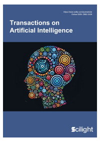
Scilight

Transactions on Artificial Intelligence
https://www.sciltp.com/journals/tai



Review

T-Cell Receptor Repertoire in Autoimmune Diseases and Their Machine Learning-Based Prediction Analysis

Tongfei Shen, Miaozhe Huo and Shuaicheng Li *

Department of Computer Science, City University of Hong Kong, Kowloon, Hong Kong, China
* Correspondence: shuaicli@cityu.edu.hk

How To Cite: Shen, T.; Huo, M.; Li, S. T-cell Receptor Repertoire in Autoimmune Diseases and Their Machine Learning-Based Prediction Analysis. *Transactions on Artificial Intelligence* **2026**, *2*(1), 78–102. https://doi.org/10.53941/tai.2026.100006

Received: 11 January 2026
Revised: 3 January 2026
Accepted: 28 February 2026
Published: 12 Mach 2026

Abstract: The T-cell receptor (TCR) is a fundamental component of the adaptive immune system, playing a crucial role in the development and progression of autoimmune diseases through its remarkable diversity and antigen specificity. Advances in high-throughput sequencing technologies and multi-omics data integration have revolutionized the ability to characterize TCR repertoires at unprecedented resolution. Coupled with emerging machine learning methodologies, these advances have opened new avenues for unraveling the complex immunopathology underlying autoimmune disorders. This review comprehensively summarizes current knowledge on the dynamic regulation of TCR repertoires in autoimmune diseases, highlighting key processes such as central tolerance failure, clonal expansion of autoreactive T cells, and regulatory T cell dysfunction, as well as the influences of genetic predisposition and immunosenescence on shaping TCR diversity. This review also provides a 3 that demonstrates how to analyze publicly available TCR repertoire datasets. We compare V and J gene usage profiles and CDR3 summary features across clinical labels to characterize between-group variation and to inform feature engineering for downstream machine learning models. Furthermore, we detail various machine learning-based diagnostic models that utilize gene usage patterns and CDR3 sequence features to accurately classify autoimmune disease status, alongside recent breakthroughs in predicting TCR-epitope binding specificity. These computational approaches not only enhance diagnostic precision but also provide mechanistic insights into immune recognition and autoreactivity. By integrating immunological principles with data-driven techniques, this work aims to offer a robust theoretical framework and practical guidance for future research in immunology and precision medicine. Ultimately, the convergence of TCR repertoire profiling and machine learning promises to drive innovative strategies for early diagnosis, personalized therapy, and improved clinical management of autoimmune diseases, enabling the transition to antigen-specific tolerogenic therapies.

Keywords: T-cell receptor; autoimmune diseases (ADs); systemic lupus erythematosus (SLE); machine learning model; immune repertoire

1. Introduction

T-cell receptors (TCRs) are membrane proteins expressed on T lymphocytes that mediate antigen specificity by recognizing peptide–MHC complexes [1–4]. The TCR-MHC–peptide interaction initiates downstream signaling pathways in T cells through molecular components that include enzymes and adapter proteins, processes which appear to have considerable implications for human disease [1,2,4–6]. The assembly of TCRs with CD3 chains on the cell surface forms a signal transduction complex, and differences in the molecular architecture of these complexes are thought to underlie the distinct antigen recognition and signaling properties of $\alpha\beta$ T cells versus $\gamma\delta$ T cells [4–6]. Each receptor comprises two chains, TCRα and TCRβ, generated via somatic V(D)J recombination.

Among the variable segments, the complementarity-determining region 3 (CDR3) serves as the primary contact site with antigen and is considered a major determinant of recognition specificity [1–4,7–9]. Through recognition of antigenic peptides presented within major histocompatibility complex MHC grooves, TCRs orchestrate adaptive immune responses that contribute to pathogen clearance, tumor surveillance, and the regulation of self-tolerance [10–14]. Consequently, scalable analysis of TCR repertoires, especially when integrated with machine learning approaches, has become an important method for assessing immune status and investigating functional immunological relationships [15–21].

Autoimmune diseases (ADs) constitute a heterogeneous group of chronic disorders in which immune responses are misdirected against host tissues, producing variable degrees of inflammation and organ dysfunction. Their etiology reflects a multifactorial interplay among genetic susceptibility, environmental exposures, and failures of central or peripheral tolerance; clinically they span systemic entities such as systemic lupus erythematosus (SLE) and organ-restricted diseases including rheumatoid arthritis (RA) and multiple sclerosis (MS) [22–24]. Therapeutic regimens remain centered on immunosuppression and immune modulation, yet mounting evidence implicates alterations in the TCR repertoire—loss of diversity and selective clonal expansions of autoreactive T cells—as central contributors to pathogenesis in many ADs [25–27]. The advent of high-throughput sequencing and multi-omics platforms has substantially refined our capacity to profile these repertoire perturbations, and structural and functional analyses of TCRs have in turn informed biomarker discovery and precision therapeutic design [27–29].

Machine learning (ML) comprises algorithmic approaches that infer predictive relationships from complex, high-dimensional datasets, and has shown considerable promise across biomedical applications, including the diagnosis and prognosis of autoimmune conditions [30,31]. For example, proteomic datasets analyzed with ML produced diagnostic classifiers for SLE with reported accuracies of 78.1%, 85.8%, and 90.0% across successive models [32]; in a separate application, an XGBoost algorithm was trained to predict relapse of lupus nephritis, with similar 5-year recurrence estimates in derivation and validation cohorts [33]. While earlier work has utilized clinical, imaging, and spectroscopic inputs, integrating features derived from TCR repertoires into ML pipelines is a comparatively recent development. Over the past five years, researchers have increasingly exploited repertoire differences between healthy donors and patients to build ML classifiers for autoimmune phenotypes [34, 35]. Although this subfield remains emergent and a comprehensive systematic review is not yet available, TCR-based ML approaches are attractive because they can provide efficient, reproducible, and less resource-intensive predictive tools relative to extensive clinical workups [36–38].

This review provides a structured account of machine learning–driven TCR prediction in autoimmune disease research, integrating basic immunology with applied computational practice. We first describe the immunobiological principles of TCR recognition and the mechanisms by which repertoire abnormalities may foster autoimmunity. We then synthesize empirical comparisons of TCR sequencing profiles from patients and matched controls, and present exploratory analyses of publicly available datasets that illustrate feature extraction and model construction. Next, we survey representative ML models that employ diverse algorithms and feature representations to classify autoimmune states. Finally, we discuss methodological challenges, potential trajectories for the field, and opportunities for combining deep learning with systems-level data to yield mechanistic insight and to accelerate translational advances aimed at improving clinical outcomes.

2. Dynamic Regulation of TCR Repertoires and Their Roles in Autoimmune Pathogenesis

2.1. Normal Function: TCR Antigen Recognition and Autoimmunity Prevention

TCRs are generated by stochastic V(D)J recombination with junctional diversity and thymic selection, producing vast specificity for peptide–MHC complexes that defends against pathogens and tumours but inevitably yields some self-reactive specificities [5,6,39–41]. The heterodimeric receptor (α/β chains) uses V regions and a recombined CDR3 to determine peptide specificity, while CD3 and CD4/CD8 modulate signalling (Figure 1A). Central tolerance—successive positive and negative selection and diversion into regulatory lineages—removes or redirects strongly self-reactive thymocytes and thereby limits peripheral export of high-affinity autoreactive clones (Figure 1B) [42]. When central tolerance fails, self-reactive clones survive, clonally expand, show skewed V/J usage and reduced diversity; peripheral inflammation and homeostatic defects further select and maintain these pathogenic expansions, cumulatively fostering autoimmune-prone repertoires [22,40,43,44].

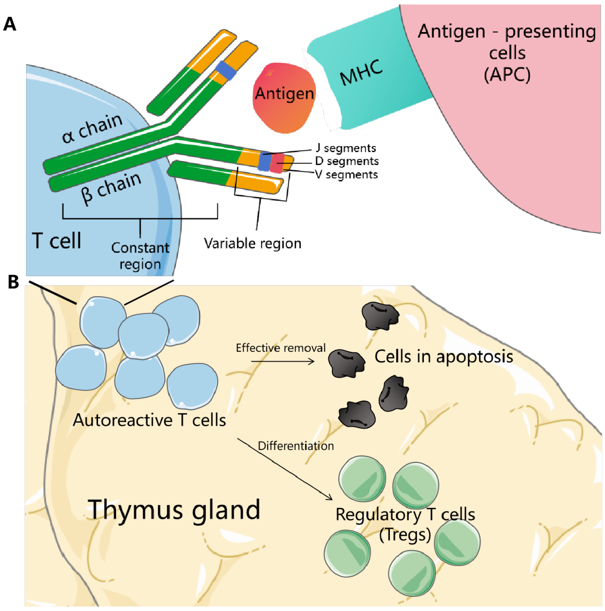


Figure 1. TCR architecture and thymic central tolerance mechanisms. (**A**) Structural features of the T cell receptor (TCR): α and β chains each have variable (V) and constant (C) regions, with CDR3 formed by V-(D)-J recombination determining antigen specificity. TCR V regions bind antigen-MHC complexes on antigen-presenting cells (APCs), while CD3 transduces activation signals and CD4/CD8 co-receptors enhance T cell activation. (**B**) Thymic central tolerance enforces self-tolerance by deleting self-reactive T cells via positive and negative selection or promoting their differentiation into regulatory T cells (Tregs), preventing autoreactive cells from entering peripheral circulation.

2.2. Autoimmunity Mechanism I: Central Tolerance Failure and Defective Thymic Selection

mTECs present tissue-restricted antigens (TRAs), many induced stochastically by *AIRE* interacting with chromatin and elongation machinery, enabling negative selection or Treg diversion that preserves tolerance [45–49]. Mutations in *AIRE*, *FOXP3* or NF-κB components disrupt these outcomes, permitting survival of self-reactive thymocytes [50]. Defective selection yields patient repertoires with biased V–J usage and shortened CDR3s—contracted diversity that spares pathogenic specificities [40, 51]. Escaped clones undergo peripheral expansion after triggers (molecular mimicry, infection, inflammation), producing oligoclonality and dominance by a few clones (Figure 2A); experimental studies link imp*AIRE*d mTEC presentation, reduced *AIRE* or IKKα mutations to inefficient deletion and constrained, pathogenic peripheral repertoires [52–54].

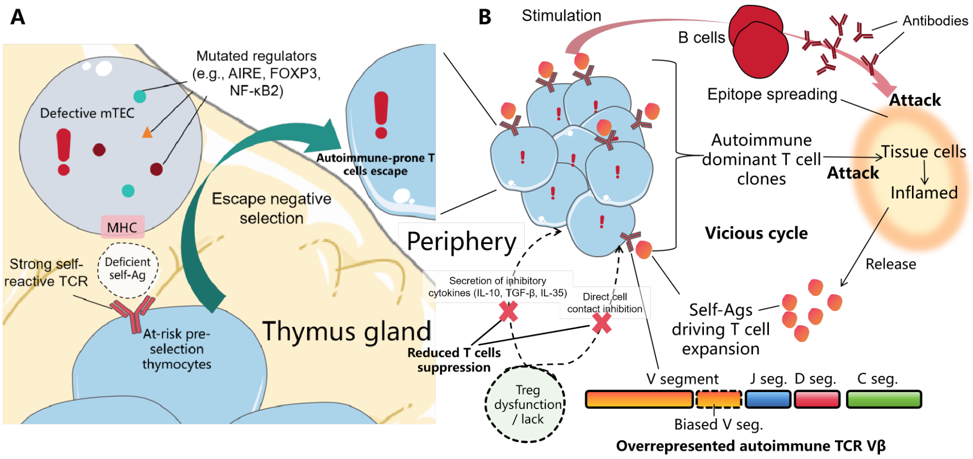

A
Mutated regulators (e.g., AIRE, FOXP3, NF-κB2)
Defective mTEC
MHC
Strong self-reactive TCR
Deficient self-Ag
At-risk pre-selection thymocytes
Escape negative selection
Autoimmune-prone T cells escape
Periphery
Thymus gland
B
Stimulation
B cells
Antibodies
Epitope spreading
Attack
Autoimmune dominant T cell clones
Tissue cells
Attack
Inflamed
Vicious cycle
Release
Secretion of inhibitory cytokines (IL-10, TGF-β, IL-35)
Direct cell contact inhibition
Reduced T cells suppression
Treg dysfunction / lack
Self-Ags driving T cell expansion
V segment
J seg.
D seg.
C seg.
Biased V seg.
Overrepresented autoimmune TCR Vβ

Figure 2. Tolerance failure and peripheral clonal expansion. (**A**) Central tolerance failure due to defective negative selection in the thymus. Mutations in regulators like *AIRE*, *FOXP3*, and *NF-κB2* impair mTEC function, reducing self-antigen presentation. This allows self-reactive thymocytes to escape deletion and exit the thymus, increasing autoimmunity risk. (**B**) Clonal expansion and bias in the pathological TCR repertoire combined with imp*AIRE*d peripheral tolerance. Treg dysfunction weakens the suppression of autoimmune T cells. Autoreactive T cells stimulate autoantibody production, causing tissue damage and inflammation. Released self-antigens drive biased T cell clonal expansion, sustaining a vicious cycle and promoting epitope spreading that worsens autoimmunity.

2.3. Autoimmunity Mechanism II: Clonal Expansion and Bias in TCR Repertoires

Peripheral antigen encounter drives selective clonal proliferation and memory formation in health, but in autoimmunity similar mechanisms expand self-reactive clones, producing oligoclonal repertoires with overrepresented *TRBV* families and recurrent CDR3 motifs; many expanded specificities are "public," shared across individuals [40,44,55,56] (Figure 2B). Empirical examples include *TRBV*13-2–biased memory CD4+ expansion in diabetic mice and a few high-frequency autoreactive CD4+ clones driving arthritis [57, 58], observations consistent with imp*AIRE*d thymic selection enabling peripheral antigen-driven oligoclonality [59]. Loss of Treg control (*FOXP3* defects or inflammatory impairment) permits unchecked expansion, tissue damage and antigen release, forming a self-sustaining loop; epitope spreading recruits new clonotypes, further narrowing diversity and worsening organ-specific disease [51,56–58,60].

