SHERPA™ Bioinformatics Engine

Systematic HLA Epitope Ranking Pan Algorithm

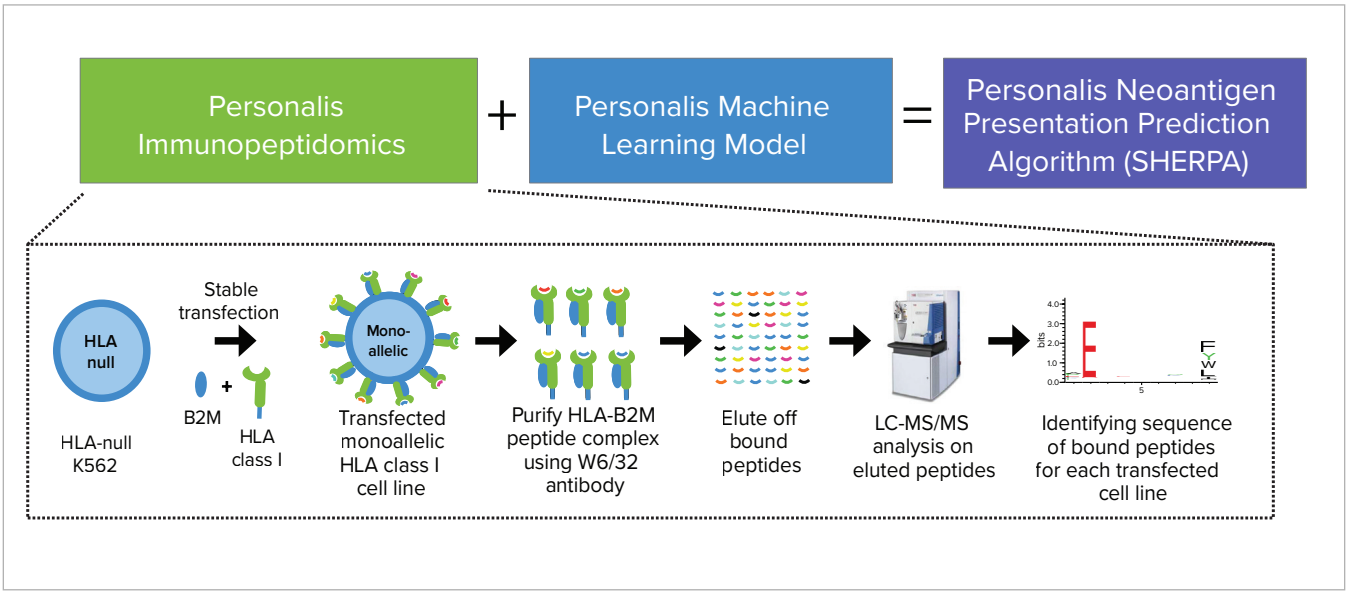
Introduction

Technologies for neoantigen discovery are critical for developing personalized cancer vaccines and neoantigen-based biomarkers. Precision neoantigen discovery entails the comprehensive detection of tumor-specific genomic variants and accurate prediction of MHC presentation of epitopes originating from such variants. The ImmunoID NeXT Platform® enables a comprehensive survey of putative neoantigens by combining highly sensitive exome scale DNA and RNA sequencing with advanced analytics. We utilized the Systematic HLA Epitope Ranking Pan Algorithm (SHERPA), our in-house developed machine learning model, for predicting MHC class I presentation and identifying potentially immunogenic patient-specific neoantigens.

