Somatic variant detection using long read RNA sequencing reveals HLA-restricted targets for immunotherapy

Felipe Batalini[1], Skylar Henry[1,2], Yining Zhang[2], Rachel Foster[3], Jessica Gayton[1], Anna Hostal[4], Eric Wilson[5], Abhishek Singharoy[2], Michael Gustafson[1], Barbara Pockaj[1], Jin Park[2], Karen S. Anderson[1,2,3]

[1]Mayo Clinic Comprehensive Cancer Center, Phoenix-AZ, [2]School of Life Sciences, Arizona State University, [3]Phoenix Children's Hospital, [4]Creighton School of Medicine, Phoenix, [5]Icahn School of Medicine at Mount Sinai, New York, [6]Virginia G. Piper Center for Personalized Diagnostics, Arizona State University

ABSTRACT

INTRODUCTION Accurate detection and prioritization of tumor-specific somatic variants is foundational to the development of personalized immunotherapies. Traditional pipelines for neoepitope identification rely heavily on short-read next-generation sequencing (NGS) and DNA-based variant calling, followed by RNA expression filtering and HLA-binding prediction. However, this approach is limited by time-intensive workflows, potential false positives from DNA-only calls, and poor capture of complex isoforms or RNA-level editing. Moreover, the accuracy of neoepitope prediction is constrained by conventional sequence-based HLA binding algorithms that do not account for structural and biophysical features of peptide-MHC interactions, particularly across underrepresented HLA alleles.

METHODS To overcome these limitations, we developed a streamlined, sequencing-to-product framework that leverages long-read RNA sequencing for direct detection of expressed somatic variants, including fusions, indels, and single nucleotide variants. Using Oxford Nanopore Technologies' cDNA-PCR approach, we enable comprehensive, full-length transcript characterization from both fresh and frozen breast tumor samples. Our bioinformatic pipeline, NeoTarget, integrates alignment, variant calling, and neoepitope inference from tumor RNA in a single, modular workflow. A core innovation of this pipeline is HLA-Inception, a deep convolutional neural network trained to predict peptide-MHC class I binding using features derived from molecular dynamics–based modeling of electrostatic potential at the peptide-HLA interface, enhancing generalizability across diverse HLA subtypes.

RESULTS Expanding on previous results, the current NeoTarget pipeline identifies, evaluates, and prioritizes single nucleotide variants, indels, and fusions. To validate this system, we used a set of 9 human breast cancer tumors benchmarked with short read sequencing. Identified neoantigens included most variants shown by clinical short read sequencing and novel structural variants (i.e., fusions) only revealed with long read sequencing with high confidence.

CONCLUSION Here, we describe the design and performance of the NeoTarget pipeline and show proof-of-concept feasibility of applying this fast and cost-effective workflow for HLA-restricted epitope discovery in breast tumors using long-read RNA sequencing. Overall, these methods can be applied to rapidly identify and prioritize neoantigens for personalized neoantigen-based immune therapies.

PROJECT OVERVIEW

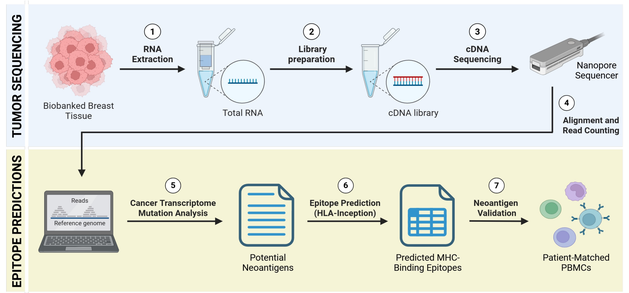
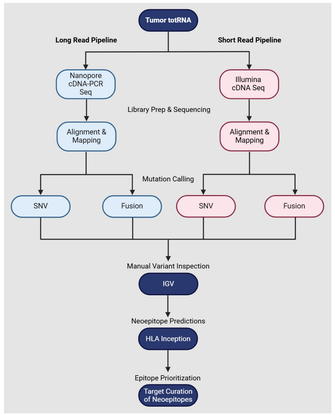


Figure 1. HLA-restricted epitope discovery workflow. Frozen tumor tissue from the Mayo Clinic biobank or fresh from surgery is disrupted and homogenized prior to RNA extraction. Total tumor RNA is then used to prepare a cDNA library for sequencing with a nanopore sequencer. Mutations in the resultant reads are identified via NeoTarget, generating a list of candidate neoepitopes for prediction by HLA-Inception. Predicted neoepitopes will then be validated in vitro with patient-matched PBMCs. Validated neoepitopes may be prioritized for personalized neoantigen-based immunotherapies.