2.4. Autoimmunity Mechanism III: Peripheral Tolerance Failure and Regulatory T Cell Dysfunction

Peripheral tolerance via Tregs (contact-dependent suppression and IL-10, TGF-β, IL-35) normally restrains autoreactive activation; quantitative or qualitative Treg defects—*FOXP3* mutations, chronic inflammation or other perturbations—erode suppression and permit pathogenic effector expansion and tissue injury [22, 42, 60, 61]. Clinical and experimental data link *FOXP3* abnormalities and reduced suppressive factor secretion to unchecked autoreactivity, where even modest Treg potency losses enable excessive self-reactive responses [62–64]. The resulting autoreactive T cell–B cell interactions generate autoantibodies that amplify inflammation, antigen release and further repertoire entrenchment [40,44].

2.5. Autoimmunity Mechanism IV: Genetic Factors and HLA Influence

Genetic background strongly shapes maladaptive repertoires: HLA alleles alter peptide presentation and hence selection and peripheral activation of autoreactive T cells, explaining HLA-associated autoimmune risk [50,51,65] (Figure 3A). Non-HLA variants (e.g., *PTPN22* R620W) modify signalling thresholds, lowering activation requirements and facilitating expansion of low-affinity autoreactive clones [50,66,67]. Thus host genotype biases both repertoire generation and selection dynamics, influencing autoimmune susceptibility [40].

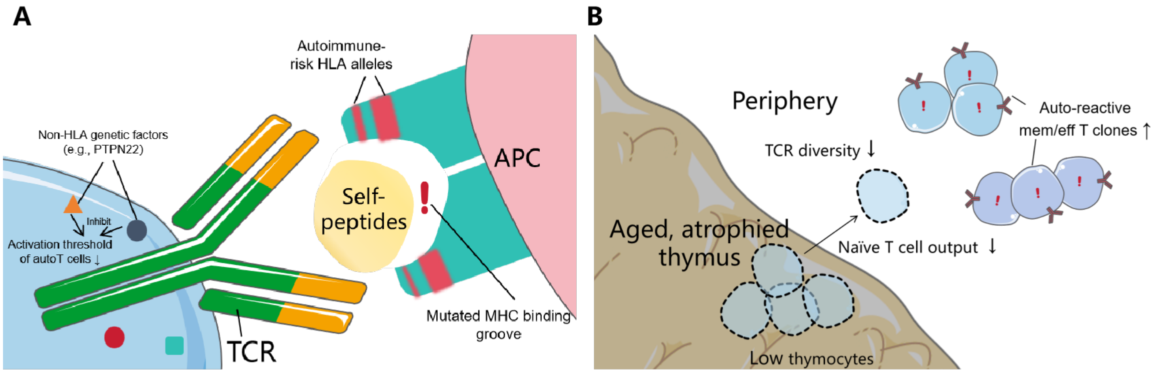

A
Autoimmune-risk HLA alleles
Non-HLA genetic factors (e.g., PTPN22)
Inhibit
Activation threshold of autoT cells ↓
Self-peptides
APC
TCR
Mutated MHC binding groove
B
Periphery
TCR diversity ↓
Auto-reactive mem/eff T clones ↑
Aged, atrophied thymus
Naïve T cell output ↓
Low thymocytes

Figure 3. Genetic and age-related influences on the autoimmune TCR repertoire. (**A**) HLA alleles modify peptide presentation via MHC groove changes, promoting autoreactive T cell survival. Non-HLA genes (e.g., *PTPN22*) lower activation thresholds, aiding autoreactive T cell expansion. (**B**) Aging causes thymic atrophy and fewer thymocytes, reducing naïve T cell output and TCR diversity. Autoreactive memory/effector T cells expand, driven by chronic infections, promoting immune senescence and chronic inflammation.

2.6. Autoimmunity Mechanism V: Age-Related Changes and Immunosenescence in TCR Repertoires

Aging causes thymic involution, reduced naïve export and contracted TCR diversity, increasing reliance on peripheral homeostatic proliferation and expanding memory/effector clones with autoreactive potential [43,68,69]. Chronic infections (e.g., CMV) and persistent dominant clones further skew repertoires toward oligoclonality and autoreactivity, compromising responses to new antigens and promoting "inflammaging" [41,53,70,71]. Within this compressed repertoire, modest genetic or environmental insults more readily breach tolerance and sustain chronic inflammation via dominant autoreactive clones [43].

These interconnected mechanisms collectively shape the autoimmune-prone TCR repertoire, as summarized in Table 1.

Table 1. Summary of mechanisms and factors contributing to TCR-mediated autoimmunity

Mechanism/Factor	Description and Principle	Key Findings and Supporting Evidence
Central tolerance failure and thymic defects	• Mutations (e.g., *AIRE*, *FOXP3*) impair negative selection • Thymic atrophy reduces deletion efficiency • Self-reactive T cells escape, narrowing repertoire and increasing autoreactivity	• Agapiou et al. showed imp*AIRE*d mTEC self-peptide expression reduces negative selection [52] • Coder et al. found thymic atrophy lowers deletion of high-affinity clones [53] • Bainter et al. demonstrated *IKK*α mutation disrupts mTEC development and nearly abolishes *AIRE* expression, enabling escape of biased self-reactive T cells [54]
Clonal expansion and repertoire bias	• Chronic self-antigen stimulation drives oligoclonal expansion • V/J gene usage bias observed • Epitope spreading recruits new autoreactive clones	• Marrero et al. observed biased *TRBV* usage and clonal expansion in diabetic mice [57] • Oh et al. reported dominant autoreactive clones expand in arthritis models [58] • Layzell et al. revealed thymic selection failure promotes repeated peripheral proliferation, generating oligoclonal, self-antigen biased TCR repertoires [59] • Rojas et al. and Prinz et al. highlighted epitope spreading broadens autoreactive clones [51,56]
Regulatory T cell dysfunction and peripheral tolerance loss	• Treg deficiency or functional impairment disrupts peripheral tolerance • Allows autoreactive T cell activation and inflammation	• Heimli et al. found *FOXP3* defects reduce Treg numbers and function [62] • Shokeen et al. linked reduced Treg cytokine secretion to pathogenic T cell activation [63] • Huang et al. emphasized slight Treg impairment fuels excessive effector T cell responses [64]
Genetic factors and HLA alleles	• Specific HLA alleles alter peptide presentation favoring autoreactive T cells • Non-HLA genes (e.g., *PTPN22*) modulate TCR signaling thresholds	• Bayley et al. showed HLA-DRB1 alleles bias antigen presentation toward self-reactivity [65] • Pratigya et al. found *PTPN22* variant lowers activation threshold, promoting autoreactive clones [67]
Age-related thymic involution and immunosenescence	• Thymic atrophy reduces naïve T cell output causing repertoire narrowing • Memory autoreactive clones expand, increasing autoimmunity risk	• Macaulay et al. observed thymic output decline narrows TCR repertoire and increases self-reactivity [68] • Coder et al. reported thymic microenvironment damage linked to inflammation and autoreactive T cell increase [53] • Müller et al. and Naumova et al. showed immunosenescence reduces diversity and increases autoreactive clones [70,71]

3. Tcr Repertoire Alterations as Potential ML Features in Autoimmune Diseases

In autoimmune disorders, perturbations in antigen-driven selection and failures of immune regulation manifest as characteristic alterations of the T cell receptor (TCR) repertoire: prominent clonal expansion, skewed gene-segment usage and modified complementarity-determining region 3 (CDR3) architecture. The CDR3 loop—whose length and amino-acid composition determine most of the peptide–MHC contact surface—therefore serves as a critical determinant of specificity and diversity, and is frequently interrogated when TCR features are used as inputs for machine-learning classifiers [72,73]. Accumulating evidence documents consistent repertoire perturbations across systemic lupus erythematosus (SLE), rheumatoid arthritis (RA), multiple sclerosis (MS) and type 1 diabetes (T1D), alterations that have been leveraged as salient ML features in several diagnostic and prognostic studies. The following sections synthesize the principal observations regarding diversity, gene segment bias and CDR3 properties, and thereby set the stage for later examples of ML-based predictive modelling [72–75].

3.1. Systemic Lupus Erythematosus (SLE)

3.1.1. Diversity and Clonal Expansion

Work on SLE has repeatedly highlighted substantial clonal skewing within the TCR β-chain compartment. Next-generation sequencing of the TCRβ CDR3 repertoire typically reveals a long tail of low-frequency clonotypes together with a small number of markedly expanded clones that contribute disproportionately to the repertoire composition [72]. Quantitative analyses using Shannon–Wiener and Simpson diversity indices indicate that, in many cohorts, overall diversity is reduced and highly expanded clones occur at elevated frequencies—findings interpreted as signatures of chronic antigenic stimulation and expansion of autoreactive T cells [76]. That said, the literature is not entirely uniform: Hou et al. report minimal net change in conventional diversity metrics in their SLE cohort, a dissenting observation that highlights heterogeneity between studies and the influence of cohort selection, sequencing depth and analytic choices [75].

3.1.2. Biased Gene Usage and Conserved Structural Features

Beyond clonality, biased usage of V, D and J gene segments is a recurrent theme in SLE repertoires. Large-scale sequence surveys show that while the underlying mechanisms of V(D)J recombination and junctional insertion remain operative, the relative frequencies of particular *TRBV* and *TRBJ* segments depart from control distributions in disease, implying selection for certain germline-encoded scaffolds during autoreactive expansions [73, 76]. Hou et al. specifically noted over- or under-representation of defined *TRBV* segments, suggesting that particular V genes may predispose to or facilitate expansion of autoreactive clonotypes. At the nucleotide level, insertion biases—such as GC-rich insertions driven by terminal deoxynucleotidyl transferase (TdT)—appear preserved in many datasets, indicating that recombinational processes remain partially constrained even when selection reshapes the final repertoires [73].

3.1.3. Cdr3 Length Distribution

Alterations in CDR3 length in SLE have proven more contentious. A subset of studies reports a shift toward shorter CDR3 amino-acid lengths among disease-associated clonotypes—Liu et al., for example, propose that autoimmune TCRs in SLE show reduced long-insertion events in junctional regions, yielding comparatively shorter CDR3s [44]. Conversely, several investigations (including Ye et al., Sui et al. and Hou et al.) found no meaningful difference in median CDR3 length or in V–D and D–J insertion/deletion profiles when comparing SLE cases to controls [75–77]. Taken together, SLE current data suggest that clonal expansion and V/J bias are more consistent findings than CDR3 length shifts, though the discordant reports underscore the need for standardized cohort definitions and uniform analytic pipelines to resolve remaining uncertainty.

3.2. Rheumatoid Arthritis (RA)

3.2.1. Repertoire Diversity and Clonality

A consistent finding in RA is contraction of the TCR repertoire, particularly within inflamed joints. Garrido-Mesa et al. reported that diversity metrics—Shannon entropy and clonality measures among them—are reduced in synovial fluid relative to p*AIRE*d peripheral blood, with pronounced oligoclonal expansions indicative of strong, local antigen-driven selection [78]. Larger AIRR-seq investigations by Aterido et al. refined this picture by identifying disease-associated meta-clones (clusters of highly similar TCRs) whose frequencies correlate with clinical serology such as rheumatoid factor (RF) and anti-citrullinated protein antibody (ACPA) status, supporting a link between repertoire skewing and disease phenotype [79].

3.2.2. V/J Gene Segment Usage Bias

Multiple independent studies document departures from control distributions in V and J gene usage in RA. Aterido et al. observed altered frequencies across dozens of V segments, multiple J segments and numerous V–J pairings—effects that were most evident within particular T cell compartments [79]. Turcinov et al. further detected enrichment of *TRBV*20-1 among CD4+ cells from ACPA-positive individuals, pointing to subset-specific biases [80]. Zhang et al. reported significant variations in V(D)J combinations (for example, *TRBV*7-2, *TRBV*11-1, *TRBV*13, *TRBV*15 and *TRBJ*2-3) in RA patients classified with kidney-deficiency syndromes, illustrating how clinical subphenotypes may map onto distinct rearrangement patterns [81]. Although exact segment preferences vary by cohort and method, the recurrent observation of over- and under-representation of particular germline segments argues that biased gene rearrangement and selection are integral to RA immunopathology.

3.2.3. Cdr3 Physicochemical Properties

Alterations extend beyond gene usage to the physicochemical characteristics of the CDR3 loop, the principal determinant of peptide contact. Early work showed that TCR clones modulated by *TNF*-inhibitor therapy differ in biochemical traits—isoelectric point, hydrophobicity and amino-acid composition—within their CDR3s, changes consistent with selective pressures imposed by antigenic targets or treatment-induced remodeling of the repertoire [79]. On CDR3 length, Zhang et al. reported a relative shortening and an increased number of dominant clonotypes in RA patients with kidney deficiency compared with healthy controls, a pattern that could reflect convergent recombination and selection for shorter, functionally convergent loops [81]. Together, these observations suggest that both sequence composition and physicochemical properties of CDR3 contribute to antigen specificity in RA.

3.2.4. Clonal Expansion Dynamics and Public TCR Sequences

Clonal expansions in RA are most apparent within the joint but frequently overlap with peripheral compartments. Amoriello et al. documented substantial sharing of expanded clonotypes between synovial fluid and blood, consistent with either recirculation of autoreactive clones or broad systemic exposure to joint antigens [82]. Public TCRs—identical or highly similar clonotypes observed across different patients—have been detected, though private expansions remain more common. Dunlap et al. repeatedly identified shared clonotypes, implicating common autoantigenic drivers (for example, citrullinated peptides or other joint-restricted epitopes) in convergent T cell responses among RA patients [83].

