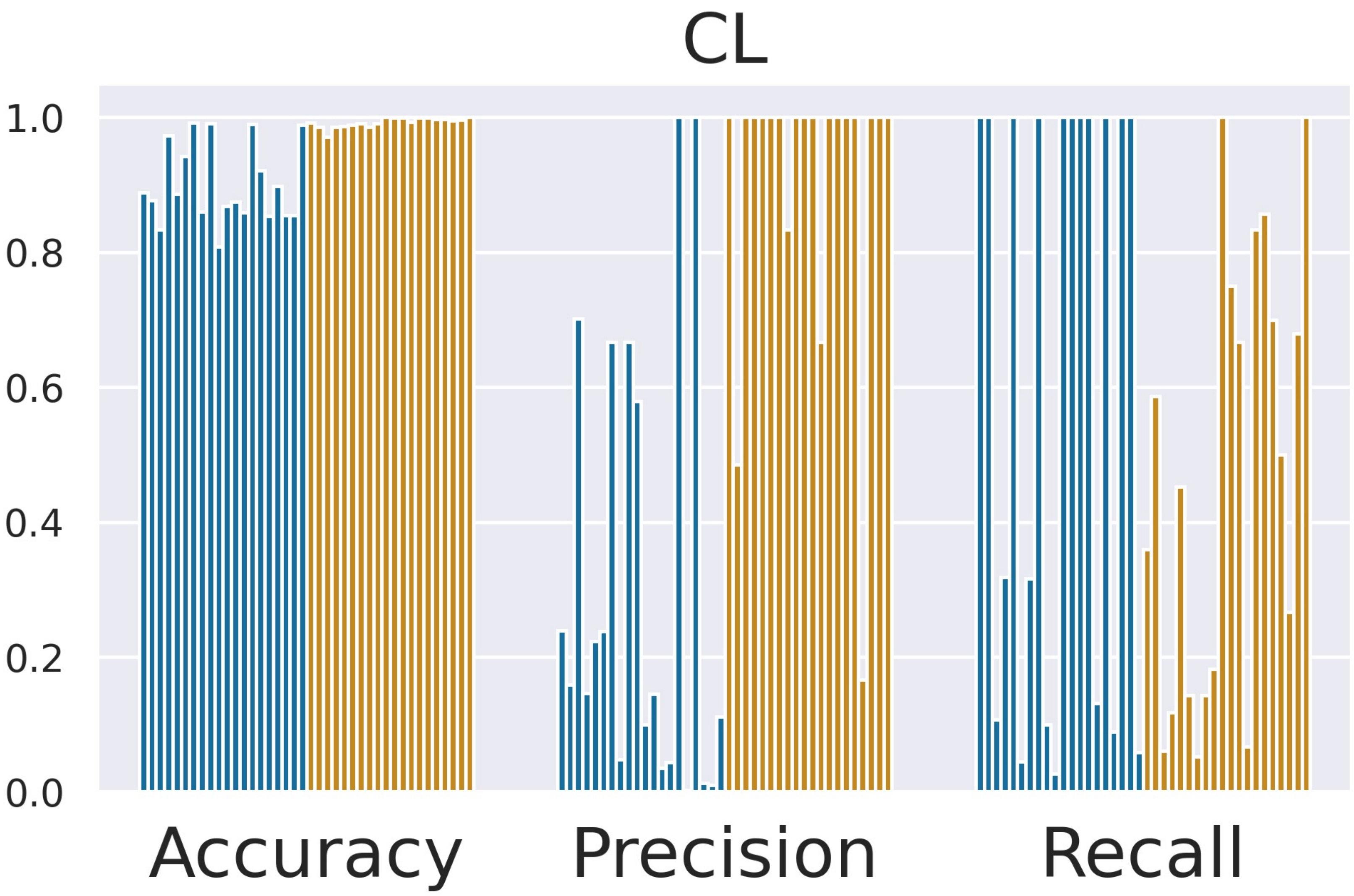
pMTnet Anchor CDR3 Region

Papadopoulou et al, Bioinformatics, 2022 (in press).