NEOTARGET PIPELINE



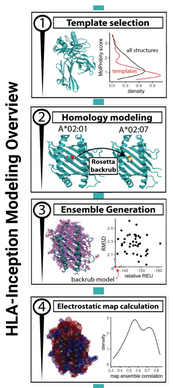


Figure 2. NeoTarget pipeline. A) Computational pipeline used to evaluate the quality of long read sequencing. Long-read sequences were mapped to the reference genome using minimap2, followed by parallel small variant (SNVs and indels) and fusion detection using ClairS-TO and CTAT-LR-Fusion, respectively. B) HLA-Inception modeling overview. MHC-I structural templates are selected based on the Molprobity score or modeled using the best aligning template structure. Then, an ensemble of binding pockets is generated and the electrostatic potential of the MHC-I binding pocket is calculated for each ensemble member.[1-2]

NANOPORE-SEQ CALLS MORE FUSIONS

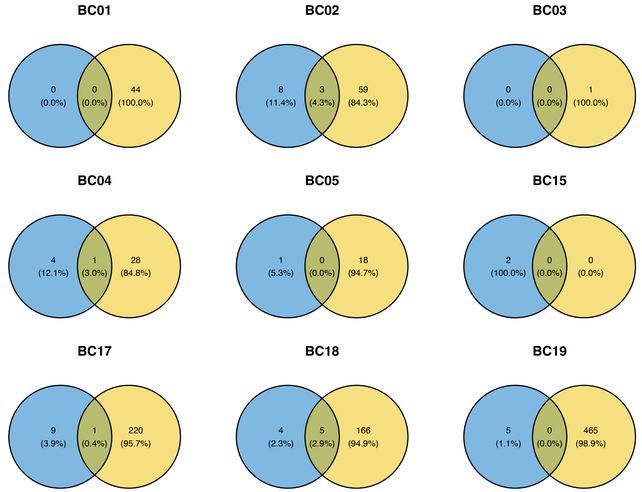


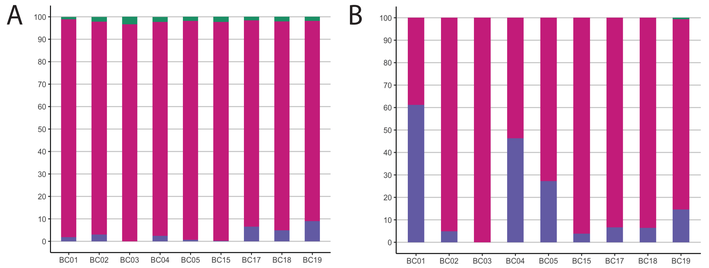
Figure 3. Comparison of fusion transcripts detected by Illumina and nanopore sequencing. Venn diagrams depict the number of unique fusion calls by short-read Illumina (blue) and long-read Nanopore (yellow) sequencing after filtering. Transcripts detected by both methods are represented by the overlapping region. 9 breast tumor samples are show. STAR-Fusion was used to call fusions from mapped short-read sequencing and CTAT-LR-Fusion was used to call fusions from mapped long-read sequences.

BC17: PREDICTED HLA-MATCHED NEOEPITOPES

Protein	AIFM2	CCNG1	MCM4	COL3A1	ATG4B
Variant	p. Met135Thr	p. Ser206Pro	p. Leu369Met	p. His1320Gln	p. Leu354Gln
Peptide	AYEDTVRQV	VLALPIIAL	RFDLIFLML	VLDVQLAFL	LVEQQPSHL

Figure 5. Examples of HLA-matched neoepitope predictions. SNV-derived neoepitope predictions from long read sequencing of sample BC17 were analyzed for matches to patient's HLA type (A*02:01:01, A*31:01:02, B*07:02:01, B*41:01:01, C*07:01:02, C*07:02:01).

