Precision neoantigen discovery using a pan-predictive machine learning model integrated into ImmunoID NeXT Platform®

#2085

Dattatreya Mellacheruvu*, Rachel Marty Pyke*, Charles Abbott, Nick Phillips, Rena McClory, John West, Richard Chen and Sean Michael Boyle

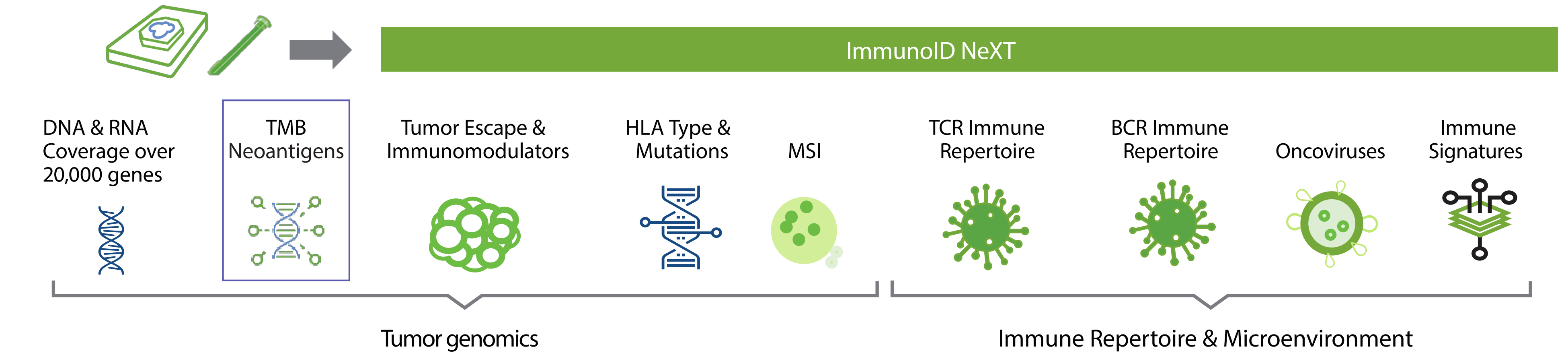
Personalis, Inc. | 1330 O'Brien Dr., Menlo Park, CA 94025
* co-authors

Contact:
Datta.Mellacheruvu@personalis.com
Rachel.Pyke@personalis.com
Sean.Boyle@personalis.com