3.2.5. T Cell Subset-Specific Repertoire Features

Finally, repertoire features in RA are subset-dependent. Amoriello et al. described peripheral helper T cells (Tph), which promote intra-articular B cell activity, as carrying distinctive TCR signatures: pronounced clonal expansions accompanied by expression of PDCD1 and other inhibitory markers, a phenotype consistent with persistent antigen stimulation and exhaustion-like regulation [82]. By contrast, regulatory T cells (Tregs) in RA tend to show weaker clonal dominance and different V/J usage patterns, consistent with a disturbed balance between effector and regulatory compartments that likely contributes to unchecked autoreactivity [79].

3.3. Multiple Sclerosis (MS)

3.3.1. Diversity Metrics and Clonal Expansion

Unlike several autoimmune conditions that exhibit overt contraction of the TCR repertoire, MS cohort analyses frequently report no clear reduction in overall diversity of sorted CD4$^+$ and CD8$^+$ populations despite the presence of pronounced clonal expansions. Alves Sousa et al. observed greater TCRβ diversity in MS samples relative to controls, a result subsequently noted by Hayashi, underscoring that bulk diversity measures can mask focused oligoclonal responses [84,85]. More targeted investigations by Amoriello et al. demonstrated substantial clonal expansions of CD8$^+$ T cells within cerebrospinal fluid and lesional brain tissue, with some clonotypes shared between compartments—evidence compatible with antigen-driven, compartment-specific selection processes that contribute to local pathology [86]. Collectively, these observations argue that MS pathology is characterized less by wholesale repertoire collapse and more by regionalized clonal dominance.

3.3.2. V/J Gene Usage Bias and *TRBV*/ *TRBJ* Preferences

High-resolution repertoire profiling in MS has revealed skewing of particular V-gene families. Several reports document expansion of defined *TRBV* groups in lesion-infiltrating T cells, consistent with selective pressures favoring particular rearrangements. Valkiers et al. provided evidence for biased *TRBV*/*TRBJ* recombination patterns in CD8$^+$ T cells, further supporting the model that antigenic selection sculpts TCR specificity in MS lesional compartments [87].

3.3.3. Cdr3 Physicochemical Characteristics

Analyses focused on the CDR3 region report subtle but reproducible physicochemical biases. Massey et al. described shifts in length distributions and amino-acid motif composition within CDR3 sequences, findings that point toward repertoire tuning against specific autoantigens or viral peptides—notably Epstein–Barr virus (EBV) candidate epitopes that have been implicated in MS pathogenesis [88]. Thus, while global diversity metrics may remain near-normal, CDR3-level features reveal selection signals with likely functional relevance.

3.4. Type 1 Diabetes (T1D)

3.4.1. Reduced Diversity and Dominant Clonal Expansions

In contrast to MS, high-throughput TCR sequencing in T1D commonly shows a net reduction in repertoire diversity relative to type 2 diabetes and non-diabetic controls. Tong et al. reported lower normalized Shannon entropy values in T1D cohorts, reflecting repertoires dominated by a small number of highly expanded clonotypes—an architecture consistent with antigen-driven selection and the presence of pathogenic, islet-targeting T cells [89]. The practical implication is that, although thousands of unique sequences are present, disease processes may be driven by a restricted set of autoreactive clones.

3.4.2. V/J Gene Usage and CDR3 Features

V and J segment usage in T1D shows considerable heterogeneity, yet certain recurrent patterns emerge. Patient repertoires differ from healthy controls in *TRBV*/*TRBJ* composition, and distinctive CDR3 length and amino-acid signatures have been reported. Eugster et al. found that GAD65-specific T cells display broad *TRAV* and *TRBV* usage, with enrichment of *TRBV*5.1 among tetramer-positive cells—evidence that antigen specificity can arise across diverse germline usages while still converging on functional CDR3 motifs [90]. These data suggest that convergent sequence features, rather than single dominant germline usages, underpin autoreactivity in T1D.

3.4.3. Antigen-Driven Selection and Public TCRs

Longitudinal and chain-level analyses support antigen-driven selection in T1D. Eugster et al. reported persistent, highly expanded clones, and while public (shared) TCRs are relatively uncommon in T1D compared with some other autoimmune diseases, instances of chain-level convergence occur: identical TCR chains arise p*AIRE*d with variable partners across individuals, indicating recurrent recombination/selection events that produce functionally analogous receptors [90]. Such partial convergence implies that common antigenic pressures can shape similar solutions in different patients despite broad interindividual diversity.

Collectively, these disease-specific TCR repertoire alterations provide a rich set of quantifiable features that can be leveraged for machine learning-based classification and monitoring of autoimmune conditions (Table 2).

Table 2. Summary of TCR repertoire alterations in autoimmune diseases

Conclusion	Study and Findings
Systemic lupus erythematosus (SLE)	
• Marked clonal expansion • Biased *TRBV*/*TRBJ* gene usage • Conserved G/C nucleotide insertions • Controversial CDR3 length alteration • Pathogenesis linked to clonal selection over CDR3 length	• Attaf et al. reported clonal expansions [72] • Sui et al. observed reduced diversity and clonal proliferation [76] • Hou et al. identified biased gene usage and conserved insertions [73,75] • Liu et al. found shortened CDR3 length (minority view) [44] • Ye et al., Sui et al., Hou et al. found no significant CDR3 length change [75–77]
Rheumatoid arthritis (RA)	
• Reduced diversity, especially synovial fluid • Strong clonal expansions and antigen selection • Marked V/J gene bias • Altered CDR3 physicochemical features • Shorter CDR3 in kidney deficiency subtype • Expanded clones overlap joint and blood • Public TCRs less frequent than private • Distinct repertoires in Tph and Tregs	• Garrido-Mesa et al. showed reduced synovial diversity and expansions [78] • Aterido et al. discovered disease-specific meta-clones linked to clinical markers [79] • Turcinov et al. detected *TRBV*20-1 bias in CD4+ ACPA+ cells [80] • Zhang et al. reported shortened CDR3 length in kidney deficiency RA [81] • Amoriello et al. found overlapping expanded clones in joint and blood [82] • Dunlap et al. identified shared public clonotypes [83] • Amoriello et al. characterized Tph-specific TCR signatures [82] • Aterido et al. observed weaker clonal dominance in Tregs [79]
Multiple sclerosis (MS)	
• Overall diversity similar or slightly increased • Significant clonal expansions in CNS • Biased *TRBV*/*TRBJ* usage • Altered CDR3 physicochemical traits • Oligoclonal expansions target CNS/viral antigens	• Alves Sousa et al. reported increased diversity [84] • Amoriello et al. found clonal expansions in cerebrospinal fluid and brain [86] • Valkiers et al. observed biased V/J gene usage [87] • Massey et al. reported CDR3 length and motif biases [88]
Type 1 diabetes (T1D)	
• Reduced diversity dominated by few clones • High V/J gene heterogeneity • Distinctive CDR3 length and composition • Low frequency public TCRs with chain-level convergence	• Tong et al. showed reduced diversity and dominant clones [89] • Eugster et al. described GAD65-specific TCR features and convergence [90]

3.5. Preliminary Small-Scale Case Study on Feature Extraction Methods for Alterations

To address ongoing uncertainties regarding potential TCR repertoire variations across autoimmune diseases, we conducted this small exploratory case study as a methodological illustration alongside the main review. Detailed methods are documented in the supplementary materials. This investigation should be regarded as a preliminary case study focused on feature extraction methodology rather than a comprehensive empirical analysis. We utilized publicly available datasets [44,74,91–93] and performed a standardized comparative analysis. Twenty samples per group were randomly selected from multiple sclerosis, rheumatoid arthritis, systemic lupus erythematosus, type 1 diabetes, and healthy controls, recognizing that this limited sample size of $n = 20$ per group constrains statistical power and generalizability. We evaluated V and J segment usage frequencies alongside their Shannon diversity indices, CDR3 amino acid length distributions with associated Shannon diversity metrics, and proportions of conserved G/C nucleotide insertions in underlying DNA sequences. The objective was to generate a harmonized perspective on repertoire features across distinct autoimmune diseases using consistent metrics.

3.5.1. Methods

Data Sources

Public TCR β repertoire datasets [44,74,91–93] were used for this exploratory analysis (downloaded from the ImmuneAccess database). Twenty samples per group were randomly selected from multiple sclerosis (MS), rheumatoid arthritis (RA), systemic lupus erythematosus (SLE), type 1 diabetes (T1D), and healthy control ckeck (CK) groups, with the latter provided by Martinez et al. [93]. Data processing followed the original study pipelines (see Supplementary Material for details).

Feature Extraction

The following features were extracted: V and J gene usage frequencies; Shannon diversity indices of V and J usage; CDR3 amino acid length distributions and their Shannon diversity; and conserved G/C nucleotide insertion proportions. Repertoires were downsampled to control for sequencing depth. Analyses were conducted per sample, with group results reported as mean ± standard deviation.

Statistical Methodology

Comparisons between autoimmune disease groups and healthy controls were performed using two-tailed Student's t-tests. Statistical testing was conducted at the per-sample level, and the p-values reported in the text reflect uncorrected significance testing. Group-level results are summarized as mean ± standard deviation.

3.5.2. Results

Analysis of V gene usage frequency suggested potential disease-associated skewing in data involved in this case. For instance, TCRBV03 in MS, TCRBV04 in RA, TCRBV07 in both RA and SLE, TCRBV20 in T1D, and TCRBV27 in SLE appeared overrepresented relative to healthy controls (all comparisons $p < 0.005$; t-test). While these observations may be consistent with selective clonal expansion in particular disease contexts, the limited sample size and the absence of multiple-comparison correction necessitate cautious interpretation, suggesting that such findings should be regarded as exploratory rather than definitive. Shannon diversity based on V gene usage was highest in controls (mean 2.8997) and lowest in MS (mean 2.8627), yielding the order CK $>$ RA $>$ T1D $>$ SLE $>$ MS. Statistical testing indicated a modest reduction in MS relative to controls ($p = 0.0227$; t-test), whereas RA, SLE, and T1D trended lower without reaching the conventional significance threshold ($p = 0.8056$, 0.2202, 0.5484; t-test, respectively) (Figure 4A, C). This limited re-analysis therefore points to a possible reduction in V gene diversity in MS compared with controls, although prior studies have reported equal or even higher diversity. Such discrepancies may reflect differences in sample selection and analytical approaches. Overall, based on the limited sample analyzed, we observed a trend of modest contraction in V gene diversity, particularly in MS, but emphasize that these results remain preliminary.

J gene usage patterns also suggested potential disease-linked biases: TCRBJ01-02 * 01 was enriched in MS ($p < 0.005$; t-test), whereas TCRBJ02-01 * 01 and TCRBJ02-07 * 01 were elevated in RA, SLE, and T1D versus controls (all $p < 0.005$; t-test). These findings may indicate selective J-segment utilization associated with autoimmune states, though again larger cohorts and more rigorous statistical controls will be needed to confirm the presence and significance of such trends (Figure 4B).

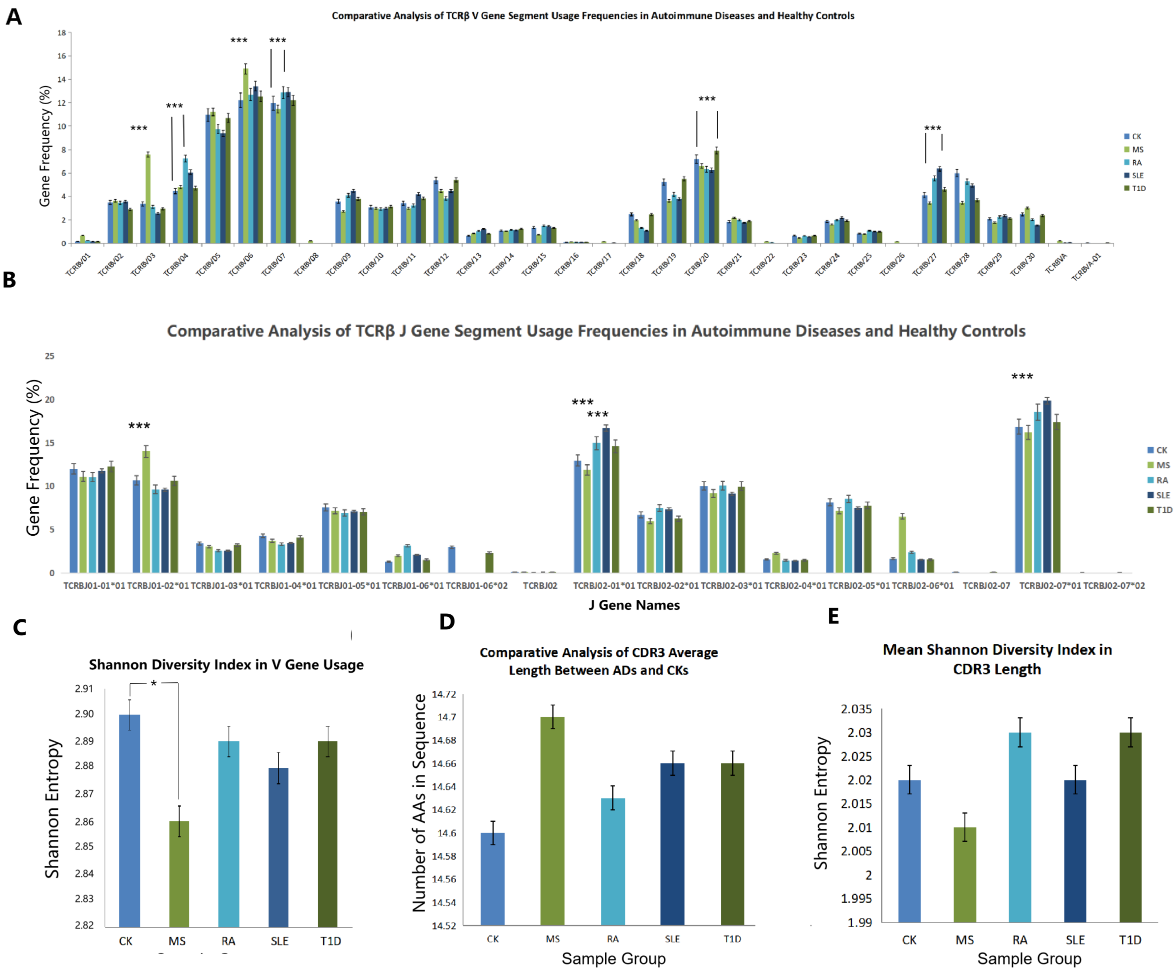


Figure 4. Preliminary small-scale case study of TCR β repertoire features across autoimmune disease groups and healthy control groups. (**A**) Comparative usage frequencies of TCR β V gene segments demonstrate V gene biases in autoimmune diseases relative to controls. (**B**) Shannon diversity indices of V gene usage reveal reduced diversity in autoimmune diseases, particularly in multiple sclerosis. (**C**) Comparative usage frequencies of TCR β J gene segments show selective J gene biases associated with autoimmune diseases. (**D**) Average CDR3 length comparison indicates marginal length increases in autoimmune diseases without reaching conventional statistical thresholds. (**E**) Shannon diversity indices of CDR3 length reveal slight diversity decreases in multiple sclerosis and modest increases in other autoimmune diseases.

