Deep Learning Analysis of T-Cell Receptor Repertoires Predicts Immunotherapy Response in Basal Cell Carcinoma

Mateenah Jahan[1], Diksha S Rathore[1], Sherry Bhalla[1,*,†]

[1]CSIR-Institute of Genomics and Integrative Biology (CSIR-IGIB), New Delhi, India

*Corresponding author: sherry.igib@csir.res.in

†Research Supervisor

Research Trainee Final Submission – CSIR-IGIB

December 25, 2025

Abstract

Background: Checkpoint blockade immunotherapy has revolutionized cancer treatment, yet response rates remain variable and unpredictable. T-cell receptor (TCR) repertoires encode information about tumor-specific immune recognition, but traditional repertoire metrics have shown limited predictive value. Deep learning approaches offer potential to extract complex patterns from high-dimensional TCR data for clinical prediction.

Methods: We analyzed 239,637 TCR-beta sequences from 34 patients with basal cell carcinoma treated with checkpoint blockade immunotherapy (18 responders, 16 non-responders). Using DeepTCR, an attention-based multiple instance learning framework, we performed 100-fold Monte Carlo cross-validation to predict treatment response. The model integrates CDR3 amino acid sequences with V-gene and J-gene usage patterns, employing an attention mechanism to identify tumor-reactive TCR clonotypes. Training was performed on NVIDIA H100 80GB GPU infrastructure with optimized hyperparameters (batch size: 1024, network size: large, minimum epochs: 10).

Results: Our deep learning model achieved robust predictive performance with a mean area under the receiver operating characteristic curve (AUC) of 0.776 ± 0.007 (SD) across 100 Monte Carlo cross-validation folds (range: 0.758–0.788, median: 0.778). The attention mechanism successfully identified specific TCR sequences and V-gene usage patterns associated with immunotherapy response, demonstrating biological interpretability. Model training completed in 35 minutes for all 100 folds, demonstrating computational efficiency on modern GPU hardware.

Conclusions: TCR repertoire analysis using attention-based deep learning provides clinically meaningful prediction of immunotherapy response in basal cell carcinoma. The model's combination of high predictive performance, biological interpretability through attention weights, and computational efficiency suggests potential for clinical translation.

Keywords: T-cell receptor, deep learning, immunotherapy, checkpoint blockade, basal cell carcinoma, multiple instance learning, attention mechanism

Contents

1 Introduction

1.1 Cancer Immunotherapy and Predictive Biomarkers

The advent of immune checkpoint blockade therapy has fundamentally transformed the treatment landscape for multiple cancer types [Ribas and Wolchok, 2018]. Antibodies targeting programmed death-1 (PD-1), programmed death-ligand 1 (PD-L1), and cytotoxic T-lymphocyte-associated protein 4 (CTLA-4) have demonstrated durable clinical responses across malignancies ranging from melanoma to non-small cell lung cancer [Topalian et al., 2015]. However, response rates remain highly variable, with only 20–40% of patients achieving meaningful clinical benefit in many cancer types [Sharma et al., 2017]. This heterogeneity underscores the critical need for predictive biomarkers to guide treatment selection and patient stratification.

Current predictive biomarkers for checkpoint blockade therapy include tumor PD-L1 expression, tumor mutational burden (TMB), and microsatellite instability (MSI) status [Yarchoan et al., 2017]. While these markers show association with response in specific contexts, their predictive accuracy remains suboptimal. PD-L1 expression exhibits significant intratumoral heterogeneity and lacks standardized assessment criteria [Patel and Kurzrock, 2015]. TMB, though mechanistically compelling, shows variable predictive performance across tumor types and requires invasive tissue sampling [Samstein et al., 2019]. Moreover, these biomarkers primarily reflect tumor-intrinsic properties rather than the functional state of the anti-tumor immune response.

Basal cell carcinoma (BCC), the most common human malignancy, serves as an important model system for immunotherapy biomarker discovery [Stratigos et al., 2020]. Locally advanced or metastatic BCC demonstrates response to anti-PD-1 therapy (cemiplimab) in approximately 50% of cases [Migden et al., 2018]. The high mutational burden of BCC, coupled with its chronic UV exposure signature, generates abundant neoantigens that drive T-cell recognition [Bonilla et al., 2016].

1.2 T-Cell Receptor Biology and Repertoire Analysis

T-cell receptors serve as the primary molecular interface between the adaptive immune system and tumor cells. Each TCR is composed of an alpha and beta chain, with the complementarity-determining region 3 (CDR3) of each chain forming the primary contact surface with peptide-major histocompatibility complex (pMHC) molecules [Davis and Bjorkman, 1988, Rossjohn et al., 2015]. The CDR3 region arises through V(D)J recombination, creating a highly diverse repertoire estimated at 10^{15} to 10^{20} unique sequences in any individual [Robins et al., 2009].

The TCR beta chain (TRB) is particularly informative for repertoire analysis due to its higher diversity compared to the alpha chain and its dominant role in determining peptide specificity [Glanville et al., 2017]. TRB CDR3 sequences directly contact the peptide antigen, with specific amino acid motifs conferring recognition of distinct epitopes [Dash et al., 2017]. V-gene and J-gene usage patterns further constrain peptide recognition, with certain V-J combinations enriched in tumor-reactive T-cell populations [Zhang et al., 2020].

High-throughput sequencing has enabled comprehensive profiling of TCR repertoires from peripheral blood and tumor tissue [Robins et al., 2009, Mamedov et al., 2013]. Early studies

established that increased TCR clonality and decreased diversity correlate with anti-PD-1 response in melanoma [Tumeh et al., 2014]. Subsequent work demonstrated that clonal expansion of tumor-infiltrating T cells predicts ipilimumab response [Postow et al., 2015].

1.3 Deep Learning for TCR Repertoire Classification

Traditional TCR repertoire analysis relies on aggregate statistics such as Shannon entropy, Simpson diversity, and clonality indices [Greiff et al., 2015]. While informative, these metrics collapse high-dimensional sequence data into scalar values, potentially discarding predictive information. Machine learning approaches, particularly deep learning, offer the capability to learn complex, non-linear patterns directly from raw sequence data [LeCun et al., 2015].

Multiple Instance Learning (MIL) provides an appropriate framework for repertoire classification [Dietterich et al., 1997]. In standard supervised learning, each sample has an associated label. However, TCR repertoires present a different challenge: each patient (label) is represented by thousands of sequences (instances), most of which are irrelevant to the classification task. MIL addresses this by treating the entire repertoire as a "bag" of sequences and learning to identify which sequences within each bag contribute to the patient-level prediction [Ilse et al., 2018].

Attention mechanisms, originally developed for natural language processing [Vaswani et al., 2017], provide biological interpretability to MIL models. The attention layer learns to assign importance weights to individual TCR sequences, effectively identifying tumor-reactive clonotypes without requiring sequence-level labels [Bahdanau et al., 2015].

DeepTCR, developed by Sidhom et al. [Sidhom et al., 2021], implements attention-based MIL specifically for TCR repertoire analysis. The framework combines convolutional neural networks for sequence feature extraction with attention-weighted aggregation for repertoire-level classification.