I. Introduction

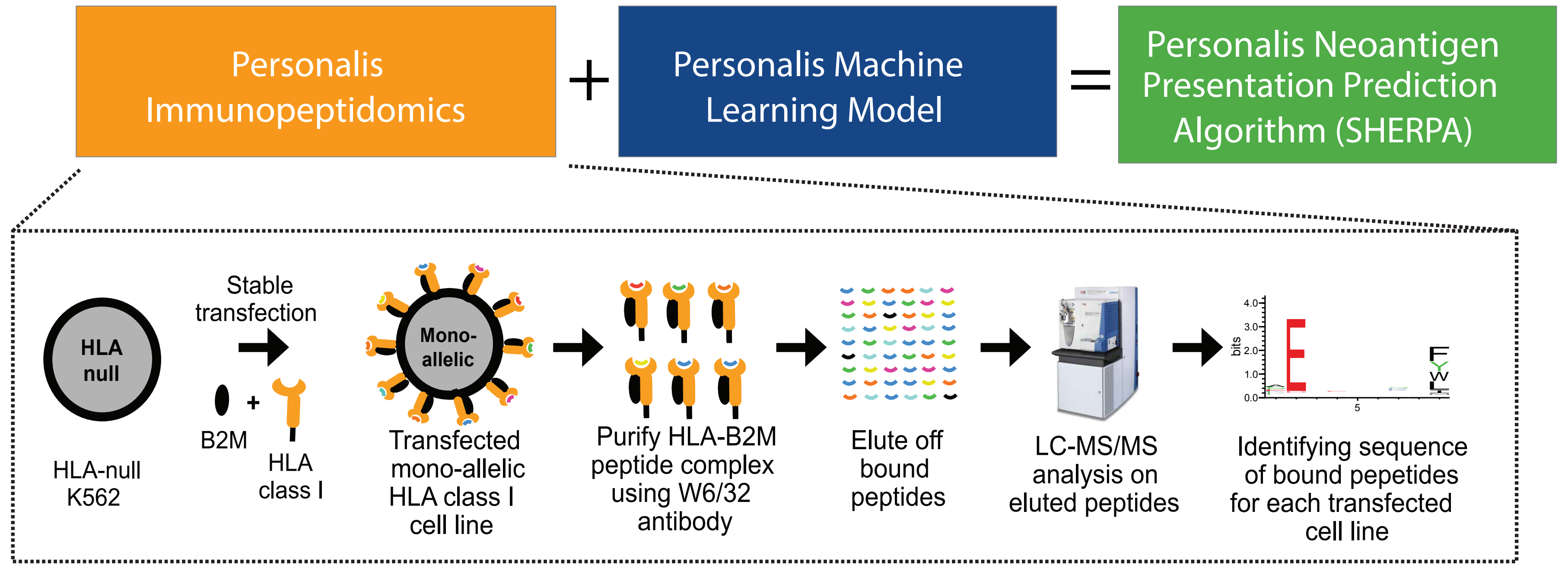
Technologies for neoantigen discovery are critical for developing personalized cancer vaccines and neoantigen-based biomarkers. Precision neoantigen discovery entails comprehensive detection of tumor-specific genomic variants and accurate prediction of MHC presentation of epitopes originating from such variants. Our ImmunoID NeXT Platform enables a comprehensive survey of putative neoantigens by combining highly sensitive and exome scale DNA and RNA sequencing with advanced analytics. Here, we present **S**ystematic **H**LA **E**pitope **R**anking **P**an **A**lgorithm (SHERPA), our pan-predictive machine learning model for predicting MHC class I presentation and identifying potentially immunogenic patient-specific neoantigens.

II. Comprehensive neoantigen profiling using ImmunoID NeXT



Our immuno-oncology platform (ImmunoID NeXT) enables researchers to analyze both a tumor and its microenvironment from a single tumor sample. In-depth interrogation of tumor and normal samples and identification of tumor-specific genomic events allows us to comprehensively profile the landscape of potential neoantigens, a critical aspect of precision neoantigen discovery.