DECODE: Interpretable pipeline for TCR binding

Explanations can be visualized and evaluated:

→ Cluster level: a sample is a true positive if it fulfills *only* the anchor rule generated from the medoid of its cluster

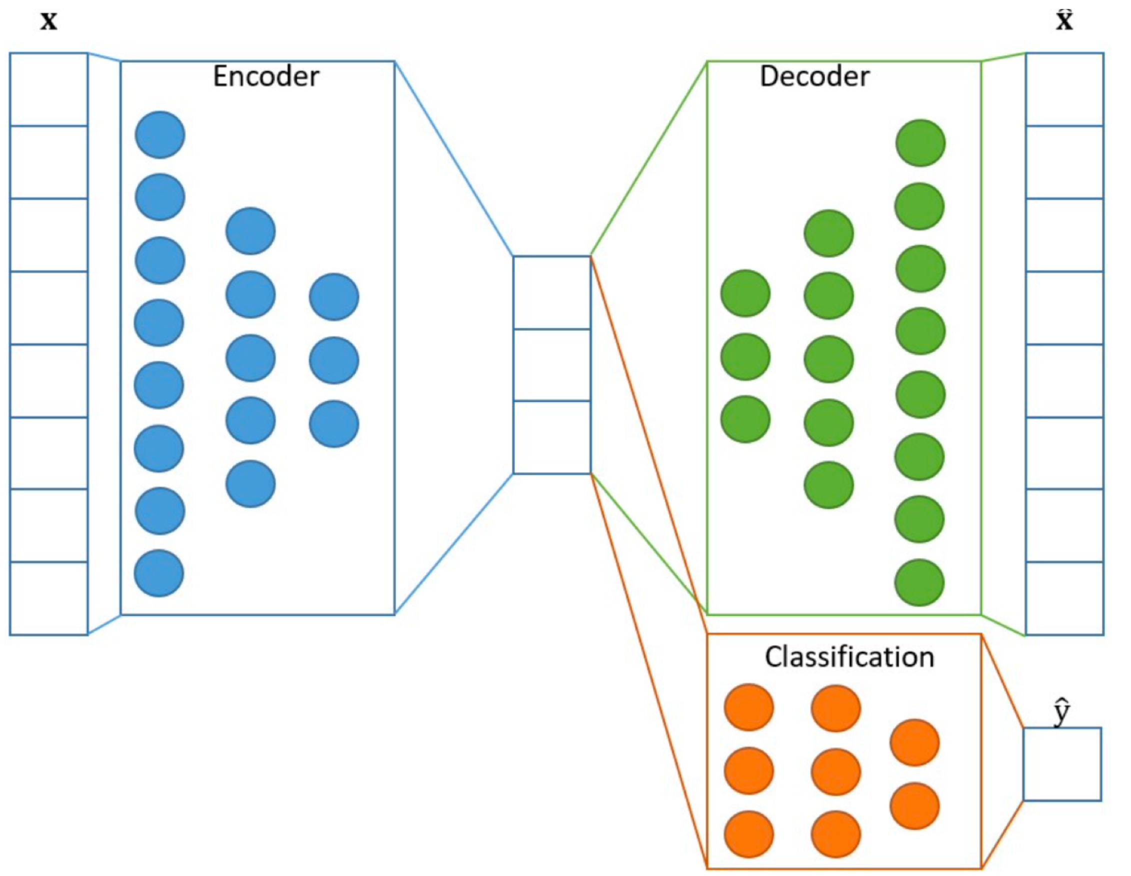


Blue – non-binding rules
Orange – binding rules

Ongoing work

Increasing the amount of epitope data

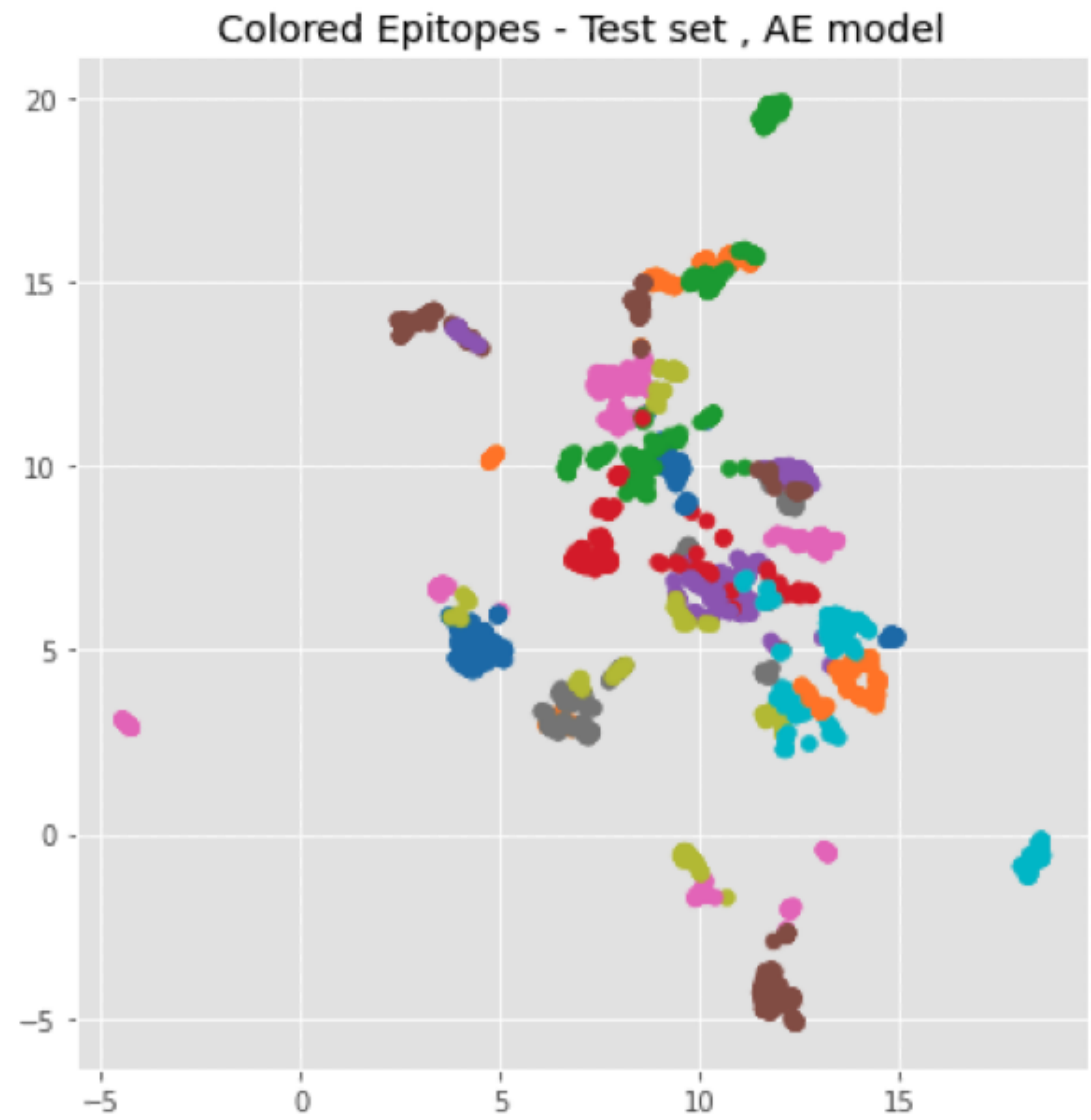
- Semi-supervised learning approaches
  - ➢ Exploit labelled and unlabelled data



Increasing the amount of epitope data

- Semi-supervised learning approaches
  - ➢ Exploit labelled and unlabelled data

- **Future work**: reinforcement learning for domain adaptation



Conclusions

- Goal
  - Predicting TCR binding is essential to understand the adaptive immune system and to engineer immune interventions
    - Optimization of cancer immunotherapies
- TITAN
  - Predicts the binding of TCRs and a small set of epitopes
    - Good accuracy when epitope is fixed
    - Interpretability through attention mechanisms
- DECODE
  - Interpretable pipeline to generate biochemical rules to explain TCR-epitope binding
- Future work
  - Predicting cross-reactivity when little epitope data
    - Semi-supervised learning approaches
    - Combination of molecular dynamics + AI

Thank you for your attention!

Systems Biology team
@ IBM Research
and
many external
collaborators.

FNSNF
Fonds national suisse
Schweizerischer Nationalfonds
Fondo nazionale svizzero
Swiss National Science Foundation

SystemsX.ch
The Swiss Initiative in Systems Biology

PrECISE*

IBM
University of Zurich UZH
ETH zürich
Technische Universität Darmstadt
Joint Research Center for Computational Biomedicine
RWTH Aachen University
institut Curie
TECHNIKON
Baylor College of Medicine
astridbio

European Commission
HORIZON 2020

European Union's Horizon 2020,
grant agreements:
No 668858, 826121, 765158,
813545.

@MariaRoCompBio