

Decoding Convergent Evolution in the Human T Cell Repertoire: A Systems Approach to Autoimmune Biomarkers

Supervisors:

Encarnita Mariotti-Ferrandiz, i3 Lab, Sorbonne Université / INSERM

Andreas Tiffeau-Mayer, Q-Immuno Lab, University College London

The adaptive immune system functions as a vast, distributed population of cells with unique receptors that evolve in real-time to recognize pathogens. This project addresses a fundamental interdisciplinary challenge in systems biology: how does a sequence space of astronomical diversity ($>10^{15}$ possible combinations) repeatedly converge on specific functional outcomes? In autoimmune diseases, this evolutionary process goes awry, yet identifying the specific T cell populations underlying disease remains a "needle in a haystack" problem in this context. Traditional approaches often fail because they look for identical sequences shared across patients. However, much like convergent evolution in nature, where different species evolve similar traits to adapt to the same environmental pressure, the immune response often involves "metaclones": distinct TCR sequences that converge on the same structural solution to recognize a disease-associated antigen. This project will combine population genetics principles with state-of-the-art machine learning to map these patterns of convergence. Using massive datasets from patients with autoimmune disorders (including rheumatoid arthritis, systemic lupus erythematosus, spondyloarthritis or osteoarthritis (non-autoimmune disease control) recruited as part of the TRANSIMMUNOM trial (NCT02466217), we will quantify how host factors like age, genetics, and cellular "fate" (Treg vs. Teff) constrain the available sequence space. Building on this statistical understanding of the sequence space, we will apply novel bioinformatics pipelines to identify HLA-restricted clusters of TCRs that serve as "evolutionary signatures" of disease. Finally, we will use contrastive learning to fine-tune protein language models to identify these signatures with high specificity in external validation cohorts. By shifting the focus from individual sequences to functional clusters, this project aims to deliver a new generation of interpretable, generalizable biomarkers for early diagnosis and therapeutic monitoring in autoimmunity.