III. Generating mono-allelic immunopeptidomics training data

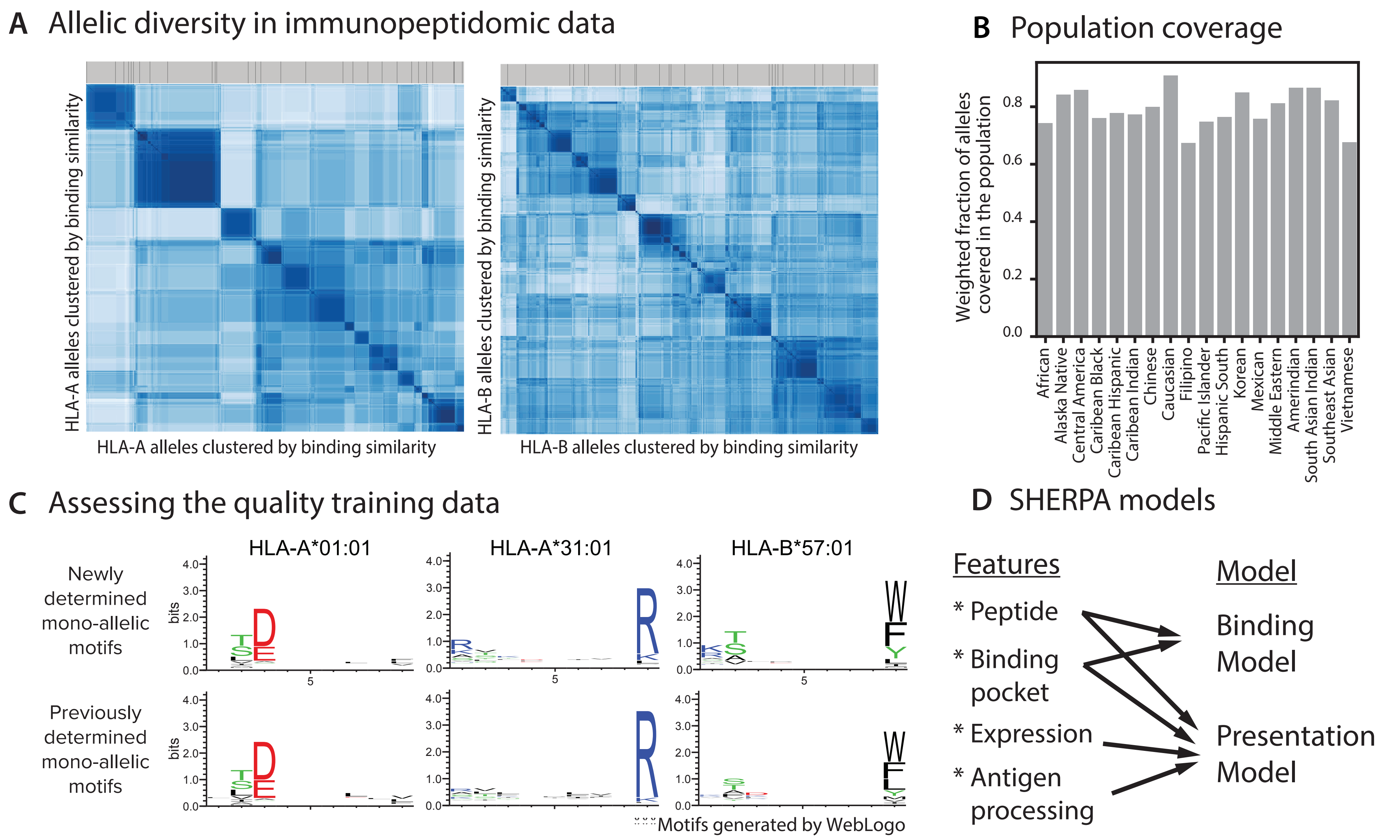


Personalis Immunopeptidomics + Personalis Machine Learning Model = Personalis Neoantigen Presentation Prediction Algorithm (SHERPA)

HLA null
HLA-null K562
Stable transfection
B2M
HLA class I
Mono-allelic
Transfected mono-allelic HLA class I cell line
Purify HLA-B2M peptide complex using W6/32 antibody
Elute off bound peptides
LC-MS/MS analysis on eluted peptides
Identifying sequence of bound pepetides for each transfected cell line
bits
4.0
3.0
2.0
1.0
0.0
5

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 advanced machine learning models (SHERPA) using high-quality immunopeptidomics data generated from 58 genetically engineered K562 cell lines that express a single allele of interest. MHC-peptide complexes were immunoprecipitated using W6/32 antibody followed by peptide elution and peptide sequencing using tandem mass spectrometry.