Generating monoallelic immunopeptidomics training data

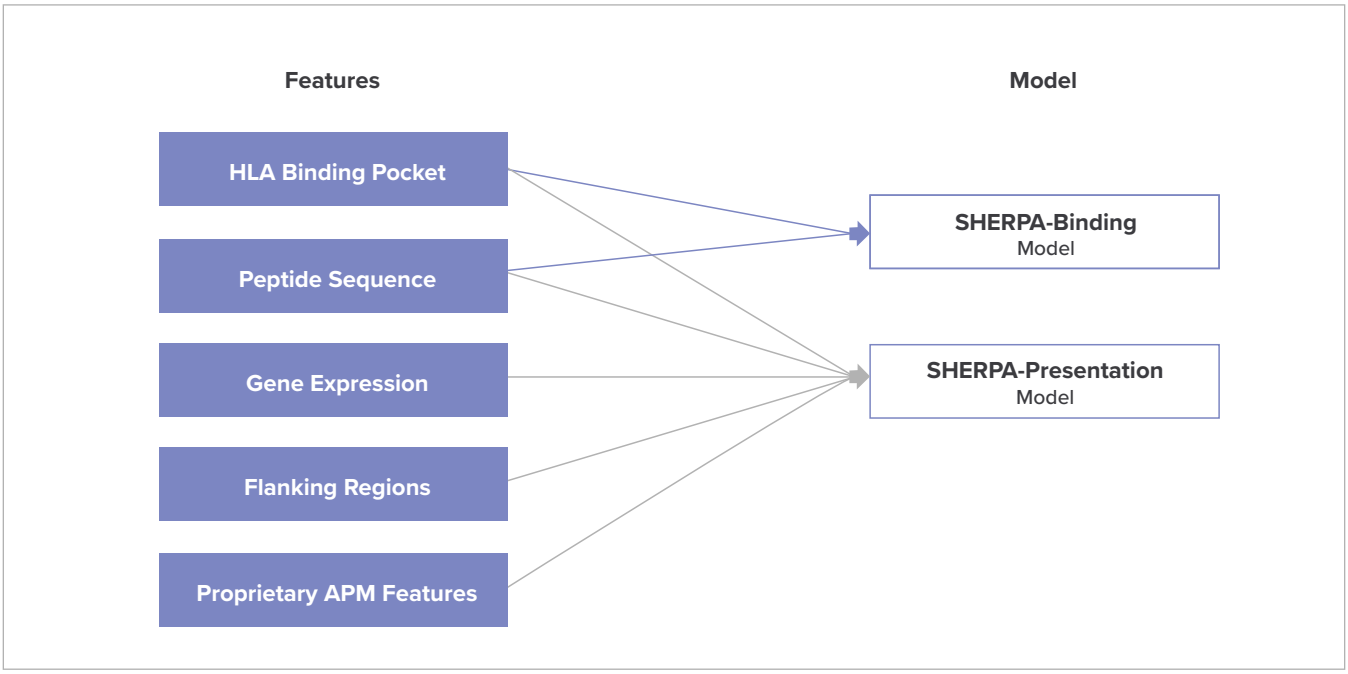
Only a small fraction of putative neoantigens are successfully presented by the MHC molecules. Accordingly, accurate prediction of antigen presentation is critical for neoantigen discovery. We trained SHERPA, our advanced machine learning model, using high-quality immunopeptidomics data generated from genetically engineered K562 cell lines that express a single allele of interest.

Figure 1A: Overview of SHERPA Training Data Set



Personalis®

Figure 1B: SHERPA Models



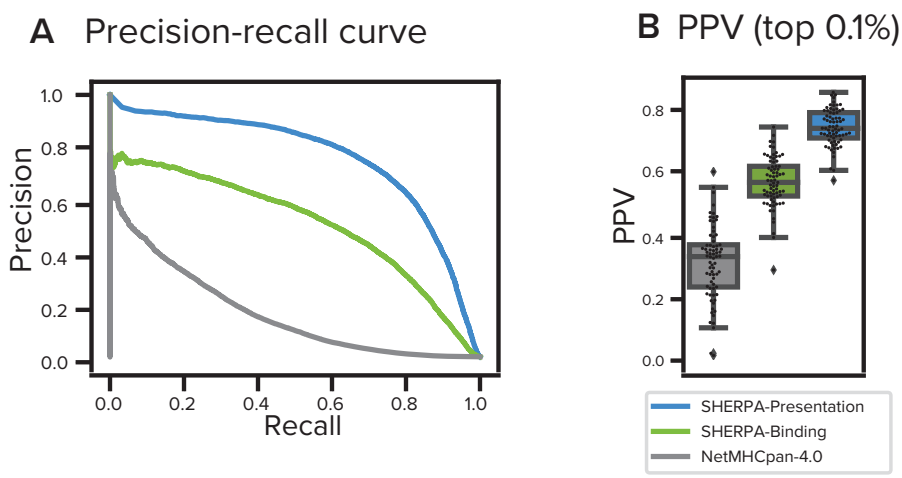
A defined set of HLA alleles were used to optimize both allelic diversity and population coverage to enable accurate and comprehensive modeling of the MHC-peptide complex processing and presentation. MHC-peptide complexes were immunoprecipitated using W6/32 antibody followed by peptide elution and peptide sequencing using tandem mass spectrometry (**Figure 1A**).

We trained two algorithms that model MHC-peptide binding and presentation using in-house generated data and publicly available data for HLA-A, HLA-B, and HLA-C alleles, with standard and proprietary features (**Figure 1B**).

Performance of SHERPA on held-out monoallelic data

The performance of SHERPA was first evaluated using 10% of the monoallelic immunopeptidomics data (held-out from training) mixed with synthetic negative examples in a 1:999 ratio (commonly assumed prevalence). SHERPA models have higher precision over all recall values compared to NetMHCPan-4.0, the state-of-the-art publicly available tool (**Figure 2A**), and significantly higher positive predictive values among the top 0.1% peptides in the test data (**Figure 2B**).