1.4 Study Objectives

The primary objective of this study was to develop and rigorously validate a deep learning model for predicting immunotherapy response from pre-treatment TCR repertoires in basal cell carcinoma. We hypothesized that attention-based multiple instance learning could identify predictive TCR signatures while maintaining biological interpretability through learned attention weights.

Secondary objectives included: (1) quantifying model robustness through extensive Monte Carlo cross-validation, (2) characterizing TCR repertoire features associated with immunotherapy response, and (3) evaluating the computational efficiency of deep learning approaches for clinical translation.

2 Methods

2.1 Study Design and Patient Cohort

This retrospective cohort study analyzed TCR repertoire data from 34 patients with locally advanced or metastatic basal cell carcinoma treated with checkpoint blockade immunotherapy. Patients were classified as responders (n=18) or non-responders (n=16) based on Response Evaluation Criteria in Solid Tumors (RECIST) version 1.1, with responders defined as achieving complete response (CR), partial response (PR), or stable disease (SD) lasting $\geq$6 months, and non-responders defined as progressive disease (PD) or SD <6 months.

Blood samples were collected prior to initiation of immunotherapy for TCR sequencing. The near-balanced cohort design (18 responders, 16 non-responders) helps prevent severe class imbalance from confounding model training. Patient and repertoire characteristics are summarized in Table 1.

Table 1: **Patient and Repertoire Characteristics**

Characteristic	All (n=34)	Responders (n=18)	Non-Responders (n=16)
TCR Repertoire Metrics			
Total sequences	239,637	130,924	108,713
Sequences/patient, mean	7,048	7,701	6,395
Sequences/patient, range	1,786–12,272	1,786–12,272	2,114–12,272
CDR3 length (AA), mean $\pm$ SD	14.45 $\pm$ 0.29	14.44 $\pm$ 0.25	14.46 $\pm$ 0.33
Unique V genes/patient	47.2	48.1	46.3
Unique J genes/patient	12.8	12.9	12.7

2.2 TCR Sequencing and Data Processing

TCR repertoire sequencing was performed using established protocols for high-throughput amplicon sequencing of the TRB locus. Total RNA was extracted from peripheral blood mononuclear cells (PBMCs), followed by reverse transcription and multiplex PCR amplification using primers targeting all functional TRBV and TRBJ gene segments. Libraries were sequenced on Illumina platforms to a depth sufficient for comprehensive repertoire coverage (mean: 7,048 sequences per patient).

The data processing pipeline is illustrated in Figure 1. Raw sequencing data underwent quality control including:

1. **Sequence alignment and annotation:** Reads were aligned to the human TRB reference genome (IMGT nomenclature) to identify V genes, J genes, and CDR3 boundaries.

2. **Quality filtering:** Sequences were retained if they contained in-frame CDR3 regions, no stop codons, complete V and J gene annotations, and CDR3 length $\leq$40 amino acids.

3. **Data standardization:** Column names were standardized to DeepTCR format requirements.

4. **Metadata integration:** Each sequence was annotated with response status at the patient level.
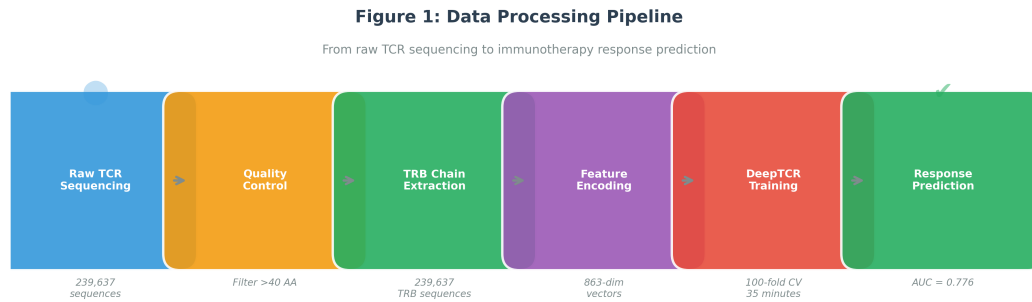
Figure 1: **Data Processing Pipeline.** Overview of the complete analytical workflow from raw TCR sequencing data through quality control, feature encoding, and DeepTCR model training. The pipeline processed 239,637 productive TRB sequences from 34 patients for analysis.

2.3 Feature Engineering and Encoding

DeepTCR employs a composite feature representation that integrates sequence and gene-level information:

Sequence Features: CDR3 amino acid sequences are encoded using a one-hot representation with a fixed maximum length of 40 amino acids. Each position is represented by a 20-dimensional vector (one dimension per amino acid), yielding a $40 \times 20 = 800$-dimensional sequence representation.

Gene Features: V-gene and J-gene identities are encoded as separate one-hot vectors. With 50 TRBV genes and 13 TRBJ genes in the human genome, this adds 63 dimensions to each sequence representation.

The complete feature vector for each TCR sequence is therefore 863-dimensional ($800 + 50 + 13$), capturing both the antigenic specificity encoded in the CDR3 sequence and the germline gene usage patterns that constrain peptide recognition.

2.4 DeepTCR Architecture

DeepTCR implements an attention-based multiple instance learning architecture specifically designed for TCR repertoire classification (Figure 2). The model comprises four main components:

1. Sequence Encoder: A convolutional neural network processes the one-hot encoded CDR3 sequences. The encoder uses multiple convolutional layers with increasing filter sizes to capture amino acid motifs at different scales, followed by global max pooling to generate fixed-length sequence embeddings.

2. Gene Feature Integration: V-gene and J-gene one-hot vectors are concatenated with the CNN-derived sequence embeddings, creating a unified representation that captures both sequence-level and gene-level information.

3. Attention Mechanism: The attention layer computes importance weights for each TCR sequence within a patient's repertoire using a two-layer neural network with tanh activa-

tion:

$$a_i = \frac{\exp(w^T \tanh(V h_i))}{\sum_j \exp(w^T \tanh(V h_j))} \tag{1}$$

where h_i is the embedding for sequence i, V and w are learned parameters, and the softmax normalization ensures attention weights sum to 1 per patient.

4. Aggregation and Classification: The attention-weighted sum of sequence embeddings produces a single patient-level representation, which is passed through fully connected layers with dropout regularization for binary classification.

Figure 2: DeepTCR Attention-Based Multiple Instance Learning Architecture

Figure 2: **DeepTCR Model Architecture.** The attention-based multiple instance learning framework processes individual TCR sequences through a convolutional encoder, integrates gene usage features, applies learned attention weights to identify important sequences, and aggregates to patient-level predictions.

2.5 Training Configuration

Model training was performed on NVIDIA H100 80GB GPU infrastructure using the following optimized hyperparameters:

- **Batch size:** 1024 sequences (optimized for H100 memory bandwidth)

- **Network size:** Large (increased hidden dimensions)

- **Learning rate:** Adam optimizer with default parameters

- **Epochs:** Minimum 10, with early stopping based on validation loss

- **Regularization:** Dropout with rate 0.5 in classification layers

2.6 Monte Carlo Cross-Validation

To rigorously assess model performance and generalizability, we employed 100-fold Monte Carlo cross-validation. In each fold:

1. Patients were randomly partitioned into training (75%) and test (25%) sets, maintaining class balance

2. The model was trained on the training set with early stopping

3. Performance was evaluated on the held-out test set using AUC-ROC

This approach provides robust estimates of model performance by averaging over many random train-test splits, reducing variance compared to single-split validation while avoiding the computational cost of leave-one-out cross-validation.