IV. Overview of training data and prediction models



A Allelic diversity in immunopeptidomic data

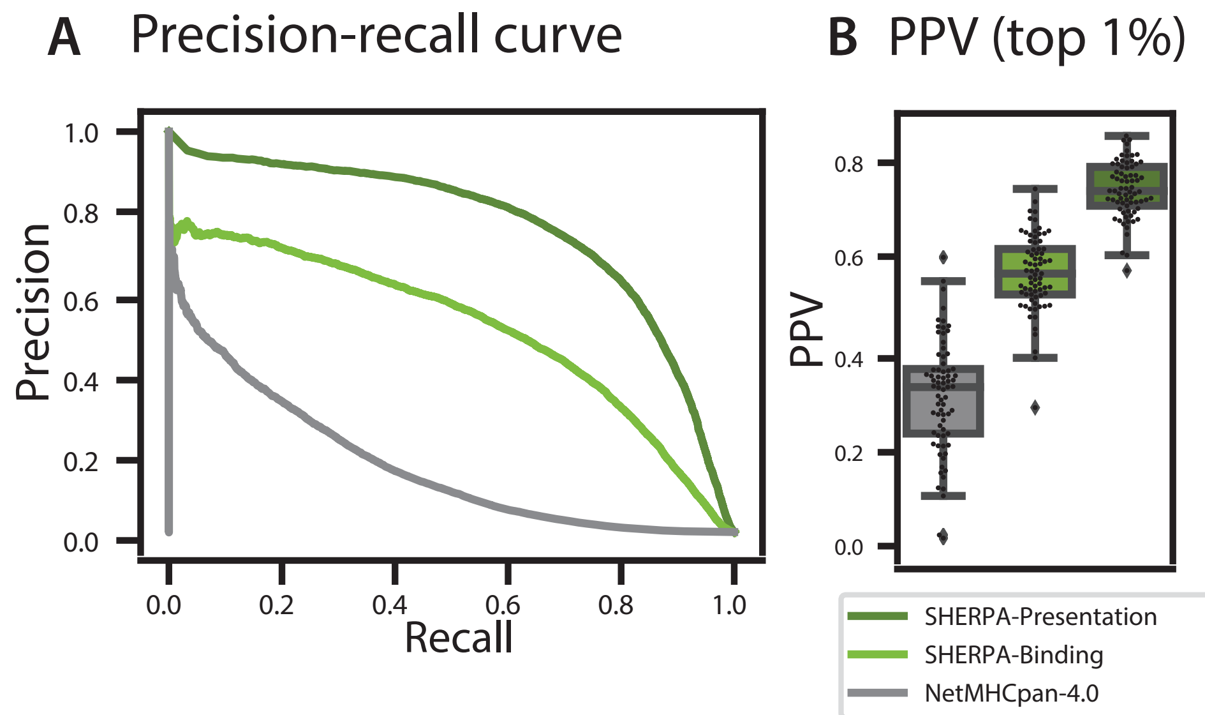
B Population coverage

C Assessing the quality training data

D SHERPA models

The performance of prediction models heavily depends on the size and representativeness of training data. Given the wide variability of HLA loci among populations, we demonstrate that the 58 alleles selected for immunopeptidomics are diverse in their binding characteristics (panel A) and prevalent among various ethnicities (panel B). The alleles we profiled represent many distinct sub-clusters (dark bars in panel A) on a heat map of alleles clustered by their similarity. Based on population frequency data from the National Marrow Donor Program, we estimate ~80% allelic coverage among various populations. Our raw data has strong motifs as shown in panel C. We trained two algorithms that model MHC-peptide binding and presentation using 58 in-house HLA-A and HLA-B alleles and 15 publicly available HLA-C alleles, with standard and proprietary features (panel D).

V. Performance of SHERPA on held-out mono-allelic data

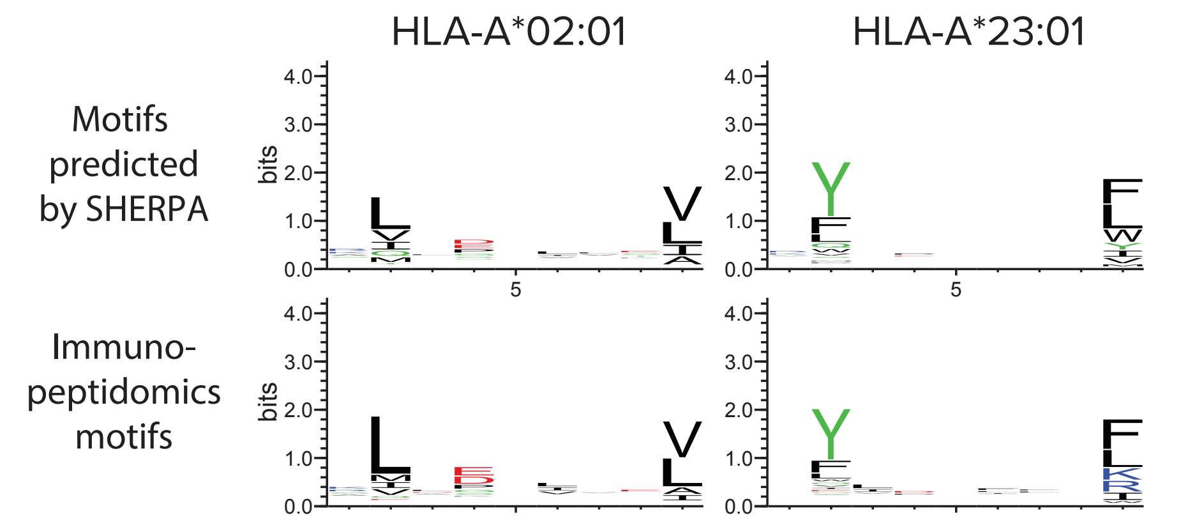


A Precision-recall curve

B PPV (top 1%)

