EBERHARD KARLS
UNIVERSITÄT
TÜBINGEN

Department of Computer Science
Chair for Biomedical Data Science
Quantitative Biology Center (QBiC)
Department for Peptide-based Immunotherapy

QBiC

Streamlining T-cell Receptor repertoire research by integrating competitive state-of-the-art tools in nf-core/airrflow

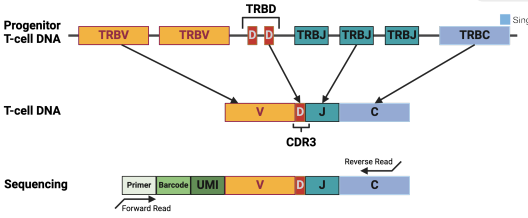
Mark Polster[1,2,3,4], Marissa Dubbelaar[1,2,4,5,6], Gisela Gabernet[7], Juliane S. Walz[2,5,6,8,9], Sven Nahnsen[1,3,4,10]

[1]Quantitative Biology Center (QBiC), University of Tübingen, Tübingen, Germany, [2]Department of Peptide-based Immunotherapy, University Hospital of Tübingen, Tübingen, Germany, [3]Department of Computer Science, University of Tübingen, Tübingen, Germany, [4]M3 Research Center, University Hospital of Tübingen, Tübingen, Germany, [5]Institute for Cell Biology, Department of Immunology, University of Tübingen, Germany, [6]Cluster of Excellence iFIT (EXC2180) "Image-Guided and Functionally Instructed Tumor Therapies", University of Tübingen, Tübingen, Germany, [7]Department of Pathology, Yale School of Medicine, New Haven, Connecticut, [8]Clinical Collaboration Unit Translational Immunology, Department of Internal Medicine, University Hospital Tübingen, Tübingen, Germany, [9]German Cancer Consortium (DKTK) and German Cancer Research Center (DKFZ), partner site Tübingen, Germany, [10]Institute for Bioinformatics and Medical Informatics (IBMI), University of Tübingen, Tübingen, Germany

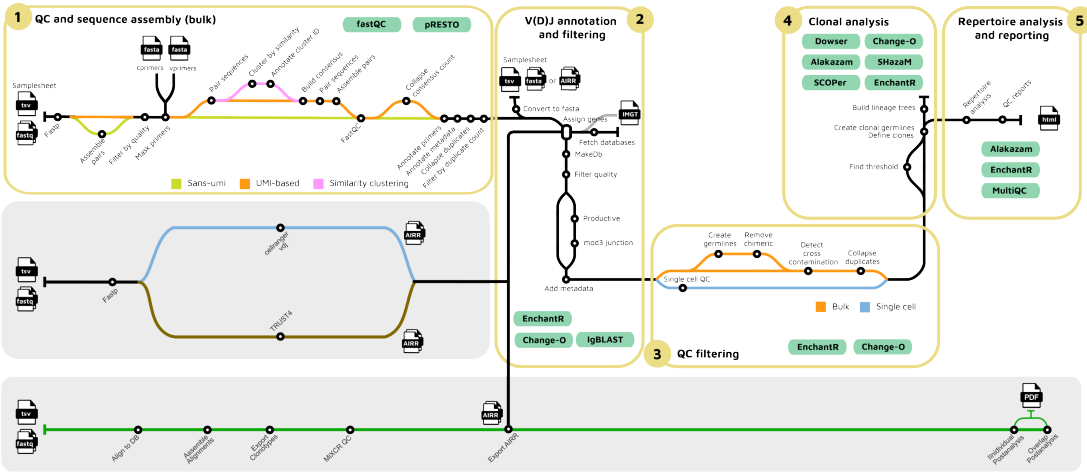
T-Cell Receptors

In our immune system, peptides are presented via the Major Histocompatibility Complex (MHC). T-cell Receptors (TCRs) can recognize and bind to this pMHC complex. Each person has a unique TCR repertoire that comprises 10^6-10^8 clonal groups that can recognize a wide array of potential antigens presented by the MHCs on antigen-presenting cells. After forming the TCR-pHLA complex, T-cells either destroy the infected cell or help triggering the immune response. Altogether, analyzing TCR repertoires can help to understand the immune response to antigens from various sources.