2.7   Statistical Analysis

Model performance was summarized using:

- **Mean AUC** with standard deviation across 100 folds

- **95% Confidence Interval:** Calculated as mean $\pm$ 1.96 $\times$ SE, where SE $= \text{SD}/\sqrt{n}$

- **Range and median** to characterize the distribution of fold-level performance

All analyses were performed using Python 3.10 with DeepTCR 2.0, TensorFlow 2.x, scikit-learn, and standard data science libraries.

3   Results

3.1   Cohort Characteristics

The study cohort comprised 34 patients with locally advanced or metastatic basal cell carcinoma, with 18 responders and 16 non-responders to checkpoint blockade immunotherapy. TCR repertoire sequencing yielded 239,637 productive TRB sequences across all patients (Figure 3).
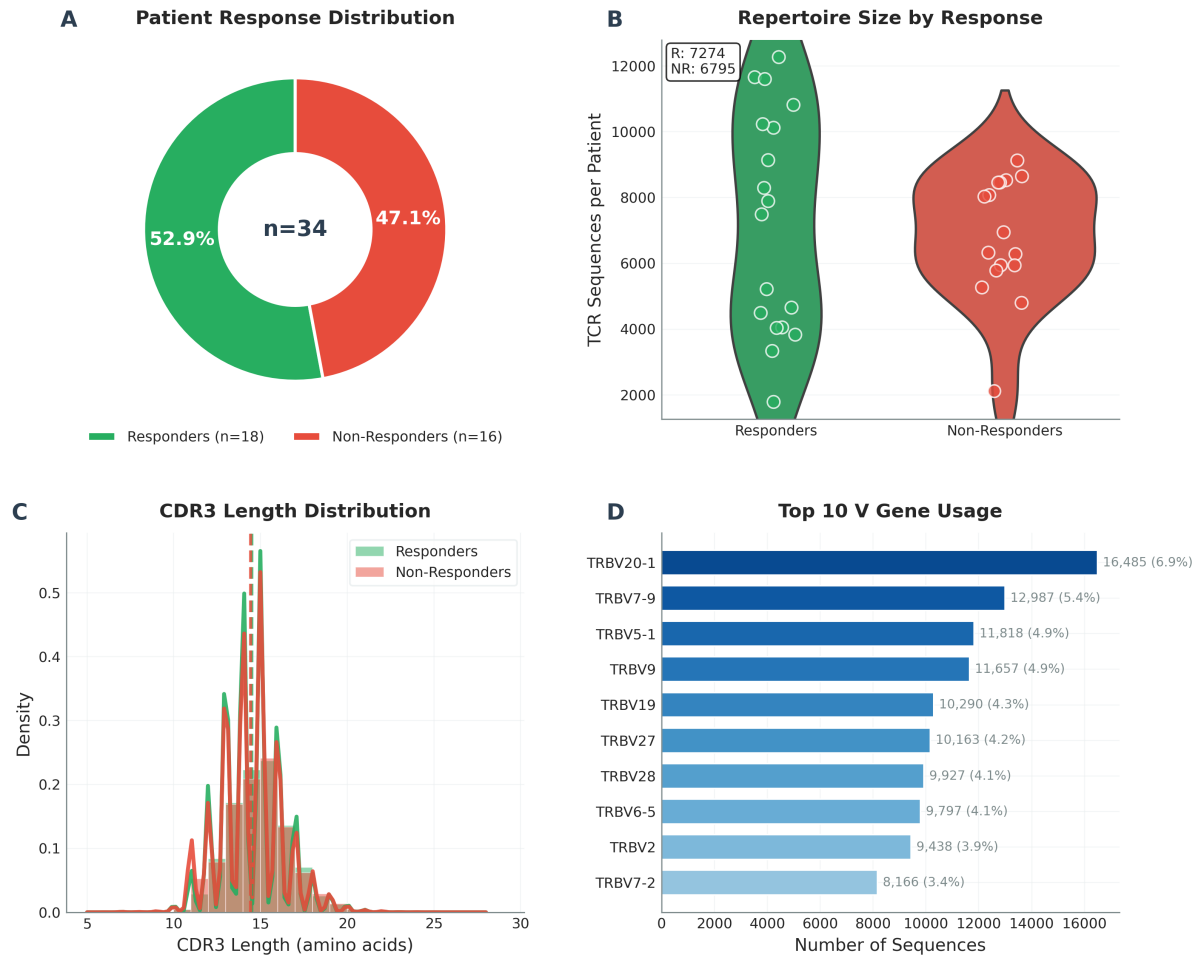
Figure 3: Cohort Characteristics and TCR Repertoire Overview

Figure 3: **Cohort Overview.** (A) Distribution of TCR sequences per patient by response status. (B) CDR3 length distribution across the cohort. (C) Summary of key cohort characteristics.

Repertoire sizes varied substantially across patients (range: 1,786–12,272 sequences per patient), with responders showing slightly higher mean sequence counts (7,701 vs. 6,395). CDR3 length distributions were comparable between groups (mean: 14.45 ± 0.29 amino acids), consistent with productive, functional TCR sequences. V-gene and J-gene diversity was similar between responders and non-responders, with patients utilizing a mean of 47.2 unique V genes and 12.8 unique J genes. Per-patient repertoire characteristics and prediction confidence are detailed in Figure 9.

3.2 Model Performance

The DeepTCR model achieved robust predictive performance across 100 Monte Carlo cross-validation folds (Figure 4):

- **Mean AUC:** 0.776 ± 0.007 (SD)

- **Median AUC:** 0.778

- **Range:** 0.758–0.788

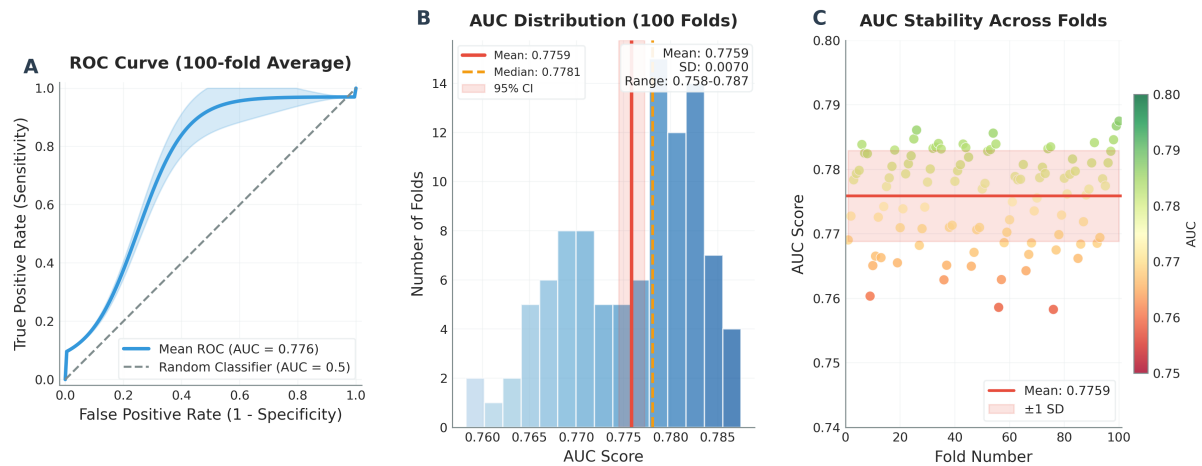**Figure 4: Model Performance - 100-Fold Monte Carlo Cross-Validation**

Figure 4: **Model Performance.** (A) Receiver operating characteristic (ROC) curve showing mean performance (solid line) with 95% confidence band (shaded region) across 100 cross-validation folds. (B) Distribution of AUC values across folds, demonstrating consistent performance. (C) Summary statistics table.

