## **Munich Health Foundation Model Symposium**



Contribution ID: 16

Type: Poster

## **Unraveling the foundation of T cell receptors**

Deciphering the interaction between T-cell receptors (TCRs) and disease-derived epitopes has been declared as one of nine Cancer Grand Challenges in 2023. This interaction plays a pivotal role in the immune system's ability to recognize and eliminate infectious pathogens and cancerous cells. Sequence-based machine learning models for predicting TCR-epitope binding are constrained by limited data available in current databases. These currently contain less than 200,000 unique TCRs and typically only partial information, predominantly focusing on the  $\square$ -chain or even just its Complementarity Determining Region 3 (CDR3). However, recent studies have demonstrated the benefits of leveraging full TCR sequence information for accurate binding prediction.

This project proposes utilizing the ever-growing repertoire of TCR sequences from single-cell studies deposited in major public sequencing archives such as the Gene Expression Omnibus, the Sequence Read Archive, and the European Bioinformatics Institute. By querying these databases, we will gather millions of full TCR sequences with paired α-Ø-chains to train a language-model-based foundation model. To leverage the advantage of paired and full TCR sequences in downstream tasks, our methodology will be adapted to facilitate their reconstruction from partially available information.

By learning the foundations of TCRs, we envision a major advancement in TCR representation and generation, thereby, enabling a wide field of applications including the investigation of single-cell TCR studies, database queries, TCR-epitope prediction, scaffolding CDR3s to full sequences, and antigen-conditioned generation. Therefore, this project represents a major step towards developing efficient vaccines, facilitating candidate selection in immunotherapy, and eventually enabling fully personalized treatment through generative modeling of TCRs specific to tumor-derived neoantigens. By unlocking the potential of TCR sequence data, our foundation model will open new avenues for advancing immunological research and clinical applications.

**Primary authors:** Mr CHERNYSHEVA, Anna; SCHUBERT, Benjamin; DROST, Felix Rupert; Mr THOMAS, Julius

Session Classification: Break + Posters session

Track Classification: Poster