CDR3 amino acid length distributions were broadly comparable across cohorts; autoimmune disease groups exhibited slightly longer mean CDR3 lengths than controls, with MS showing the largest mean (14.703 $\pm$ 0.237) versus controls (14.595 $\pm$ 0.182). These differences did not reach conventional statistical thresholds (all $p > 0.05$; t-test), and variability measures were similar across groups (mean SD $\approx$ 1.85–1.88), suggesting limited influence of autoimmune disease status on overall CDR3 length architecture (Figure 4D, supplementary Figure 1A).

Shannon indices derived from CDR3 length distributions displayed a small decrease in MS but modest increases in other autoimmune diseases, reflecting heterogeneous influences of autoimmune pathology on length diversity that may depend on individual-level factors (Figure 4F).

In summary, this multi-metric analysis provides an exploratory perspective on disease-associated remodeling of the TCR β repertoire: V and J gene usage may exhibit biases in specific autoimmune diseases, most notably in MS, whereas CDR3 length distributions and G/C insertion conservation appear largely stable in these datasets. These findings suggest that standardized metrics can help highlight repertoire features potentially linked to disease, though larger cohorts and more stringent statistical approaches will be important to refine and validate these observations.

4. Machine Learning-Based Prediction Using TCR Repertoires and Databases

4.1. Tcr Databases and Repertoire Contents

T-cell receptor (TCR) databases are indispensable for both experimental immunology and computational modelling efforts; they underpin studies of autoimmune pathology and serve as training/validation sources for machine learning approaches. The principal public resources in current use include IEDB, VDJdb, McPAS-TCR, AIRR and iReceptor. For example, IEDB provides an extensive compendium of experimentally validated immune epitopes and associated binding data that researchers routinely exploit for epitope prediction and immune monitoring tasks [94]. VDJdb curates TCR sequences linked to known antigens, reporting CDR3 sequences together with V/J annotations—information that is particularly useful when constructing supervised classifiers or motif discovery pipelines [95]. McPAS-TCR collects disease-associated TCRs and annotated epitopes, making it well suited to investigations of infection- and autoimmunity-related repertoires [96]. In addition, AIRR and iReceptor act as community platforms for standardized sharing and large-scale integration of immune receptor sequencing data, thereby enabling cross-cohort mining and reproducible repertoire analyses [97,98].

Typical TCR repertoire records contain multiple layers of information: the amino-acid sequence of the CDR3 region (the principal determinant of antigen contact), V, D (for the β chain), and J gene assignments, chain identity (α, β, γ, δ), clone abundance or frequency, and metadata such as donor disease status or tissue of origin. Because CDR3 mediates antigen recognition, it is the primary focus in most sequence-based studies, while V(D)J usage patterns provide complementary signals about recombinational biases and potential publicities. Clone frequency reports reveal expansion dynamics and immune activation. Several databases additionally offer experimentally validated TCR–epitope pairings and functional annotations, which help to connect sequence patterns with biological roles and disease associations. High-quality sequencing accompanied by rich, standardized metadata (sample provenance, sequencing depth, quality metrics) is therefore essential for downstream bioinformatics and for training robust machine learning models.

The standardized frameworks offered by these databases facilitate systematic analysis of immune repertoire features across different disease contexts and experimental conditions (Tables 3 and 4).

Table 3. Summary of TCR and Epitope Related Databases

Database Name	Data Format	Brief Description	Significances and Applications
IEDB [94]	Epitope data: Extensive curated experimental data. Contains millions of entries. Data type: Experimentally validated T cell and B cell epitopes.	A comprehensive resource for antibody and T cell epitopes. Covers infectious diseases, allergies, autoimmunity, and transplantation. Provides prediction and analysis tools.	The primary repository for epitope data. Essential for epitope discovery and vaccine design. Useful for studying immune responses. Includes integrated prediction tools.
VDJdb [95]	TCR–epitope pairs: Curated TCR sequences with known antigen specificity. Most are single-chain records. Data type: High-confidence, manually curated pairs.	Focuses on TCR sequences paired with epitopes and MHC molecules. Includes CDR3 sequences and V/J gene usage.	Designed for studying TCR–antigen interactions. High-quality data for training TCR specificity models. Suitable for machine learning applications.
McPAS-TCR [96]	TCR–disease associations: TCR sequences linked to specific diseases. Data type: Manually curated from published literature.	Contains TCR sequences associated with pathological conditions. Includes infectious diseases, autoimmune disorders, and cancer.	Valuable for immunopathology research. Enables exploration of TCR repertoires in disease contexts. Helps identify public TCRs in infections like COVID-19.
AIRR Community [97]	Standardized RepSeq data: Framework for sharing immune receptor data. Data type: Raw and processed data from numerous studies.	An international community establishing data standards. Focuses on TCR and BCR repertoire sequencing data.	Promotes data reproducibility and open access. Standardized format enables large-scale meta-analyses. Facilitates comparisons across labs and projects.
iReceptor [98]	Integrated RepSeq data: Gateway federating multiple AIRR-compliant repositories. Data type: Unified query interface for large-scale data.	Integrates millions of TCR/BCR sequences from distributed sources. Provides a unified search interface.	Enables cross-database queries and large-scale mining. Users can search without downloading individual datasets. Accelerates scientific discovery.

Table 4. TCR repertoire feature categories

Content Category	Description
CDR3 sequence	The complementarity-determining region 3 amino acid sequence, critical for antigen specificity
V gene	Variable gene segment involved in recombination
D gene	Diversity gene segment (β chain only)
J gene	Joining gene segment
Chain type	α (TRA), β (TRB), or other chains (γ , δ)
Clone frequency	Abundance or frequency of specific TCR clones in the sample
Sample source	Donor information (disease status, tissue origin, time point)
Pairing information	α-β chain pairing where available
Antigen epitope information	Recognized antigen peptide sequences or epitopes (experimentally validated)
Functional annotation	Immune phenotype, cell type, disease relevance
Sequencing quality metrics	Read length, depth, error rate

4.2. Tcr-Based Diagnosis of Autoimmune Disease States

Using repertoire data to classify disease versus health is an active area of research. Broadly speaking, models fall into two families depending on input design: those driven by gene-level features (V/J usage, V–J pairing) and those built on CDR3 sequence or physicochemical representations.

4.2.1. Gene Feature-Driven Diagnostic Models

Several groups have demonstrated that gene-level patterns can provide discriminative signals for autoimmune phenotypes. Liu et al. analyzed TCRβ repertoires from a large cohort—877 SLE patients, 206 RA patients, and 439 healthy controls—and trained a random forest classifier leveraging V and J gene usage together with V–J pairing frequencies. Their pipeline identified 198 SLE-specific and 53 RA-specific clones and reported very high sensitivity and specificity under their evaluation framework, underscoring the potential of high-dimensional gene features for clinical discrimination [44]. While these findings are encouraging, they were derived from a single large cohort under particular technical settings, and further confirmation in independent datasets would be valuable.

Ye et al. constructed a peripheral blood repertoire from 10 lupus nephritis patients and 10 controls, then trained a random forest on V–J frequency features to create TCR-LupusDetect. The model combined diversity indices with specific V–J motifs and attained a cross-validated AUC of 0.89, illustrating the feasibility of non-invasive blood-based repertoire diagnostics in LN [77].

By contrast, Dibble et al. evaluated TCR diversity across 160 samples from the UK ME/CFS biobank (including ME/CFS cases, MS disease controls and healthy donors) using a potential SVM (P-SVM) on CDR3 and full VDJ rearrangement features. Their findings underscore practical limitations: despite encouraging in silico performance, clinical sample classification was not significant, highlighting current limitations in sample size, cohort heterogeneity and technical variability [99].

Taken together, these studies emphasize both the promise of gene feature models and the need for rigorous validation across larger, better-annotated cohorts before translation to clinical assays (Table 5).

Table 5. Gene feature-based diagnostic models summary

Reference	Date	Dataset (Pts/CKs)	Base Model	Summary Description
Liu [44]	2019	SLE 877, RA 206, CK 439	Random Forest	V/J gene usage and pairing features; achieved 100% sensitivity and specificity for SLE/RA diagnosis, identifying disease-specific clones.
Ye [77]	2020	LN 10, CK 10	Random Forest	Utilized V-J gene frequencies and diversity indices; AUC 0.89 for lupus nephritis detection from blood samples.
Dibble [99]	2024	ME/CFS 160, MS, CK	P-SVM	Used CDR3 sequences and V-D-J rearrangements; found no significant classification power for ME/CFS/MS, highlighting sample and technical limitations.

https://doi.org/10.53941/tai.2026.100006

4.2.2. Cdr3 Sequence and Physicochemical Feature-Based Diagnostic Models

A second class of methods encodes the CDR3 amino-acid sequence and derived physicochemical descriptors for classification. Fowler et al. developed GlutenDetect by leveraging known gluten-specific TCRα and TCRβ sequences and frequency thresholds; trained on intestinal CD4$^+$ repertoires (20 celiac patients under gluten and gluten-free diets plus controls) and tested on an independent set, the model reached perfect accuracy on training and 80% on the held-out set—an encouraging result given the absence of an oral gluten challenge [100].

Ma et al. studied 662 newly diagnosed pediatric ITP patients and applied classical ML algorithms (logistic regression, SVM, random forest, XGBoost) to predict chronicity. Their XGBoost-ITP predictor combined immunological measures (Th17/Treg ratios, TCR$\gamma\delta^+$ counts) with demographic features and achieved an AUROC of 0.85 and 80% accuracy on test data, demonstrating the benefit of integrating repertoire signals with immunophenotyping [101].

Shen et al. reported DeepTAPE, a convolutional neural network long short term memory hybrid architecture with residual connections trained on TCRα CDR3 sequences and V gene frequencies from 877 systemic lupus erythematosus patients and 439 controls. They derived an Autoimmune Risk Score as the mean of sequence level risk estimates. Cross validated performance metrics included an area under the curve (AUC) of 97.99% and accuracy of 93.97%. The Autoimmune Risk Score demonstrated correlation with clinical activity measures, suggesting potential diagnostic and prognostic utility. It should be noted that these performance metrics were obtained from internal cross validation within the original dataset. When evaluated on an independent external dataset from a different source than the ImmuneAccess database, the model showed some generalization capacity for juvenile idiopathic arthritis and autoimmune arthritis, though with reduced AUC values of up to 82.67% and 90.33%, respectively. This performance decline in external validation highlights a limitation regarding generalizability beyond the original cohort and underscores the need for further validation in diverse populations [102].

He et al. proposed SLEpitopeNet, a hybrid CNN + BiLSTM architecture with a scaled dot-product attention fusion layer. Using 4456 SLE-related epitopes (1116 positives, 3340 negatives), they encoded amino-acid composition, dipeptide composition and spectrum descriptors and computed attention weights as

$$A = \mathrm{softmax}\left(\frac{H_{\mathrm{LSTM}}\, W_q \left(H_{\mathrm{CNN}}\, W_k\right)^T}{\sqrt{d_k}}\right)$$

where H_{LSTM} and H_{CNN} are the BiLSTM and CNN feature matrices, W_q, W_k are learned projections, and $\sqrt{d_k}$ is the scaling dimension. The model attained ROC-AUC 0.9506 and F1 0.8333, outperforming several comparator methods by combining local pattern recognition with long-range contextual modelling [103].

Rawat et al. evaluated T1D-associated repertoires across 2250 peripheral blood samples (patients, relatives, controls) and applied DeepRC, a CNN within a multiple-instance learning framework that extracts positional amino-acid frequency features from CDR3β sequences. DeepRC reached AUC 0.77 and balanced accuracy 72.9%, exceeding k-mer logistic regression and public-clone frequency baselines and demonstrating automated motif discovery with interpretable outputs [104].

Yang et al. presented AutoY, a convolutional neural network, and LSTMY, an attention-enhanced bidirectional long short-term memory network, trained on Adaptive Biotechnologies repertoire data encompassing rheumatoid arthritis, type 1 diabetes, multiple sclerosis, and insulin autoantibody positivity. The prediction was formulated within a multiple instance learning framework defined by

$$\check{Y} = P(Y = 1 \mid \{M_1, \ldots, M_k\}) = \sigma' \left(W^{L'\mathrm{T}} [\tilde{y}_1, \ldots, \tilde{y}_k]^{\mathrm{T}} + b^{L'} \right),$$

where M_k denotes the k-th TCR feature matrix, $P(Y = 1 \mid \{M_1, \ldots, M_k\})$ represents the probability of autoimmune disease for the library, $\sigma'(x)$ is a sigmoid activation function, and $W^{L'} \in \mathbb{R}^k$ and $b^{L'} \in \mathbb{R}$ are the weight matrix and bias term, respectively. The AutoY model demonstrated notably high discriminatory performance for type 1 diabetes and multiple sclerosis, with AUC values of 0.9991 and 0.9961, respectively, and exhibited greater stability compared to LSTMY. These results suggest that integrating local and long-range sequence representations could be beneficial for non-invasive autoimmune diagnostics [35]. However, it should be noted that these exceptional performance metrics were obtained from internal validation within the original dataset. The study did not include external validation on completely independent cohorts from different sources or technical platforms, which represents a limitation regarding the assessment of model generalizability. Near-perfect classification performance is uncommon in biological prediction tasks and may indicate that the model is leveraging training-set-specific signals rather than more generalizable disease-associated features [105]. Therefore, additional validation in independent and clinically diverse populations would be valuable to confirm the robustness and broader applicability of these findings (Table 6).

