# EpiDab: Self-supervised transfer learning-based AI platform for neoantigen predictions

Youngmahn Han[1*], Aeri Lee[2]

[1] Korea Institute of Science and Technology Information
[2] BioBrain Inc.

**Luseed Bio**

Cancer immunotherapy has become a principal pillar of treatment for several cancer types over the past few decades because of its fewer side effects and long-term and wide-ranging anti-cancer effects, and is being optimized for combinations with immune checkpoint inhibitors and other therapies including neoantigen-based cancer vaccines. Artificial Intelligence (AI) is a key tool for dramatically reducing the development costs of cancer vaccines through predicting immunogenic neo-peptide candidates. There are technical limitations such as lack of true immunogenic peptide data and the peptide length variability in constructing a reliable predictive model. In order to overcoming these huddles, here, we present a self-supervised transfer learning-based AI platform to provide user-specific models for predicting immunogenic neo-peptides.
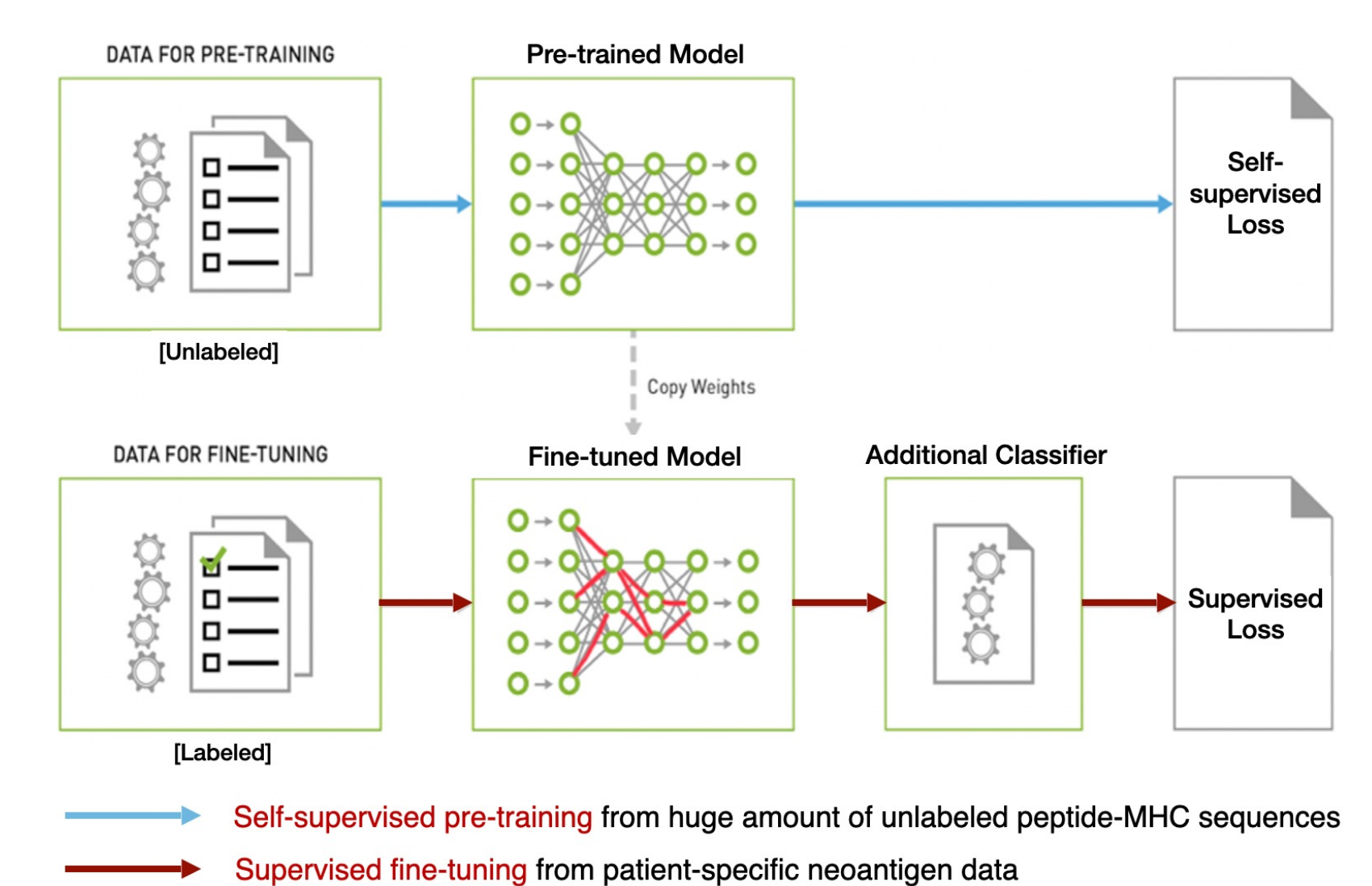
## Introduction

