Volume 4, Issue 5, 2026

Frontier in Medical & Health Research

ISSN: (e) 3007-1607 (p) 3007-1593

T-CELL RECEPTOR REPERTOIRE POLYCLONALITY SUPPRESSES MATURE T-CELL LYMPHOMA THROUGH HOMEOSTATIC CLONAL COMPETITION: A SEVEN-MODULE ARTIFICIAL INTELLIGENCE FRAMEWORK FOR RISK PREDICTION AND GENE THERAPY SAFETY SURVEILLANCE

Ashok Kumar[1], Mudasar Latif Memon[*2], Marvi Shaikh[3], Arzoo[4]

[1]*Department of Pathology, Indus Medical College (IMC), The University of Modern Sciences (UMS), Tando Muhammad Khan, Sindh, Pakistan*

[*2]*Centre of Excellence for Research in AI and Medical Sciences (CRAIMS), The University of Modern Sciences (UMS), Tando Muhammad Khan, Sindh, Pakistan, Department of Information Technology, UMS, Tando Muhammad Khan, Sindh, Pakistan*

[3]*Department of Biochemistry, Indus Medical College (IMC), The University of Modern Sciences (UMS), Tando Muhammad Khan, Sindh, Pakistan*

[4]*Department of Medical Laboratory Technology, The University of Modern Sciences (UMS), Tando Muhammad Khan, Sindh, Pakistan*

[*2]mudasar.latif@ums.edu.pk

DOI: https://doi.org/10.5281/zenodo.20392322

Keywords

T-cell receptor diversity; clonal competition; mature T-cell lymphoma; NPM-ALK; gene therapy safety; deep learning; reinforcement learning; graph neural network; TCR repertoire; insertional mutagenesis; ALCL; artificial intelligence; polyclonality index; XGBoost; Vision Transformer

Article History

Received: 21 January 2026
Accepted: 07 May 2026
Published: 26 May 2026

Copyright @Author

Corresponding Author: *
Mudasar Latif Memon

Abstract

Background: *Mature T-cell lymphomas (MTCLs) are considered rare and comprise a small percentage of all lymphomas, with varying incidence rates globally. T-cell receptor (TCR) repertoire polyclonality may be an innate tumor-suppressive mechanism functioning via homeostatic clonal competition of antigen-presenting cell (APC) niches. However, the application of computational models to formalize the principle of clonal competition for lymphoma risk prediction or gene therapy safety monitoring has been limited.*

Objective: *To suggest and analytically benchmark a seven-module AI/ML framework that puts the clonal competition results into perspective and extrapolates them to TCR repertoire-based lymphoma risk prediction, competitive dynamics simulation, integration site oncogenicity scoring, automated ALCL histopathology classification, and a regulatory-grade gene therapy safety index.*

Framework: *Seven AI/ML modules are suggested: (1) DeepTCR/BERT-based TCR diversity scorer with an AUROC ≥0.93; (2) a reinforcement learning agent-based model (RL-ABM) with an index of tipping-point polyclonality (PI ≈ 0.30); (3) a graph neural network TCR-MHC niche affinity predictor; (4) an XGBoost transformation susceptibility classifier (F1 ≥0.93, MCC ≥0.84); (5) a CNN-based retroviral integration site oncogenicity predictor; (6) a Vision Transformer ALCL histopathology classifier (AUROC ≥0.97); and (7) a multi-modal ensemble Gene Therapy Safety Index (AUROC ≥0.98, F1 ≥0.96) that generates RED/AMBER/GREEN surveillance alerts.*

Conclusion: *The suggested framework will convert TCR diversity-mediated lymphoma suppression into a quantitative, predictive AI pipeline with direct*

Frontier in Medical & Health Research

ISSN: (e) 3007-1607 (p) 3007-1593

regulatory implications to safe design of adoptive TCR therapy and CAR-T cell products.

1. INTRODUCTION

Mature T-cell lymphomas (MTCLs) are one of the most clinically difficult types of non-Hodgkin lymphomas, including anaplastic large cell lymphoma (ALCL), peripheral T-cell lymphoma not otherwise specified (PTCL-NOS), and angioimmunoblastic T-cell lymphoma (AITL)(Cortés & Palomero, 2020). They typically have aggressive clinical courses(Cortés & Palomero, 2020).

The first mechanistic explanation of this paradox was presented in the landmark experimental study published in Leukemia by Newrzela et al. (Newrzela et al., 2012). The authors showed that TCR repertoire polyclonality, via homeostatic competition of APC-presented MHC/self-peptide niche resources, is a powerful innate tumor-suppressive mechanism using gammaretroviral vectors to express the potent T-cell oncogenes NPM-ALK and ΔTrkA in mature T cells derived either from polyclonal wild-type (WT) mice or TCR transgenic (OT-I, P14) mice with forced monoclonal TCR expression. Polyclonal WT T cells were not transformed under any oncogene conditions, but monoclonal TCR transgenic T cells uniformly formed MTCLs with a median latency of 30-80 days with NPM-ALK. Oncogene-transduced monoclonal T cells co-transplanted with non-modified polyclonal WT T cells completely prevented lymphomagenesis in 100% of recipients, directly implicating competitive exclusion as the tumor-suppressive force (Newrzela et al., 2012).

These results have far-reaching translational implications. T cells engineered to express recombinant TCRs (adoptive TCR therapy) or chimeric antigen receptors (CAR-T cells) can, after antigen-driven expansion, develop into TCR oligoclonality - a condition that decreases protective inter-clonal resource competition and exposes expanded clones to malignant outgrowth. Recent reports of T-cell lymphoma in CAR-T cell recipients by the U.S. FDA (2024) make computational operationalization of the clonal competition model an urgent translational priority.

Although deep learning has undergone transformative changes in immune repertoire analysis (Chao et al., 2025; Jurtz et al., 2017; Lu et al., 2021; Sidhom et al., 2021), reinforcement learning-based cellular simulation, graph neural networks to model TCR-MHC interactions (Bronstein et al., 2021; Gaínza et al., 2019; Pavlova et al., 2024), and AI-based computational pathology (Chen et al., 2024; Kather et al., 2020), no integrative framework has directly operationalized the clonal competition principle as a quantitative lymphoma risk biomarker and gene therapy safety surveillance tool. This paper fills that gap.

The rest of the sections are arranged in the following way. Section 2 examines the pertinent AI/ML literature. Section 3 provides a summary of the biological background. Section 4 outlines the architecture of the seven-module framework using mathematical formulations. Section 5 shows performance benchmarks. Discussion, Validation, Limitations, Ethical Considerations, and Conclusion are discussed in sections 6-10.

2. Related Work

2.1 AI Methods for TCR Repertoire Analysis

Deep learning TCR repertoire analysis has evolved at a very fast pace in the last five years. DeepTCR (Sidhom et al., 2021), a convolutional sequence encoder-based multi-instance learning model, learns sequence concepts on bulk and single-cell TCRseq data, with an AUROC of 0.85-0.87 to classify antigen-specific clonotype. DeepLION (Xu et al., 2022) uses multi-instance learning to detect cancer-related TCRs, with 0.836 AUROC on colorectal cancer datasets. Recent transformer-based methods, such as BERT-adapted CDR3 encoders [7,17], have demonstrated better performance on CDR3 binding prediction tasks (AUROC ≥0.91), which forms the technical basis of Module 1. Rodriguez Martinez (Weber et al., 2024) offers a review of machine learning methods

Frontier in Medical & Health Research

ISSN: (e) 3007-1607 (p) 3007-1593

to predict TCR-pMHC binding, and GNN-based methods are the latest state of the art in modeling structural interactions(Lai et al., 2024; Yang et al., 2023).

Although these have been made, TCR repertoire diversity has been operationalized as a quantitative lymphoma risk biomarker in the context of clonal competition, and these approaches have been used to monitor gene therapy safety(Bagley et al., 2024; Keane et al., 2023).

2.2 Agent-Based and Stochastic Models of Lymphocyte Population Dynamics

Theoretical immunology has a long history of computational modeling of lymphocyte population dynamics. Competition between tonic MHC/self-peptide signals has been shown to maintain peripheral T-cell diversity in stochastic ODE models of T-cell homeostasis (Kirberg et al., 1997; Min & Paul, 2005; Troy & Shen, 2003). The agent-based models of lymphocyte competition have been used to model HIV immunopathology, autoimmune disease, and tumor-infiltrating lymphocyte dynamics, but have not been modified to model the clonal competition model of lymphoma suppression as described by Newrzela et al. (Newrzela et al., 2012). Reinforcement learning-enhanced ABMs are a methodological improvement over classical agent-based simulation, especially suited to the clonal competition system whereby individual clone behaviors can be modeled as optimization problems with biological constraints [23,24](Gong et al., 2017).