Table 6. Summary of diagnostic models based on CDR3 sequence and physicochemical features

Reference	Year	Dataset	Model	Key Characteristics, Performance and Limitations
Fowler et al. [100]	2023	Celiac 21, HC 14	Interpretable ML	100% training, 80% test accuracy; limited by small sample size and absence of gluten challenge.
Ma et al. [101]	2023	ITP 662	XGBoost	AUROC 0.85, 80% accuracy; single-center data may limit generalizability.
Shen et al. [102]	2024	SLE 877, HC 439	CNN+LSTM+Residual	AUC 97.99%, accuracy 93.97%; external validation showed reduced performance (AUC 82.67–90.33%).
Rawat et al. [104]	2024	T1D 2250	CNN, MIL	AUC 0.77, balanced accuracy 72.9%; moderate performance may reflect disease heterogeneity.
He et al. [103]	2025	SLE epitopes 4456	CNN+BiLSTM+Attention	ROC-AUC 0.9506, F1 0.8333; epitope-based approach requires further clinical validation.
Yang et al. [35]	2025	RA, T1D, MS, IAA	CNN & Attention-BiLSTM	AUC 0.9991 (T1D), 0.9961 (MS); exceptional performance requires independent cohort validation.

In conclusion, when evaluating the comparative performance of immune receptor-based machine learning approaches for autoimmune disease diagnosis, several methodological considerations warrant careful attention. Direct performance comparisons across studies remain challenging due to fundamental differences in dataset compositions, technical platforms, and evaluation frameworks. The limited number of directly comparable studies employing standardized evaluation frameworks, particularly before 2020, restricts definitive conclusions. Within this constrained analytical landscape, the BCR-based classifier developed by Ostmeyer et al. in 2017 [106], while not TCR-based, provides a useful reference point for multiple sclerosis diagnosis, achieving 87% leave-one-out cross-validation accuracy and 72% validation accuracy on independent data. This intermediate performance level may be contextualized against the more recent TCR-based deep learning approach by Yang et al. [35], It reported an AUC of 0.9961 for multiple sclerosis. Similarly, for systemic lupus erythematosus, the gene feature-driven methodology proposed by Liu et al. in 2019 [44] can be compared with the CDR3-based approach of Shen et al. [102]. with the latter appearing to demonstrate enhanced diagnostic capability. The progression from earlier methods to contemporary TCR-based deep learning approaches suggests a general trend toward improved diagnostic performance over time. However, the specialized nature of immune receptor-based machine learning for autoimmune diagnostics, combined with the relatively small number of high-quality comparable studies, necessitates cautious interpretation of these observations. The field would benefit from additional rigorously validated studies to establish more definitive performance benchmarks across different autoimmune conditions and methodological approaches.

4.3. Immune Epitope Binding Prediction Models

Predicting TCR–epitope binding remains a central challenge and a rapidly progressing subfield. Rajitha Rajeshwar T. et al. constructed TCR-H, an SVM-based predictor trained on an assembled set of 107,000 positive and 147,000 negative TCR–epitope pairs drawn from IEDB, VDJdb and McPAS-TCR. By using full-sequence physicochemical encodings of both CDR3β and peptide sequences and relying on experimentally validated negatives, their model yielded AUC-ROC values in the 0.87–0.92 range under hard-split evaluations (unseen epitopes and TCRs), and exhibited advantages over ensemble tree methods in interpretability and negative-sample handling [107].

Weber et al. profiled TCRβ repertoires from 73 giant cell arteritis patients and 69 age-matched controls, applying the tcrdist3 metric and a K-nearest neighbor classifier to nominate 1526 GCA-associated sequences; their TITAN framework further encodes TCR and epitope sequences with dual-modality attention and fuses them via structured attention where

$$\alpha_i = \frac{\exp(u_i)}{\sum_{j=1}^{T}\exp(u_j)}, \quad \text{where} \quad \vec{u} = \tanh(\mathbf{X}_1\mathbf{W}_1 + \mathbf{W}_3(\mathbf{X}_2\mathbf{W}_2))\vec{v}.$$

Here, $\mathbf{X}_1$ and $\mathbf{X}_2$ are convolutional features and $\mathbf{W}_1, \mathbf{W}_2, \mathbf{W}_3, \vec{v}$ are learnable parameters. TITAN demonstrated improved generalization and interpretability through attention visualization, enhancing autoimmune epitope recognition performance [69].

Darmawan et al. developed MITNet and MITNet-Fusion, Transformer plus CNN hybrids, trained on public repositories (IEDB, VDJdb, McPAS) and benchmarked on dominant epitopes such as GILGFVFTL, GLCTLVAML and NLVPMVATV. By combining AAC, DIP, SPC and composite descriptors (AADIP), MITNet-Fusion achieved an AUC up to 0.87 for those epitopes, outperforming prior methods by capturing both local motifs and long-range dependencies [108].

Wang et al. introduced TPepRet, a BiGRU + RetNet hybrid that augments positional encoding with a decay matrix D and computes retention as

$$S_n = \gamma S_{n-1} + K_n^T V_n, \quad \text{Retention}(X) = (QK^T \odot D)V,$$

training on over 83,000 positives from public databases. TPepRet emphasizes physicochemical encodings of CDR3β residues and outputs binding probabilities via a sigmoid; it reports AUC 0.87 on known peptides and displays favorable generalization and robustness [109].

Overall, the field has moved from simple motif or k-mer methods toward architectures that blend local pattern detection with global sequence context and attention mechanisms; these hybrids tend to show better generalization to unseen TCR–epitope pairs while also offering interpretable elements that can be mapped back to biologically meaningful motifs (Table 7).

Table 7. Summary of TCR-epitope binding prediction models

Reference	Date	Dataset (Pts/CKs)	Base Model	Key Contributions and Findings
Weber et al. [69]	2023	GCA 73, CK 69	KNN + Attention	Employs dual attention mechanism and transfer learning for autoimmune epitope recognition, demonstrating high accuracy on unseen data.
Darmawan et al. [108]	2023	IEDB, VDJdb, McPAS epitopes	Transformer + CNN	Combines AAC, DIP, and SPC features to capture long-range and local motifs, achieving peak AUC 0.87 in epitope prediction.
Demerdash et al. [107]	2024	107k pos, 147k neg TCR-epitope pairs	SVM	Uses full-sequence physicochemical features with validated negatives, achieving AUC 0.87–0.92 and providing interpretable predictions.
Wang et al. [109]	2025	83k positive TCR-peptide pairs	BiGRU + RetNet LLM	Utilizes physicochemical CDR3β encoding with decay matrix for positional information, achieving AUC 0.87 in binding prediction.

5. Future Perspectives and Discussion

The convergence of T-cell receptor (TCR) repertoire analysis with advanced machine learning techniques offers a promising direction for advancing research, diagnosis, and treatment of autoimmune diseases. As our understanding of TCR diversity, clonal expansion, and antigen-specific immune mechanisms deepens, these insights hold significant promise for translation into precision immunomedicine. Recent advances in strategic reasoning frameworks, particularly game theory-inspired approaches like the Game-Theoretic Artificial Intelligence (GTAI) framework proposed by Guo and Wu [110], provide novel methodologies for translating complex biological insights into actionable clinical strategies. The GTAI framework formalizes strategic reasoning through four iterative phases—observation and diagnosis, treatment planning, execution, and outcome evaluation—offering a structured approach for navigating the complex decision landscapes in autoimmune disease management. While these strategic frameworks show particular relevance for TCR-based diagnostic and therapeutic strategies, their application to autoimmune diseases remains largely conceptual and requires empirical validation to establish practical utility.

Moving forward, research must systematically integrate longitudinal multi-omics datasets—including genomics, transcriptomics, proteomics, and epigenomics—to enable a comprehensive, temporal dissection of immune dysregulation and TCR repertoire dynamics, thereby elucidating critical immunoregulatory nodes throughout disease progression [111]. In practice, this means designing studies with repeated sampling, standardized processing pipelines and harmonized metadata so that temporal trajectories, rather than static snapshots, drive model development and mechanistic inference.

5.1. Advances and Challenges in Machine Learning for TCR Analysis

Machine learning has demonstrated tremendous potential in analyzing TCR data; however, existing models still face notable challenges in terms of generalizability, interpretability, and clinical translation [112,113]. In this regard, the establishment of large-scale, high-quality, cross-disease standardized TCR databases is foundational, as it facilitates the unification of model training and validation protocols, thereby accelerating the deployment of ML applications [114]. Equally important are transparent benchmarking practices and the routine reporting of cohort composition, preprocessing steps and evaluation splits—details that materially affect reproducibility but are often under-reported.

One particularly important frontier lies in developing interpretable and generalizable ML models capable of parsing the high-dimensional complexity inherent in TCR sequences. Advances in explainable artificial intelligence (XAI) are crucial for identifying biologically meaningful autoreactive features embedded within CDR3 motifs and V(D)J gene usage biases [115,116]. Such interpretability not only enhances clinician trust and adoption but also fosters the discovery of novel pathogenic mechanisms, thus informing targeted immunomodulatory strategies. For example, SHAP (Shapley Additive Explanations), though not originally applied to autoimmune diseases, has contributed significantly to predictive and therapeutic approaches in immune-related disorders. Tan et al. utilized explainable ML models (e.g., XGBoost) on single-cell data to predict tumor-reactive TCRs in TCR-T cell therapy, markedly improving prediction accuracy (e.g., the geometric mean of the predicTCR tool increased from 0.38 to 0.74) [117]. Similarly, Sun et al. applied SHAP analysis in gastrointestinal cancers to identify key genes such as *CXCL13*, constructing a high-accuracy (AUC = 0.99) neoantigen T cell recognition model and revealing its differentiation mechanisms [118]. It is anticipated that such approaches will extend beyond oncology to autoimmune disease prediction, whereby attributing model predictions to specific physicochemical features of CDR3 or gene usage patterns can begin to decode the "black-box" nature of these models and yield biological insights into determinants of self-reactivity. This interpretability is indispensable not only for refining predictive algorithms but also for translating findings into biologically meaningful insights that guide therapeutic interventions. In short, interpretability is both a scientific goal and a prerequisite for clinical uptake.

Moreover, the accurate and large-scale capture of p*AIRE*d TCR α and β chains remains a technical bottleneck. The future implementation of stable and efficient chain pairing recognition through high-throughput single-cell sequencing technologies will generate richer datasets, thereby facilitating the development of models that more faithfully reflect TCR antigen recognition specificity [1]. In conjunction with improved epitope mapping and structural biology modeling, these advances will drive precise prediction of TCR–epitope binding affinities, opening new avenues for personalized vaccine design and TCR engineering therapies [119–121]. Practically speaking, researchers should prioritise experimental designs that preserve pairing information and report pairing recovery rates, so that downstream model limitations are transparent and interpretable.

5.2. Data Leakage and Overfitting in TCR Repertoire Analysis

TCR repertoire analysis faces significant methodological challenges concerning data leakage and overfitting, particularly when sequence sharing occurs between training and test sets from subjects with public TCR clones. Investigations such as those by Yang et al., who developed AutoY and LSTMY models using Adaptive Biotechnologies repertoire data [35], and Shen et al., who created the DeepTAPE architecture [102], have reported exceptionally high performance metrics. However, these outcomes typically derive from internal validation within the original datasets, while their performance declines on external datasets. The observed AUC values approaching 1.0 in certain cases may reflect specific characteristics of the training data and could indicate susceptibility to overfitting. Related repertoire-classification studies have similarly noted that models can show strong in-dataset performance yet degrade under distribution shifts, consistent with limited cross-cohort generalizability [122,123]. Performance degradation during external validation highlights limitations in model generalizability beyond the original cohorts.

This phenomenon may be attributed to methodological and material similarities when testing data originates from the same source. Although positive and negative samples maintain correct labeling, they might share common clones and characteristics, potentially leading to inflated performance estimates during model evaluation. This scenario could provide an unfair advantage as models encounter familiar patterns during testing. While researchers typically employ external independent test sets honestly, which reveal performance declines on unfamiliar data, users should interpret these apparently elevated metrics obtained from internal data-based evaluations with considerable caution. These observations emphasize the necessity for additional validation across diverse populations and the implementation of rigorous data partitioning strategies that account for public clonotypes to prevent artificially enhanced performance assessments.

The development of unified, large-scale databases with consistent formatting represents another critical consideration. Ensuring that real-world clinical data maintain format consistency with training and testing datasets could help mitigate potential advantages models might possess on their original training data. Utilizing databases with uniform sources and formats during both model development and application might reduce discrepancies, though models may still demonstrate reduced adaptability to novel data compared to their performance on familiar training sets. The absence of standardized benchmarks and consistent data formats across different studies complicates fair comparisons between methodological approaches. Consequently, direct comparisons of AUC values without accounting for variations in dataset usage, preprocessing pipelines, and validation schemes risk misinterpretation. Establishing unified, large-scale databases with consistent formatting appears crucial for enabling robust benchmarking and promoting reproducible research within this field.

