NeoantigenID™

An analytical module of ImmunoID NeXT™

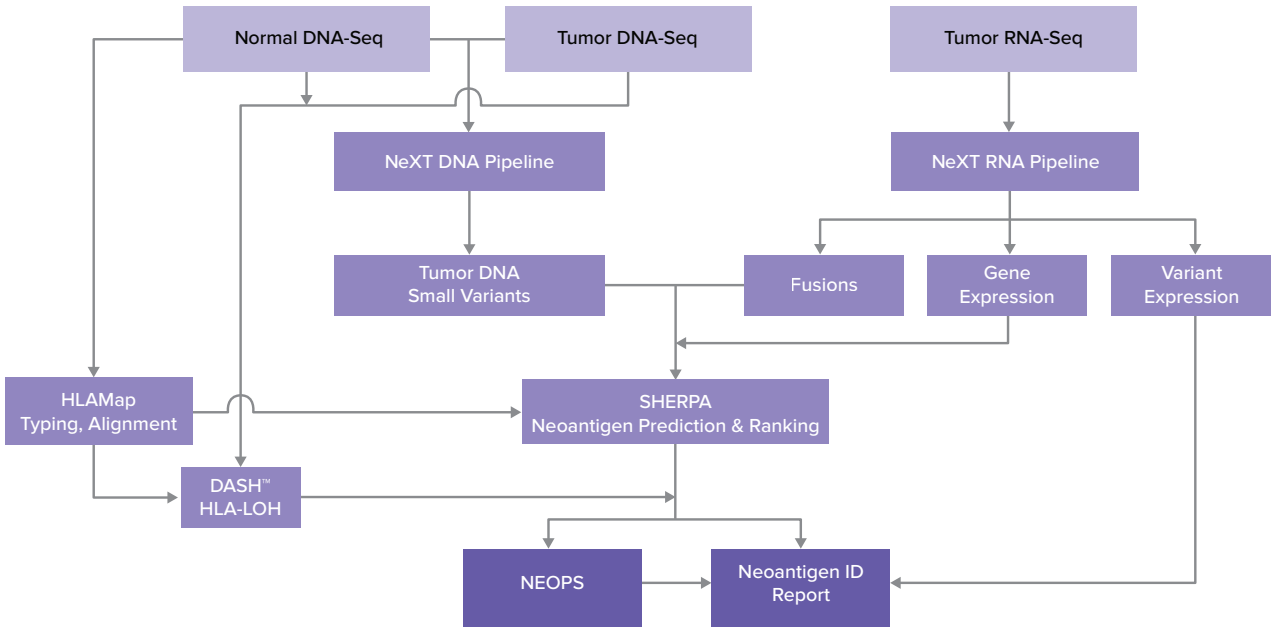
Comprehensive Identification of Putative Neoantigens

Neoantigens are mutated peptides that are expressed by tumor cells, but not by normal tissue. As a result, they can be recognized as foreign antigens by cells of the immune system — making them promising targets for personalized cancer immunotherapies such as vaccines and adoptive cell therapies. However, because neoantigens can arise from somatic mutations occurring in any gene in the genome, neoantigen identification requires both exome-scale DNA and RNA sequencing.

NeoantigenID, a suite of advanced analytics available via ImmunoID NeXT, leverages Personalis' augmented exome and transcriptome for accurate identification of tumor-specific somatic mutations (**Figure 1**). These mutations are a rich source of putative neoantigens. Current computational neoantigen prediction algorithms suffer from limited training datasets and poor performance. To overcome these challenges, Personalis has developed a Systematic HLA Epitope Ranking Pan Algorithm (SHERPA™) to improve neoantigen presentation prediction based on MHC Class I binding potential, level of gene expression, and other key parameters. SHERPA is integrated into the NeoantigenID analytics engine to deliver a comprehensive characterization of putative neoantigens.

Accurate neoantigen prediction with SHERPA also enables the determination of neoantigen burden as well as the generation of the Personalis Composite Neoantigen Presentation Score (NEOPS™) which has the potential to improve the predictive and/or prognostic utility of these biomarkers for precision oncology and immunotherapy applications.