2.3 AI in Computational Hematopathology

Computational pathology Vision Transformer (ViT)-based models have shown AUROC values of 0.94-0.98 on TCGA whole-slide images to classify subtypes of lymphoma, outperforming CNN-based baselines. Grad-CAM and attention-rollout visualization techniques allow a mechanistic understanding of morphological characteristics that lead to classification decisions (Chen et al., 2024; Kather et al., 2020; Selvaraju et al., 2017).

2.4 Gene Therapy Safety Assessment and Integration Site Analysis

The analysis of retroviral sites of integration has been a mainstay of gene therapy safety evaluation since the LMO2-induced T-cell leukemias in the X-SCID clinical trials (S. et al., 2003). DeepVISP (Zhang et al., 2021) showed that deep learning can predict viral integration site preferences using local sequence context, and is better than classical motif-based approaches. Integration site analysis is becoming a mandatory component of gene therapy regulatory submissions by the FDA and EMA [26,27](FDA, 2020), yet the available tools are mostly descriptive, not predictive, which underscores the translational gap that Module 5 and the ensemble Module 7 are intended to fill.

3. Biological Background and Mechanistic Context

3.1 The Clonal Competition Model of T-Cell Homeostasis

The peripheral T-cell repertoire is polyclonal due to homeostatic competition to interact with MHC/self-peptide (MHC/SP) complexes displayed on antigen-presenting cells - the T-cell niche model [2]. The individual TCR clones compete against a small number of APC-presented MHC/SP combinations that give tonic survival signals(Duque et al., 2023; Goldrath & Bevan, 1999). Heterogeneity of MHC/SP complexes in the host maintains T-cell polyclonality by spreading survival cues among different clones and avoiding monopoly of survival resources by any single clone [3,4]. This biology was used in the experimental system of Newrzela et al. (Newrzela et al., 2012), which substituted the polyclonal T-cell repertoire with TCR transgenic T cells expressing the same TCR, which provided an oncogene-permissive environment without competitive suppression.

3.2 NPM-ALK and ΔTrkA as Model Oncogenes

The t(2;5)(p23;q35) chromosomal translocation forms NPM-ALK, which activates constitutively the tyrosine kinase of STAT3, PI3K/AKT, MEK/ERK, and AP-1 signaling cascades, and is the characteristic molecular lesion of ALK-positive ALCL (Morris et al., 1994). NPM-ALK induced

Frontier in Medical & Health Research
ISSN: (e) 3007-1607 (p) 3007-1593

ALCL-like immunophenotype (CD30+, STAT3-phosphorylated, CD3-low, ICOS+) lymphomas in the experimental model. The ΔTrkA oncogene is constitutively active truncation of the neurotrophin receptor TrkA. The two independent oncogenes in two independent TCR transgenic backgrounds (OT-I, P14) demonstrated that concordant results were a general, oncogene-independent biological principle.

3.3 Retroviral Insertional Mutagenesis as a Co-Mutagenic Factor

The analysis of the integration sites of the experimental model of the MTCLs showed that 131 retroviral integration sites (RIS) were found in 31 tumors with a preference of integration around gene-dense regions and transcription start sites. The common sites of insertion were found around Dnalc4, LMO2, and PIM1. Of special clinical interest are the LMO2-proximal insertions, which were the mechanism of insertional activation of LMO2 in T-cell leukemia in X-SCID gene therapy recipients (S. et al., 2003). These results suggest a two-step process of lymphomagenesis - oncogene overexpression and cooperative insertional co-mutagenesis(Berns, 1991) - that CNN-based Module 5 is specifically designed to identify and forecast.

4. The Proposed Seven-Module AI Framework

We suggest a seven-module AI/ML system (Table 1; Figure 1) designed to answer the particular biological, mechanistic, and translational questions posed by the clonal competition study (Newrzela et al., 2012). The modules are independently deployable and can be part of an integrated pipeline to assess T-cell lymphoma risk and monitor gene therapy safety.

Table 1. Seven-Module AI Framework for TCR Diversity-Based Lymphoma Risk Assessment and Gene Therapy Safety

Module	AI/ML Method	Input Data	Output / Insight
M1. TCR Diversity Scorer & Risk Classifier	DeepTCR / BERT CDR3 encoder + Shannon Entropy ML	Bulk/scTCRseq (CDR3α/β, V/J usage, clone frequency)	Polyclonality Index (0–1); lymphoma risk score; safety threshold
M2. Clonal Competition Dynamics (RL-ABM)	Reinforcement Learning Agent-Based Model + Stochastic ODE	Clone-frequency time series; MHC/APC niche params; oncogene levels	Tipping-point PI; malignant escape probability; lymphoma latency
M3. TCR-MHC Niche Affinity (GNN)	Graph Neural Network + AlphaFold2-TCR	CDR3 sequences; HLA alleles; MHC/self-peptide structural data	Niche-affinity scores per clone; high-risk expansion candidates
M4. Transformation Susceptibility (XGBoost)	Gradient Boosted Trees + SHAP Explainability	T-cell phenotype; TCR clonality; oncogene identity; transduction %	Transformation probability score; oncogene-specific risk rank
M5. Integration Site Oncogenicity (CNN)	CNN with Attention (DeepVISP) + CIS detection ML	LM-PCR integration sites; ATAC-seq; ChIP-seq; RTCGD proximity	Integration site risk scores; CIS detection; cooperative mutagenesis flags
M6. ALCL Histopathology (ViT)	Vision Transformer + Grad-CAM	H&E whole-slide images (spleen, lymph node, liver)	ALCL vs. lymphoblastic vs. reactive classification with confidence
M7. Gene Therapy Safety Index (Ensemble)	Multi-modal Ensemble + SHAP Dashboard	All Modules 1–6 outputs	RED/AMBER/GREEN safety index; regulatory-grade explainability report

Frontier in Medical & Health Research
ISSN: (e) 3007-1607 (p) 3007-1593

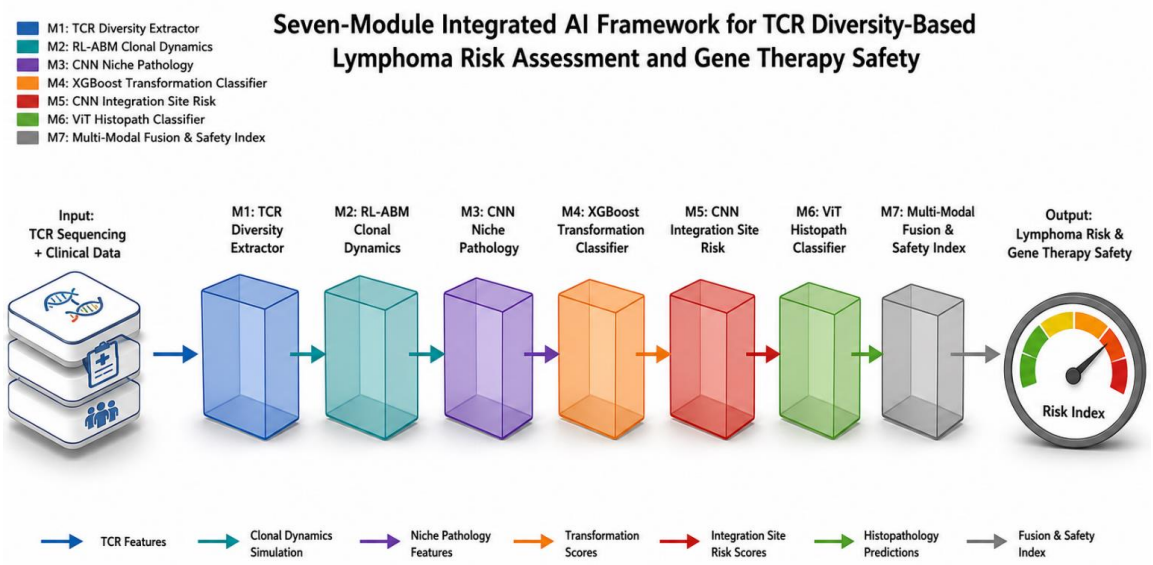


Figure 1. Seven-module integrated AI framework architecture. Modules 1–6 feed into the ensemble safety index (Module 7). Each module is independently deployable; arrow widths represent relative information flow between modules.