5.3. Integrative Multi-Modal Modeling and Genetic Context

Another highly promising direction is the integration of patient-specific genetic background information into predictive models, with particular emphasis on *HLA* polymorphisms and non-*HLA* immune regulatory gene variants. Such integrative strategies will enable ML models to capture the influence of host immunogenetic context on shaping the TCR repertoire, thereby enhancing the accuracy of disease risk assessment and therapeutic response prediction [124]. Combining these genomic layers with transcriptomic readouts and proteomic measures can resolve whether observed repertoire shifts reflect intrinsic selection pressures or transient activation states. Furthermore, combining these data with environmental and lifestyle factors could substantially improve model performance and facilitate early interventions based on the individual's comprehensive immune landscape.

From a clinical standpoint, TCR repertoire–based diagnostic and prognostic models exhibit great potential as non-invasive tools for early screening and personalized treatment of autoimmune diseases [125]. Strengthening interdisciplinary collaborations that integrate clinical phenotyping, imaging, and immune functional assays will further accelerate the construction of multimodal, fused diagnostic systems [126]. Additionally, TCR repertoire analyses not only serve diagnostic purposes but also inform the development of TCR-T cell therapies by precisely targeting pathogenic clones to enable personalized immunomodulation. Given the complexity of immune networks, translating TCR repertoire research findings into effective clinical interventions remains a continuous challenge, necessitating further exploration and clinical validation [127,128]. Concurrently, developing cost-effective and widely accessible sequencing platforms is critical for broad clinical adoption of these models; without attention to affordability and scalability, even the most accurate models will struggle to achieve real-world impact.

5.4. Critical Synthesis: Framework, Taxonomy and Theoretical Bottlenecks

We previously reviewed characteristic features of TCR repertoires across autoimmune diseases and a range of machine learning approaches that make use of those features [35,44,77,99–104]. To translate dispersed empirical findings into an actionable research agenda, we propose a pragmatic framework that maps common TCR feature categories to appropriate machine learning model families and identifies attendant theoretical bottlenecks and priority directions for investigation. The framework is arranged along a first dimension of data granularity: static sequence features CDR3 sequences, k-mers and physicochemical descriptors; gene level features V/D/J usage and pairing; repertoire level quantitative features clone frequencies and Shannon/Simpson diversity; and clinical and genetic context HLA, phenotype labels and longitudinal information [72,73,75,76,86–90]. The second dimension groups algorithms by modelling capability: sequence pattern models CNN and Transformer, sample level multiple instance models MIL and DeepRC, similarity and graph based approaches tcrdist and GNN, and multimodal fusion models that incorporate genetic and phenotypic covariates [69,104]. From this mapping a set of practical observations emerges. When the objective is detection of short sequence motifs or local patterns, convolutional filters or Transformer architectures tend to perform better than shallow models that rely largely on V and J frequency features [102,103]. When labels are available only at the sample level rather than the sequence level, multiple instance learning and pooling strategies exemplified by DeepRC and MIL are generally more appropriate [69, 104]. When paired alpha beta chain or single cell data are available, joint modelling via paired chain multimodal networks appears likely to improve TCR epitope prediction accuracy [125–128].

Despite the pragmatic guidance afforded by this mapping, several theoretical and practical constraints persist. Principal challenges include scarce chain pairing and epitope annotation that induce weak label regimes and hinder out of distribution generalization, sample imbalance and database biases across studies populations and technologies that constrain cross cohort reproducibility, limited model interpretability that complicates biological validation and a weak structural linkage between motifs inferred by deep networks and actual MHC binding sites, and insufficient formal modelling of longitudinal trajectories and immune dynamics [129–132] (Figure 5). To mitigate these constraints we propose a staged research roadmap. Short term priorities should emphasize creation of public paired chain and epitope annotated benchmark datasets together with standardized evaluation protocols. Medium term efforts should advance multimodal joint learning that incorporates HLA, transcriptomic and clinical covariates. Long term priorities include development of interpretable hybrid models that combine structural modelling with data driven attention constraints to better connect sequence patterns and molecular interaction mechanisms. Collectively this taxonomy and the enumerated bottlenecks provide evidence based guidance for method selection and indicate data priorities for funders and database curators [130, 131].

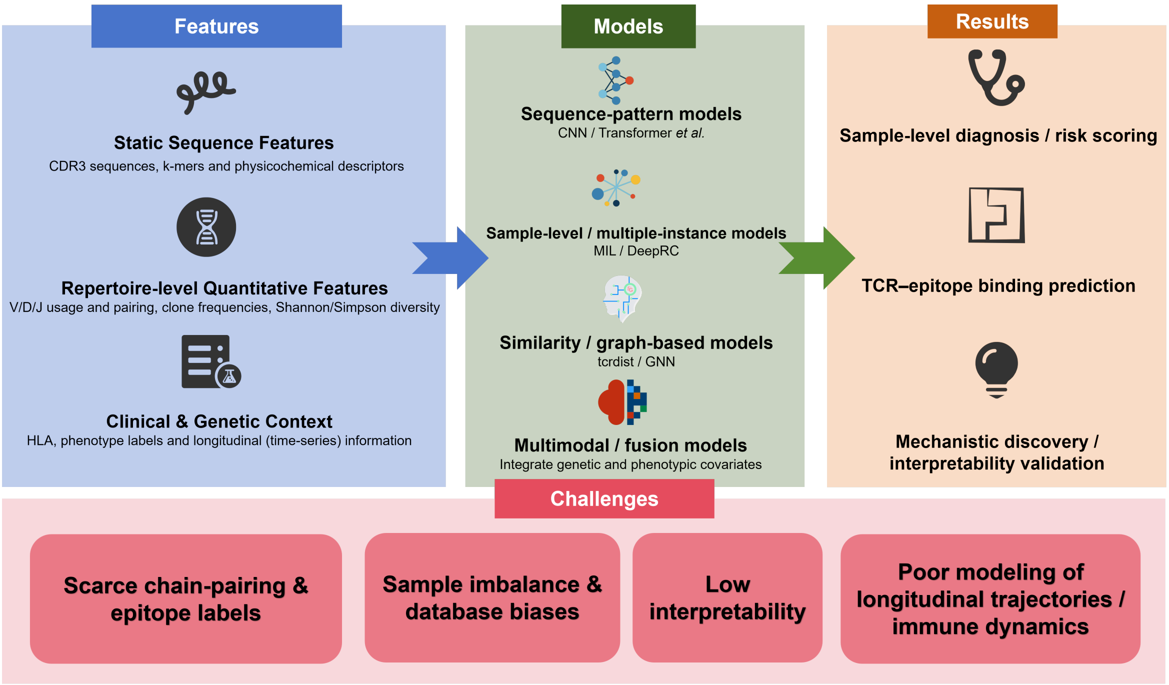

Features
Static Sequence Features
CDR3 sequences, k-mers and physicochemical descriptors
Repertoire-level Quantitative Features
V/D/J usage and pairing, clone frequencies, Shannon/Simpson diversity
Clinical & Genetic Context
HLA, phenotype labels and longitudinal (time-series) information
Models
Sequence-pattern models
CNN / Transformer et al.
Sample-level / multiple-instance models
MIL / DeepRC
Similarity / graph-based models
tcrdist / GNN
Multimodal / fusion models
Integrate genetic and phenotypic covariates
Results
Sample-level diagnosis / risk scoring
TCR–epitope binding prediction
Mechanistic discovery / interpretability validation
Challenges
Scarce chain-pairing & epitope labels
Sample imbalance & database biases
Low interpretability
Poor modeling of longitudinal trajectories / immune dynamics

Figure 5. Overview of feature types, modeling approaches, results and challenges for repertoire-based TCR analysis. **Left**: feature types (sequence, repertoire-level and clinical/genetic). **Middle**: modeling approaches (sequence-pattern, MIL/sample-level, similarity/graph-based, multimodal). **Right**: results (diagnosis/risk scoring, epitope binding, interpretability). Bottom: key challenges.

5.5. Ethical Considerations and Data Governance

Immune-repertoire sequencing and ML-driven diagnostics pose heightened privacy risks because repertoires can reveal genetic predispositions, immune status, and exposure history. Robust data protection — including strong anonymization, encryption, documented de-identification and threat models — is essential across collection, storage and analysis. Clear governance policies on data access, provenance and usage rights are needed to ensure transparency and accountability. Privacy-preserving computational methods (e.g., federated learning, secure multi-party computation) can enable collaborative model building without raw data sharing, but require interoperable infrastructure and legal agreements. Informed consent processes should cover potential secondary and commercial uses, and governance must promote equitable benefit sharing to avoid disadvantaging vulnerable groups. Finally, international ethical guidelines and regulatory frameworks tailored to repertoire research and ML applications, developed with broad stakeholder engagement, are urgently required to detect and mitigate harms and bias.

6. Conclusions

In summary, the intricate interplay between TCR repertoire dynamics and autoimmune pathogenesis underscores the profound complexity inherent in the breakdown of immune self-tolerance. Moreover, numerous disease-specific alterations manifest within the TCR repertoire, providing valuable signatures that can be harnessed for machine learning (ML)–based prediction. Indeed, recent advances have demonstrated tangible successes in employing ML models to predict autoimmune disorders. Our review has emphasized that failures in both central and peripheral tolerance—evidenced by defects in thymic selection, dysfunctions in regulatory T cells, and underlying genetic predispositions—collectively shape a pathogenic TCR landscape. This landscape is characterized by biased gene segment usage, prominent clonal expansions, and altered repertoire diversity. Importantly, these immunological insights establish a robust foundation for leveraging high-throughput sequencing datasets to train ML algorithms capable of discerning subtle, disease-specific TCR features. Specifically, the integration of deep learning architectures that combine complementarity-determining region 3 (CDR3) sequence motifs with gene usage patterns has yielded unprecedented predictive accuracy. Concurrently, the emergence of interpretable artificial intelligence approaches is beginning to elucidate the biological mechanisms underlying autoreactivity encoded within TCR repertoires. Nevertheless, significant challenges persist. These include the pressing need for larger and more standardized datasets, improved approaches for p*AIRE*d $\alpha\beta$-chain sequencing, and enhanced model generalizability alongside translational feasibility in clinical settings. Looking ahead, future research integrating multi-omics data, patient-specific genetic profiles, and longitudinal immune monitoring holds considerable promise to refine these predictive frameworks, thereby advancing the field of precision immunomedicine. Ultimately, the convergence of immunogenomics and machine learning points toward new opportunities for early, non-invasive diagnosis and personalized therapeutic strategies in autoimmune diseases, while also contributing to a deeper understanding of immune dysregulation. This integrated approach may open promising avenues for therapeutic innovation and improved patient outcomes.

Supplementary Materials

The following supporting information can be downloaded at: https://media.sciltp.com/articles/others/2603121434363717/TAI-26010046-Supplementary.pdf, Figure S1: Preliminary small-scale case study of TCR β repertoire features across autoimmune disease groups and healthy control groups.

Author Contributions

T.S.: Literature research, writing, and revision. M.H.: Literature research, writing, and revision. S.L.: supervision. All authors have read and agreed to the published version of the manuscript.

Acknowledgements and Funding

This work was supported by the National Natural Science Foundation of China (32270687).

Institutional Review Board Statement

The study did not involve humans or animals, and ethical review and approval were therefore not required for this research.

Informed Consent Statement

This study is a secondary analysis of publicly available anonymized datasets. The requirement for informed consent was therefore not applicable to this work, as the authors did not have any direct interaction with human subjects.

Data Availability Statement

All datasets analyzed in this study are publicly available from the immuneACCESS repository (Adaptive Biotechnologies). The datasets used are listed below:

- Mitchell et al.'s study [74]—https://clients.adaptivebiotech.com/pub/mitchell-2022-jcii
- Mustjoki et al.'s study [91]—https://clients.adaptivebiotech.com/pub/mustjoki-2017-natcomms
- Gold et al.'s study [92]—https://clients.adaptivebiotech.com/pub/gold-2019-cr
- Martinez et al.'s study [93]—https://clients.adaptivebiotech.com/pub/martinez-2025-s

These are previously published, publicly archived immunosequencing datasets; no new primary data were generated in this study. Processed analysis files and code used to produce the results are available from the corresponding author upon reasonable request.

Conflicts of Interest

The authors declare no conflict of interest.

Use of AI and AI-Assisted Technologies

During the preparation of this work, the authors used a large language model (ChatGPT, developed by OpenAI) and Grammarly to assist with language polishing and grammatical corrections. The authors are not native English speakers, and these tools were employed solely to improve the clarity and readability of the manuscript. After using these services, the authors thoroughly reviewed, revised, and take full responsibility for the final content of the publication.

Abbreviations

Abbreviation	Meaning
AD	Autoimmune diseases
CK	Control(s) check
SLE	Systemic lupus erythematosus
RA	Rheumatoid arthritis
MS	Multiple sclerosis
T1D	Type 1 diabetes
TCR	T-cell receptor
CDR3	Complementarity-determining region 3
V(D)J	Variable (Diversity) Joining recombination
mTEC	Medullary thymic epithelial cell
Treg	Regulatory T cell
ML	Machine learning
AUC	Area under the ROC curve
AIRR	Adaptive Immune Receptor Repertoire community
IEDB	Immune Epitope Database
VDJdb	Curated TCR–epitope database (VDJdb)
McPAS-TCR	Manually curated pathology-associated TCR catalogue
GTAI	Game-Theoretic Artificial Intelligence
SHAP	Shapley Additive ExPlanations

References

1. Pai, J.A.; Satpathy, A.T. High-throughput and single-cell T cell receptor sequencing technologies. *Nat. Methods* **2021**, *18*, 881–892.
2. Chuang, H.C.; Li, R.; Huang, H.; et al. Single-cell sequencing of full-length transcripts and T-cell receptors with automated high-throughput smart-seq3. *BMC Genom.* **2024**, *25*, 1127.
3. Zhang, Y.; Xu, Q.; Gao, Z.; et al. High-throughput screening for optimizing adoptive T cell therapies. *Exp. Hematol. Oncol.* **2024**, *13*, 113.
4. Qu, H.Q.; Kao, C.; Hakonarson, H. Single-cell RNA sequencing technology landscape in 2023. *Stem Cells* **2024**, *42*, 1–12.
5. Shah, K.; Al-Haidari, A.; Sun, J.; et al. T cell receptor (TCR) signaling in health and disease. *Signal Transduct. Target. Ther.* **2021**, *6*, 412.
6. Pageon, S.V.; Tabarin, T.; Yamamoto, Y.; et al. Functional role of T-cell receptor nanoclusters in signal initiation and antigen discrimination. *Proc. Natl. Acad. Sci. USA* **2016**, *113*, E5454–E5463.
7. Wang, Y.; Li, R.; Tong, R.; et al. Integrating single-cell RNA and T cell/B cell receptor sequencing with mass cytometry reveals dynamic trajectories of human peripheral immune cells from birth to old age. *Nat. Immunol.* **2025**, *26*, 308–322.
8. Zhao, L.; Wang, Q.; Yang, C.; et al. Application of single-cell sequencing technology in research on colorectal cancer. *J. Pers. Med.* **2024**, *14*, 108.
9. Huang, S.; Shi, W.; Li, S.; et al. Advanced sequencing-based high-throughput and long-read single-cell transcriptome analysis. *Lab Chip* **2024**, *24*, 2601–2621.
10. Qin, R.; Zhang, Y.; Shi, J.; et al. TCR catch bonds nonlinearly control CD8 cooperation to shape T cell specificity. *Cell Res.* **2025**, *35*, 265–283.

