Applying immunopeptidomics and machine learning to improve neoantigen prediction for therapeutic and diagnostic use

#4465

Dattatreya Mellacheruvu, Nick Phillips, Rachel Marty Pyke, Gabor Bartha, Jason Harris, Robert Power, Rena McClory, John West, Richard Chen, Sean Michael Boyle

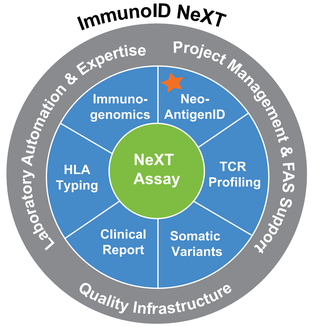
Personalis, Inc. 1330 O'Brien Drive, Menlo Park, CA 94025

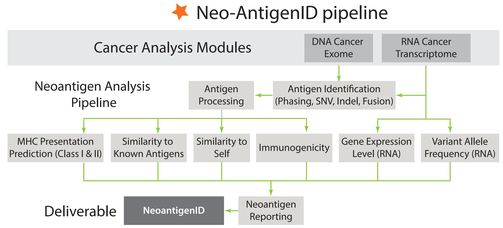
Contact: sean.boyle@personalis.com

Background

Neoantigens are increasingly critical in immuno-oncology as therapeutic targets for neoantigen-based personalized cancer vaccines (PCVs) and as potential biomarkers for immunotherapy response. However, optimizing technologies for identifying neoepitopes that are more likely to provoke an immune response remains an important challenge. Current major histocompatibility complex (MHC) presentation prediction algorithms are primarily trained using *in vitro* MHC binding data, which does not encompass certain important factors for neoantigen presentation such as proteasomal cleavage and transport. Recent advances in immuno-affinity purification and mass spectrometry technology make it possible to identify processed cell surface MHC bound peptides in an *in vivo* setting, providing the opportunity for the development of improved neoantigen prediction pipelines.

ImmunoID NeXT platform & the neoantigen analytics engine





Within our neoantigen pipeline, variants that are detected by DNA and RNA cancer analysis pipelines are processed for antigen identification, including SNVs, indels, and fusion events. Collectively, our ImmunoID product provides a comprehensive assessment of features that may be used for identifying and ranking potentially immunogenic neoantigens.

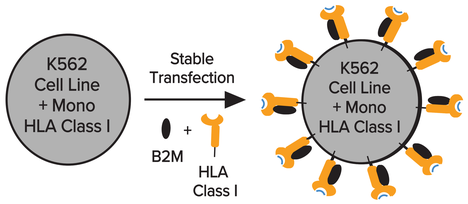
Prediction models for MHC presentation

Generating high-quality training data → **Building accurate prediction models**

Generating mono-allelic cell lines → Immuno-peptidomics for identifying MHC-bound peptides → Training machine learning models → Evaluation & benchmarking

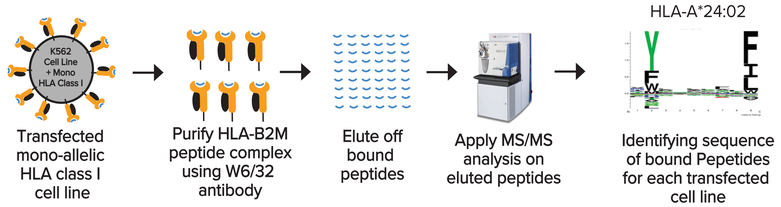
Generating mono-allelic HLA class I cell lines

Mono-allelic human leukocyte antigens (HLA) class I cell lines were generated by transfecting individual class I HLA alleles into the HLA class I null K562 cell line, prioritizing HLA class I alleles that are of high frequency across different populations.

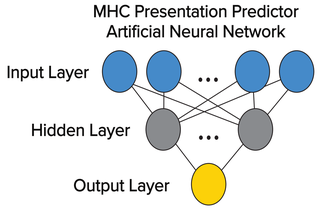


Immuno-peptidomics: identifying HLA-bound peptides

Immuno-peptidomics was applied to isolate and identify HLA class I bound peptides. As all cells in each stably transfected population expressed only one HLA allele, all immuno-precipitated peptides from a pool of cells were known to be bound to that specific HLA allele.

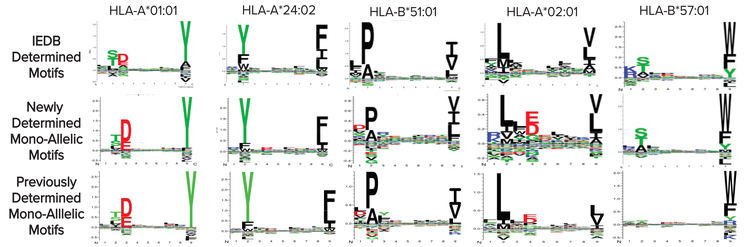


Prediction model



We developed and trained neural networks to predict MHC class I presentation for each assayed HLA class I allele. Our recent advances in training data generation, including our mono-allelic MHC class I cell line generation and immuno-peptidomics, provide an opportunity to accurately model MHC-peptide presentation. This new approach to generating binding data takes into consideration cleavage and transportation, which are critically important for presentation assessment. Leveraging these advancements, we developed a brand-new prediction algorithm which outperforms publicly available prediction tools.

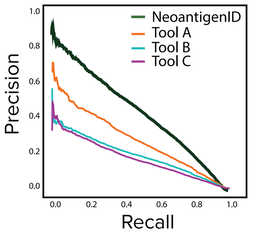
Evaluating newly generated data by comparing binding motifs

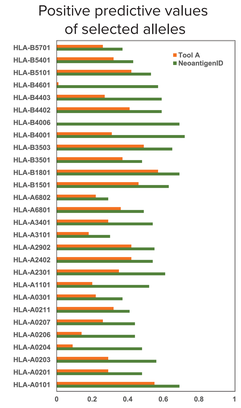


We examined the similarities and distinctions between both IEDB and mono-allelic derived published motifs for the above five HLA class I alleles. For example, our A*02:01 peptides contain a similar signature at position 4 to mono-allelic motifs. However, we retain the stronger VL pattern at position 9, indicating a more conserved binding preference. As peptides derived from mono-allelic lines do not require deconvolution and include proteasomal processing, we have increased confidence in our newly identified peptide motifs over multi-allelic results.

Assessing the performance of our prediction algorithm

Our neoantigen prediction algorithm (**NeoantigenID**), trained on our own *in vivo* peptide data, consistently achieves a higher overall sensitivity and specificity than other commercially available tools based on either *in vitro* MHC binding or immuno-peptidomics derived data.





Conclusions

* ImmunoID NeXT platform generates a comprehensive and accurate list of candidate neo-antigens.

* Our immuno-peptidomics platform using mono-allelic cell lines generates high quality training data.

* Our MHC presentation prediction platform out-performs other publicly available tools, with significant implications for therapeutic and diagnostic use.

Personalis®