Evaluating APOBEC-Driven Mutations as a Source of Neoantigens for Cancer Immunotherapy

Kyle Nguyen[1], Darren Taylor[2], Mark Yarmarkovich[3]
University of Texas at Austin[1] | NYU Grossman School of Medicine[1,2,3]

Yarmarkovich Lab
Vilcek Institute of Graduate Biomedical Sciences

ABSTRACT

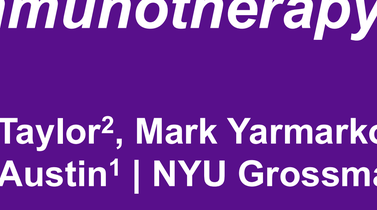
Neoantigens—peptides uniquely presented on tumor cells—are emerging as targets for next-generation immunotherapies, including T-cell receptor (TCR) and Peptide-Centric CAR-T cell therapies. APOBEC3A, a cytidine deaminase frequently upregulated in cancer, introduces C>T mutations that can give rise to novel neoantigens. We developed a computational pipeline to simulate APOBEC3A-like mutations across 35,608 coding transcripts and identify high-confidence 9-mer peptides suitable for immune targeting.

From 69,247 mutation sites, we generated 58,591 unique mutant peptides. We used 77 breast cancer immunopeptidomics datasets to detect their presence (TesorAI) narrowing this to 2,530; DepMap CERES scores prioritized 322 from essential genes. NetMHCpan 4.1 predicted strong HLA binding for 69 peptides, and NetMHCStabPan further refined this to 13 stable candidates.

We also began extending this pipeline to transposable elements (TEs), which are frequent APOBEC substrates and may harbor overlooked neoantigen sources. This work provides a framework for identifying mutation-derived peptides with strong MHC presentation potential—critical for expanding the scope of CAR-T cell therapies beyond shared antigens.

BACKGROUND / INTRODUCTION

Figure 1.(Right) CAR-T cells recognize tumor-specific peptides presented on MHC complexes. Tumor cells present neoantigen peptides on their surface via MHC molecules. CAR-T or TCR-engineered T cells can recognize these peptide–MHC complexes, triggering targeted immune responses. Identifying tumor-specific peptides is essential for designing effective and selective immunotherapies.



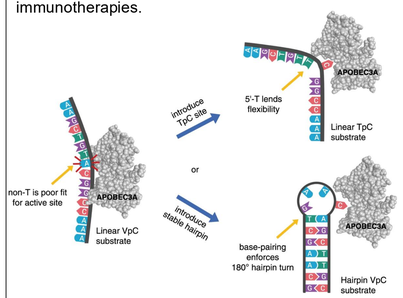


Figure 2. (Left) APOBEC3A activity depends on DNA sequence and structure. APOBEC3A preferentially targets single-stranded DNA regions enriched for TpC motifs. Structural features like hairpins or loops increase accessibility and stabilize the enzyme–substrate interaction, promoting localized C-to-T deamination at hotspot motifs.

[1]Caval, V., Suspène, R., Shapira, M. et al. Structural and sequence determinants of APOBEC3A substrate specificity. *Nat Chem Biol* 18, 1420–1428 (2022).

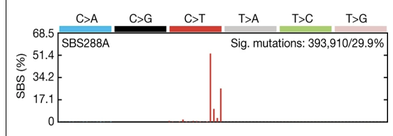


Figure 3. APOBEC3 activity produces C>T mutations in cancer, seen in signature SBS288A.

[2]Petljak, M., Li, Y., Casswell, G. et al. Characterizing mutational signatures in human cancer cell lines reveals pervasive APOBEC mutagenesis. *Nature* 594, 438–443 (2021).

Funding: NYU Grossman School of Medicine Summer Undergraduate Research Program

PIPELINE

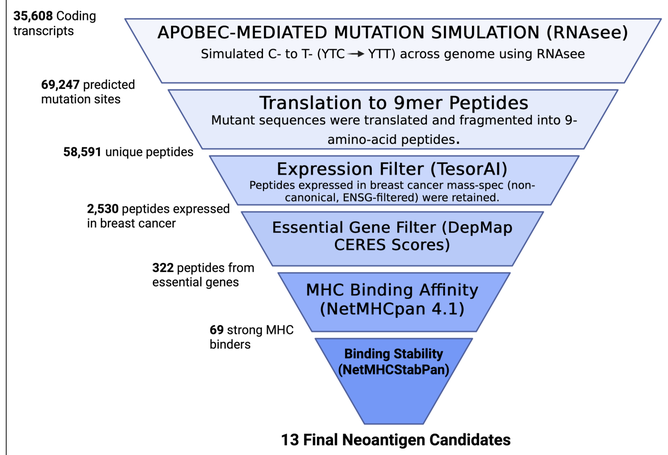


Figure 6. Pipeline for identifying neoantigens candidates from APOBEC-simulated mutations in coding genes.