Figure 1: NeoantigenID Pipeline Diagram



Personalis®

It Starts with a Superior Assay

Robust neoantigen identification requires more than standard sequencing and informatics methods. Personalis' NeXT™ assay combines DNA and RNA sequencing across >20,000 genes. Features include:

- **Augmented Coverage:** Proprietary Accuracy and Content Enhanced (ACE) Technology augments coverage of difficult-to-sequence regions (e.g. areas of high-GC content) across all >20,000 genes that are typically poorly covered or missed with conventional sequencing approaches.
- **Deep Sequencing:** ~300X mean whole exome sequencing coverage and 200M total reads of RNA sequencing of the tumor specimen.
- **High Sensitivity:** Accurate detection of low-abundance single nucleotide variants (SNVs), insertions/deletions (indels), and gene fusions.
- **Improved HLA:** Specific targeting of human leukocyte antigens (HLA) loci for accurate HLA typing.

HLA genotyping is an essential component of the neoantigen prediction process. Special attention was paid to the HLA loci in the NeXT Exome™ design including targeting alternative reference sequences to maximize coverage of alleles. Personalis' HLA typing tool has been integrated into the NeoantigenID analytics engine; enabling the highly-accurate *in silico* typing of all HLA Class I and Class II loci, which is critical for ensuring the precision of downstream peptide-MHC-binding predictions.

The HLA typing validation study was performed on a total of 15 proficiency testing samples with known, but blinded HLA genotype profiles. Ten of these samples were sourced from the American Society of Histocompatibility and Immunogenetics (ASHI) and five additional samples were obtained from the College of American Pathologists (CAP). Each of these samples had previously been independently genotyped via various orthogonal clinical tests, and these results were compared against Personalis' data for HLA Class I (A, B, C) and Class II alleles (DRB1, DPA1, DPB1,DQA1, DQB1, DRB3, DRB4, DRB5).

As demonstrated in **Table 1**, Personalis' HLA typing tool performed exceptionally well in accurately genotyping not only the HLA Class I loci, but also the more challenging HLA Class II loci.

Neoantigen Identification

Somatic mutations, including SNVs, indels, and fusions, each produce different protein products. While SNVs result in single amino acid changes, indels and fusions can create multiple frame-shift protein products through alternative splicing. As genes also have many transcripts, a single somatic mutation can result in numerous protein

Table 1: HLA Genotyping Validation Study

HLA Loci	# Alleles	# Correct Calls	HLA Genotyping Concordance
All Class I	**90**	**90**	**100.0%**
All Class II	**180**	**177**	**98.3%**
All Class I + II	**270**	**267**	**98.9%**

products. Collectively, accurate peptide usage relies on inclusion of frame-shift events, proper transcript selection, application of variant phasing, and consideration of variant expression.

NeoantigenID incorporates information from different variant types beyond SNVs, to also include indels and fusions in antigen considerations. This is critical as frame-shift indels could prove to be a rich source of immunogenic neoantigens as has been shown by Turajlic et al.[1]

Neoantigen Processing

Personalis' NeoantigenID analytics engine is used for peptide sequence determination (**Figure 1**). The peptide sequences listed in the report are generated from DNA sequences containing the somatic mutations detected in the tumor. The altered amino acid residue is positioned at the center and flanked by the remaining residues on each side to make up a full length 9-11mer peptide. Each peptide is then further characterized based on multiple parameters.

The peptide sequence processing workflow begins with the integration of accurate exome-based *in silico* HLA typing results for both MHC Class I and Class II loci (**Figure 2**) into the respective MHC binding prediction modules described below.

Class I MHC Binding Prediction

For Class I MHC binding prediction, NeoantigenID utilizes Personalis' proprietary algorithm, SHERPA, which incorporates multiple key parameters in defining potential antigenicity. SHERPA utilizes in-house generated immunopeptidomics data, publicly available & curated mono- and multi-allelic data, as well as binding affinity data as a training set (**Figure 3**). The publicly available, multi-allelic data from several tissue types were systematically reprocessed and deconvoluted to capture the diverse facets of antigen processing and presentation. This combined approach resulted in one of the largest training datasets consisting of 180 unique human alleles and >1.4 million positive peptides. The integration of different training data types resulted into decreased bias, increased generalizability, and improved performance of SHERPA.