- AI is a key tool in neoantigen-based cancer vaccine development through predicting immunogenic neo-epitopes



Aldous, A. R, *Bioorg. & Med. Chem*, 2018
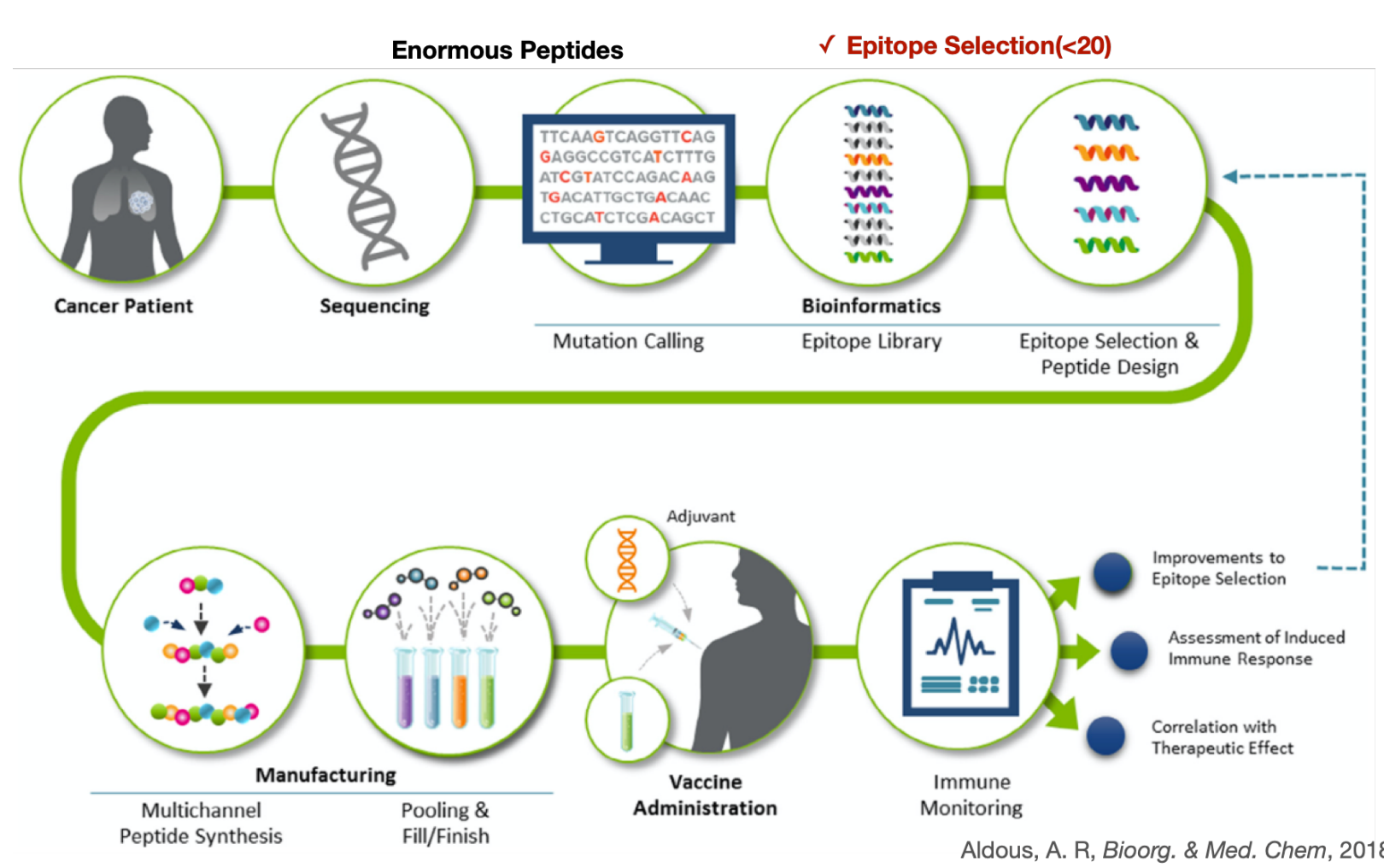
- Peptide length variability leads to complex interaction patterns in peptide-MHC bindings
- Bidirectional self-attention-based protein embeddings can learn contextually such patterns