11. Gao, L.; Zhang, Y.; Ge, F.; et al. Structure-directed pan-specific T-cell receptor–peptide-major histocompatibility complex interaction prediction. *J. Chem. Inf. Model.* **2025**, *65*, 4674–4686.
12. Baker, T.C. Improving Detection and Quantification of Major Histocompatibility Complex (MHC)-Presented Immunopeptides for Vaccine Development. Ph.D. Thesis, University of British Columbia, Vancouver, BC, Canada, 2024.
13. Shi, Y. Comparative Analysis of TCR and TCR-pMHC Complex Structure Prediction Tools. Ph.D. Thesis, University of Tennessee, Knoxville, UK, 2024.
14. de Wit, A.S.; Bianchi, F.; van den Bogaart, G. Antigen presentation of post-translationally modified peptides in major histocompatibility complexes. *Immunol. Cell Biol.* **2025**, *103*, 161–177.
15. Barbosa, C.R.; Barton, J.; Shepherd, A.J.; et al. Mechanistic diversity in MHC class I antigen recognition. *Biochem. J.* **2021**, *478*, 4187–4202.
16. Aran, A.; Garrigós, L.; Curigliano, G.; et al. Evaluation of the TCR repertoire as a predictive and prognostic biomarker in cancer: Diversity or clonality? *Cancers* **2022**, *14*, 1771.
17. Joglekar, A.V.; Li, G. T cell antigen discovery. *Nat. Methods* **2021**, *18*, 873–880.
18. Malviya, M.; Aretz, Z.E.; Molvi, Z.; et al. Challenges and solutions for therapeutic TCR-based agents. *Immunol. Rev.* **2023**, *320*, 58–82.
19. Li, J.; Xiao, Z.; Wang, D.; et al. The screening, identification, design and clinical application of tumor-specific neoantigens for TCR-T cells. *Mol. Cancer* **2023**, *22*, 141.
20. Sidhom, J.W.; Larman, H.B.; Pardoll, D.M.; et al. DeepTCR is a deep learning framework for revealing sequence concepts within T-cell repertoires. *Nat. Commun.* **2021**, *12*, 1605.
21. Lu, T.; Zhang, Z.; Zhu, J.; et al. Deep learning-based prediction of the T cell receptor–antigen binding specificity. *Nat. Mach. Intell.* **2021**, *3*, 864–875.
22. Pisetsky, D.S. Pathogenesis of autoimmune disease. *Nat. Rev. Nephrol.* **2023**, *19*, 509–524.
23. Porsch, F.; Binder, C.J. Autoimmune diseases and atherosclerotic cardiovascular disease. *Nat. Rev. Cardiol.* **2024**, *21*, 780–807.
24. Shah, H.; Liu, Z.; Guo, W.; et al. Immune-regulating extracellular vesicles: A new frontier in autoimmune disease therapy. *Essays Biochem.* **2025**, *69*, 161–168.
25. Kumar, S.; Kaushik, D.; Sharma, S.K. Autoimmune Disorders: Types, Symptoms, and Risk Factors. In *Artificial Intelligence and Autoimmune Diseases: Applications in the Diagnosis, Prognosis, and Therapeutics*; Springer: Singapore, 2024; pp. 3–31.
26. Song, R.; Jia, X.; Zhao, J.; et al. T cell receptor revision and immune repertoire changes in autoimmune diseases. *Int. Rev. Immunol.* **2022**, *41*, 517–533.
27. He, J.; Shen, J.; Luo, W.; et al. Research progress on application of single-cell TCR/BCR sequencing technology to the tumor immune microenvironment, autoimmune diseases, and infectious diseases. *Front. Immunol.* **2022**, *13*, 969808.
28. Field, M.A. Detecting pathogenic variants in autoimmune diseases using high-throughput sequencing. *Immunol. Cell Biol.* **2021**, *99*, 146–156.
29. Jia, X.; Zhai, T.Y.; Wang, B.; et al. High-throughput T cell receptor sequencing reveals differential immune repertoires in autoimmune thyroid diseases. *Mol. Cell. Endocrinol.* **2022**, *550*, 111644.
30. Binson, V.; Thomas, S.; Subramoniam, M.; et al. A review of machine learning algorithms for biomedical applications. *Ann. Biomed. Eng.* **2024**, *52*, 1159–1183.
31. Strzelecki, M.; Badura, P. Machine learning for biomedical application. *Appl. Sci.* **2022**, *12*, 2022
32. Huang, Z.; Shi, Y.; Cai, B.; et al. MALDI-TOF MS combined with magnetic beads for detecting serum protein biomarkers and establishment of boosting decision tree model for diagnosis of systemic lupus erythematosus. *Rheumatology* **2009**, *48*, 626–631.
33. Chen, Y.; Huang, S.; Chen, T.; et al. Machine learning for prediction and risk stratification of lupus nephritis renal flare. *Am. J. Nephrol.* **2021**, *52*, 152–160.
34. Kockelbergh, H. Machine Learning Approaches for Diagnosis of Autoimmune Disease with the T-Cell Receptor Repertoire. Ph.D. Thesis, University of Liverpool, Liverpool, UK, 2024.
35. Yang, D.; Peng, X.; Zheng, S.; et al. Deep learning-based prediction of autoimmune diseases. *Sci. Rep.* **2025**, *15*, 4576.
36. Danieli, M.G.; Brunetto, S.; Gammeri, L.; et al. Machine learning application in autoimmune diseases: State of art and future prospectives. *Autoimmun. Rev.* **2024**, *23*, 103496.
37. Zhao, Q.; Jiang, Y.; Xiang, S.; et al. Engineered TCR-T cell immunotherapy in anticancer precision medicine: Pros and cons. *Front. Immunol.* **2021**, *12*, 658753.
38. Lin, P.; Lin, Y.; Mai, Z.; et al. Targeting cancer with precision: Strategical insights into TCR-engineered T cell therapies. *Theranostics* **2025**, *15*, 300.
39. Levi, R.; Louzoun, Y. Two step selection for bias in β chain VJ pairing. *Front. Immunol.* **2022**, *13*, 906217.
40. Mitchell, A.M.; Michels, A.W. T cell receptor sequencing in autoimmunity. *J. Life Sci.* **2020**, *2*, 38.
41. Foth, S.; Völkel, S.; Bauersachs, D.; et al. T cell repertoire during ontogeny and characteristics in inflammatory disorders in adults and childhood. *Front. Immunol.* **2021**, *11*, 611573.

42. Boehncke, W.H.; Brembilla, N.C. Autoreactive T-lymphocytes in inflammatory skin diseases. *Front. Immunol.* **2019**, *10*, 1198.
43. Amoriello, R.; Mariottini, A.; Ballerini, C. Immunosenescence and autoimmunity: Exploiting the T-cell receptor repertoire to investigate the impact of aging on multiple sclerosis. *Front. Immunol.* **2021**, *12*, 799380.
44. Liu, X.; Zhang, W.; Zhao, M.; et al. T cell receptor β repertoires as novel diagnostic markers for systemic lupus erythematosus and rheumatoid arthritis. *Ann. Rheum. Dis.* **2019**, *78*, 1070–1078.
45. Sansom, S.N.; Shikama-Dorn, N.; Zhanybekova, S.; et al. Population and single-cell genomics reveal the Aire dependency, relief from polycomb silencing, and distribution of self-antigen expression in thymic epithelia. *Genome Res.* **2014**, *24*, 1918–1931.
46. Yano, M.; Kuroda, N.; Han, H.; et al. Aire controls the differentiation program of thymic epithelial cells in the medulla for the establishment of self-tolerance. *J. Exp. Med.* **2008**, *205*, 2827–2838.
47. Akiyama, T.; Shinzawa, M.; Qin, J.; et al. Regulations of gene expression in medullary thymic epithelial cells required for preventing the onset of autoimmune diseases. *Front. Immunol.* **2013**, *4*, 249.
48. Wang, J.; Chitsaz, F.; Derbyshire, M.K.; et al. The conserved domain database in 2023. *Nucleic Acids Res.* **2023**, *51*, D384–D388.
49. ElTanbouly, M.A.; Noelle, R.J. Rethinking peripheral T cell tolerance: Checkpoints across a T cell's journey. *Nat. Rev. Immunol.* **2021**, *21*, 257–267.
50. Hatano, H.; Ishigaki, K. Functional genetics to understand the etiology of autoimmunity. *Genes* **2023**, *14*, 572.
51. Prinz, J.C. Immunogenic self-peptides-the great unknowns in autoimmunity: Identifying T-cell epitopes driving the autoimmune response in autoimmune diseases. *Front. Immunol.* **2023**, *13*, 1097871.
52. Agapiou, M. Revisiting Development and Homeostasis of Thymic Regulatory T Cells in Type 1 Diabetes. Ph.D. Thesis, University of York, York, UK, 2017.
53. Coder, B. Thymic Involution Perturbs Negative Selection and Leads to Chronic Inflammation. Ph.D. Thesis, University of North Texas Health Science Center at Fort Worth, Fort Worth, TX, USA, 2015.
54. Bainter, W.; Lougaris, V.; Wallace, J.G.; et al. Combined immunodeficiency with autoimmunity caused by a homozygous missense mutation in inhibitor of nuclear factor κB kinase alpha (IKKα). *Sci. Immunol.* **2021**, *6*, eabf6723.
55. Xue, Z.; Wu, L.; Tian, R.; et al. Integrative mapping of human CD8+ T cells in inflammation and cancer. *Nat. Methods* **2025**, *22*, 435–445.
56. Rojas, M.; Acosta-Ampudia, Y.; Heuer, L.S.; et al. Antigen-specific T cells and autoimmunity. *J. Autoimmun.* **2024**, *148*, 103303.
57. Marrero, I.; Aguilera, C.; Hamm, D.E.; et al. High-throughput sequencing reveals restricted TCR Vβ usage and public TCRβ clonotypes among pancreatic lymph node memory CD4+ T cells and their involvement in autoimmune diabetes. *Mol. Immunol.* **2016**, *74*, 82–95.
58. Oh, S. The Effect of T Cell Receptor Specificity on CD4+ CD25+ Regulatory T Cell Function in an Autoimmune Setting. Ph.D. Thesis, University of Pennsylvania, Philadelphia, PA, USA, 2010.
59. Layzell, S.J. The Role of IKK Dignalling in T Cells. Ph.D. Thesis, University College London, London, UK, 2022.
60. Sogkas, G.; Atschekzei, F.; Adriawan, I.R.; et al. Cellular and molecular mechanisms breaking immune tolerance in inborn errors of immunity. *Cell. Mol. Immunol.* **2021**, *18*, 1122–1140.
61. Sundaresan, B.; Shirafkan, F.; Ripperger, K.; et al. The role of viral infections in the onset of autoimmune diseases. *Viruses* **2023**, *15*, 782.
62. Heimli, M. Characterization of Regulatory T Cells in Autoimmune Polyendocrine Syndrome Type I, a Model Disease for Autoimmunity. Master's Thesis, The University of Bergen, Bergen, Norway, 2018.
63. Shokeen, N.; Saini, C.; Sapra, L.; et al. Role of Regulatory T Lymphocytes in Health and Disease. In *Systems and Synthetic Immunology*; Springer: Singapore, 2020; pp. 201–243.
64. Huang, F.; Sattler, S. *Regulatory T Cell Deficiency in Systemic Autoimmune Disorders-Causal Relationship and Underlying Immunological Mechanisms*; InTech: Rijeka, Croatia, 2011.
65. Bayley, R. Altered Leukocyte Signalling Thresholds in Rheumatoid Arthritis through Changes in the Function of the Protein Tyrosine Phosphatase PTPN22/LYP. Ph.D. Thesis, University of Birmingham, Birmingham, UK, 2014.
66. Li, Y.; Jiang, W.; Mellins, E.D. TCR-like antibodies targeting autoantigen-MHC complexes: A mini-review. *Front. Immunol.* **2022**, *13*, 968432.
67. Pratigya, G. Deciphering the Link between PTPN22 and Autoimmunity. Ph.D. Thesis, University of Birmingham, Birmingham, UK, 2011.
68. Macaulay, R. The Role of Immune Inhibitory Receptors in Age-Associated Immune Decline. Ph.D. Thesis, University College London, London, UK, 2011.
69. Weber, A. T Cell Receptor Specificity Profiling: A Machine Learning Approach. Ph.D. Thesis, ETH Zurich, Zurich, Switzerland, 2023.
70. Müller, L.; Pawelec, G.; Derhovanessian, E. The Immune System during Ageing. In *Diet, Immunity and Inflammation*; Elsevier: Amsterdam, The Netherlands, 2013, pp. 631–651.