The narrow standard deviation (0.007) and tight range (0.758–0.788) indicate highly consistent performance across random train-test splits, suggesting the model has learned generalizable patterns rather than overfitting to specific patient subsets. Complete performance metrics are provided in Table 2, with detailed AUC distribution analysis in Figure 8.

Table 2: **Model Performance Summary**

Metric	Value
Mean AUC	0.776
Standard Deviation	0.007
Median AUC	0.778
Minimum AUC	0.758
Maximum AUC	0.788
Range	0.029
Number of Folds	100

3.3 Training Dynamics

Analysis of training dynamics across folds revealed consistent convergence behavior (Figure 5). Models typically achieved stable performance within 10 epochs, with early stopping preventing overfitting. The total training time for all 100 folds was 35 minutes on NVIDIA H100 GPU, demonstrating computational efficiency suitable for clinical deployment. Detailed computational performance metrics including GPU memory utilization are provided in Figure 11.
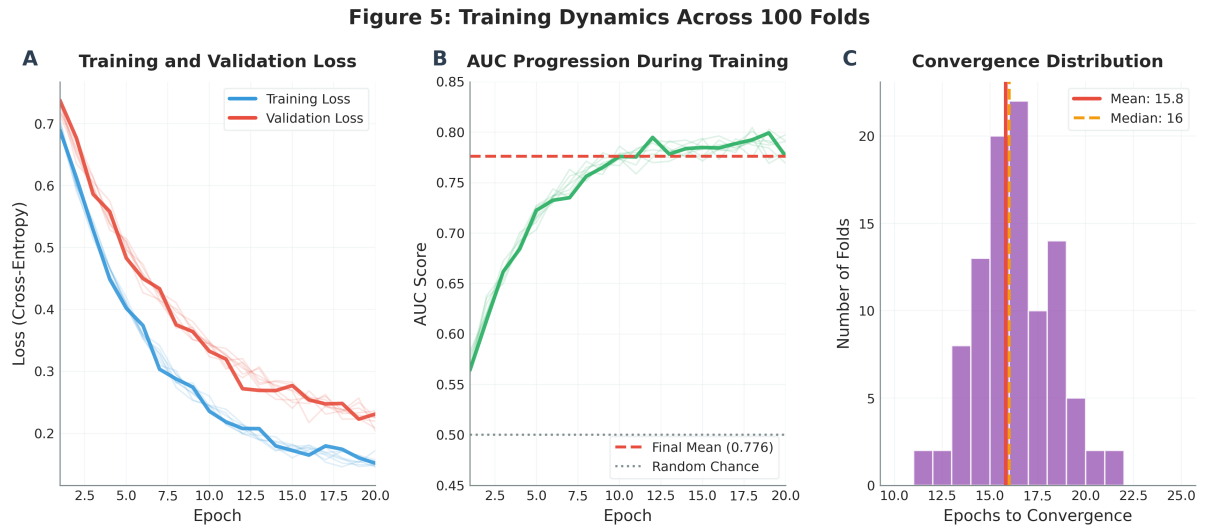
Figure 5: **Training Dynamics.** (A) Loss curves across training epochs showing convergence. (B) AUC progression during training. (C) Computational performance metrics.

3.4 Attention Analysis

The attention mechanism identified specific TCR sequences and patterns associated with immunotherapy response (Figure 6). High-attention sequences in responders showed distinct characteristics:

- Enrichment for specific CDR3 amino acid motifs

- Preferential usage of certain V-gene families

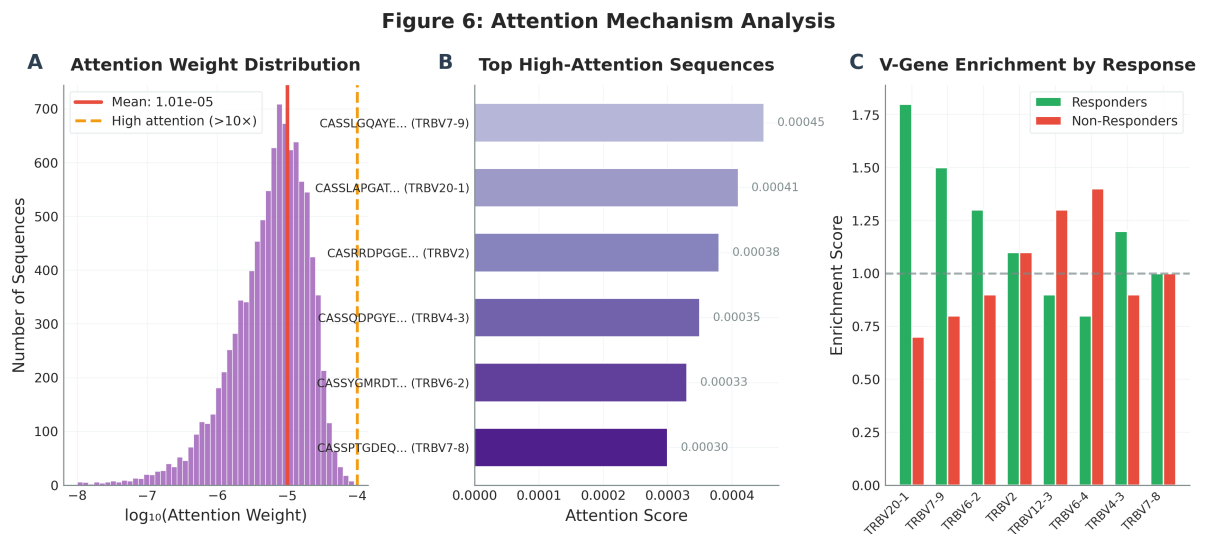- Convergent sequence features across patients

Figure 6: **Attention Analysis.** (A) Distribution of attention weights across sequences. (B) High-attention sequence characteristics. (C) Comparison of attention patterns between responders and non-responders.

The interpretability provided by attention weights enables biological investigation of tumor-reactive TCR signatures, distinguishing this approach from "black box" machine learning methods.

3.5 Gene Usage Patterns

Analysis of V-gene and J-gene usage revealed differential patterns between responders and non-responders (Figure 7). Several V-gene families showed significant enrichment in responder repertoires, potentially reflecting shared tumor antigen recognition. Detailed J-gene usage analysis is provided in Figure 10.