| Model \ Pattern | Local motif | Long-range interaction | Bidirectional interaction |
|---|---|---|---|
| Simple NN | X | X | X |
| CNN | O | X | X |
| RNN/LSTM | X | △ | △ |
| Uni-directional | X | △ | △ |
| Bi-directional self-attention | △ | O | O |

- Insufficient training data(true neo-epitopes)
- Self-supervised transfer learning is a powerful tool for extracting useful information from unlabeled amino acid sequences



→ Self-supervised pre-training from huge amount of unlabeled peptide-MHC sequences
→ Supervised fine-tuning from patient-specific neoantigen data

## Results

- **EpiDab** is a supercomputing-based AI platform employing the self-supervised transfer learning
- User-friendly web interfaces for generating and managing user-specific AI model from the pre-trained master model
- Training server based on high-performance computing resources and massive storages
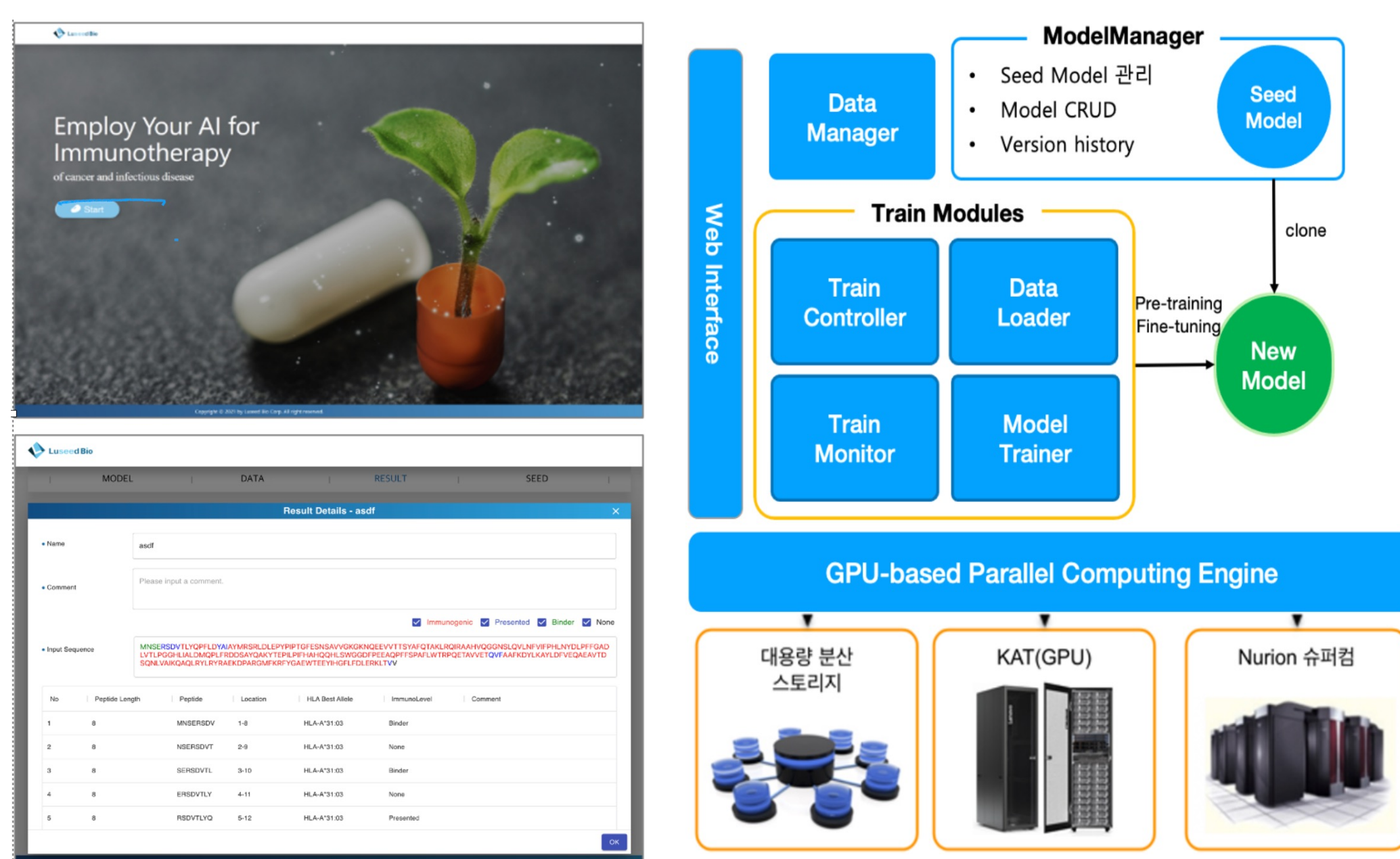