VARIANT CONTRIBUTIONS TO INCEPTION



C

	BC01	BC02	BC03	BC04	BC05	BC15	BC17	BC18	BC19
Source	biobank	biobank	biobank	biobank	biobank	fresh	fresh	fresh	fresh
RIN	5.6	4.5	5.1	5	7.3	9.2	8.8	8.5	7.1
SNV	26	39	2	22	8	***	69	44	122
Indel	0	0	0	0	0	***	0	0	1
Fusion	41	2	0	19	3	***	5	3	21

Figure 4. Representation of variant sources in HLA-Inception workflow. A) Relative contribution of variant type (SNV, Indel, or Fusion) to HLA-Inception input. B) Relative proportion of predicted HLA-matched neoepitopes derived from SNV, Indel, or Fusion. Mutations were called from cDNA-PCR nanopore sequencing using CTAT-LR-Fusion (fusions) and a customized workflow for ClairS-TO (SNVs and indels). Epitopes were predicted from HLA-Inception using a 99.5% binding score threshold.

FUTURE PHASE I TRIAL DESIGN

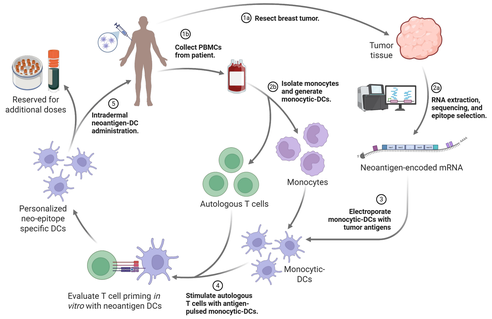


Figure 5. Schematic of proposed future phase I clinical trial for mRNA neoepitope-DCs. The NeoTarget pipeline will be used to identify candidate epitopes for multiepitope mRNA. After validation, mRNA will be electroporated into patient monocytic-DCs. Personalized, neo-epitope specific DCs will be delivered to the patient.

CONCLUSIONS

- The NeoTarget pipeline effectively and rapidly identifies neoantigens for personalized identification of neoantigens that may be used for personalized, neoantigen-based immunotherapies.
- Long-read RNA sequencing identifies relevant variants also noted in short-read sequencing as well as novel high confidence fusions not identified with short-read sequencing.
- Neoantigens derived from fusions have increased representation in HLA-Inception output compared to input in analyzed breast cancer samples.

ACKNOWLEDGEMENTS

Supported in part by Mayo Clinic Breast Cancer SPORE grant P50 CA116201 from the National Institutes of Health (FB, MG, KA). We also acknowledge support from the Breast Cancer Research Foundation (FB), Brendan Mann Foundation (FB), the Desert Mountain CARE Program (FB), and the Arizona State University-Mayo Clinic Alliance for Health Care SEED program (MG, KA).

REFERENCES

1. Kennedy, R. et al. J. Immunol. Baltim. Md 1950 2006, 177 (5), 2862–2872.
2. Kreiter, et al. Nature 2015, 520 (7549), 692–696.
3. Sahin, U. et al. Nature 2017, 547 (7662), 222–226.
4. Ott, P. A. et al. Nature 2017, 547 (7662), 217–221.
5. Keskin, D. B. et al. Nature 2019, 565 (7738), 234–239.
6. Taylor, B. C. et al. Cancer Discov. 2024, 14 (2), 290–307.
7. Hao, X. et al. Nat. Cell Biol. 2024, 26 (11), 1971–1983.
8. Hao, X. et al. Cell Stem Cell 2023, 30 (5), 648-664.e8.
9. Gonzalez-Ericsson, P. I. et al. Cancer Immunol. Res. 2025, 13 (4), 602–609.
10. Laborde, R. R. et al. Front. Immunol. 2014, 5.
11. Wilson, E. A. et al. Cell Rep. Med. 2021, 2 (3), 100221.
12. Wilson, E. et al. Cell Syst. 2024, 15 (4), 362-373.e7.

© Poster Template by Genigraphics® 1.800.790.4001 www.genigraphics.com