Unravelling the dynamics and the specificity of the TCR repertoire of regulatory T cells in patients with autoimmune diseases treated with low-dose IL-2

Scientific background

T lymphocytes (TL) play a key role in the immune response. Despite a strong selection process occurring in the thymus, precluding from the generation of autoreactive effector T-cells, it is known that a fraction of cells escape that selection process (*1*), contributing to the development of autoimmune diseases (ADs). Regulatory T cells (Tregs) that are known to physiologically control autoreactive T-cells (*2*) have been shown to be quantitatively or qualitatively insufficient in ADs (*3*).

A key component of T-cells is their antigen-specific receptor, the T-cell receptor (TCR), composed of an α chain and a β chain (TCRαβ). Generated by a somatic rearrangement mechanism between multiple genes followed by random nucleotide addition/deletion, the TCRαβ forms a theoretical repertoire of up to 10^{19} unique TCRs with an estimation of the size of the actual circulating repertoire in humans around 10^{8} unique TCRs (*4*). The composition of this TCR repertoire varies with the environment, across time and in space, at the organism and population levels, mirroring each individual health status. Next generation sequencing (NGS) now offers the unprecedented depth to accurately study TCR repertoires. Therefore, the analysis of the TCR repertoire in diseases has the power to provide a better understanding of pathophysiology. **As a proof of concept of the power of machine learning to identify TCR signatures**, we studied TCR repertoires of CD4 Teffs and CD4 Tregs from patients with type-1 diabetes (T1D) or rheumatoid arthritis (RA), as well as healthy volunteers (HV). Combining sparse Partial Least Squares Discriminant Analysis (sPLSDA)(*5*) together with a classification strategy, we identified a TCRαβ signature that can classify i) patients compared to HV, and importantly ii) T1D versus RA patients, supporting the disease specificity of the signatures (Barennes et al, in preparation).

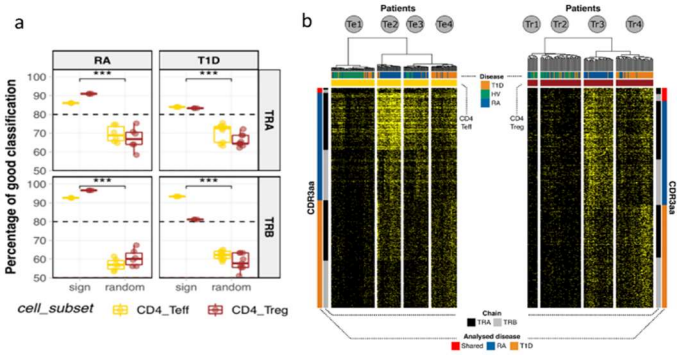

Figure 1 : **a. Classification power of Teff and Treg CDR3 signature**. Percentage of good classification. Each boxplot represents the mean of classification results over a Leave-One-Out strategy applied by disease (RA, left; T1D, right), cell subset (Treg, brown; Teff, yellow) and TCR chain (TRA, top; TRB, bottom) as compared with the HVs («sign») and with a random selection of CDR3s («random»)). **b. Signatures' clustering power.** Heatmaps represent each CDR3aa (x axis) present (yellow) and absent (black) from the Teff (left) and Treg (right) signatures in HV (green), T1D (orange) and RA (blue) patients (y axis). TRA (black) and TRB (gray) CDR3s are included. CDR3aa shared between groups are shown in red.

We now aim to evaluate the benefit of such approach to study the biological efficacy of treatments in AD. Indeed, more than 50% of AD patients are not responders to the first line treatment, and assessment of novel therapies often lack biomarkers of biological activity of the treatment. To address that question, we will take advantage of a dataset of TCR repertoires obtained as part of a double-blinded phase II trial during which systemic lupus erythematosus (SLE) patients have been treated with an immunotherapy targeting Tregs, IL-2 at low-doses (LUPIL-2, NCT02955615). SLE is a systemic autoimmune disease involving Tregs deficiency (*6*). IL-2 is the essential cytokine for the survival and suppressive function of Tregs (*7*). Clinical efficacy was observed in LUPIL-2 (*8*) ; notably, responders are characterized by an increased Tregs (numbers and marker of activation) (*9*). Blood Teffs and Tregs from treated and placebo patients have been collected and sequenced for their TCR offering the unique opportunity to search for TCR biomarkers of clinical response to IL-2.