- General pre-trained master model from a large amount of unlabeled protein sequences and peptide-MHC binding data
- Specialized user-specific model can fine-tuned from the pre-trained master model using specific user data
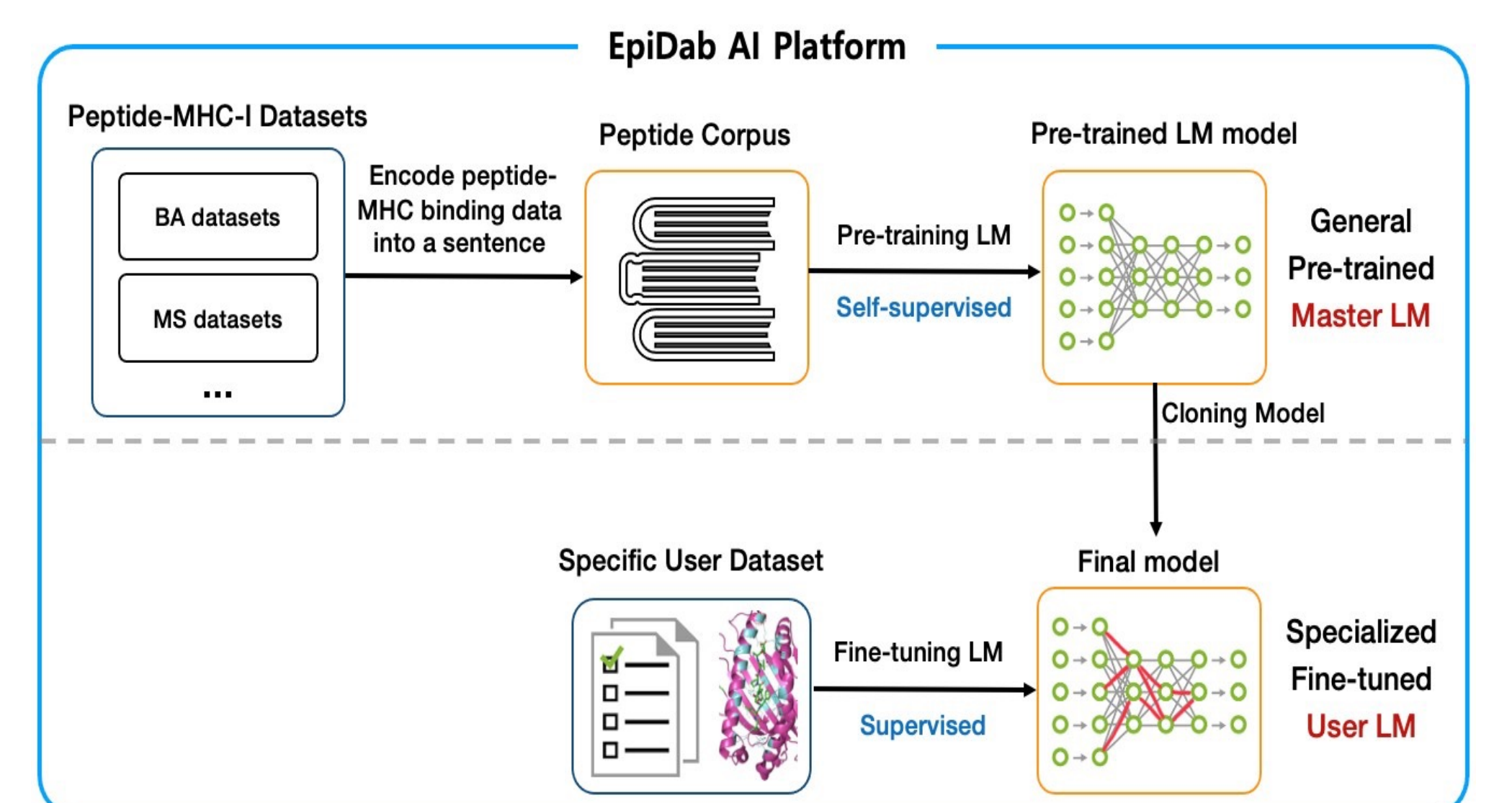


### Proof of Concept

- The user model was fine-tuned from the pre-trained master model using SARS-CoV-2 epitope-specific TCR recognition dataset
- The fine-tuned model was evaluated on SARS-CoV-2 epitope-specific TCR recognition datasets from recent studies(Immunity and ImmuneCODE)



