

## EDITORIAL



# Structural constraints in T-cell repertoire selection predicted by machine learning

© The Author(s), under exclusive licence to Springer Nature Limited 2021

*Genes & Immunity* (2021) 22:203–204; <https://doi.org/10.1038/s41435-021-00147-3>

The T-cell receptor (TCR) repertoire of the peripheral T cells is sculpted from stochastically generated TCRs on developing thymocytes in the thymus by a series of selection steps that delete cells bearing TCRs not recognizing self-peptide–MHC (pMHC) complexes or recognizing them with high affinity, known as positive selection or negative selection. There are two models to explain how TCR signals determine selection outcome. In the threshold model, thymocytes bearing TCRs that signal above a negative selection threshold will be deleted, while T cells experiencing low to intermediate TCR signaling strengths will survive through positive selection. In the sustained signaling model, high-affinity and low-affinity interactions between TCRs and pMHC complexes trigger biochemically different signaling cascades; low-affinity TCRs induce sustained signaling while TCR signaling after high-affinity stimulation is intense but short [1]. Little is known about how the TCR sequence determines the outcomes of selection. In the current issue of *genes & immunity*, Ostmeier et al. [2] develop an approach to identify how TCR protein sequences influence the selection fate using machine learning.

A large number of functional T lymphocytes in the periphery contain both productive and non-productive TCR genes [3]. The authors use those productive and non-productive TCRB genes from mature T cells to define unselected and selected repertoires, assuming that the sequence of the non-productive TCR protein is closely related to a non-selected TCR. Thus, the authors develop an algorithm to computationally repair non-productive TCR genes to obtain productive copies with the fewest alteration, which maximally preserve the original biological sequences. This approach allows to exclude known biases from VDJ recombination in the unselected repertoire from the model. Moreover, it does not rely on obtaining the repertoire of thymocytes expressing solely the TCRB gene, giving the approach potentially broader applicability. The authors used both sets of TCR protein sequences to train a machine-learning model. The model returns a probability of  $P_{\text{SURVIVE}}$  to any TCRB sequence; with  $P_{\text{SURVIVE}} > 0.5$ , the TCRB gene predicted to be a productive one and  $P_{\text{SURVIVE}} < 0.5$ , the TCRB gene predicted to be repaired.

To test the model, the authors use TCR genes from developing thymocytes, which include pre-selected and post-selected thymocytes. Distribution of  $P_{\text{SURVIVE}}$  is bimodal for thymic TCRB genes. In contrast, the distribution of productive TCRB genes from splenocytes was consistent with all splenocytes having survived selection. Similar unimodal were obtained for TCRB genes from blood and colon. This approach is therefore very powerful to identify pre-selected or post-selected TCR genes.

The authors propose that their approach might find applications in personalized medicine by predicting a T-cell repertoire prone to increased autoreactivity, eventually resulting in an autoimmune disease. Indeed, central tolerance is a key checkpoint and defects in positive and negative selection in animal models

have been shown to cause autoimmunity. With the exception of AIRE mutations causing the autoimmune polyendocrinopathy-candidiasis-ectodermal dystrophy or APECED, there is little evidence for a role of negative selection in human disease [4]. However, more complex mechanisms in thymic selection rather than the simple failed deletion of a peptide-specific autoreactive T cells may contribute to autoimmune diseases such as Type 1 diabetes, multiple sclerosis, and rheumatoid arthritis [5, 6]. There has been a decade-long, ongoing debate on the role of positive selection, for example, in HLA-DRB1 homozygosity conferring more severe disease in rheumatoid arthritis [7]. The rheumatoid arthritis-like manifestations in the SKG mice, which carry a mutation in the ZAP70 signaling molecule, also appear to be a consequence of faulty positive selection [8].

The approach described in this paper provides a composite assessment of positive and negative selection, and it cannot easily be envisioned how these two processes can be separated. Whether both processes are equally represented in the algorithm remains to be clarified. It appears to be likely that most of the algorithm is driven by positive selection, namely to describe whether the two proteins of TCR and MHC fit to each other based on their 3-D structure. This fitting may go beyond MHC differences, i.e., there may be universal structural principles. If the genetic predisposition of HLA polymorphisms to autoimmune disease comes from a better HLA-TCR fit, such as described above, the tool developed here may be very informative. It appears less likely that the algorithm can predict failure in negative selection, the importance of which has been mainly concluded from genetic manipulation in murine models such as the K/BxN strain but is less evident for spontaneous human disease.

In summary, reconstituting TCR selection in patients in-silico by using PBMCs will be an important tool for understanding TCR selection processes in human autoimmune diseases. So far, our test armamentarium to examining and quantifying disease processes has been very limited. The approach developed here may eventually provide valuable diagnostic information. Whether and how these insights will lead to new preventive and therapeutic interventions remains to be seen.

Wenqiang Cao <sup>1,2</sup>✉ and Jörg J. Goronzy <sup>1,2</sup>✉

<sup>1</sup>Division of Immunology and Rheumatology, Department of Medicine, Stanford University, Stanford, CA, USA. <sup>2</sup>Department of Medicine, Palo Alto Veterans Administration Healthcare System, Palo Alto, CA, USA. ✉email: [wenqcao@stanford.edu](mailto:wenqcao@stanford.edu); [jgoronzy@stanford.edu](mailto:jgoronzy@stanford.edu)

## REFERENCES

1. Moran AE, Hogquist KA. T-cell receptor affinity in thymic development. *Immunology* 2012;135:261–7.
2. Jared O, Lindsay C, Benjamin G, Scott C. Reconstituting T cell receptor selection in-silico. *Genes Immun.* 2021;in press.

Received: 4 June 2021 Revised: 15 June 2021 Accepted: 29 June 2021  
Published online: 12 July 2021

3. Li S, Wilkinson MF. Nonsense surveillance in lymphocytes? *Immunity* 1998;8:135–41.
4. Cheng M, Anderson MS. Thymic tolerance as a key brake on autoimmunity. *Nat Immunol.* 2018;19:659–64.
5. Bluestone JA, Bour-Jordan H, Cheng M, Anderson M. T cells in the control of organ-specific autoimmunity. *J Clin Investig.* 2015;125:2250–60.
6. Walscher-Kuntz DR, Weyand CM, Weaver AJ, O'Fallon WM, Goronzy JJ. Mechanisms underlying the formation of the T cell receptor repertoire in rheumatoid arthritis. *Immunity* 1995;2:597–605.
7. Goronzy JJ, Weyand CM. Developments in the scientific understanding of rheumatoid arthritis. *Arthritis Res Ther.* 2009;11:249.
8. Sakaguchi N, Takahashi T, Hata H, Nomura T, Tagami T, Yamazaki S, et al. Altered thymic T-cell selection due to a mutation of the ZAP-70 gene causes autoimmune arthritis in mice. *Nature.* 2003;426:454–60.

### COMPETING INTERESTS

The authors declare no competing interests.

### ADDITIONAL INFORMATION

**Correspondence** and requests for materials should be addressed to W.C. or J.J.G.

**Reprints and permission information** is available at <http://www.nature.com/reprints>

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.