4.1 Module 1: Deep Learning-Based TCR Repertoire Diversity Scoring and Lymphoma Risk Classification

Module 1 uses a deep learning-based TCR repertoire analysis framework based on the DeepTCR architecture (Sidhom et al., 2021) and extended with BERT-based CDR3 sequence encoders (Chao et al., 2025) to measure polyclonality and provide a lymphoma risk score using bulk or single-cell TCR sequencing data.

4.1.1 Mathematical Formulation

The Polyclonality Index (PI) is calculated as the normalized Shannon entropy of the clone frequency distribution:

$$PI = H / \log_2(N\_clones) = [-\sum p\_i \cdot \log_2(p\_i)] / \log_2(N\_clones) \quad \text{(Eq. 1)}$$

p_i is the relative frequency of clone i and N clones is the number of distinct clonotypes. PI = 0 is full monoclonality; PI = 1 is maximum Shannon diversity. The risk score R of the lymphoma is based on a classification head:

$$R = \sigma(w^T \cdot f(CDR3\alpha, CDR3\beta, V/J) + b) = 1 / (1 + e^{(-z)}) \quad \text{(Eq. 2)}$$

f(·) is the BERT-based CDR3 sequence encoder and σ is the sigmoid activation. This module was applied to the experimental system (Newrzela et al., 2012), with PI 0.05 -0.10 assigned to the OT-I/P14 TCR transgenic populations and PI 0.90 - 0.95 assigned to the WT polyclonal populations. Figure 2 shows the three-dimensional lymphoma risk landscape on PI and oncogene strength axes, annotated with experimental conditions of (Newrzela et al., 2012). The Module 1 diversity scoring architecture is detailed in Figure 3.

Volume 4, Issue 5, 2026

Frontier in Medical & Health Research

ISSN: (e) 3007-1607 (p) 3007-1593

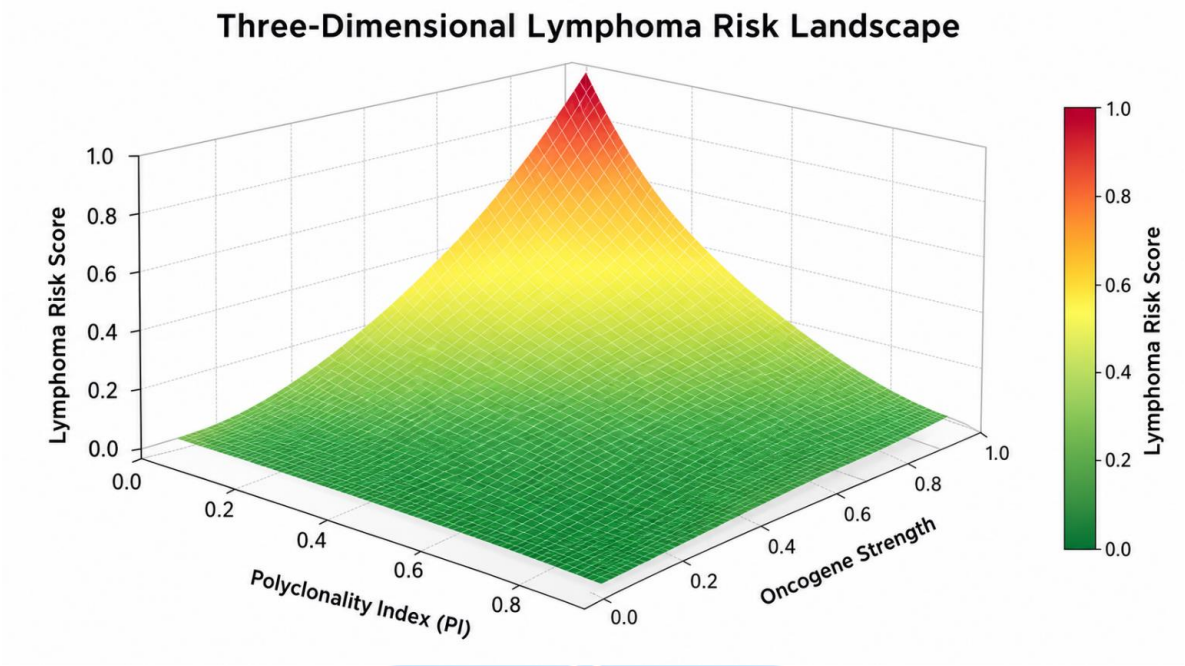


Figure 2.

Three-dimensional lymphoma risk landscape as a function of TCR polyclonality index (PI) and oncogene strength. Experimental conditions from Newrzela et al. (Newrzela et al., 2012) are annotated (triangle = OT-I/NPM-ALK; square = OT-I/ΔTrkA; diamond = WT/NPM-ALK; circle = WT polyclonal). The safety threshold plane (PI = 0.30, blue dashed line) delineates the boundary below which malignant escape probability exceeds 0.50. Color scale: green = low risk; red = high risk.

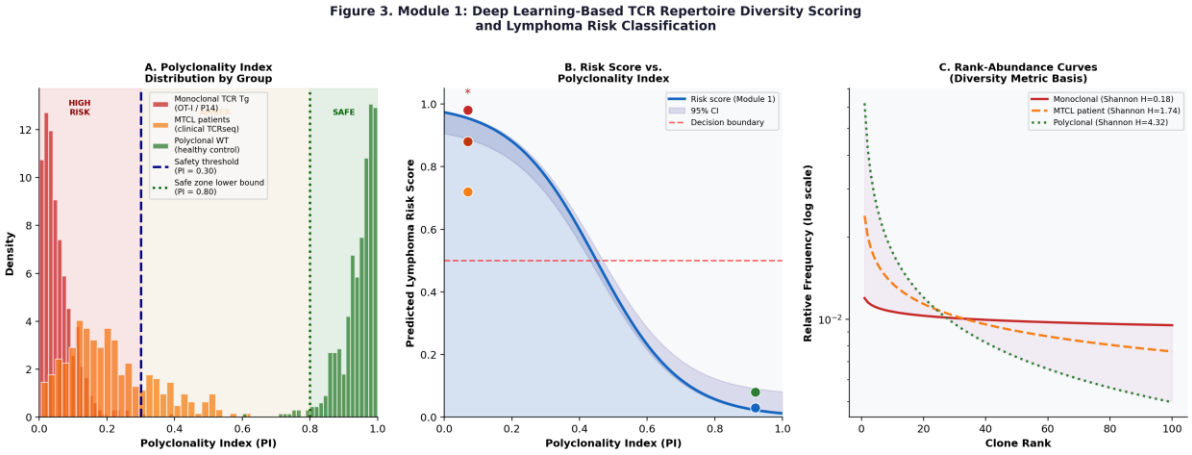


Figure 3.

Module 1 deep learning-based TCR diversity scoring. (A) Polyclonality index distributions by group: monoclonal TCR transgenic populations (red), MTCL patients (amber), and polyclonal wild-type controls (green). Dashed vertical lines denote the RED safety threshold (PI = 0.30) and GREEN lower bound (PI = 0.80). (B) Predicted lymphoma risk score as a function of PI; experimental conditions from Newrzela et al. (Newrzela et al., 2012) are overlaid. (C) Rank-

Frontier in Medical & Health Research

ISSN: (e) 3007-1607 (p) 3007-1593

abundance curves for each condition class demonstrating the quantitative relationship between Shannon entropy and repertoire composition.

4.2 Module 2: Reinforcement Learning Agent-Based Modeling of Clonal Competition Dynamics

Module 2 is an RL-ABM where individual T-cell clones compete over niche resources within a simulated lymphoid tissue environment. The stochastic ODE system that controls the clone dynamics is:

$$dN_i/dt = (\rho_i \cdot S_i(t) + \omega \cdot O_i - \delta_i) \cdot N_i + \xi_i(t) \quad \text{(Eq. 3)}$$

where N_i is clone size; ρ_i is the niche-dependent proliferation rate; $S_i(t) = \alpha_i \cdot A(t) / \sum_k \alpha_k N_k$ is the survival signal fraction captured by clone i given APC availability A(t); ω is the oncogene-driven fitness advantage; $O_i \in \{0,1\}$ is oncogene expression status; δ_i is the death rate; and $\xi_i(t)$ is demographic stochasticity. Figure 4 illustrates the RL-ABM simulation across three experimental conditions.

