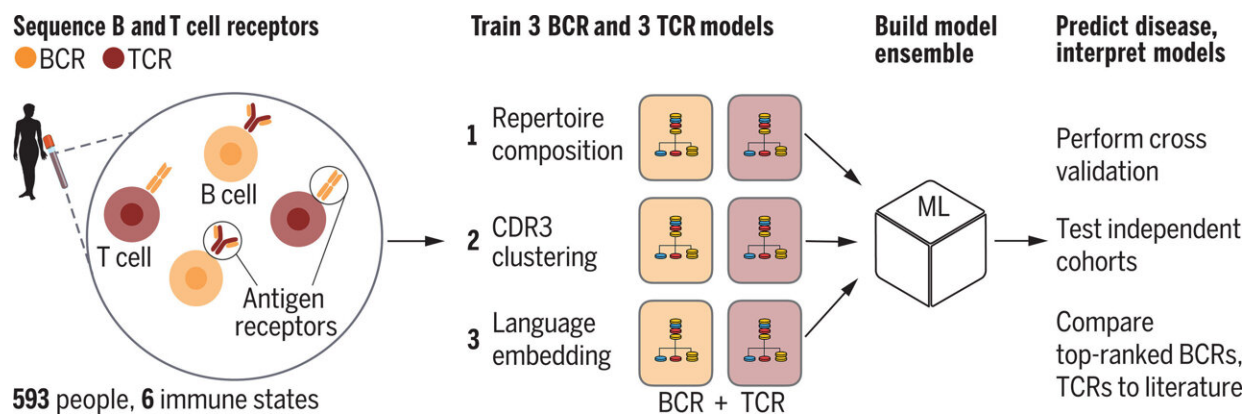


Machine learning tool decodes immune receptor sequences to diagnose multiple diseases

February 24 2025, by Justin Jackson



From blood to disease classification with immune receptor sequencing. Credit: *Science* (2025). DOI: 10.1126/science.adp2407

Stanford University researchers developed a machine learning-based method capable of diagnosing multiple diseases using B cell and T cell receptor sequences. The model, called Machine learning for Immunological Diagnosis (Mal-ID), distinguished between COVID-19, HIV, lupus, type 1 diabetes, influenza vaccination response, and healthy states, achieving near-perfect classification.

Conventional diagnostics rely on patient history, physical examinations, and [laboratory tests](#), often requiring multiple rounds to diagnose

complex diseases like autoimmune conditions.

B cell receptors (BCRs) and T cell receptors (TCRs) are generated through random recombination processes and change after infections, vaccinations, or in [autoimmune diseases](#), offering potential as biomarkers for immune activity. Leveraging receptor sequence data could allow simultaneous assessment of various diseases.

In the study, "Disease diagnostics using [machine learning](#) of B cell and T cell receptor sequences," [published](#) in *Science*, researchers analyzed BCR heavy chain and TCR beta chain sequences from 593 individuals.

Participants included 63 with COVID-19, 95 with HIV, 86 with lupus, 92 with type 1 diabetes, 37 who received influenza vaccination, and 220 healthy controls. Paired BCR and TCR data were available for 542 individuals.

Mal-ID correctly classified immune status from [blood samples](#) of 542 individuals with both BCR and TCR data. High classification performance was achieved with BCR data alone with an area under the receiver operating characteristic curve (AUROC) of 0.959 in the full 593 cohort.

Lupus was accurately distinguished from other conditions with 93% sensitivity and 90% specificity. External datasets validated the model's generalizability, achieving up to 1.0 AUROC on independent BCR cohorts and 0.99 AUROC on TCR cohorts after threshold adjustments.

Results showed that combined B and T cell analyses outperformed single-locus approaches. Some immunoglobulin heavy-chain V genes were associated with [viral infections](#) or autoimmune status, aligning with existing immunological knowledge.

SARS-CoV-2-specific BCR sequences from external databases received higher COVID-19 association scores compared to healthy controls. Minimal and non-significant batch effects or demographic factors such as age, sex, or ancestry influenced classification performance.

Mal-ID uses three models per receptor type: 1) repertoire composition assessing gene segment usage and somatic hypermutation rates, 2) clustering of complementarity-determining region 3 (CDR3) sequences to identify disease-associated patterns, and 3) embeddings from protein language models to capture structural similarities. An ensemble model integrated these approaches to predict disease states.

Results indicate that immune receptor sequencing can potentially serve as a versatile diagnostic tool for a range of infections, autoimmune conditions, and vaccine responses. Future efforts may validate its broader clinical potential.

More information: Maxim E. Zaslavsky et al, Disease diagnostics using machine learning of B cell and T cell receptor sequences, *Science* (2025). [DOI: 10.1126/science.adp2407](https://doi.org/10.1126/science.adp2407)

© 2025 Science X Network

Citation: Machine learning tool decodes immune receptor sequences to diagnose multiple diseases (2025, February 24) retrieved 25 February 2025 from <https://medicalxpress.com/news/2025-02-machine-tool-decodes-immune-receptor.html>

This document is subject to copyright. Apart from any fair dealing for the purpose of private study or research, no part may be reproduced without the written permission. The content is provided for information purposes only.