Given the high diversity of the TCR repertoire, i) knowing the distribution laws that governs TCR repertoire generation and selection as well as ii) identifying antigen-specificity at the population level is required to envision TCR as a biomarker and/or a novel therapeutic target. In this line, the Greiff lab (Laboratory for Computational and Systems Immunology) developed innovative immune-receptor-adapted machine learning approaches for comparing, predicting and generating antigen-specific immune receptors(*10*, *11*). While progress in immune receptor machine learning has been made in

recent years, it remains unclear how to identify antigen-specific-specific sequence motifs and then link them to biological function. Importantly, improved antigen specificity prediction would allow the in silico annotation of publicly repertoires (>5 Billion immune receptors are now stored in public databases) and thus an even broader understanding of the prevalence of autoantigen-specific receptors in the human population in health and disease.

Aims of the program

Combining deep immunology, computational biology and in-vitro assays, the general objective of this project **is to study the dynamics and specificity of Tregs and Teffs TCRs obtained from SLE patients** included in the LUPIL-2 trial. The specific objectives of the project are to:

- characterize the TCR repertoire of the patients' Tregs and Teffs before and during treatment;
- identify disease and/or treatment specific TCR signatures;
- infer and validate the TCR specificities of these signatures.

Material and Methods:

LUPIL-2 trial*:* a Phase II, multicenter, randomized, double-blind clinical trial involving patients with moderate to severe SLE. Fifty patients received ld-IL2 and fifty received placebo over 6-monthes.

Samples*:* Blood samples were taken at five follow-up points: before treatment, at day 5 of treatment when the proportion of Treg is at its maximum, at mid-treatment (month 3), at the end of treatment (month 6) and 3 months after the end of treatment. The TCR sequences from Tregs and Teffs of 1000 samples have been generated by the i3 host laboratory using a benchmarked method (*12*) together with comparable TCR sequences obtained from several clinical trials evaluating ld-IL2 efficacy: MS-IL2 (NCT02424396), FACIL-2 (NCT03970954), TRANSREG (NCT01988506), HEALTHIL-2 (NCT03837093) as well as from 19 AD as part of the Transimmunom trial (NCT02466217). *All the data are available*.

Repertoire exploration*:* Diversity, VJ gene usage, clonotype frequency and other criteria will be used to characterize patient TCR repertoires using an Adaptive Immune Receptor Repertoire sequencing (AIRR-seq) data analysis strategy available in both labs.

Antigen specificity signature*:* Various pipelines already implemented in the i3 lab will allow the identification of disease-specific TCRs.

Inference of specificity*:* An important question in this study is the identification of the specificity of the TCRs of interest. The Greiff Systems immunology host lab developed an ecosystem to address that challenging question (Immune ML, (*13*)). Specifically, they developed an approach to predict antibody-epitope binding, which led to the first identification of a common vocabulary (or rules) for antibody specificity(*14*). As part of the project, the candidate will adapt this strategy to predicting TCR specificity that includes the need to predict the interaction of TCR with a peptide forming a complex with MHC molecules. Specifically, we will take advantage of existing TCR-epitope sequencing and structural data as well as the generated antigen specificity data in order to develop ML methods for TCR-epitope prediction that both incorporate sequence and structure. We will then apply these methods to entire TCR repertoire to understand the frequency of autoantigen-specific sequences in health and disease.

In-vitro screening*:* Inferred specificities will be tested and validated by in vitro assays established in the i3 host-lab based on a published method (*15*). Briefly, core TCRs with identified specificities will be engineered into lentivirus constructs that express a reporter gene for transduction efficacy evaluation. These lentiviruses will be used to transduce 5KC T-hybridoma cells expressing an NFAT-driven fluorescent reporter ZsGreen-1 along with human CD8/CD4. Transduced cells are incubated with increasing concentration of antigens presented by K562 cells transduced with lentiviral vectors encoding appropriate HLA Class-II molecules, the fluorescent activity is measured. We anticipate 50 core TCRs for 10 to 20 antigens.

Expected results

The analysis strategy will characterize the TCR repertoire of SLE patients, during the course of the treatment and in comparison with other ADs. The identification of SLE specific signatures as well as its dynamics and relation with the response to treatment will provide clinico-biological correlates and biomarkers of response to treatment. Specificity inference and epitope prediction will serves as a platform to test and validate in-vitro the predicted specificities and provide new therapeutic targets.

Candidate profile:

The expected candidate will have a training in biology and/or bioinformatics with strong interest for Systems Biology and immunology. The candidate will contribute to the modelling of this data collection using the available workflows developed in both laboratories as well as implementing new strategies. In-vitro experiment will be covered by the student depending on her/his profile, with help from i3 staff.