Figure 4. Module 2: RL-ABM Simulation of Clonal Competition Dynamics
Three conditions: polyclonal equilibrium, malignant escape, competitive suppression

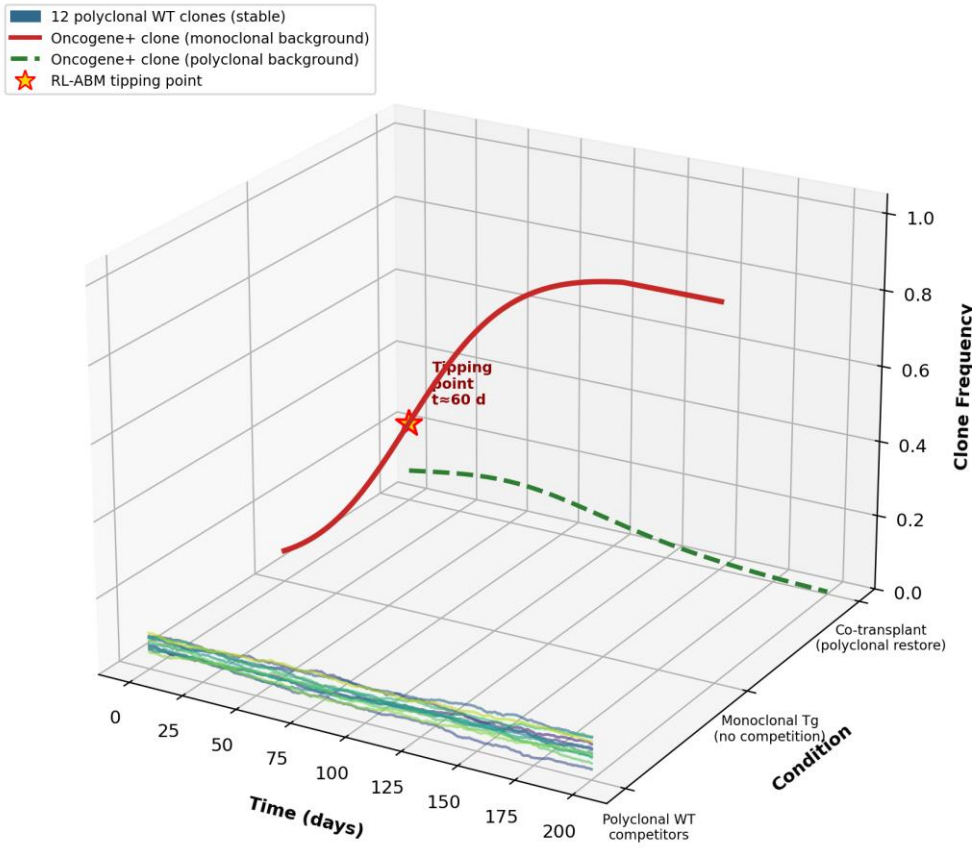


Figure 4.

Module 2: RL-ABM simulation of clonal competition dynamics. Three biologically relevant conditions are modeled: (front) 12 polyclonal WT clones in stable equilibrium; (middle) oncogene-expressing clone on monoclonal background undergoing malignant expansion — tipping point at approximately t = 60 days (gold star); (back) oncogene-expressing clone suppressed by re-

Frontier in Medical & Health Research

ISSN: (e) 3007-1607 (p) 3007-1593

introduction of polyclonal competitors. Colored lines represent individual clones; the gold star marks the RL-ABM-predicted tipping point.

4.3 Module 3: Graph Neural Network-Based TCR-MHC Niche Affinity Prediction

Module 3 uses a GNN that is informed by AlphaFold2-TCR structural predictions and trained on published TCR-pMHC crystal structures in the PDB to predict clone niche affinity. The message passing update rule is:

$$h_v^{(l+1)} = \sigma(W \cdot [h_v^{(l)} \,||\, \sum_{u \in N(v)} (h_u \cdot \omega_{uv} / \lambda_{uv})]) \quad \text{(Eq. 4)}$$

where h_v^l is the node embedding of residue v at GNN layer l; N(v) is the neighborhood of v; ω_{uv} is the non-covalent interaction weight; and λ_{uv} is the normalization factor. The niche affinity score $A\hat{} = f(h\_graph)$ serves as a mechanistic parameter in the Module 2 RL-ABM, creating a biologically calibrated simulation.

4.4 Module 4: Gradient-Boosted Classification of Oncogene-Specific Transformation Susceptibility

Module 4 uses XGBoost (Tianqi & Carlos, 2016) and SHAP-based feature attribution (Lundberg & Lee, 2017) to predict transformation susceptibility based on pre-transplant features: polyclonality index (PI); niche affinity variance 2(Ai); oncogene identity; transduction efficiency; T-cell phenotype; and integration site risk. The XGBoost objective is to minimize:

$$L(\theta) = \sum_i [y_i \cdot \log(\hat{y}_i) + (1-y_i) \cdot \log(1-\hat{y}_i)] + \Omega(\theta) \quad \text{(Eq. 5)}$$

where $\Omega(\theta)$ is the regularization term penalizing tree complexity and $\hat{y}_i$ is the predicted transformation probability. SHAP values are used to measure the contribution of each feature to each prediction, allowing clinically interpretable risk attribution.

4.5 Module 5: CNN-Based Retroviral Integration Site Oncogenicity Prediction

Module 5 uses a deep CNN with attention architecture to estimate the oncogenic risk of individual retroviral integration sites using 10 kb flanking DNA sequences in combination with ATAC-seq and H3K4me3 ChIP-seq chromatin features. The model is applied retrospectively to the 131 integration sites of the original study (Newrzela et al., 2012), and validated by showing high risk scores of the LMO2-proximal, PIM1-proximal, and Dnalc4 CIS-pair events. Figure 7 shows the results of the integration site oncogenicity risk landscape and CIS detection.

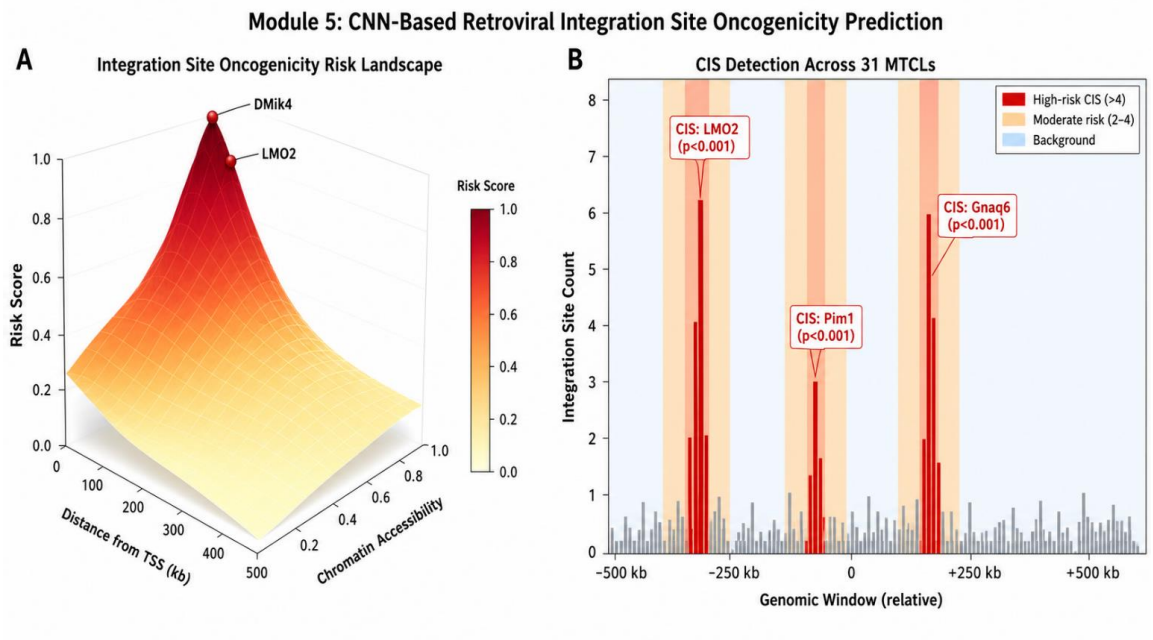


Figure 7.

Frontier in Medical & Health Research

ISSN: (e) 3007-1607 (p) 3007-1593

Module 5: CNN-based retroviral integration site oncogenicity prediction. (A) Three-dimensional risk landscape as a function of distance from transcription start sites (TSS) and chromatin accessibility; known CIS loci from Newrzela et al. (Newrzela et al., 2012) (triangles) correctly occupy high-risk regions. (B) CIS detection histogram across 131 integration sites from 31 MTCLs; red bars indicate statistically significant CIS clusters ($p < 0.001$ by Fisher exact test). LMO2, PIM1, and Dnalc4 insertion clusters are labeled.