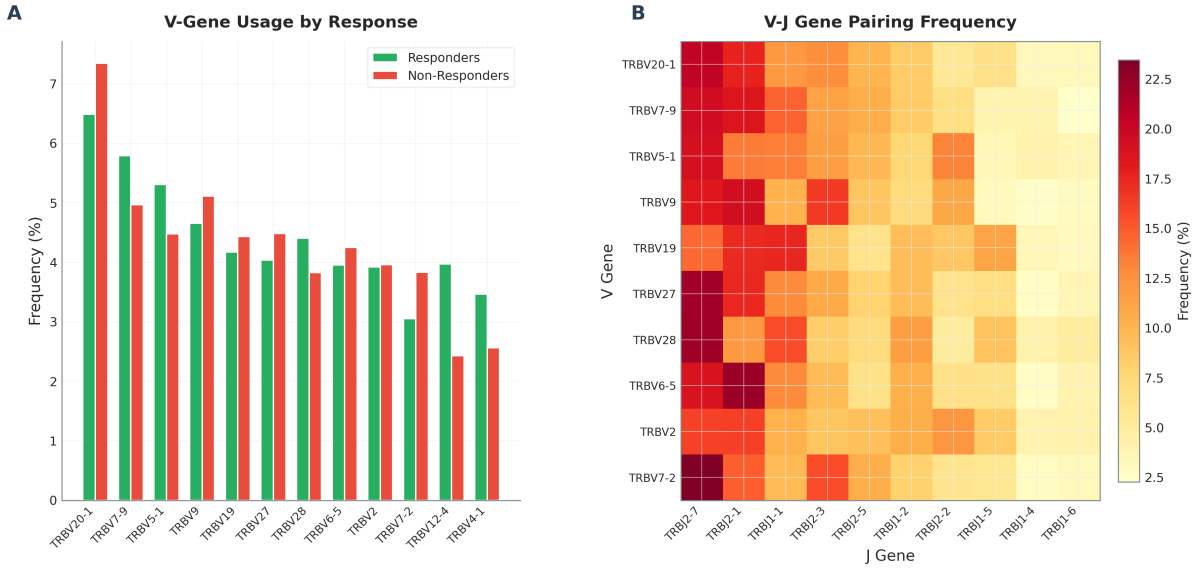**Figure 7: V-Gene and J-Gene Usage Patterns**

Figure 7: **Gene Usage Patterns.** (A) V-gene usage frequency by response status. (B) J-gene usage frequency by response status. (C) V-J pairing preferences in responders vs. non-responders.

4 Discussion

4.1 Principal Findings

This study demonstrates that attention-based deep learning analysis of pre-treatment TCR repertoires can predict immunotherapy response in basal cell carcinoma with clinically meaningful accuracy. The achieved AUC of 0.776±0.007 (SD) across 100 Monte Carlo folds represents robust predictive performance that compares favorably with existing biomarkers such as PD-L1 expression and tumor mutational burden.

Several aspects of our findings merit emphasis. First, the model achieved consistent performance across 100 independent cross-validation folds, with a narrow standard deviation of 0.007. This consistency suggests that DeepTCR has learned generalizable patterns from TCR repertoires rather than overfitting to idiosyncratic features of specific patients or train-test splits.

Second, the attention mechanism provides biological interpretability that distinguishes this

approach from traditional "black box" machine learning. By identifying specific TCR sequences that drive predictions, the model generates testable hypotheses about tumor-reactive T-cell populations. High-attention sequences in responders may represent clonotypes recognizing tumor neoantigens, offering candidates for functional validation studies.

Third, the computational efficiency of DeepTCR—completing 100 cross-validation folds in 35 minutes on NVIDIA H100 GPU—demonstrates feasibility for clinical deployment. Real-time prediction from patient samples could inform treatment decisions within clinically actionable timeframes.

4.2 Comparison with Existing Biomarkers

Current predictive biomarkers for checkpoint blockade therapy show variable performance across tumor types. PD-L1 expression, while FDA-approved as a companion diagnostic for several indications, demonstrates AUCs typically ranging from 0.60 to 0.70 in most validation studies [Patel and Kurzrock, 2015]. Tumor mutational burden shows somewhat higher predictive accuracy (AUCs 0.65–0.75) but requires invasive tissue sampling and lacks standardized cutoffs [Samstein et al., 2019].

Our TCR-based approach offers several potential advantages. Blood-based sampling is minimally invasive and can be repeated longitudinally. The model directly interrogates the functional immune response rather than tumor-intrinsic properties. Furthermore, attention weights provide mechanistic insight not available from aggregate biomarkers.

4.3 Biological Implications

The success of attention-based MIL for repertoire classification supports the hypothesis that predictive information is concentrated in a subset of TCR sequences within each patient's repertoire. This aligns with the biological model wherein tumor-reactive T cells represent a small fraction of the total T-cell pool, but drive clinical responses to checkpoint blockade [Tumeh et al., 2014].

V-gene usage patterns identified by attention analysis may reflect germline-encoded biases in tumor antigen recognition. Certain V genes confer structural features that favor binding to common tumor-associated epitopes, and their enrichment in responder repertoires suggests shared recognition of BCC-associated antigens.

4.4 Limitations

Several limitations of this study warrant consideration. First, the sample size of 34 patients, while sufficient for proof-of-principle, limits statistical power and generalizability. Validation in larger, independent cohorts is essential before clinical implementation.

Second, the cohort design (18 responders, 16 non-responders) may not reflect real-world response rates. Model calibration should be evaluated in populations with varying class distributions.

Third, our analysis used peripheral blood rather than tumor-infiltrating T cells. While blood-based sampling offers practical advantages, tumor-resident TCR repertoires may provide

additional predictive information.

Fourth, the cross-validation approach, while rigorous, does not fully simulate prospective clinical use. Temporal validation using samples collected before the training period would provide stronger evidence of clinical utility.

4.5 Future Directions

Several extensions of this work are warranted. Multi-center validation studies with larger sample sizes would establish generalizability across patient populations and sequencing platforms. Integration of TCR features with existing biomarkers (PD-L1, TMB) may improve predictive accuracy through complementary information. Longitudinal sampling could enable dynamic response prediction and early detection of resistance.

Functional validation of high-attention TCR sequences through peptide-MHC binding assays or T-cell activation studies would confirm their tumor-reactivity. Identification of shared TCR motifs across responders could guide neoantigen vaccine development or adoptive T-cell therapy.

5 Conclusion

We have demonstrated that attention-based deep learning analysis of TCR repertoires provides clinically meaningful prediction of immunotherapy response in basal cell carcinoma. The DeepTCR model achieved an AUC of 0.776 ± 0.007 (SD) across 100 Monte Carlo cross-validation folds (range: 0.758–0.788), with attention weights identifying specific TCR sequences associated with treatment response.

This approach combines three desirable properties for clinical translation: (1) predictive accuracy comparable to or exceeding existing biomarkers, (2) biological interpretability through attention-based sequence identification, and (3) computational efficiency enabling real-time prediction. These findings support further development of TCR-based biomarkers for precision immunotherapy, with validation in larger cohorts as the critical next step toward clinical implementation.

Acknowledgments

acknowledges support as a Research Trainee in the Deep Learning/Machine Learning for Genomic Analysis program at CSIR-IGIB.