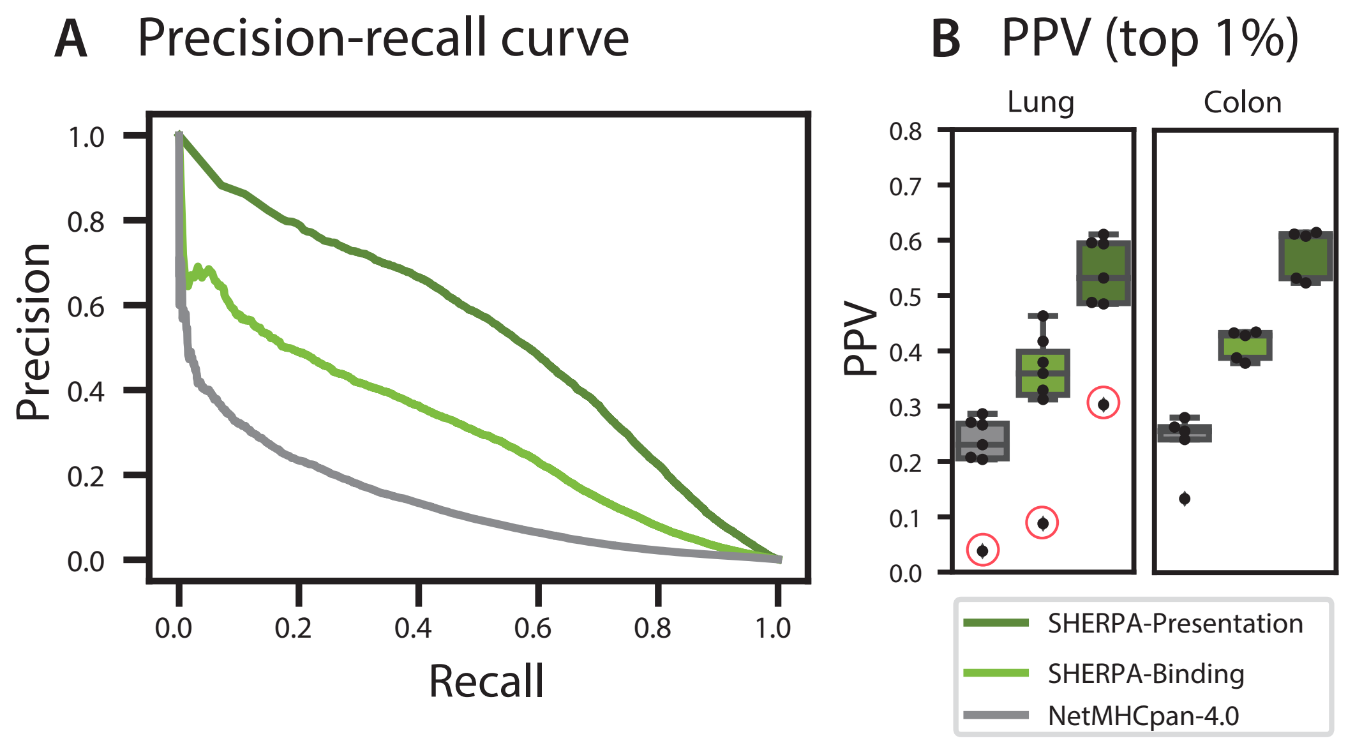
The performance of SHERPA was first evaluated using 10% of the immunopeptidomics data (held-out from training) mixed with synthetic negative examples in a 1:999 ratio (commonly assumed prevalence). SHERPA models have higher precision over all recall values compared to NetMHCPan-4.0, the state-of-the-art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

VI. 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 are not seen by the model. The examples shown here indicate a high degree of agreement between predicted motifs of 'unseen' alleles and corresponding raw data, demonstrating strong pan-allelic performance.