Along with Class I neoantigen prediction from SHERPA, the NeoantigenID report also provides default output from NetMHCpan. Further, the characterization of Class II MHC binding prediction is provided by the defaut output from NetMHCIIpan.

Figure 2: HLA Class I and Class II Structures

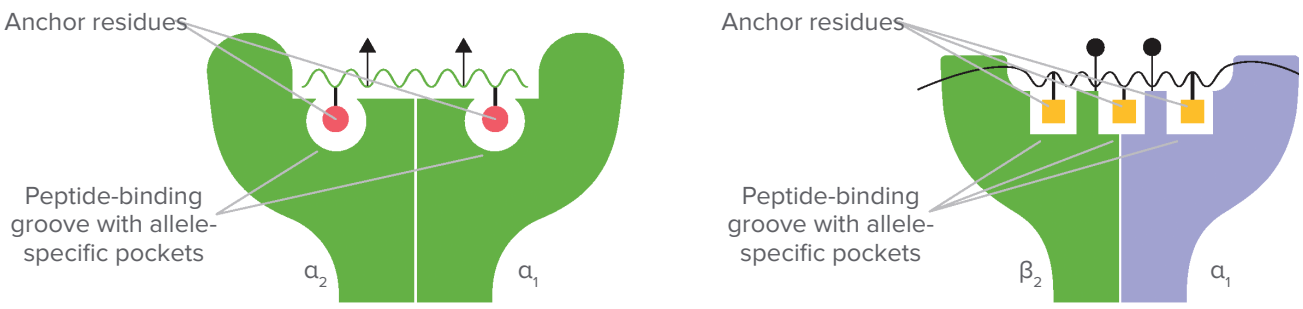
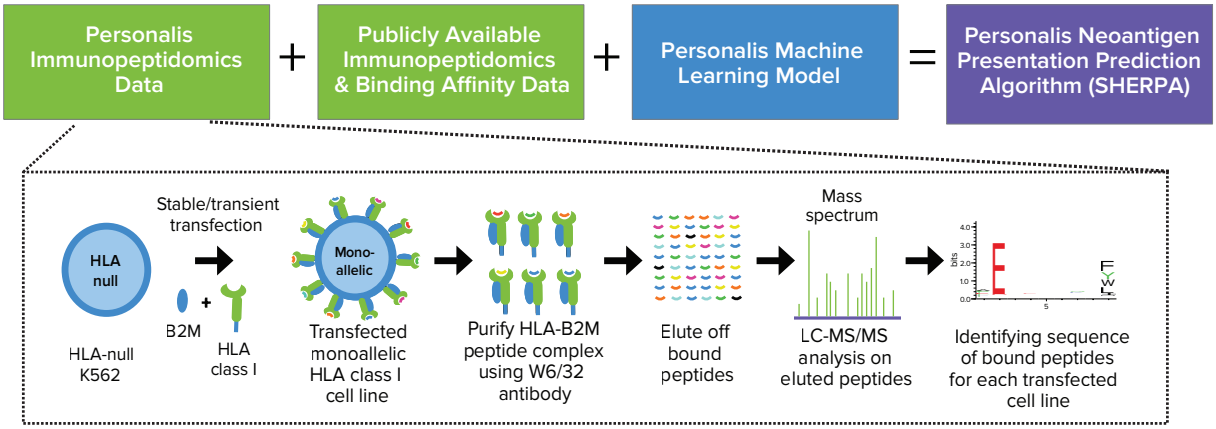


Figure 3: Overview of SHERPA Machine Learning Algorithm



Additional Neoantigen Discovery Secondary Metrics

Personalis also provides the following additional features as secondary metrics to enable further prioritization of the predicted putative neoantigen candidates.

