Tool/Model Summary Template

##### 1. Tool/Model Name

The model is referred to as "Deep Generative Selection Models," with mentions of "sONNia" or "soNHNs" in the text, likely a shorthand for a deep neural network-based approach. For consistency, we will use "Deep Generative Selection Models (sONNia)" as the name, acknowledging potential typographical variations in the document.

##### 2. References

The primary reference is:

- Lasschint et al., 2021, "Deep Generative Selection Models of T and B Cell Receptor Repertoires with soNNia," \*Proceedings of the National Academy of Sciences (PNAS)\*, DOI: https://doi.org/10.1073/pnas.2023411118.

Additional cited works include:

- Ref. 41: Mentioned for single-cell datasets of TCR αβ pairs from the human primary T cell receptor repertoire (n=1,000 cells, PAGE4).

- Refs. 34-38: Cited for previous analyses of TCRs and selection processes (PAGE5).

- Ref. 43: Used as an empirical baseline in validation (Fig. 4F, PAGE7).

- SI Appendix: Contains supplementary figures (e.g., Figs. S3, S4, S7D, S11) and Table S1, providing additional data and methods.

##### 3. Category

The model is categorized under "Computational Biology" and "Immunology," specifically as a "Deep Learning" tool for modeling immune receptor repertoires. It aligns with V(D)J simulation and full repertoire simulator categories, given its focus on selection processes post-recombination.

##### 4. Main Purpose/Function

The primary purpose is to model and infer functional selection processes of T cell receptor (TCR) and B cell receptor (BCR) repertoires using deep generative models. It addresses the problem of understanding how these receptors are shaped by selection pressures, including clonal and negative selection, to ensure immune tolerance and effective pathogen recognition. The model focuses on intra- and interchain interactions, predicting properties such as epitope specificity and cell type classification (e.g., CD8+ vs. CD8- TCRs).

##### 5. Key Features

The key features of the model include:

- \*\*Probabilistic:\*\* Yes, it uses probabilistic frameworks like exponential functions and mutual information to model selection processes.

- \*\*ML-based:\*\* Yes, it employs deep neural networks (DNNs) to capture complex correlations between receptor chains.

- \*\*Modular:\*\* Yes, it includes sub-models such as linear models, deep independent models, and deep joint models, which can be used separately or combined.

- \*\*AIRR-compliant:\*\* Not explicitly stated, but it uses standard immunological data, suggesting potential compatibility with Adaptive Immune Receptor Repertoire (AIRR) standards.

- \*\*Supports SHM:\*\* No, the model focuses on post-V(D)J recombination selection, not somatic hypermutation (SHM).

- Other notable features:

- Analyzes mutual information to assess intra- and interchain dependencies.

- Models selection for binding to self-MHC and foreign MHC peptides.

- Supports computational sorting and classification of receptors using log-likelihood ratios.

##### 6. Input Requirements

The model requires the following inputs:

- Single-cell datasets of TCR and BCR repertoires, including:

- TCR α and β chain sequences.

- BCR immunoglobulin heavy (IgH) and light (IgL) chain sequences.

- Features such as gene choice, CDR (complementarity-determining region) length, and amino acid usage (PAGE4, PAGE6).

- Data on receptor interactions with ligands, such as CD8α and MHC peptides.

- Parameters for modeling, including selection factors and interchain correlation data.

##### 7. Output

The outputs generated by the model include:

- Post-selection distributions of TCR and BCR repertoires, such as \(P\_{TCR}(x\_1, x\_2)\) and \(P\_{BCR}(x\_1, x\_2)\) (PAGE6).

- Mutual information metrics showing correlations between receptor features, both intra- and interchain.

- Classification probabilities, such as the likelihood of a receptor being CD8+ or CD8-, using log-likelihood ratios.

- Shannon divergences and selection factors for amino acid usage and other features (PAGE6).

- Predictions of epitope specificity and cell type-specific selection patterns (PAGE8).

##### 8. Validation/Evaluation

The model was validated and evaluated as follows:

- Trained and tested on real-world datasets, including TCR αβ pairs from 1,000 human primary T cells (Ref. 41, PAGE4).

- Compared with empirical baselines and prior studies, such as Ref. 43 (Fig. 4F, PAGE7).

- Evaluation metrics include accuracy in reconstructing interchain correlations, sensitivity, and specificity for classification tasks (e.g., CD8+ vs. CD8- TCRs, PAGE7).

- The model's consistency with observed biological data, such as mutual information and Shannon divergences, was assessed.

- Given its publication in PNAS, it is likely widely cited and recognized in the scientific community, though explicit usage statistics are not provided.