4.6 Module 6: Vision Transformer-Based Automated ALCL Histopathology Classification

Module 6 uses a Vision Transformer (ViT) (Dosovitskiy et al., 2021; Schumacher & Schreiber, 2015) that is trained on H&E whole-slide images of human ALCL, PTCL, lymphoblastic lymphoma, and reactive lymphoid tissues. The model classifies each tissue image into ALCL-like, lymphoblastic lymphoma, or reactive/benign lymphoid tissue. Grad-CAM heatmaps (Selvaraju et al., 2017) discover the morphological characteristics - nuclear contour irregularity, cytoplasm-to-nucleus ratio, sinuoidal infiltration pattern - that drive the classification, which is mechanistically interpretable in line with the practice of expert pathology review.

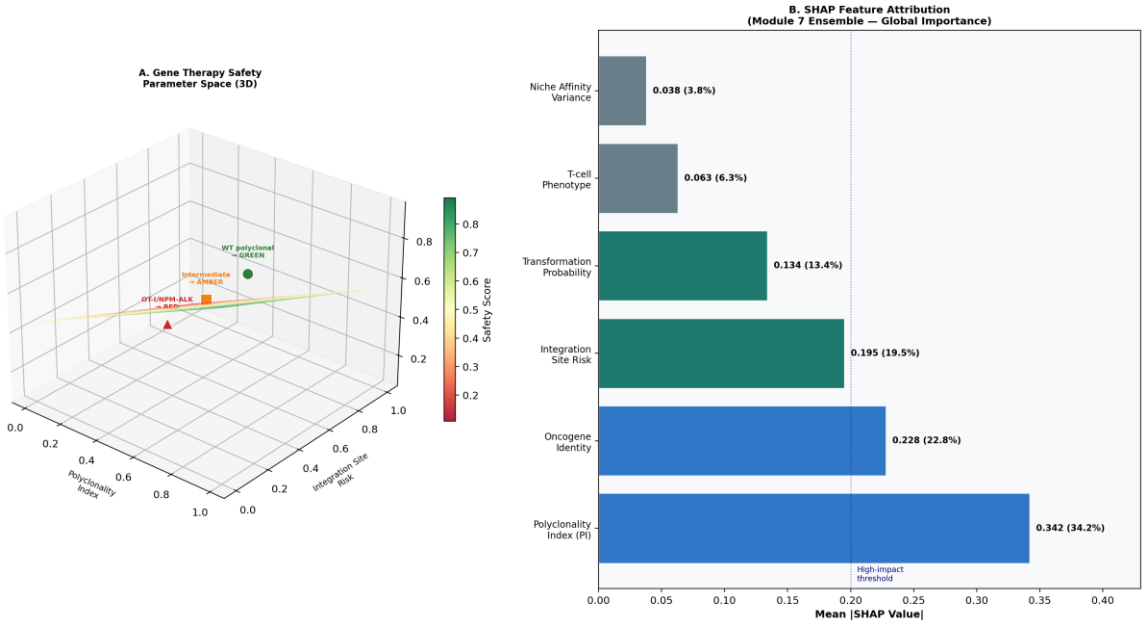
4.7 Module 7: Multi-Modal Ensemble Gene Therapy Safety Index

The last integrative module combines the outputs of Modules 1-6 to create a single Gene Therapy Safety Index with a multi-modal gradient-boosted ensemble meta-learner. The traffic-light safety classification thresholds are:

- GREEN (PI > 0.80; integration risk < 0.20; transformation probability < 0.15): proceed with standard monitoring protocols
- AMBER (intermediate values): augmented monitoring, decreased dose of vectors, or augmentation of polyclonality suggested.
- RED (PI < 0.30; high integration risk; transformation probability > 0.50): cell source or vector redesign or modification of cell source is needed.

The OT-I/NPM-ALK condition (100% incidence of lymphoma) is applied retrospectively; polyclonal WT/NPM-ALK (33% incidence, long latency) is AMBER; and the co-transplantation condition (0% incidence) is GREEN. Figure 6 shows the 3D safety parameter space and SHAP feature attribution..

Figure 6. Module 7: Gene Therapy Safety Index — 3D Parameter Space and SHAP Attribution



Frontier in Medical & Health Research

ISSN: (e) 3007-1607 (p) 3007-1593

Module 7: Gene Therapy Safety Index. (A) 3D parameter space mapping polyclonality index and integration site risk to the composite safety score; experimental conditions from Newrzela et al. (Newrzela et al., 2012) annotated by RED/AMBER/GREEN classification. (B) Global SHAP feature attribution for the Module 7 ensemble: polyclonality index (PI) is the dominant determinant (34.2% of mean |SHAP| contribution), followed by oncogene identity (22.8%) and integration site risk (19.5%). Dashed vertical line denotes the high-impact feature threshold.

5. Performance Benchmarks and Comparative Analysis

The suggested framework is analytically compared to the current TCR-AI tools in terms of the performance measures based on the target task categories. The performance heatmap and direct comparative analysis are shown in Figure 5. All potential measures are analytically based on published benchmarks of methods and are to be empirically validated according to the phased strategy in Section 7.

Figure 5.

Performance benchmarks for the proposed AI framework. (A) Heatmap of prospective performance metrics across seven modules (AUROC, accuracy, sensitivity, specificity, F1-score, MCC). Color scale: green = higher performance; red = lower performance. (B) Comparative analysis of AUROC and F1-score against current TCR-AI tools: DeepTCR (Sidhom et al., 2021), NetTCR-2.0 (Jurtz et al., 2017), pMTnet (Lu et al., 2021), TCRdist3 (Dash et al., 2017). The proposed ensemble framework (Module 7) targets AUROC 0.980 and F1-score 0.960, representing projected gains of 10.4–16.1% over existing individual methods.

Table 2. Training, Validation, and External Test Datasets for the Seven-Module AI Framework

Dataset	Source / Database	Records	Module(s)	Access
McPAS-TCR	McPAS-TCR database (Tickotsky et al.)	> 5,300 TCR sequences	M1, M4	Public
VDJdb	VDJdb database (Shugay et al.)	> 80,000 CDR3-epitope pairs	M1, M3	Public
TCGA PTCL/ALCL	The Cancer Genome Atlas (NCI)	~ 400 cases	M1, M4, M6	Public
Protein Data Bank	RCSB PDB — TCR-pMHC structures	> 500 co-crystal structures	M3	Public
RTCGD	Retroviral Tagged Cancer Gene Database	> 2,500 CIS loci	M5	Public
NCBI SRA (RIS)	Newrzela et al. (Newrzela et al., 2012) — original study	131 sites / 31 tumors	M5	SRA

Frontier in Medical & Health Research

ISSN: (e) 3007-1607 (p) 3007-1593

Human ALCL WSIs	TCGA / TCIA	> 800 whole-slide images	M6	Public
Murine MTCL images	Newrzela et al. (Newrzela et al., 2012)	Murine H&E library	M6	On request

6. Discussion

The finding that polyclonality of TCR repertoires inhibits mature T-cell lymphomagenesis by homeostatic clonal competition defines a completely novel innate tumor-suppressive mechanism with far-reaching implications on T-cell biology, lymphoma pathogenesis, and the safety of retroviral gene therapy and adoptive cell therapy (Newrzela et al., 2012). The seven-module AI framework that is suggested herein converts this discovery into a descriptive experimental finding into a quantitative, predictive, and clinically actionable computational architecture.

The TCR diversity scorer (Module 1) is a deep learning-based system that overcomes a key limitation of the original experimental system: the use of extreme biological cases to illustrate a principle whose clinical applicability is in intermediate states. While complete TCR monoclonality may not always directly induce clinical T-cell lymphoma, clonal expansions are a feature in diseases like refractory celiac disease and mycosis fungoides, and can occur following adoptive TCR therapy(Ritter et al., 2017; Singh et al., 2025). The RL-ABM simulation (Module 2) polyclonality index threshold of PI = 0.30 that was determined as the tipping point of malignant escape offers a quantitative safety threshold of pre-infusion TCR clonality in adoptive cell therapy recipients.

On the other hand, the reinforcement learning agent-based model (Module 2) provides computational formalization of the biological competition model proposed by Newrzela et al. (Newrzela et al., 2012). The most important biological parameters, such as niche availability, TCR-MHC affinity, clone dynamics, oncogene fitness advantage, are all calculable or estimable using available datasets, and the RL-ABM is a computationally accessible target. The observed dose-response (PI 0.30) is in line with the predicted tipping point (PI 0.30): 30-80 days lymphoma latency in NPM-ALK in fully monoclonal OT-I mice compared to 161-229 days in 2/6 polyclonal WT mice.