71. Naumova, E.N.; Naumov, Y.N.; Gorski, J. Measuring Immunological Age: From T Cell Repertoires to Populations. In *Handbook of Immunosenescence*; Springer: Berlin/Heidelberg, Germany, 2018, pp. 1–62.
72. Attaf, M.; Huseby, E.; Sewell, A.K. $\alpha\beta$ T cell receptors as predictors of health and disease. *Cell. Mol. Immunol.* **2015**, *12*, 391–399.
73. Hou, X.; Chen, J.; Lu, C.; et al. The conserved T cell receptor repertoire observed in patients with systemic lupus erythematosus. *Int. J. Clin. Exp. Med.* **2017**, *10*, 2053–2065.
74. Mitchell, A.M.; Baschal, E.E.; McDaniel, K.A.; et al. Temporal development of T cell receptor repertoires during childhood in health and disease. *JCI Insight* **2022**, *7*, e161885.
75. Hou, X.; Wei, W.; Zhang, J.; et al. Characterisation of T and B cell receptor repertoire in patients with systemic lupus erythematosus. *Clin. Exp. Rheumatol.* **2023**, *41*, 2216–2223.
76. Sui, W.; Hou, X.; Zou, G.; et al. Composition and variation analysis of the TCR β-chain CDR3 repertoire in systemic lupus erythematosus using high-throughput sequencing. *Mol. Immunol.* **2015**, *67*, 455–464.
77. Ye, X.; Wang, Z.; Ye, Q.; et al. High-throughput sequencing-based analysis of T cell repertoire in lupus nephritis. *Front. Immunol.* **2020**, *11*, 1618.
78. Garrido-Mesa, J.; Brown, M.A. Antigen-driven T cell responses in rheumatic diseases: Insights from T cell receptor repertoire studies. *Nat. Rev. Rheumatol.* **2025**, *21*, 157–173.
79. Aterido, A.; López-Lasanta, M.; Blanco, F.; et al. Seven-chain adaptive immune receptor repertoire analysis in rheumatoid arthritis reveals novel features associated with disease and clinically relevant phenotypes. *Genome Biol.* **2024**, *25*, 68.
80. Turcinov, S.; af Klint, E.; Van Schoubroeck, B.; et al. Diversity and clonality of T cell receptor repertoire and antigen specificities in small joints of early rheumatoid arthritis. *Arthritis Rheumatol.* **2023**, *75*, 673–684.
81. Zhang, L.; Jiao, W.; Deng, H.; et al. High-throughput Treg cell receptor sequencing reveals differential immune repertoires in rheumatoid arthritis with kidney deficiency. *PeerJ* **2023**, *11*, e14837.
82. Amoriello, R. T-Cell Response in Relapsing-Remitting Multiple Sclerosis: A Computational Approach to T-Cell Receptor Repertoire Diversity before and during Disease-Modifying Therapies. Ph.D. Thesis, University of Florence, Firenze, Italy, 2020.
83. Dunlap, G.; Wagner, A.; Meednu, N.; et al. Clonal associations of lymphocyte subsets and functional states revealed by single cell. *bioRxiv* **2023**, 2023.03.18.533282.
84. Alves Sousa, A.P.; Johnson, K.R.; Ohayon, J.; et al. Comprehensive analysis of TCR-β repertoire in patients with neurological immune-mediated disorders. *Sci. Rep.* **2019**, *9*, 344.
85. Hayashi, F.; Isobe, N.; Glanville, J.; et al. A new clustering method identifies multiple sclerosis-specific T-cell receptors. *Ann. Clin. Transl. Neurol.* **2021**, *8*, 163–176.
86. Amoriello, R.; Chernigovskaya, M.; Greiff, V.; et al. TCR repertoire diversity in multiple sclerosis: High-dimensional bioinformatics analysis of sequences from brain, cerebrospinal fluid and peripheral blood. *EBioMedicine* **2021**, *68*, 103429.
87. Valkiers, S.; Dams, A.; Kuznetsova, M.; et al. Linking myelin and Epstein-Barr virus specific immune responses in multiple sclerosis: Insights from integrated public T cell receptor repertoires. *bioRxiv* **2024**, 2024-10.
88. Massey, J. Extensive Reshaping of the T Cell Repertoire Following Autologous Haematopoietic Stem Cell Transplantation in Multiple Sclerosis. Ph.D. Thesis, UNSW Sydney, Sydney, Australia, 2021.
89. Tong, Y.; Li, Z.; Zhang, H.; et al. T cell repertoire diversity is decreased in type 1 diabetes patients. *Genom. Proteom. Bioinform.* **2016**, *14*, 338–348.
90. Eugster, A.; Lindner, A.; Catani, M.; et al. High diversity in the TCR repertoire of GAD65 autoantigen-specific human CD4+ T cells. *J. Immunol.* **2015**, *194*, 2531–2538.
91. Savola, P.; Kelkka, T.; Rajala, H.; et al. Somatic mutations in clonally expanded cytotoxic T lymphocytes in patients with newly diagnosed rheumatoid arthritis. *Nat. Commun.* **2017**, *8*, 15869.
92. Ramien, C.; Yusko, E.C.; Engler, J.B.; et al. T cell repertoire dynamics during pregnancy in multiple sclerosis. *Cell Rep.* **2019**, *29*, 810–815.
93. Martinez Carmona, K.; Lothert, P.K.; Fedyshyn, B.; et al. Characterization of maternal and fetal immunity following in utero spina bifida repair and surgery-induced preterm birth. *Prenat. Diagn.* **2025**, *45*, 1816–1826.
94. Vita, R.; Mahajan, S.; Overton, J.A.; et al. The immune epitope database (IEDB): 2018 update. *Nucleic Acids Res.* **2019**, *47*, D339–D343.
95. Bagaev, D.V.; Vroomans, R.M.; Samir, J.; et al. VDJdb in 2019: Database extension, new analysis infrastructure and a T-cell receptor motif compendium. *Nucleic Acids Res.* **2020**, *48*, D1057–D1062.
96. Tickotsky, N.; Sagiv, T.; Prilusky, J.; et al. McPAS-TCR: A manually curated catalogue of pathology-associated T cell receptor sequences. *Bioinformatics* **2017**, *33*, 2924–2929.
97. Vander Heiden, J.A.; Marquez, S.; Marthandan, N.; et al. AIRR community standardized representations for annotated immune repertoires. *Front. Immunol.* **2018**, *9*, 2206.
98. Corrie, B.D.; Marthandan, N.; Zimonja, B.; et al. iReceptor: A platform for querying and analyzing antibody/B-cell and T-cell receptor repertoire data across federated repositories. *Immunol. Rev.* **2018**, *284*, 24–41.

99. Dibble, J.J.; Ferneyhough, B.; Roddis, M.; et al. Comparison of T-cell receptor diversity of people with myalgic encephalomyelitis versus controls. *BMC Res. Notes* **2024**, *17*, 17.
100. Fowler, A.; FitzPatrick, M.; Shanmugarasa, A.; et al. An interpretable classification model using gluten-specific TCR sequences shows diagnostic potential in coeliac disease. *Biomolecules* **2023**, *13*, 1707.
101. Ma, J.; Cui, C.; Tang, Y.; et al. Machine learning models developed and internally validated for predicting chronicity in pediatric immune thrombocytopenia. *J. Thromb. Haemost.* **2024**, *22*, 1167–1178.
102. Shen, T.; Huo, M.; Nie, W.; et al. DeepTAPE: Enhancing systemic lupus erythematosus diagnosis with deep learning based on TCRβ CDR3 sequences. In Proceedings of the 2024 IEEE International Conference on Bioinformatics and Biomedicine (BIBM), Lisboa, Portugal, 3–6 December 2024; pp. 1149–1154.
103. He, J.; Liu, Z.; Tang, X. A deep learning model for predicting systemic lupus erythematosus-associated epitopes. *BMC Med. Inform. Decis. Mak.* **2025**, *25*, 230.
104. Rawat, P.; Shapiro, M.R.; Peters, L.D.; et al. Identification of a type 1 diabetes-associated T cell receptor repertoire signature from the human peripheral blood. *Sci. Adv.* **2026**, *12*, eadx7448.
105. Geirhos, R.; Jacobsen, J.H.; Michaelis, C.; et al. Shortcut learning in deep neural networks. *Nat. Mach. Intell.* **2020**, *2*, 665–673.
106. Ostmeyer, J.; Christley, S.; Rounds, W.H.; et al. Statistical classifiers for diagnosing disease from immune repertoires: A case study using multiple sclerosis. *BMC Bioinform.* **2017**, *18*, 401.
107. Demerdash, O.N.; Smith, J.C. TCR-H: Explainable machine learning prediction of T-cell receptor epitope binding on unseen datasets. *Front. Immunol.* **2024**, *15*, 1426173.
108. Darmawan, J.T.; Leu, J.S.; Avian, C.; et al. MITNet: A fusion transformer and convolutional neural network architecture approach for T-cell epitope prediction. *Brief. Bioinform.* **2023**, *24*, bbad202.
109. Wang, M.; Fan, W.; Wu, T.; et al. TPEPret: A deep learning model for characterizing T-cell receptors-antigen binding patterns. *Bioinformatics* **2025**, *41*, btaf022.
110. Guo, S.; Wu, D.O. Game theoretical AI for precision medicine. *Trans. Artif. Intell.* **2025**, *1*, 170–196.
111. Walsh, L.A.; Quail, D.F. Decoding the tumor microenvironment with spatial technologies. *Nat. Immunol.* **2023**, *24*, 1982–1993.
112. Nagano, Y. Overcoming Data Bottlenecks in T Cell Receptor Specificity Prediction with Effective Machine Learning. Ph.D. Thesis, University College London, London, UK, 2024.
113. Weber, A.; Pélissier, A.; Martínez, M.R. T-cell receptor binding prediction: A machine learning revolution. *ImmunoInformatics* **2024**, *15*, 100040.
114. Zeng, Y.; Gao, Y.; He, L.; et al. Smart delivery vehicles for cancer: Categories, unique roles and therapeutic strategies. *Nanoscale Adv.* **2024**, *6*, 4275–4308.
115. Harris, J.C. Explainable Machine Learning in the Field of V(D)J Recombination. Ph.D. Thesis, The University of Oklahoma Health Sciences Center, Oklahoma City, OK, USA, 2025.
116. Tu, L.; Xu, A.; Lou, H.; et al. Unveiling fundamental principles: Visualizing T cell immunity with explainable artificial intelligence. *Med. Plus* **2025**, *2*, 100072.
117. Tan, C.L.; Lindner, K.; Boschert, T.; et al. Prediction of tumor-reactive T cell receptors from scRNA-seq data for personalized T cell therapy. *Nat. Biotechnol.* **2025**, *43*, 134–142.
118. Sun, H.; Han, X.; Du, Z.; et al. Machine learning for the identification of neoantigen-reactive CD8+ T cells in gastrointestinal cancer using single-cell sequencing. *Br. J. Cancer* **2024**, *131*, 387–402.
119. Milighetti, M. Analysis of T Cell Receptor Sequence and Structure to Understand the Drivers of Antigen Specificity. Ph.D. Thesis, University College London, London, UK, 2023.
120. Whalley, T. Novel Bioinformatics Tools for Epitope-Based Peptide Vaccine Design. Ph.D. Thesis, Cardiff University, Cardiff, UK, 2022.
121. Qi, F.; Huang, Q.; Xuan, Y.; et al. A roadmap for T cell receptor-peptide-bound major histocompatibility complex binding prediction by machine learning: Glimpse and foresight. *Brief. Bioinform.* **2025**, *26*, bbaf327.
122. Katayama, Y.; Kobayashi, T.J. Comparative study of repertoire classification methods reveals data efficiency of k-mer feature extraction. *Front. Immunol.* **2022**, *13*, 797640.
123. Katayama, Y.; Yokota, R.; Akiyama, T.; et al. Machine learning approaches to TCR repertoire analysis. *Front. Immunol.* **2022**, *13*, 858057.
124. Kidd, B.; Dudley, J. Systems Immunology. In *Translational Immunology: Mechanisms and Pharmacologic Approaches*; Elsevier: Amsterdam, The Netherlands, 2015; p. 1.
125. Vivas, A.J.; Boumediene, S.; Tobón, G.J. Predicting autoimmune diseases: A comprehensive review of classic biomarkers and advances in artificial intelligence. *Autoimmun. Rev.* **2024**, *23*, 103611.
126. Xu, X.; Li, J.; Zhu, Z.; et al. A comprehensive review on synergy of multi-modal data and AI technologies in medical diagnosis. *Bioengineering* **2024**, *11*, 219.
127. Baulu, E.; Gardet, C.; Chuvin, N.; et al. TCR-engineered T cell therapy in solid tumors: State of the art and perspectives. *Sci. Adv.* **2023**, *9*, eadf3700.

128. Zhang, J.; Wang, L. The emerging world of TCR-T cell trials against cancer: A systematic review. *Technol. Cancer Res. Treat.* **2019**, *18*, 1533033819831068.
129. Dhusia, K.; Su, Z.; Wu, Y. A structural-based machine learning method to classify binding affinities between TCR and peptide-MHC complexes. *Mol. Immunol.* **2021**, *139*, 76–86.
130. Deng, L.; Ly, C.; Abdollahi, S.; et al. Performance comparison of TCR-pMHC prediction tools reveals a strong data dependency. *Front. Immunol.* **2023**, *14*, 1128326.
131. Dens, C.; Bittremieux, W.; Affaticati, F.; et al. Interpretable deep learning to uncover the molecular binding patterns determining TCR–epitope interaction predictions. *ImmunoInformatics* **2023**, *11*, 100027.
132. Parr, T.; Bhat, A.; Zeidman, P.; et al. Dynamic causal modelling of immune heterogeneity. *Sci. Rep.* **2021**, *11*, 11400.