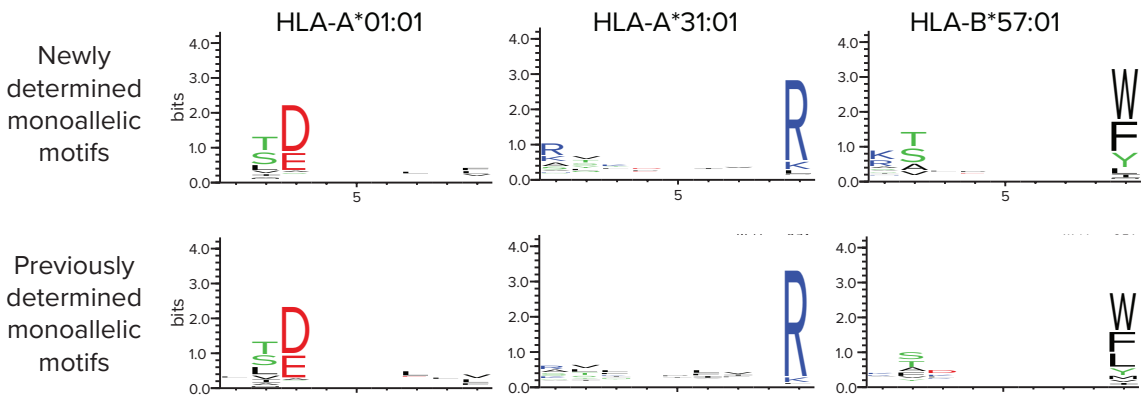
Figure 2A and 2B: SHERPA Enables Superior Neoantigen Presentation Prediction



Pan-allelic performance of SHERPA on unseen alleles

The wide genetic variability in the HLA loci necessitates the creation of a pan-allelic prediction model in order to apply it to patient samples with uncommon alleles. Joint modeling of alleles and presented peptides enables SHERPA to make predictions for alleles that have not been seen by the model. The examples shown in Figure 3 indicate a high degree of agreement between predicted motifs of 'unseen' alleles and corresponding raw immunopeptidomics data, demonstrating strong pan-allelic performance.

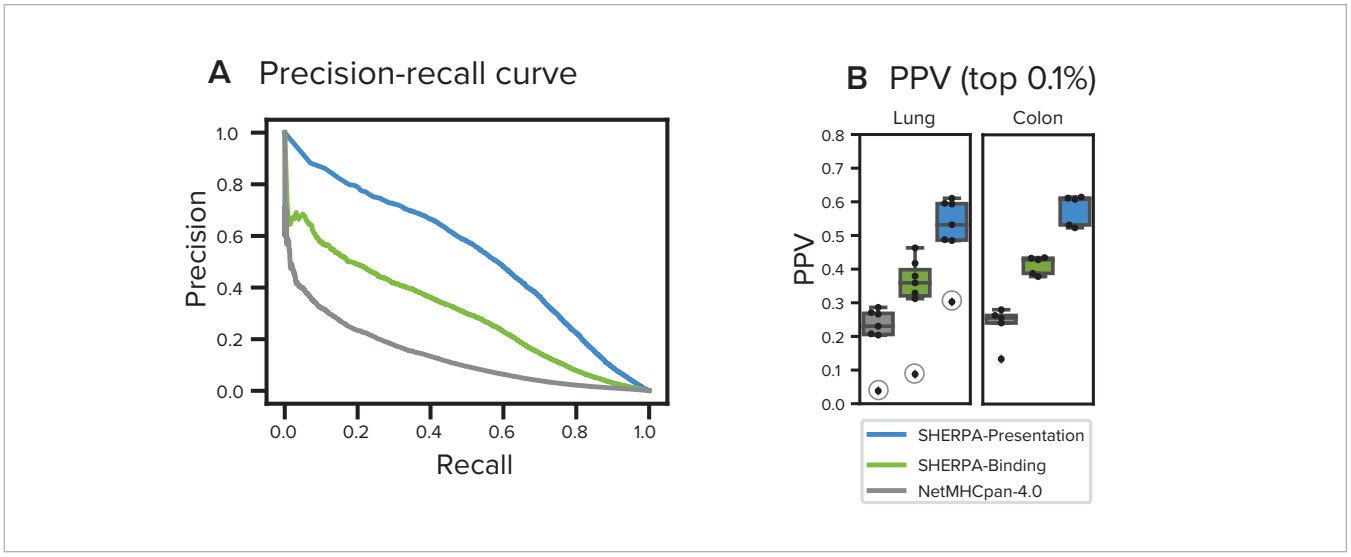
Figure 3: SHERPA Accurately Predicts Motifs Where Training Data Was Withheld