Co-direction laboratory description, complementarity, interdiciplinarityand funding:

The i3 laboratory in France and the Laboratory for Computational and Systems Immunology in Norway are both leaders in the field of AIRR, **yet with very complementary and interdisciplinary skills:**

- **The i3 host laboratory** has more than 20 years history of translational research, focusing on immunology and autoimmune diseases, with an established expertise in wet-lab experiment, including in-vitro models. **Encarnita Mariotti-Ferrandiz (EMF), the PI of the project**, is an immunologist from Sorbonne Université, with > 20 years' experience studying the TCR repertoire, from wet-lab method to bioinformatics tool development. She is a senior member of the Institut Universitaire de France (IUF 2022-2027). She obtained national and European grants (ANR, H2020) and published in high-profile journals, together with the PhD students she supervised (see ORCID: 0000-0002-8770-0186).

- **The Victor Greiff (VG)-led Systems immunology** laboratory masters machine learning approaches applied to the study of AIRR. With more than 15 years' experience, Victor Greiff has developed experimental and computational technology for mapping the interaction rules of antibody-antigen binding. VG has an outstanding publication records in high-profile publications and research funding (two EU grants (H2020, IMI), Research Council of Norway (see ORCID:0000-0003-2622-5032). VG is also an industry advisor for >10 biotech startups (e.g., immunai, Absci, LabGenius) and word-leading pharma companies (Roche/Genentech), a unique asset for trainees.

Complementarity: While EMF has focused on the understanding of TCR-based autoimmunity as well as the benchmarking and development of novel experimental assays for unbiased TCR repertoire profiling, Greiff has mostly focused on the development of machine learning approaches for antibody binding rule prediction. Here, **both groups combine their expertise to profile TCR specificity of selected T cell subpopulations to identify novel TCR-specific rules that drive autoimmunity**.

Interdiciplinarity: Both research groups are inherently interdisciplinary (TCR and antibody, experimental and computational method development, machine learning, fundamental immunology). Therefore, their combination only increases the degree of interdiciplinarity of this proposal.

Both laboratories are offering a unique interdisciplinary environment, with biologists, immunologists, clinicians, computer scientists and bioinformaticians. The candidate will be based half of the time in the i3 laboratory located on the Pitié-Salpêtrière hospital campus in Paris (France) and half of the time in the Computational and Systems immunology in Oslo (Norway).

Funding: The project is fully funded as the required data are available. The in-vitro experimental part will be covered by the IUF funding (EMF). Housing support is provided by the UiO.

References

1. F. P. Legoux *et al.*, *Immunity*. **43**, 896–908 (2015).
2. S. Sakaguchi, M. Miyara, C. M. Costantino, D. A. Hafler, *Nat Rev Immunol.* **10**, 490–500 (2010).
3. M. Noack, P. Miossec, *Autoimmun Rev*. **13**, 668–77 (2014).
4. P. Bradley, P. G. Thomas, *Annu. Rev. Immunol.* **37**, 547–570 (2019).
5. K.-A. L. Cao *et al.*, *mixOmics: Omics Data Integration Project* (2017; https://cran.r-project.org/web/packages/mixOmics/index.html).
6. S. Mellor-Pita *et al.*, *Ann. Rheum. Dis.* **65**, 553–554 (2006).
7. D. Klatzmann, A. K. Abbas, *Nat. Rev. Immunol.* **15**, 283–294 (2015).
8. M. Rosenzwajg *et al.*, *J. Autoimmun.* **58**, 48–58 (2015).
9. J. Y. Humrich *et al.*, *Ann. Rheum. Dis.* **81**, 1685–1694 (2022).
10. V. Greiff, G. Yaari, L. G. Cowell, *Curr. Opin. Syst. Biol.* **24**, 109–119 (2020).
11. C. R. Weber *et al.*, *Bioinformatics*, doi:10.1093/bioinformatics/btaa158.
12. P. Barennes *et al.*, *Nat. Biotechnol.*, 1–10 (2020).
13. M. Pavlović *et al.*, *Nat. Mach. Intell.* **3**, 936–944 (2021).
14. R. Akbar *et al.*, *Cell Rep.* **34** (2021), doi:10.1016/j.celrep.2021.108856.
15. S. E. Mann *et al.*, *Front. Immunol.* **11**, 633 (2020).