VII. Performance of SHERPA on independent tissue samples

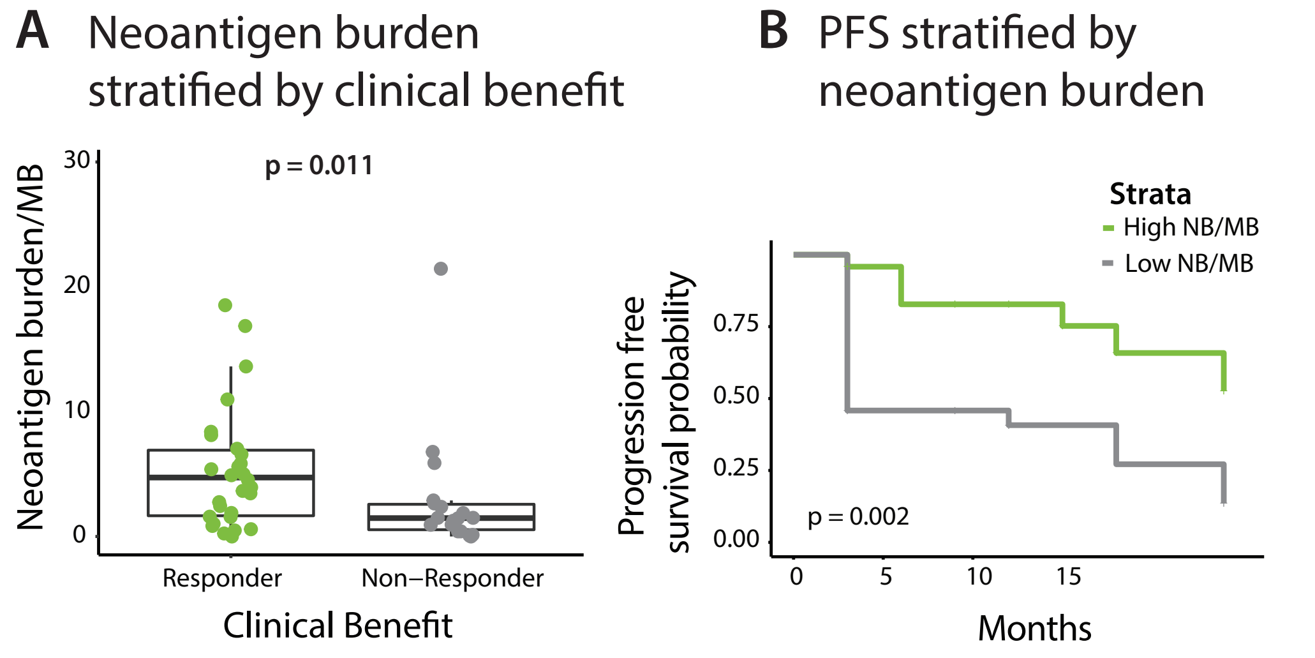


A Precision-recall curve

B PPV (top 1%)

We further evaluated the performance of SHERPA by performing both ImmunoID NeXT Platform analysis and immunopeptidomics on the same tissue samples. Patient-specific scores for each antigen were calculated by aggregating prediction scores across all HLA alleles in the sample. SHERPA has consistently higher performance both on PR curves (panel A) and positive PPV estimates (panel B). Interestingly, we observe significantly reduced PPV estimates for a sample (circled in red) with loss of heterozygosity as estimated by DASH (Personalis poster #6678), indicating the importance of HLA-LOH status in such analyses.

VIII. Developing biomarkers for immunotherapy using SHERPA



A Neoantigen burden stratified by clinical benefit

B PFS stratified by neoantigen burden

We have shown that the presence of neoantigens is a good prognostic biomarker for response to immunotherapy (Personalis poster #2478). 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 (panel A). Also, neoantigen burden is a powerful prognostic biomarker to stratify patients by progression free survival (panel B). SHERPA enables the development of more advanced biomarkers than tumor mutational burden.

IX. Summary and concluding remarks

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

Personalis®