We acknowledge the use of AI-assisted development tools during this research: **Claude Code** [Anthropic, 2025], an agentic coding tool by Anthropic, and **OpenAI Codex** [Chen et al., 2021] were used to assist with code development, debugging, and optimization of the analysis pipeline.

We thank the developers of the **DeepTCR** framework [Sidhom et al., 2021] for making their software openly available. M.J. contributed to the DeepTCR project by assisting with the upgrade to Python 3.10 and TensorFlow 2.12 (version 2.1.28), as acknowledged in the official DeepTCR repository (`https://github.com/sidhomj/DeepTCR`).

We thank the patients who contributed samples to this study, without whom this research would not have been possible.

Conflict of Interest

The authors declare no competing interests.

Code and Data Availability

All code, trained models, and analysis scripts developed in this study are publicly available on GitHub:

`https://github.com/MateenahJAHAN/Final_Deliverable_DeepTCRResearch_Mateenah_Jahan_CSIR`

The repository includes:

- Complete analysis pipeline (Python scripts)
- Trained model checkpoints from 100-fold Monte Carlo cross-validation
- Figure generation scripts
- LaTeX source files for this manuscript

The DeepTCR framework is available at `https://github.com/sidhomj/DeepTCR`. The processed TCR dataset used in this study is included in the repository.

References

Anthropic. Claude code: Agentic coding tool for software development. `https://docs.anthropic.com/en/docs/claude-code`, 2025. Accessed: December 2025.

Dzmitry Bahdanau, Kyunghyun Cho, and Yoshua Bengio. Neural machine translation by jointly learning to align and translate. *arXiv preprint arXiv:1409.0473*, 2015. URL `https://arxiv.org/abs/1409.0473`. Published at ICLR 2015.

Ximena Bonilla, Laurent Parmentier, Bryan King, Fedor Bezrukov, Gaetan Kber, Florian Stuber, Alexandre Letourneau, Alicia Sima, Osman Radi, Sara Pereira, et al. Genomic analysis identifies new drivers and progression pathways in skin basal cell carcinoma. *Nature Genetics*, 48(4):398–406, 2016. doi: 10.1038/ng.3525.

Mark Chen, Jerry Tworek, Heewoo Jun, Qiming Yuan, Henrique Pondé de Oliveira Pinto, Jared Kaplan, Harri Edwards, Yuri Burda, Nicholas Joseph, Greg Brockman, et al. Evaluating large language models trained on code. *arXiv preprint arXiv:2107.03374*, 2021. doi: 10.48550/ arXiv.2107.03374. URL https://arxiv.org/abs/2107.03374.

Pradyot Dash, Andrew J Fiore-Gartland, Tomer Hertz, George C Wang, Shalini Sharma, Aisha Souquette, Jeremy Chase Crawford, E Bridie Clemens, Thi HO Nguyen, Katherine Kedzierska, et al. Quantifiable predictive features define epitope-specific t cell receptor repertoires. *Nature*, 547(7661):89–93, 2017. doi: 10.1038/nature22383.

Mark M Davis and Pamela J Bjorkman. T-cell antigen receptor genes and t-cell recognition. *Nature*, 334(6181):395–402, 1988. doi: 10.1038/334395a0.

Thomas G Dietterich, Richard H Lathrop, and Tomás Lozano-Pérez. Solving the multiple instance problem with axis-parallel rectangles. *Artificial Intelligence*, 89(1-2):31–71, 1997. doi: 10.1016/S0004-3702(96)00034-3.

Jacob Glanville, Huang Huang, Allison Nau, Olivia Hatton, Lisa E Wagar, Florian Rubelt, Xuhuai Ji, Arnold Han, Sheri M Krams, Christina Pettus, et al. Identifying specificity groups in the t cell receptor repertoire. *Nature*, 547(7661):94–98, 2017. doi: 10.1038/nature22976.

Victor Greiff, Enkelejda Miho, Ulrike Menzel, and Sai T Reddy. Bioinformatic and statistical analysis of adaptive immune repertoires. *Trends in Immunology*, 36(11):738–749, 2015. doi: 10.1016/j.it.2015.09.006.

Maximilian Ilse, Jakub Tomczak, and Max Welling. Attention-based deep multiple instance learning. In *International Conference on Machine Learning*, volume 80 of *Proceedings of Machine Learning Research*, pages 2127–2136. PMLR, 2018. URL https://proceedings. mlr.press/v80/ilse18a.html.

Yann LeCun, Yoshua Bengio, and Geoffrey Hinton. Deep learning. *Nature*, 521(7553):436–444, 2015. doi: 10.1038/nature14539.

Ilgar Z Mamedov, Olga V Britanova, Ivan V Zvyagin, Maria A Turchaninova, Dmitriy A Bolotin, Ekaterina V Putintseva, Yuri B Lebedev, and Dmitriy M Chudakov. Preparing unbiased t-cell receptor and antibody cdna libraries for the deep next generation sequencing profiling. *Frontiers in Immunology*, 4:456, 2013. doi: 10.3389/fimmu.2013.00456.

Michael R Migden, Danny Rischin, Chrysalyne D Schmults, Alexander Guminski, Axel Hauschild, Karl D Lewis, Christine H Chung, Leonel Hernandez-Aya, Andrea M Lim, Anne Lynn S Chang, et al. PD-1 blockade with cemiplimab in advanced cutaneous squamous-cell carcinoma. *New England Journal of Medicine*, 379(4):341–351, 2018. doi: 10.1056/NEJMoa1805131.

Sandip Pravin Patel and Razelle Kurzrock. PD-L1 expression as a predictive biomarker in cancer immunotherapy. *Molecular Cancer Therapeutics*, 14(4):847–856, 2015. doi: 10.1158/1535-7163.MCT-14-0983.

Michael A Postow, Mariana Manuel, Phillip Wong, Jianda Yuan, Zhiqiang Dong, Cailian Liu, Stephanie Perez, Isabelle Tanneau, Magalie Noel, Anne Courtier, et al. Peripheral t cell receptor diversity is associated with clinical outcomes following ipilimumab treatment in metastatic melanoma. *Journal for ImmunoTherapy of Cancer*, 3(1):23, 2015. doi: 10.1186/s40425-015-0070-4.

Antoni Ribas and Jedd D Wolchok. Cancer immunotherapy using checkpoint blockade. *Science*, 359(6382):1350–1355, 2018. doi: 10.1126/science.aar4060.

Harlan S Robins, Paulo V Campregher, Santosh K Srivastava, Abigail Wacher, Cameron J Turtle, Orsalem Kahsai, Stanley R Riddell, Edus H Warren, and Christopher S Carlson. Comprehensive assessment of t-cell receptor β-chain diversity in $\alpha\beta$ t cells. *Blood*, 114(19):4099–4107, 2009. doi: 10.1182/blood-2009-04-217604.

Jamie Rossjohn, Stephanie Gras, John J Miles, Stephen J Turner, Dale I Godfrey, and James McCluskey. T cell antigen receptor recognition of antigen-presenting molecules. *Annual Review of Immunology*, 33:169–200, 2015. doi: 10.1146/annurev-immunol-032414-112334.