**F1 score: 0.94**     **F1 score: 0.97**
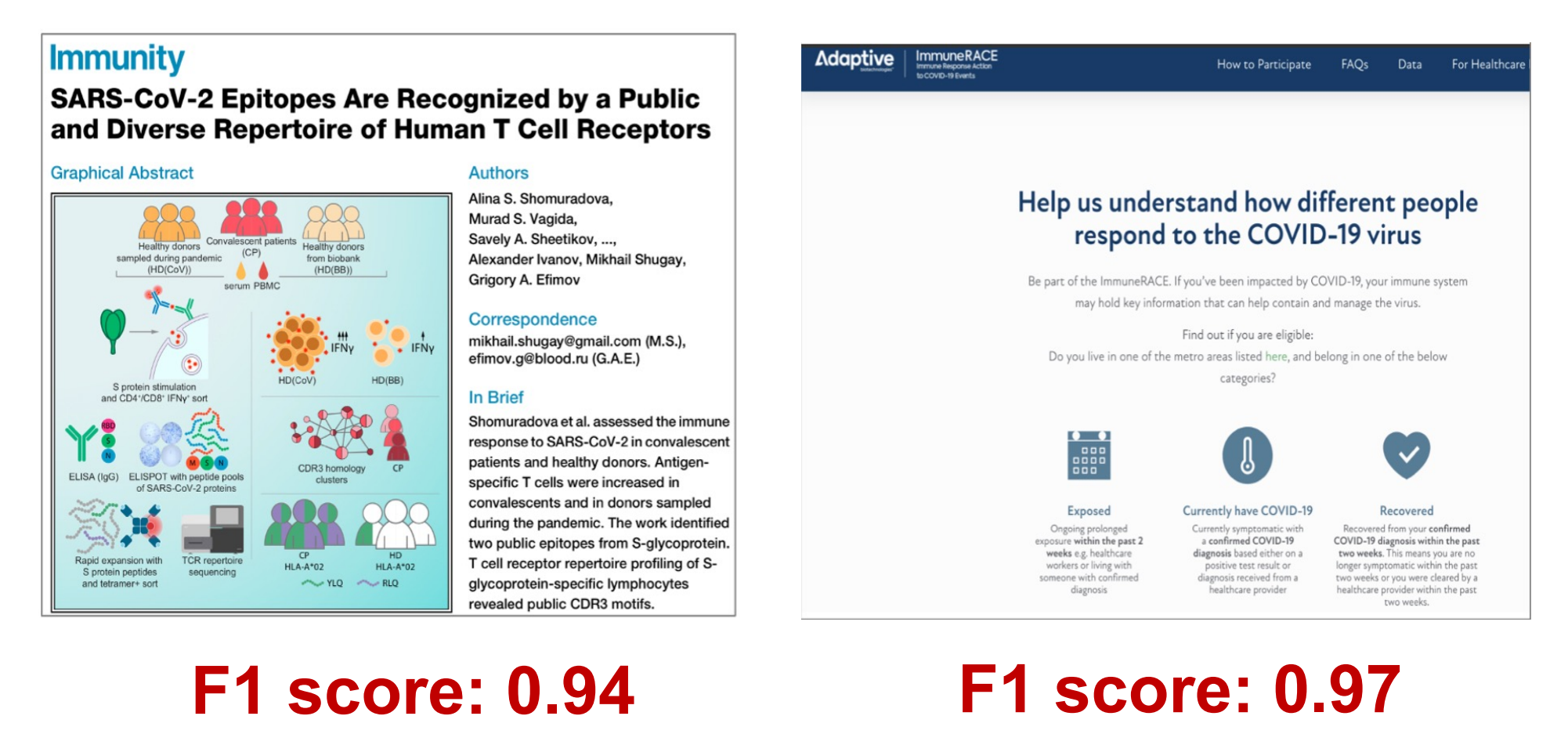
## Concluding Remarks

- **EpiDab** is a supercomputing-based AI platform employing the self-supervised transfer learning to overcome the limitations in increasing the prediction power of AI models, such as insufficient training data and peptide length variability
- We will establish the startup, **LuseedBio** for EpiDab platform services that provide user-specific AI models for predicting immunogenic neo-peptides.
- We anticipate that **EpiDab** AI services dramatically reduce the cost of developing effective vaccines of cancer or novel infectious diseases and increase the success rate of the vaccines

## References

- Vitiello, A. & Zanetti, M. (2017), Neoantigen prediction and the need for validation. *Nature Biotechnology*, **35**, pp. 815–817.
- Vaswani, A. *et al*. (2017), Attention is all you need. In *NIPS*, pp. 5998-6008.
- Aldous, A. R. & Dong, J. Z. (2018), Personalized neoantigen vaccines: A new approach to cancer immunotherapy. *Bioorganic & Medicinal Chemistry* **26**, 2842–2849
- Devlin, J. *et al*. (2018), BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding. *arXiv.org,* arXiv:1810.04805
- Rao, R. *et al*. (2019), Evaluating Protein Transfer Learning with TAPE. In *Advances in neural information processing systems* **32**, pp. 9689–9701.
- Peters, B. *et al*. (2020), T Cell Epitope Predictions. *Annu. Rev. Immunol*., **38**, pp. 123–145