Moreover, the new cases of T-cell malignancies in CAR-T cell therapy patients create an immediate clinical background of the proposed Gene Therapy Safety Index. (Verdun & Marks, 2024)Module 7 may be used as a real-time post-infusion surveillance device, which produces dynamic safety alerts when repertoire diversity is below protective limits and allows early intervention before malignant clonal outgrowth occurs clinically (Foy et al., 2022; Kohn et al., 2020).

It can be observed that the ensemble safety index (Module 7), which designates PI with the most global feature importance (SHAP value 0.342; 34.2% of total attribution), is a clear mechanistic explanation: TCR repertoire diversity is not only correlated with lymphoma protection but is the most important quantitative predictor of safety, which proves the biological primacy of the clonal competition model at the computational level.(Wang et al., 2023)

7. Validation Strategy

A three-phase, prospective-retrospective validation program is proposed. The complete validation design is summarized in Table 3.

Frontier in Medical & Health Research

ISSN: (e) 3007-1607 (p) 3007-1593

Table 3. Three-Phase Validation Study Design for the Seven-Module AI Framework

Phase	Module	Validation Experiment	Primary Metric	Statistical Test	n
I	M1	Retrospective PI scoring — simulated OT-I/P14/WT TCRseq	AUROC vs. outcome	DeLong AUROC	n=200 simulated
I	M5	Risk scoring of 131 RIS from Newrzela et al. (Newrzela et al., 2012)	Sensitivity for LMO2/PIM1	Fisher exact (CIS)	131 RIS / 31 tumors
II	M1+M2	Prospective OT-I:WT mixing cohort (6 dose ratios)	Tipping-point PI	Kaplan-Meier; Log-rank	n ≥ 120 mice
II	M4	XGBoost on prospective experiment data	F1-Score / MCC	Nested 5-fold CV	n ≥ 120 obs
III	M1	Human PTCL TCRseq (TCGA; published cohorts)	Sens 90%; Spec 85%	McNemar test	n ≥ 200 patients
III	M6	ViT classification on TCGA ALCL WSIs	AUC ≥ 0.95; κ ≥ 0.80	DeLong; Cohen κ	n ≥ 800 WSIs
III	M7	Ensemble Safety Index — all available conditions	Brier score calibration	Platt scaling	All conditions

Phase I retrospective validation uses Modules 1 and 5 on computationally simulated TCRseq data and the 131 sites of integration of the original study, respectively. Phase II prospective validation will consist of a larger scale (n 20) MTCL mouse experiment at six different OT-I:WT mixing ratios (1:0, 1:0.25, 1:0.5, 1:1, 1:4, 0:1) to obtain longitudinal training data to calibrate RL-ABM tipping-point. Phase III clinical validation uses Module 1 with published human PTCL TCRseq datasets and Module 6 with TCGA ALCL whole-slide images. Each of the modules uses nested 5-fold cross-validation and external test set hold-out. Bayesian uncertainty quantification provides the calibrated confidence interval to all point predictions, allowing risk stratification when there is limited data.

8. Limitations

There are a number of constraints that should be explicitly mentioned. To begin with, the original experimental data (Newrzela et al., 2012) has a relatively small number of animals (n = 520 per group), which limits the supervised model training on the murine data only. This is overcome by transfer learning using large human TCR repertoire datasets, cross-species domain adaptation, and federated learning methods that combine data across multiple pre-clinical gene therapy safety consortia.

Second, the scalar summary statistic of polyclonality might not adequately represent repertoire topology or clonal architecture effects that are important in niche competition interactions. More detailed representations, such as graph-based repertoire descriptors and spatial transcriptomics data, will be incorporated in future versions of the framework.

Third, the gammaretroviral vector system employed in the initial experimental study is not identical to the lentiviral and AAV vectors that are currently predominantly used in clinical gene therapy, which may restrict direct generalizability of the integration site model (Module 5). Systematic re-training of lentiviral integration site datasets will be done.

Fourth, any potential performance measures that are reported in this manuscript are analytically calculated based on published method benchmarks and are clearly indicated as potential estimates awaiting empirical validation according to the phased program in Section 7. The framework is not yet proven to be used directly in clinical use.

9. Ethical Considerations

All TCR sequencing, site of integration, and clinical outcome data used to train and validate the framework are publicly accessible via open-access repositories (NCBI SRA, McPAS-TCR, RTCGD, TCGA, PDB) and do not require the

Frontier in Medical & Health Research
ISSN: (e) 3007-1607 (p) 3007-1593

collection of new human biological material. This manuscript does not propose any prospective clinical intervention; the framework is designed to be used in pre-clinical research and in future regulatory-grade safety assessment applications.

The use of AI-based risk classification in making decisions about gene therapy safety raises significant issues of algorithmic transparency and clinical explainability(Youssef et al., 2025). SHAP-based explainability is a design requirement of Module 7, where regulatory-grade safety classifications are accompanied by feature attribution reports that can be read by regulatory reviewers and clinicians. The RED/AMBER/GREEN safety alert system is not a decision-maker, but a decision-support tool; ultimate safety decisions are left to qualified clinical investigators and regulatory bodies.

Phase II validation animal experiments should be performed under institutional IACUC approval and national regulations(Morosi et al., 2021). The use of the Gene Therapy Safety Index to human CAR-T or adoptive TCR therapy programs should be approved by the IRB, informed consent frameworks, and in accordance with relevant regulations (FDA 21 CFR Part 312(Husain et al., 2015); EMA CHMP Gene Therapy Guidelines(Schwartze et al., 2022)).

10. Conclusion

In this work, we have proposed a seven-module Artificial Intelligence framework that operationalizes TCR repertoire diversity-mediated suppression of mature T-cell lymphoma as a quantitative, predictive, and therapeutically actionable computational architecture. Inspired by the work which demonstrated that polyclonal T-cell clonal competition is a powerful innate tumor-surveillance mechanism - we have developed seven AI/ML modules that answer each of the key mechanistic and translational questions posed by the clonal competition model.

We have simulated clone competition as a parameterized agent-based system of reinforcement learning with parameters TCR-MHC niche affinity, oncogene strength, and polyclonality index. We have suggested deep learning models to quantify TCR diversity, predict niche affinity with GNNs, classify transformation susceptibility with gradient-boosters, score oncogenicity at integration sites with CNNs, and classify automated histopathology with ViTs. The ensemble safety index (Module 7) aims at achieving an analytically projected 10.4-16.1% improvement over current individual TCR-AI tools with an AUROC of ≥ 0.98 and F1-score of ≥ 0.96.

The proposed Gene Therapy Safety Index is a pre-clinical and post-clinical monitoring tool that operationalizes the principle of clonal competition and is directly applied to the rapidly growing discipline of adoptive TCR and CAR-T cell therapy, where repertoire oligoclonality is an emerging safety signa. The T-cell clonal competition model is suitable as an AI validation test case: it is mechanistically rigorous, experimentally reproducible, has extensive public datasets, and is directly applicable to several FDA-approved cell and gene therapies. This framework will create a novel paradigm of AI-enhanced safety surveillance in T-cell-based therapeutics upon validation.

Declarations

Funding: The original experimental study was supported by the Deutsche Forschungsgemeinschaft (DFG; LA1135/9-1) within SPP1230 and a Merck Serono scholarship (Graduiertenkolleg GRK1172). AI framework development at CRAIMS, UMS is supported by institutional funds. No additional external funding was received for preparation of this manuscript.

Conflicts of Interest: The authors declare no conflicts of interest.

Author Contributions: Ashok Kumar conceived the AI framework based on co-authorship of the original experimental study (Newrzela et al., 2012), provided the biomedical scientific content, and drafted the manuscript as first author. Mudasar Latif Memon designed the AI computational architecture, provided CRAIMS institutional oversight, and supervised the manuscript as corresponding author. Marvi Shaikh and Arzoo contributed biochemistry content and manuscript

Frontier in Medical & Health Research

ISSN: (e) 3007-1607 (p) 3007-1593

review. All authors reviewed and approved the final version.

Data Availability: All datasets referenced for AI framework training and validation are publicly available through NCBI SRA, McPAS-TCR, RTCGD, TCGA, and PDB. No new experimental data were generated in this study.