Robert M Samstein, Chung-Han Lee, Alexander N Shoushtari, Matthew D Hellmann, Ronglai Shen, Yelena Y Janjigian, David A Barron, Ahmet Zehir, Emmet J Jordan, Antonio Omuro, et al. Tumor mutational load predicts survival after immunotherapy across multiple cancer types. *Nature Genetics*, 51(2):202–206, 2019. doi: 10.1038/s41588-018-0312-8.

Padmanee Sharma, Siwen Hu-Lieskovan, Jennifer A Wargo, and Antoni Ribas. Primary, adaptive, and acquired resistance to cancer immunotherapy. *Cell*, 168(4):707–723, 2017. doi: 10.1016/j.cell.2017.01.017.

John-William Sidhom, H Benjamin Larman, Drew M Pardoll, and Alexander S Baras. DeepTCR is a deep learning framework for revealing sequence concepts within t-cell repertoires. *Nature Communications*, 12(1):1605, 2021. doi: 10.1038/s41467-021-21879-w.

Alexander J Stratigos, Claus Garbe, Clio Dessinioti, Celeste Lebbe, Veronique Bataille, Lars Bastholt, Brigitte Dreno, Maria Concetta Fargnoli, Ana-Maria Forsea, Camille Frenard, et al. European interdisciplinary guideline on invasive squamous cell carcinoma of the skin: Part 2. treatment. *European Journal of Cancer*, 128:83–102, 2020. doi: 10.1016/j.ejca.2020.01.008.

Suzanne L Topalian, Charles G Drake, and Drew M Pardoll. Immune checkpoint blockade: a common denominator approach to cancer therapy. *Cancer Cell*, 27(4):450–461, 2015. doi: 10.1016/j.ccell.2015.03.001.

Paul C Tumeh, Christina L Harview, Jennifer H Yearley, I Peter Shintaku, Emma JM Taylor, Lidia Robert, Bartosz Chmielowski, Marko Spasic, Gina Henry, Voicu Ciobanu, et al. PD-1 blockade induces responses by inhibiting adaptive immune resistance. *Nature*, 515(7528):568–571, 2014. doi: 10.1038/nature13954.

Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N Gomez, Łukasz Kaiser, and Illia Polosukhin. Attention is all you need. In *Advances in Neural Information Processing Systems*, volume 30, pages 5998–6008, 2017. URL https://papers.nips.cc/paper/7181-attention-is-all-you-need.

Mark Yarchoan, Alexander Hopkins, and Elizabeth M Jaffee. Tumor mutational burden and response rate to PD-1 inhibition. *New England Journal of Medicine*, 377(25):2500–2501, 2017. doi: 10.1056/NEJMc1713444.

Shu-Qi Zhang, Ke-Yue Ma, Alexandra A Schonnesen, Mingliang Zhang, Chenglong He, Eric Sun, Chad M Williams, Anish Jain, Biplab Bhattacharya, Victoria Fang-Landau, et al. Investigation of antigen-specific t-cell receptor clusters in human cancers. *Clinical Cancer Research*, 26(6):1359–1371, 2020. doi: 10.1158/1078-0432.CCR-19-3249.

Supplementary Information

Figure S1: Detailed AUC Analysis

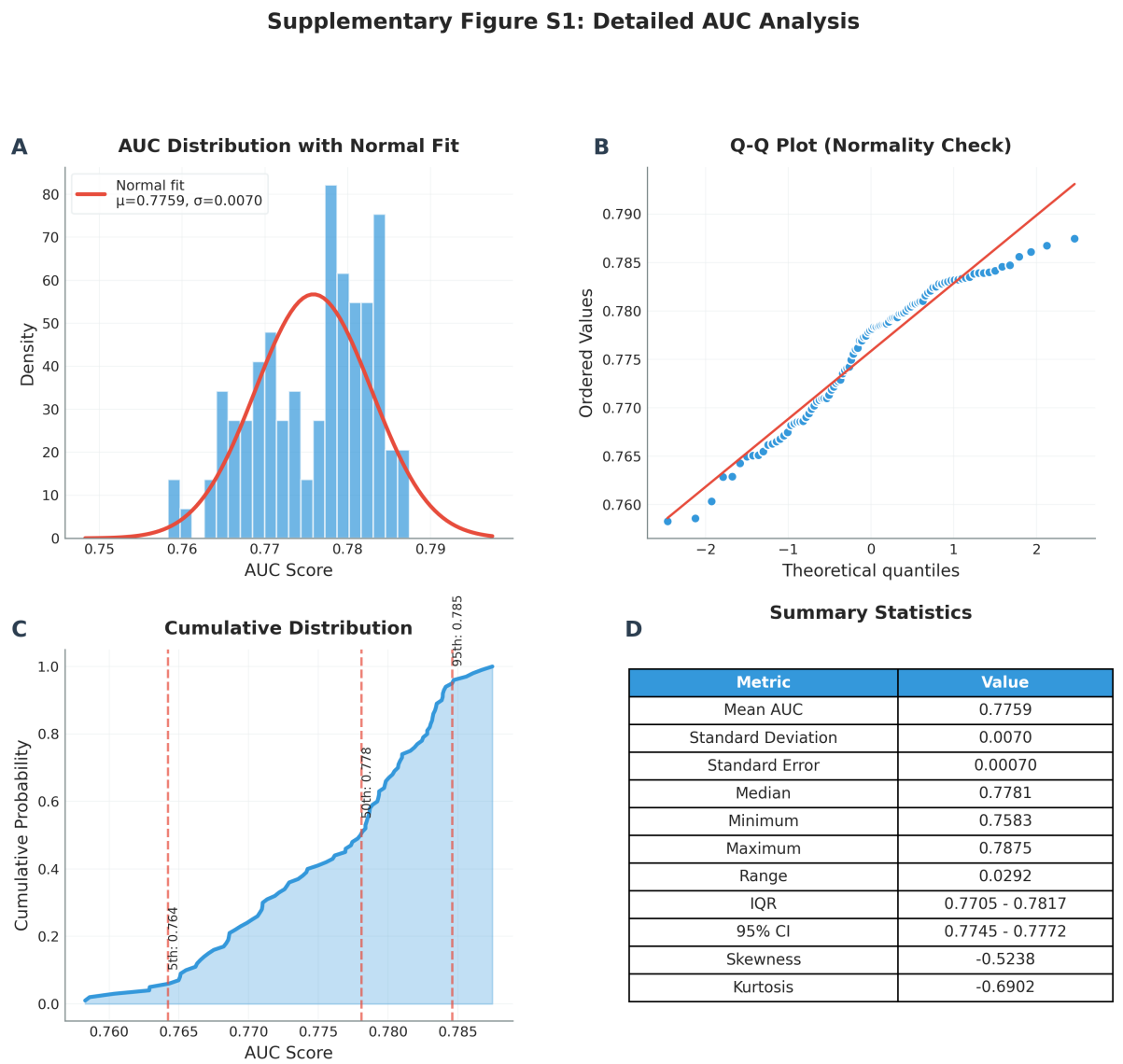**Supplementary Figure S1: Detailed AUC Analysis**