Performance of SHERPA on independent tissue samples

We further evaluated the performance of SHERPA by performing both ImmunoID NeXT™ and immunopeptidomics on the same tumor tissue samples. Patient-specific scores for each antigen were determined by taking the best score across all HLA alleles in the sample. SHERPA has consistently higher performance both on precision-recall curves (**Figure 4A**) and positive predictive value (PPV) estimates (**Figure 4B**). Interestingly, we observed significantly reduced PPV estimates for a sample (gray circles) with loss of heterozygosity (LOH)*, indicating the importance of HLA-LOH status in such analyses.

Figure 4A and 4B: Validation of Model Performance On Tumor Samples

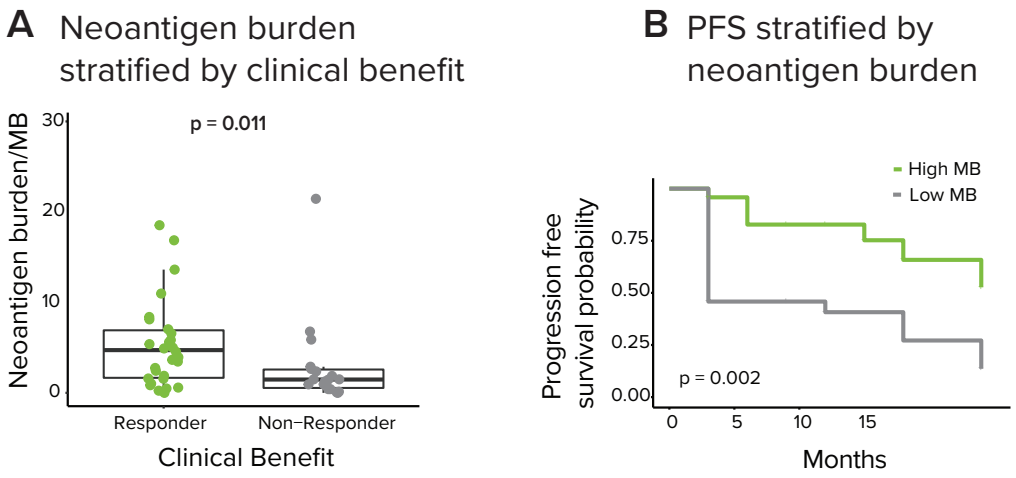


*Personalis has developed a sensitive method called Deletion of Allele-Specific HLAs (DASH) to accurately detect loss of heterozygosity (LOH) in the HLA locus.

Developing biomarkers for immunotherapy using SHERPA

Applying SHERPA to a cohort of 55 unresectable, stage III/IV melanoma patients treated with anti-PD-1 therapy, we see a clear difference in the neoantigen burden between responders and non-responders (**Figure 5A**). Also, neoantigen burden is a powerful prognostic biomarker to stratify patients by progression free survival (**Figure 5B**). SHERPA enables the development of more advanced biomarkers than tumor mutational burden.

Figure 5A and 5B: Neoantigen Burden As a Powerful Prognostic Biomarker



Summary

In conclusion, we have developed SHERPA, a machine learning-based prediction model for neoantigen discovery, using large-scale and high-quality immunopeptidomics data from genetically engineered monoallelic cell lines. SHERPA consistently shows higher performance in comparison to the widely accepted and state-of-the-art publicly available tool (NetMHCPan 4.0), both on held-out monoallelic data and tissue samples. Further, the pan-prediction capabilities of SHERPA enable accurate prediction of neoantigens from patient samples and allow the development of personalized cancer therapy and neoantigen-based biomarkers that are prognostic of response to immunotherapy.

Get in Touch

To learn more about how we can help accelerate your personalized cancer vaccine or neoantigen-based adoptive cell therapy development program, contact us at **info@personalis.com.**

Sales Contact

United States
info@personalis.com

Europe
europe@personalis.com

Other Countries
info@personalis.com

Personalis, Inc.

1330 O'Brien Drive, Menlo Park, CA 94025
+1 855-GENOME4 (436-6634) | +1 650-752-1300
www.personalis.com

© 2020 Personalis, Inc. All rights reserved. Personalis®, ACE®, ImmunoID NeXT Platform®, and SHERPA™, are trademarks of Personalis, Inc., ("Personalis") in the United States and/or other countries.

For Research Use Only. Not for use in diagnostic procedures.

Personalis®

102-326-A