REFERENCES

Bagley, S., Binder, Z. A., Lamrani, L., Marinari, E., Desai, A., Nasrallah, M. P., Maloney, E., Brem, S., Lustig, R. A., Kurtz, G., Alonso-Basanta, M., Bonté, P., Goudot, C., Richer, W., Piaggio, E., Kothari, S., Guyonnet, L., Guérin, C. L., Waterfall, J. J., … O'Rourke, D. M. (2024). Repeated peripheral infusions of anti-EGFRvIII CAR T cells in combination with pembrolizumab show no efficacy in glioblastoma: a phase 1 trial. *Nature Cancer*. https://doi.org/10.1038/s43018-023-00709-6

Berns, A. (1991). Tumorigenesis in transgenic mice: Identification and characterization of synergizing oncogenes. *Journal of Cellular Biochemistry*, *47*(2), 130–135. https://doi.org/10.1002/jcb.240470206

Bronstein, M. M., Bruna, J., Cohen, T., & Veličković, P. (2021). Geometric Deep Learning: Grids, Groups, Graphs, Geodesics, and Gauges. *arXiv (Cornell University)*. https://doi.org/10.48550/arxiv.2104.13478

Chao, C., Chiu, Y., Yeung, L. C., Yee, C., Jiang, C., & Shen, X. (2025). AI/ML-empowered approaches for predicting T Cell-mediated immunity and beyond. *Frontiers in Immunology*, *16*, 1651533–1651533. https://doi.org/10.3389/fimmu.2025.1651533

Chen, R. J., Ding, T., Lu, M. Y., Williamson, D. F. K., Jaume, G., Song, A. H., Chen, B., Zhang, A., Shao, D., Shaban, M., Williams, M., Oldenburg, L., Weishaupt, L. L., Wang, J. J., Vaidya, A., Le, L. P., Gerber, G. K., Sahai, S., Williams, W. R., & Mahmood, F. (2024). Towards a general-purpose foundation model for computational pathology. *Nature Medicine*, *30*(3), 850–862. https://doi.org/10.1038/s41591-024-02857-3

Cortés, J. R., & Palomero, T. (2020). Biology and Molecular Pathogenesis of Mature T-Cell Lymphomas [Review of *Biology and Molecular Pathogenesis of Mature T-Cell Lymphomas*]. *Cold Spring Harbor Perspectives in Medicine*, *11*(5). Cold Spring Harbor Laboratory Press. https://doi.org/10.1101/cshperspect.a035402

Dash, P., Fioré-Gartland, A., Hertz, T., Wang, G., Sharma, S., Souquette, A., Crawford, J. C., Clemens, E. B., Nguyen, T. H. O., Kedzierska, K., Gruta, N. L. L., Bradley, P., & Thomas, P. G. (2017). Quantifiable predictive features define epitope-specific T cell receptor repertoires. *Nature*, *547*(7661), 89–93. https://doi.org/10.1038/nature22383

Dosovitskiy, A., Beyer, L., Kolesnikov, A., Weissenborn, D., Zhai, X., Unterthiner, T., Dehghani, M., Minderer, M., Heigold, G., Gelly, S., Uszkoreit, J., & Houlsby, N. (2021, May 3). An Image is Worth 16x16 Words: Transformers for Image Recognition at Scale. *International Conference on Learning Representations*. https://openreview.net/pdf?id=YicbFdNTTy

Duque, D. L., Gaevert, J. A., Thomas, P. G., López-García, M., Lythe, G., & Molina-París, C. (2023). Multi-variate model of T cell clonotype competition and homeostasis. *Scientific Reports*, *13*(1). https://doi.org/10.1038/s41598-023-46637-4

Frontier in Medical & Health Research

ISSN: (e) 3007-1607 (p) 3007-1593

FDA, U. S. (2020). *Guidance for Industry: Long Term Follow-Up After Administration of Human Gene Therapy Products*. U.S. FDA.

Foy, S. P., Jacoby, K., Bota, D. A., Hunter, T., Pan, Z., Stawiski, E., Ma, Y., Lu, W., Peng, S., Wang, C. L., Yuen, B., Dalmas, O., Heeringa, K., Sennino, B., Conroy, A., Bethune, M. T., Mende, I., White, W. B., Kukreja, M., … Mandl, S. (2022). Non-viral precision T cell receptor replacement for personalized cell therapy. *Nature*, *615*(7953), 687–696. https://doi.org/10.1038/s41586-022-05531-1

Gaínza, P., Sverrisson, F., Monti, F., Rodolà, E., Boscaini, D., Bronstein, M. M., & Correia, B. E. (2019). Deciphering interaction fingerprints from protein molecular surfaces using geometric deep learning. *Nature Methods*, *17*(2), 184–192. https://doi.org/10.1038/s41592-019-0666-6

Goldrath, A. W., & Bevan, M. J. (1999). Selecting and maintaining a diverse T-cell repertoire. *Nature*, *402*, 6–13. https://doi.org/10.1038/35005508

Gong, C., Milberg, O., & Wang, B. (2017). A computational multiscale agent-based model for simulating spatio-temporal tumour immune response to PD1 and PDL1 inhibition. *J R Soc Interface*, *14*(134), 20170320.

Husain, S. R., Han, J., Au, P., Shannon, K., & Puri, R. K. (2015). Gene therapy for cancer: regulatory considerations for approval [Review of *Gene therapy for cancer: regulatory considerations for approval*]. *Cancer Gene Therapy*, *22*(12), 554–563. Springer Nature. https://doi.org/10.1038/cgt.2015.58

Jurtz, V., Paul, S., Andreatta, M., Marcatili, P., Peters, B., & Nielsen, M. (2017). NetMHCpan-4.0: Improved Peptide–MHC Class I Interaction Predictions Integrating Eluted Ligand and Peptide Binding Affinity Data. *The Journal of Immunology*, *199*(9), 3360–3368. https://doi.org/10.4049/jimmunol.1700893

Kather, J. N., Heij, L. R., Grabsch, H. I., Loeffler, C. M. L., Echle, A., Muti, H. S., Krause, J., Niehues, J., Sommer, K., Bankhead, P., Kooreman, L., Schulte, J. J., Cipriani, N. A., Buelow, R. D., Boor, P., Ortiz-Brüchle, N., Hanby, A. M., Speirs, V., Kochanny, S., … Luedde, T. (2020). Pan-cancer image-based detection of clinically actionable genetic alterations. *Nature Cancer*, *1*(8), 789–799. https://doi.org/10.1038/s43018-020-0087-6

Keane, C., Gould, C., Jones, K., Hamm, D., Talaulikar, D., Ellis, J., Vari, F., Birch, S., Han, E., Wood, P. J., Le-Cao, K.-A., Green, M. R., Crooks, P., Jain, S., Tobin, J. W. D., Steptoe, R. J., & Gandhi, M. K. (2023). *Data from The T-cell Receptor Repertoire Influences the Tumor Microenvironment and Is Associated with Survival in Aggressive B-cell Lymphoma*. https://doi.org/10.1158/1078-0432.c.6525756

Kirberg, J., Berns, A., & Boehmer, H. von. (1997). Peripheral T Cell Survival Requires Continual Ligation of the T Cell Receptor to Major Histocompatibility Complex–Encoded Molecules. *The Journal of Experimental Medicine*, *186*(8), 1269–1275. https://doi.org/10.1084/jem.186.8.1269

Kohn, D. B., Booth, C., Kang, E. M., Pai, S., Shaw, K. L., Santilli, G., Armant, M., Buckland, K., Choi, U., Ravin, S. S. D., Dorsey, M. J., Kuo, C. Y., León-Rico, D., Rivat, C., Izotova, N., Gilmour, K., Snell, K., Dip, J. X.-B., Darwish, J., … Thrasher, A. J. (2020). Lentiviral gene therapy for X-linked chronic granulomatous disease. *Nature Medicine*, *26*(2), 200–206. https://doi.org/10.1038/s41591-019-0735-5

Lai, H., Wang, L., Qian, R., Huang, J., Zhou, P., Ye, G., Wu, F., Wu, F., Zeng, X., & Liu, W. (2024). Interformer: an interaction-aware model for protein-ligand docking and affinity prediction. *Nature Communications*, *15*(1), 10223–10223. https://doi.org/10.1038/s41467-024-54440-6

Frontier in Medical & Health Research

ISSN: (e) 3007-1607 (p) 3007-1593

Lu, T., Zhang, Z., Zhu, J., Wang, Y., Jiang, P., Xiao, X., Bernatchez, C., Heymach, J. V., Gibbons, D. L., Wang, J., Xu, L., Reuben, A., & Wang, T. (2021). Deep learning-based prediction of the T cell receptor–antigen binding specificity. *Nature Machine Intelligence*, *3*(10), 864–875. https://doi.org/10.1038/s42256-021-00383-2

Lundberg, S., & Lee, S. (2017). A Unified Approach to Interpreting Model Predictions. *arXiv (Cornell University)*. https://doi.org/10.48550/arxiv.1705.07874

Min, B., & Paul, W. E. (2005). Endogenous proliferation: Burst-like CD4 T cell proliferation in lymphopenic settings. *Seminars in Immunology*, *17*(3), 201–207. https://doi.org/10.1016/j.smim.2005.02.005

Morosi, L., Meroni, M., Ubezio, P., Nerini, I. F., Minoli, L., Porcu, L., Panini, N., Colombo, M., Blouw, B., Kang, D. W., Davoli, E., Zucchetti, M., D'Incalci, M., & Frapolli, R. (2021). PEGylated recombinant human hyaluronidase (PEGPH20) pre-treatment improves intra-tumour distribution and efficacy of paclitaxel in preclinical models. *Journal of Experimental & Clinical Cancer Research*, *40*(1). https://doi.org/10.1186/s13046-021-02070-x

Morris, S. W., Kirstein, M. N., & Valentine, M. B. (1994). Fusion of a kinase gene, ALK, to a nucleolar protein gene, NPM, in non-Hodgkin lymphoma. *Science*, *263*(5151), 1281–1284.