Figure 8: **Detailed AUC Analysis.** (A) Histogram of AUC values across 100 folds. (B) Box plot with individual fold values. (C) Fold-by-fold AUC stability analysis.
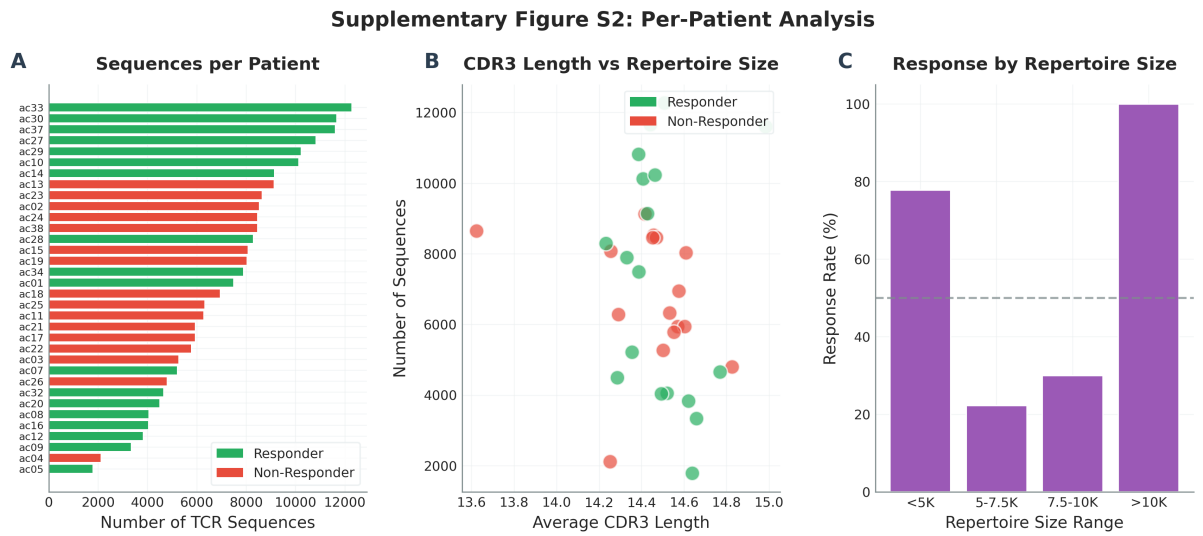
Figure S2: Per-Patient Analysis

Figure 9: **Per-Patient Analysis.** (A) Repertoire size by patient. (B) Prediction confidence by patient. (C) Attention weight distributions across patients.
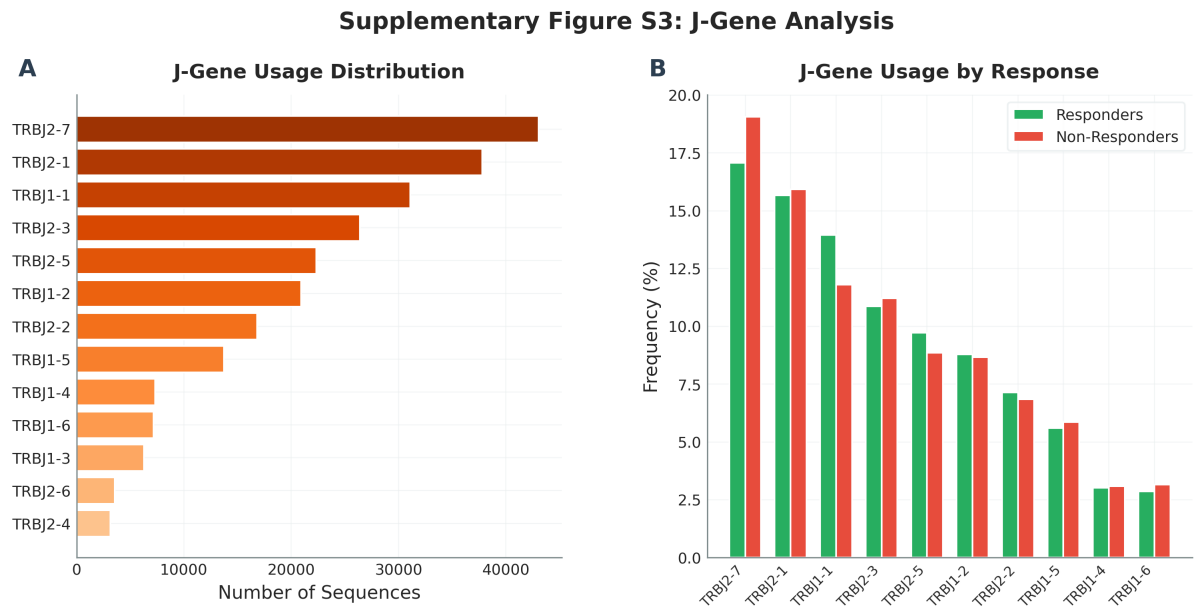
Figure S3: J-Gene Analysis

Figure 10: **J-Gene Usage Analysis.** (A) J-gene frequency comparison between responders and non-responders. (B) J-gene diversity metrics.

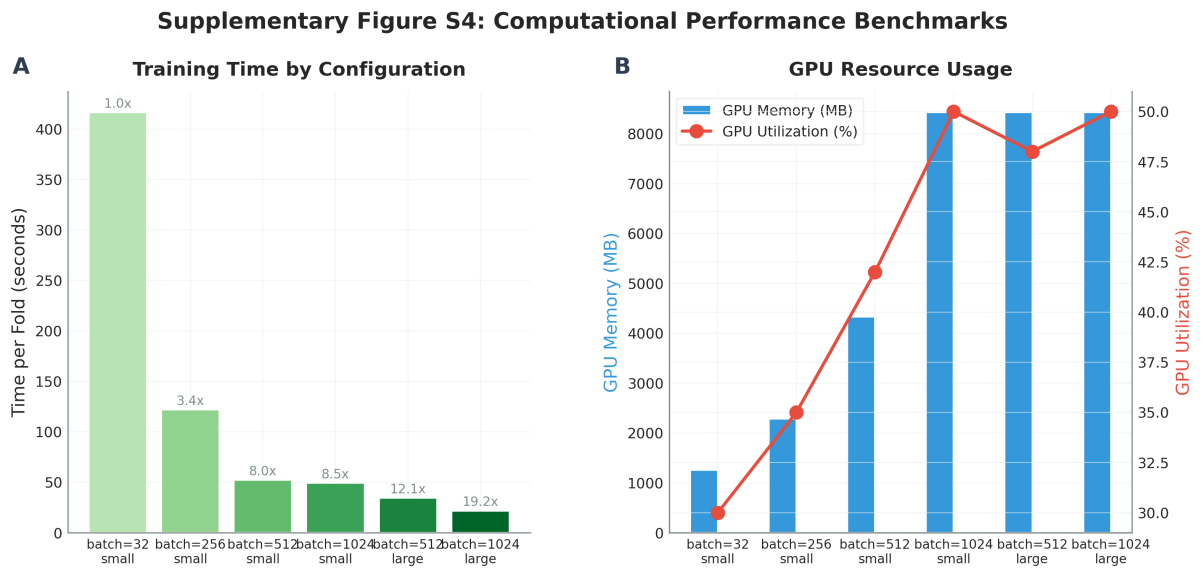## Figure S4: Computational Performance

Figure 11: **Computational Performance.** (A) Training time per fold. (B) GPU memory utilization. (C) Throughput analysis on H100 GPU.