##### 9. Strengths

The strengths of the model include:

- Ability to capture complex intra- and interchain correlations using DNNs, which simpler models might miss.

- High accuracy in modeling receptor selection and diversity, consistent with empirical observations.

- Versatility for both TCRs and BCRs, making it applicable to a broad range of immunological studies.

- Enables computational sorting and prediction of receptor properties, which is valuable for research and potential diagnostic applications.

- Provides insights into immune tolerance and selection processes, enhancing understanding of immune system dynamics.

##### 10. Limitations

The limitations and known issues include:

- May struggle with rare features or individual variability due to sampling issues or differences in HLA (human leukocyte antigen) associations (PAGE4).

- Interchain correlations have not been systematically investigated, suggesting potential gaps in understanding (PAGE4).

- Computationally intensive, requiring large datasets for training DNNs, which may limit scalability.

- No explicit mention of handling noise or incomplete data, which could affect real-world applicability.

- Does not support somatic hypermutation (SHM), limiting its use for certain B cell studies involving affinity maturation.

##### 11. Applications in Autoimmunity

The model has indirect relevance to autoimmune diseases:

- It models selection processes that ensure immune tolerance, such as negative selection to prevent self-reactivity, which is crucial for preventing autoimmune conditions (PAGE9 mentions Treg cells down-regulating effector T cells to prevent autoimmunity).

- It can help study deviations in selection processes that might lead to autoimmune responses, such as altered clonal or negative selection.

- Potential to classify receptor repertoires in autoimmune conditions to identify pathogenic T or B cell populations.

- While not explicitly used for autoimmune diseases in the paper, its ability to model tolerance mechanisms makes it suitable for such applications, particularly in studying conditions like lupus, rheumatoid arthritis, or type 1 diabetes.

##### 12. Notable Citations/Use Cases

Notable citations and use cases include:

- The primary paper itself: Lasschint et al., 2021, \*PNAS\*, DOI: https://doi.org/10.1073/pnas.2023411118.

- Cited previous work on TCRs (Refs. 34-38) for validation of selection processes (PAGE5).

- Use case: Analysis of TCR αβ pairs from 1,000 human primary T cells (Ref. 41, PAGE4).

- Use case: Decomposition of T cell subtypes (e.g., Tregs, Th cells, CD8+ cells) in synthetic mixtures, as shown in SI Appendix, Fig. S11 (PAGE7).

- The publication in PNAS suggests high visibility and credibility, with potential for further citations in immunological research.

##### 13. Reviewer’s Comments

No explicit reviewer comments are provided in the attachment. However, the publication in \*PNAS\* (a prestigious journal) implies that the paper underwent rigorous peer review and was deemed significant. Inferred positive aspects include:

- Novelty in using deep learning for immune receptor modeling, particularly for capturing complex correlations.

- Potential impact on understanding immune tolerance and autoimmune diseases.

Open questions might include:

- How well does the model generalize to larger or more diverse datasets?

- Can it be adapted to include somatic hypermutation for B cell studies?

Suggestions for improvement:

- Systematically investigate interchain correlations to address current gaps.

- Enhance documentation and accessibility, such as releasing open-source code or ensuring AIRR compliance.

- Validate with datasets specific to autoimmune diseases to demonstrate practical utility.

#### Discussion and Analysis

The "Deep Generative Selection Models (sONNia)" represents a significant advancement in modeling immune receptor repertoires, leveraging deep learning to address the complexity of selection processes. Its ability to model both TCRs and BCRs, capture intra- and interchain correlations, and predict receptor properties makes it a versatile tool for immunological research. However, limitations such as the lack of systematic investigation into interchain correlations and the computational intensity highlight areas for future development. The model's potential applications in autoimmunity, particularly in studying immune tolerance, suggest it could play a role in advancing our understanding and treatment of autoimmune diseases, though further validation is needed.

#### Conclusion

This survey note provides a detailed summary of the "Deep Generative Selection Models (sONNia)" based on the provided article, addressing all aspects of the Tool/Model Summary Template. It highlights the model's capabilities, limitations, and potential, offering insights for researchers and practitioners in immunology and computational biology.

\*\*Citations:\*\*

- Lasschint et al., 2021, "Deep Generative Selection Models of T and B Cell Receptor Repertoires with soNNia," \*Proceedings of the National Academy of Sciences (PNAS)\*, DOI: https://doi.org/10.1073/pnas.2023411118.

- Ref. 41: Single-cell datasets for TCR αβ pairs (specific details not provided due to OCR limitations).

- Refs. 34-38: Previous analyses of TCRs (specific details not provided).