Due to the wide variety of TCR sequencing libraries and kits, harmonizing openly accessible datasets is challenging due to missing information (e.g., primers, barcodes, and unique molecular identifiers).



nf-core/airrflow novelties for TCR seq



Gabernet et al. have previously demonstrated the analysis of TCR and B-cell receptor (BCR) repertoires using the Immcantation framework[1] integrated in nf-core/airrflow. Three novel workflows have been added to promote the accessibility of open datasets as well as the flexibility of using various state-of-the-art tools:

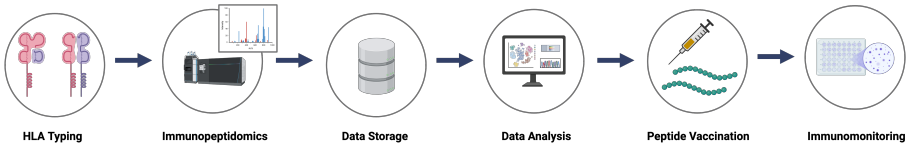
1) the cellranger sequence assembly workflow was integrated in the Immcantation framework to analyze raw single cell (sc) adaptive immune receptor repertoire (AIRR) sequencing files from 10XGenomics. 2) MiXCR allows the analysis of various commercial kits and library preparation protocols (sc and bulk) with ease. This will enable users to efficiently analyze their data and to integrate many publicly available TCR data into their research. In the end, this process includes additional postprocessing plots. 3) the integration of TRUST4 into the Immcantation framework allows users without specific TCR or BCR (sc and bulk) sequencing files to grasp receptors from unselected RNA-seq files.

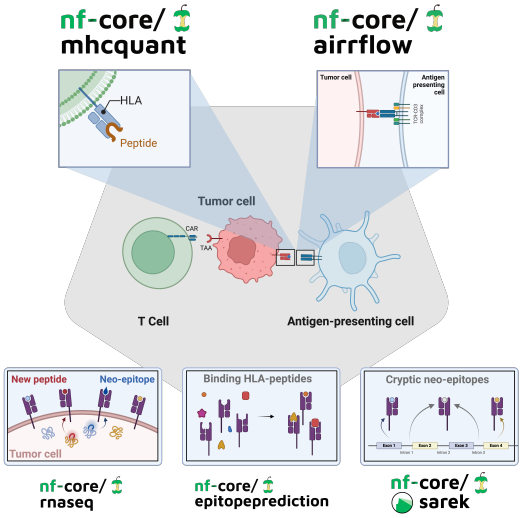
Peptide vaccination workflow

Many immunotherapies are based on the interaction between MHC-bound peptide antigens and T-cells, as seen in peptide cancer vaccinations, where synthesized peptides attract T-cells to cancer cells. Finding the optimal peptide vaccines for cancer patients involves a highly data-driven approach that combines substantial wet-lab analysis with big data consisting of genomics, transcriptomics, and immunopeptidomics data. The critical component of TCR sequencing data will allow immunologists and physicians to select suitable peptides and monitor patients' immune status before and after vaccination.

In the future, TCR sequencing might open up new possibilities for immunomonitoring, where traditional wet-lab techniques may be supported by using TCR-seq approaches to monitor TCR clones that bind vaccinated peptides over time, thus assisting in the assessment of the optimal timepoints for booster vaccinations. Also, streamlined integration of TCR sequencing data into a reproducible cancer vaccination workflow combined with label-free mass spectrometry-based immunopeptidomics data will be essential for better selecting the most promising peptides and understanding the process of TCR-pHLA binding.

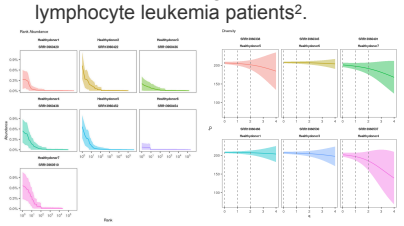
This study aims to enhance nf-core/airrflow by incorporating state-of-the-art tools to facilitate the integration of (publicly available) TCR sequence data, which can be used in multi-omics analyses.





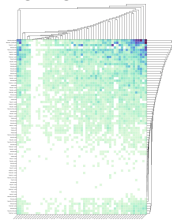
Application

TCR repertoire analysis was performed using nf-core/airrflow on a publicly available dataset of scRNA and scTCR sequencing of T-cells from large granular lymphocyte leukemia patients[2].



TCR diversity from 10X Genomics TCR-seq

TCR diversity from unselected RNA-seq files.



VJ combination barplot for an individual donor

Conclusion

- **nf-core/airrflow allows to analyze a great variety of publicly available data.**
- **More datatypes can be analyzed in a reproducible fashion using nf-core/airrflow.**
- **The combination of different state-of-the-art tools streamline TCR repertoire analysis.**
- **nf-core/airrflow will benefit analysis for peptide vaccination**

References

1. Gabernet et al. (2024), PLOS Computational Biology
2. Gao, S. et al. (2022), Nature Communications

Interested? Let's connect!

Acknowledgements

We would like to acknowledge funding under the German National Research Infrastructure for Immunology (NFDI4Immuno) [NFDI 49/1 - 501875662] by the DFG.