56 IMPROVED NEXT GENERATION SEQUENCING BASED CLASS I HLA TYPING THROUGH EXOME ENHANCEMENT

Lee McDaniel*, Rachel Pyke, Charles Abbott, Gabor Bartha, John West, Richard Chen, Sean Boyle. *Personalis, Inc, Darlington, MD, USA*

Background Precision immuno-oncology is increasingly relevant to cancer therapy given the ascendance of immunotherapy. While next-generation sequencing (NGS) based algorithms may elucidate immunotherapeutic response, many such algorithms require highly accurate Class I HLA typing. One major challenge of HLA type derivation resides in highly polymorphic HLA allelic diversity, which conventional exome sequencing technologies poorly capture. Further, accurate HLA typing requires definitive distinction between thousands of potential HLA alleles. These challenges may cause widely used NGS HLA typing tools, such as Polysolver and Optitype, to perform inaccurate HLA typing. Poor HLA coverage poses the risk of silently mistyping HLA alleles, yielding inaccurate downstream HLA loss of heterozygosity (LOH) detection and neoepitope predictions.

Methods We designed the ImmunoID NeXT Platform® to more comprehensively profile the HLA region. To evaluate the accuracy of conventional NGS-based Class I HLA typing, a widely used dbGaP project (phs000452, n=160) of melanoma NGS data was evaluated alongside a set of over 500 solid tumor cancer patient samples sequenced on the ImmunoID NeXT Platform. Read coverage was derived from both GRCh38 and HLA allele database alignments. To test whether Polysolver over represents specific HLA alleles under reduced read conditions, a Monte Carlo bootstrap approach predicted theoretical allele frequency ranges.

Results Below 20x read coverage, nearly 50% of Polysolver HLA calls (phs000452) are homozygous, representing a divergence from typical HLA homozygous rates of between 10–20%, with $p<10^{-15}$ (Fisher's Exact) compared to reference 1000 Genomes homozygous rates. Polysolver's homozygous, heterozygous, and no-calls demonstrated a statistically significant difference in coverage ($p<10^{-6}$, Kruskal-Wallis) across all Class I HLA genes per Polysolver and public exome data (phs000452). The Personalis ImmunoID NeXT™ cohort did not demonstrate such a trend despite a similar exome-wide sequencing depth. Further, sixteen rare HLA alleles were identified with sample frequencies greater than expected from the dbGaP data set, with no such alleles identified from the Personalis ImmunoID NeXT data set.

Conclusions HLA typing may silently fail in the context of reduced read coverage without HLA-specific platform augmentation. This silent failure can have large implications for accurate neoantigen prediction and HLA LOH detection, both of which are becoming increasingly important for immuno-oncology treatment modalities such as personalized cancer vaccines, adoptive cell therapies, and blockade therapy response biomarkers. Studies utilizing neoepitope and HLA LOH prediction require careful validation for HLA calls, including assessments of coverage and homozygous rates, and may benefit from increased HLA locus coverage.

http://dx.doi.org/10.1136/jitc-2020-SITC2020.0056

57 PRECISION NEOANTIGEN DISCOVERY USING NOVEL ALGORITHMS AND EXPANDED HLA-LIGANDOME DATASETS

Dattatreya Mellacheruvu*, Rachel Pyke, Charles Abbott, Nick Phillips, Sejal Desai, Rena McClory, John West, Richard Chen, Sean Boyle. *Personalis Inc., Menlo Park, CA, USA*

Background Accurately identified neoantigens can be effective therapeutic agents in both adjuvant and neoadjuvant settings. A key challenge for neoantigen discovery has been the availability of accurate prediction models for MHC peptide presentation. We have shown previously that our proprietary model based on (i) large-scale, in-house mono-allelic data, (ii) custom features that model antigen processing, and (iii) advanced machine learning algorithms has strong performance. We have extended upon our work by systematically integrating large quantities of high-quality, publicly available data, implementing new modelling algorithms, and rigorously testing our models. These extensions lead to substantial improvements in performance and generalizability. Our algorithm, named Systematic HLA Epitope Ranking Pan Algorithm (SHERPA™), is integrated into the ImmunoID NeXT Platform®, our immunogenomics and transcriptomics platform specifically designed to enable the development of immunotherapies.

Methods In-house immunopeptidomic data was generated using stably transfected HLA-null K562 cells lines that express a single HLA allele of interest, followed by immunoprecipitation using W6/32 antibody and LC-MS/MS. Public immunopeptidomics data was downloaded from repositories such as MassIVE and processed uniformly using in-house pipelines to generate peptide lists filtered at 1% false discovery rate. Other metrics (features) were either extracted from source data or generated internally by re-processing samples utilizing the ImmunoID NeXT Platform.

Results We have generated large-scale and high-quality immunopeptidomics data by using approximately 60 mono-allelic cell lines that unambiguously assign peptides to their presenting alleles to create our primary models. Briefly, our primary 'binding' algorithm models MHC-peptide binding using peptide and binding pockets while our primary 'presentation' model uses additional features to model antigen processing and presentation. Both primary models have significantly higher precision across all recall values in multiple test data sets, including mono-allelic cell lines and multi-allelic tissue samples. To further improve the performance of our model, we expanded the diversity of our training set using high-quality, publicly available mono-allelic immunopeptidomics data. Furthermore, multi-allelic data was integrated by resolving peptide-to-allele mappings using our primary models. We then trained a new model using the expanded training data and a new composite machine learning architecture. The resulting secondary model further improves performance and generalizability across several tissue samples.

Conclusions Improving technologies for neoantigen discovery is critical for many therapeutic applications, including personalized neoantigen vaccines, and neoantigen-based biomarkers for immunotherapies. Our new and improved algorithm (SHERPA) has significantly higher performance compared to a state-of-the-art public algorithm and furthers this objective.

http://dx.doi.org/10.1136/jitc-2020-SITC2020.0057