Similarity-to-known Antigens Module

This module measures how similar a peptide is to a database of known antigens. The more negative the score, the more differentiated the peptide is compared to known antigens. Peptides that are similar to known antigens are more likely to be recognized by T-cells. We use the full epitope database from IEDB[2] excluding human epitopes, thus these are mainly epitopes from viruses and other pathogens. This is an internally developed tool at Personalis.

Similarity-to-self Module

This module produces a metric for each peptide that indicates how similar the peptide is to any portion of any protein that the subject's normal tissues produce. The more negative the score, the more differentiated the peptide is from "self" peptides. If this value is "0", this means that the peptide matches a "self" protein. This is also an internally developed tool at Personalis.

Immunogenicity

Immunogenicity is a measure of the likeliness that a given peptide-MHC complex (pMHC) will be recognized by T-cells. Critical factors regarding immunogenicity prediction are 1) position of a presented peptide (P4 through P6 for Class I), and 2) type of amino acid (aromatic or non-aromatic, acidic or basic, charged or non-charged).

This module was developed internally based on published literature[3] and implements an algorithm that scores the immunogenicity of a pMHC based on an amino acid enrichment analysis of a large set of immunogenic and non-immunogenic pMHCs. Increasingly positive scores indicate higher immunogenicity. This score is independent of the similarity modules.

Footnotes

1. Turajlic et al. Lancet Oncol. 2017 Aug; 18(8):1009-1021.
2. Immune Epitope Database and Analysis Resource http://www.iedb.org.
3. Callis et al. PLoS Comput Biol. 2013 Oct:9(10):e1003266.
4. Sahin et al. Nature. 2017 Jul 13;547(7662):222-226.
5. Havel et al. Nat Rev Cancer. 2019 March; 19(3):133-150.

Gene-and Variant-level Expression

Gene-level expression changes are undeniably helping to shape the understanding of patient response to checkpoint inhibitors like pembrolizumab. Likewise, RNA variant — and gene-level — expression can be a powerful tool to filter neoantigen candidates. As witnessed in the first-in-human neoantigen-based cancer vaccine clinical trials, RNA variants have been suggested as a criteria for helping to determine potential putative neoantigen candidates[4].

Comprehensive List of Putative Neoantigens

NeoantigenID provides all detected neoantigens and their comprehensive characterization based on Personalis' state-of-the-art prediction algorithm for MHC Class I binding and presentation.

Personalis Composite Biomarker

Recent studies[5] have shown that integration of key molecular features can increase the predictive utility of these biomarkers in determining immune checkpoint inhibitor efficacy.

Utilizing SHERPA for neoantigen prediction, our analytics also provide neoantigen burden, as well as Personalis Composite Neoantigen Presentation Score (NEOPS), which further incorporates HLA loss of heterozygosity (LOH) and somatic variants from HLA and antigen presentation machinery (APM) genes (**Table 2**).

Table 2: NeoantigenID Summary Report Example

Composite Biomarker

	Score
Neoantigen Presentation Score (NEOPS)	60.76

Neoantigen Burden

	Genes/ Alleles Impacted	Total number of peptides from small variants in DNA	Total number of peptides per Mb from small variants in DNA	Total number of peptides from small variants in DNA and fusions in RNA	Number of peptides per Mb from small variants in DNA and fusions in RNA
Observed Neoantigen Burden		2,493	71.51	2,494	71.54
Corrected Neoantigen Burden (considering HLA LOH)		2,493	71.51	2,494	71.54
Corrected Neoantigen Burden (considering HLA somatic small variants)	HLA-C*07:02	2,117	60.73	2,118	60.76
Corrected Neoantigen Burden (considering APM somatic small variants)		2,493	71.51	2,494	71.54

Number of Presented Peptides

	Predicted presented small variants in DNA and fusions in RNA (<0.5% rank)
Number of peptides presented by any one HLA type	2106
Number of peptides presented by any two HLA types	337
Number of peptides presented by any three or more HLA types	51

Tumor Mutation Burden

Summary Small Variants	SNVs	Indels	Total
Non-synonymous Somatic Variants	2,659	66	2,725
Non-synonymous Somatic Variants per MB	76.32	1.89	78.21

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

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

Personalis®

DOC-000570(B)