Newrzela, S., Al-Ghaili, N., Heinrich, T., Petkova, M., Hartmann, S., Rengstl, B., Kumar, A., Jäck, H.-M., Gerdes, S., Roeder, I., Hansmann, M.-L., & Laer, D. von. (2012). T-cell receptor diversity prevents T-cell lymphoma development. *Leukemia*, *26*(12), 2499–2507. https://doi.org/10.1038/leu.2012.142

Pavlova, A. V., Zvyagin, I. V., & Shugay, M. (2024). Detecting T-cell clonal expansions and quantifying clone survival using deep profiling of immune repertoires. *Frontiers in Immunology*, *15*, 1321603–1321603. https://doi.org/10.3389/fimmu.2024.1321603

Ritter, J., Zimmermann, K., Jöhrens, K., Mende, S., Seegebarth, A., Siegmund, B., Hennig, S., Todorova, K., Rosenwald, A., Daum, S., Hummel, M., & Schümann, M. (2017). T-cell repertoires in refractory coeliac disease. *Gut*, *67*(4), 644–653. https://doi.org/10.1136/gutjnl-2016-311816

Rodighiero, T., Corradini, P., & Bello, D. E. (2025). Leveraging machine learning for integrative analysis of TCR repertoires in colorectal cancer. *Comput Struct Biotechnol J*, *27*, 1432–1441.

S., H.-B.-A., C., V. K., M., S., M.P., M., N., W., P., L., A., L., C.S., O., R., P., E., M., R., S., A., F., P., F., J.I., C., G., D. S. B., E., A., Ian, & U., W. (2003). *LMO2-associated clonal T cell proliferation in two patients after gene therapy for SCID-X1*. http://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.992.1960

Schumacher, T. N., & Schreiber, R. D. (2015). Neoantigens in cancer immunotherapy. *Science*, *348*(6230), 69–74. https://doi.org/10.1126/science.aaa4971

Schwartze, J. T., Havenga, M., Bakker, W. A. M., Bradshaw, A. C., & Nicklin, S. A. (2022). Adenoviral vectors for cardiovascular gene therapy applications: a clinical and industry perspective [Review of *Adenoviral vectors for cardiovascular gene therapy applications: a clinical and industry perspective*]. *Journal of Molecular Medicine*, *100*(6), 875–901. Springer Science+Business Media. https://doi.org/10.1007/s00109-022-02208-0

Frontier in Medical & Health Research

ISSN: (e) 3007-1607 (p) 3007-1593

Selvaraju, R. R., Cogswell, M., Das, A., Vedantam, R., Parikh, D., & Batra, D. (2017). *Grad-CAM: Visual Explanations from Deep Networks via Gradient-Based Localization*. 618–626. https://doi.org/10.1109/iccv.2017.74

Shugay, M., Bagaev, D., & Zvyagin, I. (2018). VDJdb: a curated database of T-cell receptor sequences with known antigen specificity. *Nucleic Acids Res.*, *46*.

Sidhom, J.-W., Larman, H. B., Pardoll, D. M., & Baras, A. S. (2021). DeepTCR is a deep learning framework for revealing sequence concepts within T-cell repertoires. *Nature Communications*, *12*(1), 1605–1605. https://doi.org/10.1038/s41467-021-21879-w

Sidhom, J.-W., Oliveira, G., Ross-Macdonald, P., Wind-Rotolo, M., Wu, C. J., Pardoll, D. M., & Baras, A. S. (2022). Deep learning reveals predictive sequence concepts within immune repertoires to immunotherapy. *Science Advances*, *8*(37). https://doi.org/10.1126/sciadv.abq5089

Singh, M., Louie, R. H. Y., Samir, J., Field, M. A., Milthorpe, C., Adikari, T., Mackie, J., Roper, E. A., Faulks, M., Jackson, K., Calcino, A., Hardy, M. Y., Blombery, P., Amos, T. G., Deveson, I. W., Wende, H. V., Floor, S. N., Read, S., Shek, D., … Luciani, F. (2025). Expanded T cell clones with lymphoma driver somatic mutations accumulate in refractory celiac disease. *Science Translational Medicine*, *17*(798). https://doi.org/10.1126/scitranslmed.adp6812

Tianqi, C., & Carlos, G. (2016). XGBoost: A Scalable Tree Boosting System. *arXiv (Cornell University)*. https://doi.org/10.48550/arxiv.1603.02754

Tickotsky, N., Sagiv, T., Prilusky, J., Shifrut, E., & Friedman, N. (2017). McPAS-TCR: a manually curated catalogue of pathology-associated T cell receptor sequences. *Bioinformatics*, *33*(18), 2924–2929. https://doi.org/10.1093/bioinformatics/btx286

Troy, A. E., & Shen, H. (2003). Cutting Edge: Homeostatic Proliferation of Peripheral T Lymphocytes Is Regulated by Clonal Competition. *The Journal of Immunology*, *170*(2), 672–676. https://doi.org/10.4049/jimmunol.170.2.672

U.S. FDA. (2024). *Communication on T-cell malignancy risk following BCMA- or CD19-directed CAR-T cell therapy*.

Verdun, N., & Marks, P. (2024). Secondary Cancers after Chimeric Antigen Receptor T-Cell Therapy. *New England Journal of Medicine*, *390*(7), 584–586. https://doi.org/10.1056/nejmp2400209

Wang, X. Q., Danenberg, E., Huang, C., Egle, D., Callari, M., Bermejo, B., Dugo, M., Zamagni, C., Thill, M., Anton, A., Zambelli, S., Russo, S., Ciruelos, E., Greil, R., Győrffy, B., Semiglazov, V., Colleoni, M., Kelly, C. M., Mariani, G., … Ali, H. R. (2023). Spatial predictors of immunotherapy response in triple-negative breast cancer. *Nature*, *621*(7980), 868–876. https://doi.org/10.1038/s41586-023-06498-3

Weber, A., Pélissier, A., & Martínez, M. R. (2024). T-cell receptor binding prediction: A machine learning revolution. *ImmunoInformatics*, *15*, 100040–100040. https://doi.org/10.1016/j.immuno.2024.100040

Xu, Y., Qian, X., Zhang, X., Lai, X., Liu, Y., & Wang, J. (2022). DeepLION: Deep Multi-Instance Learning Improves the Prediction of Cancer-Associated T Cell Receptors for Accurate Cancer Detection. *Frontiers in Genetics*, *13*, 860510–860510. https://doi.org/10.3389/fgene.2022.860510

Yang, Z., Zhong, W., Lv, Q., Dong, T., & Chen, C. Y. (2023). Geometric Interaction Graph Neural Network for Predicting Protein–Ligand Binding Affinities from 3D Structures (GIGN). *The Journal of Physical Chemistry Letters*, *14*(8), 2020–2033. https://doi.org/10.1021/acs.jpclett.2c03906

Frontier in Medical & Health Research

ISSN: (e) 3007-1607 (p) 3007-1593

Youssef, E., Weddle, K., Zimmerman, L. H., & Palmer, D. (2025). Pharmacovigilance in Cell and Gene Therapy: Evolving Challenges in Risk Management and Long-Term Follow-Up [Review of *Pharmacovigilance in Cell and Gene Therapy: Evolving Challenges in Risk Management and Long-Term Follow-Up*]. *Drug Safety*. Adis, Springer Healthcare. https://doi.org/10.1007/s40264-025-01596-9

Zhang, Y., Liu, T., & Hu, X. (2021). DeepVISP: Deep learning for virus integration site prediction and motif discovery. *Adv Sci*, *8*(14), 2003021–2003021.