RESULTS

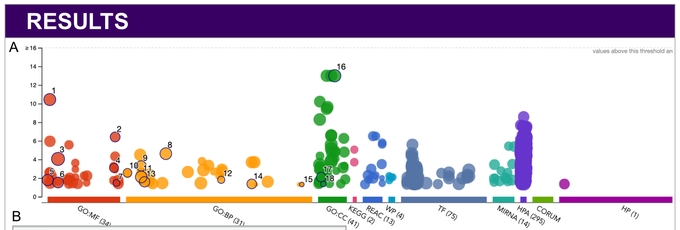


Figure 7. GO enrichment analysis of essential neoantigen source genes.
(A) Manhattan plot of enriched Gene Ontology (GO) terms from gProfiler analysis using the 322 essential neoantigen genes. Dots represent GO terms across categories including molecular function (MF), biological process (BP), and cellular component (CC), with y-axis indicating $-\log_{10}$ adjusted *p*-value.
(B) Top 18 significantly enriched GO terms, including nucleic acid binding, chromatin organization, and DNA repair processes, indicating that essential neoantigen source genes are strongly linked to genome regulation and stability.

RESULTS

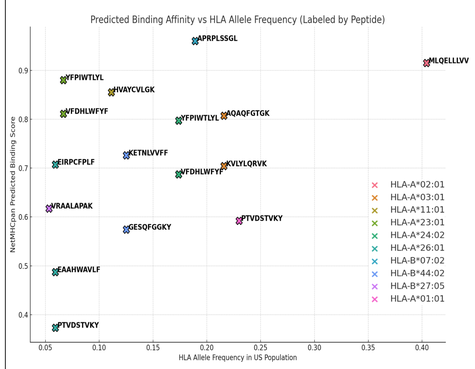


Figure 7. Predicted Neoantigen Binding vs HLA Allele Frequency
Each point represents a neoantigen candidate, labeled by peptide, plotted by predicted HLA binding affinity (NetMHCpan) and U.S. HLA allele frequency. Peptides like *MLQELLLVV* (HLA-A*02:01) combine strong binding with broad population coverage (~40%).

Gene	Functional Summary
SRCAP	Chromatin remodeler, regulates gene expression
RANBP2	Regulates nuclear transport and protein modification
EIF4A1	RNA helicase in translation initiation
RPN1	Involved in N-linked glycosylation
TRRAP	Transcription coactivator, histone acetylation
UBR4	Protein degradation, cell cycle regulation
TCOF1	rRNA processing and craniofacial development
CHD4	Epigenetic gene silencing, chromatin remodeling
SBNO1	Possibly involved in transcriptional regulation

Figure 8. Functional network and pathway roles of neoantigen-source genes. Source gene functions from Cytoscape analysis, highlighting roles in chromatin remodeling (e.g., CHD4, TRRAP, SRCAP) and other cellular processes (e.g., UBR4, EIF4A1), suggesting both shared and distinct biological origins.

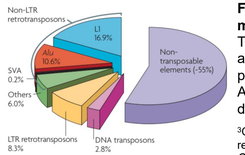


Figure 9. Transposable elements (TEs) make up ~45% of the human genome.
TEs like LINEs, SINEs, and ERVs are abundant and functionally diverse. Their prevalence presents an opportunity to expand APOBEC-driven neoantigen discovery to TE-derived peptides.

[3]Cordaux, R. & Batzer, M.A. The impact of retrotransposons on human genome evolution. *Nat Rev Genet.* 10, 691–703 (2009).

CONCLUSIONS

This study presents a computational pipeline for discovering neoantigens arising from APOBEC3A-mediated C>T mutations. By simulating mutations across 35,608 coding transcripts, we identified 13 strong and stable 9-mer peptides predicted to bind common HLA alleles, including HLA-A*02:01, which is present in over 40% of the U.S. population. These peptides originate from essential genes, supporting their potential as immunotherapy targets.

We also initiated a parallel analysis of transposable elements (TEs), which are frequent APOBEC targets. After identifying TE subfamilies with conserved YTC motifs, the next steps will mirror the gene workflow: simulate mutations, generate peptides, and apply expression, essentiality, and MHC binding filters to identify TE-derived neoantigens.

The next critical focus is validation. While we excluded canonical gene products, some simulated peptides may still arise in normal tissues through unrelated background mutations. To address this, we will screen candidates against healthy immunopeptidomics datasets to eliminate any peptides already present in the normal immunopeptidome. Longer term, integrating whole genome sequencing, RNA-seq and immunopeptidomics data from matched tumor and normal samples will allow us to confirm mutation presence, expression, and antigen presentation—ensuring that selected neoantigens are truly tumor-specific.

ACKNOWLEDGMENTS

I would like to sincerely thank my mentor, Dr. Darren Taylor, for his guidance, support, and friendship—thank you for making this an incredible summer. I'm also grateful to Dr. Mark Yarmarkovich, Minkyung Kim, and the entire lab for welcoming me and making me feel part of the team. Special thanks to Fanta Sissoko, Susanne Tranguch, Amy Wang, Elaine Camacho-Hernandez, and Teniola Idowu for creating an environment that allowed me to learn, grow, and genuinely enjoy every moment.
To my fellow SURPies—thank you for being some of the most thoughtful, funny, and inspiring people I've met. I'm so grateful for the memories we made, and I'll carry this summer with me long after it's over.