

SNAF: Accurate and compatible computational framework for identifying splicing derived neoantigens

Guangyuan Li, BS^{1,2}, Anukana Bhattacharjee PhD², Nathan Salomonis, PhD^{1,2}

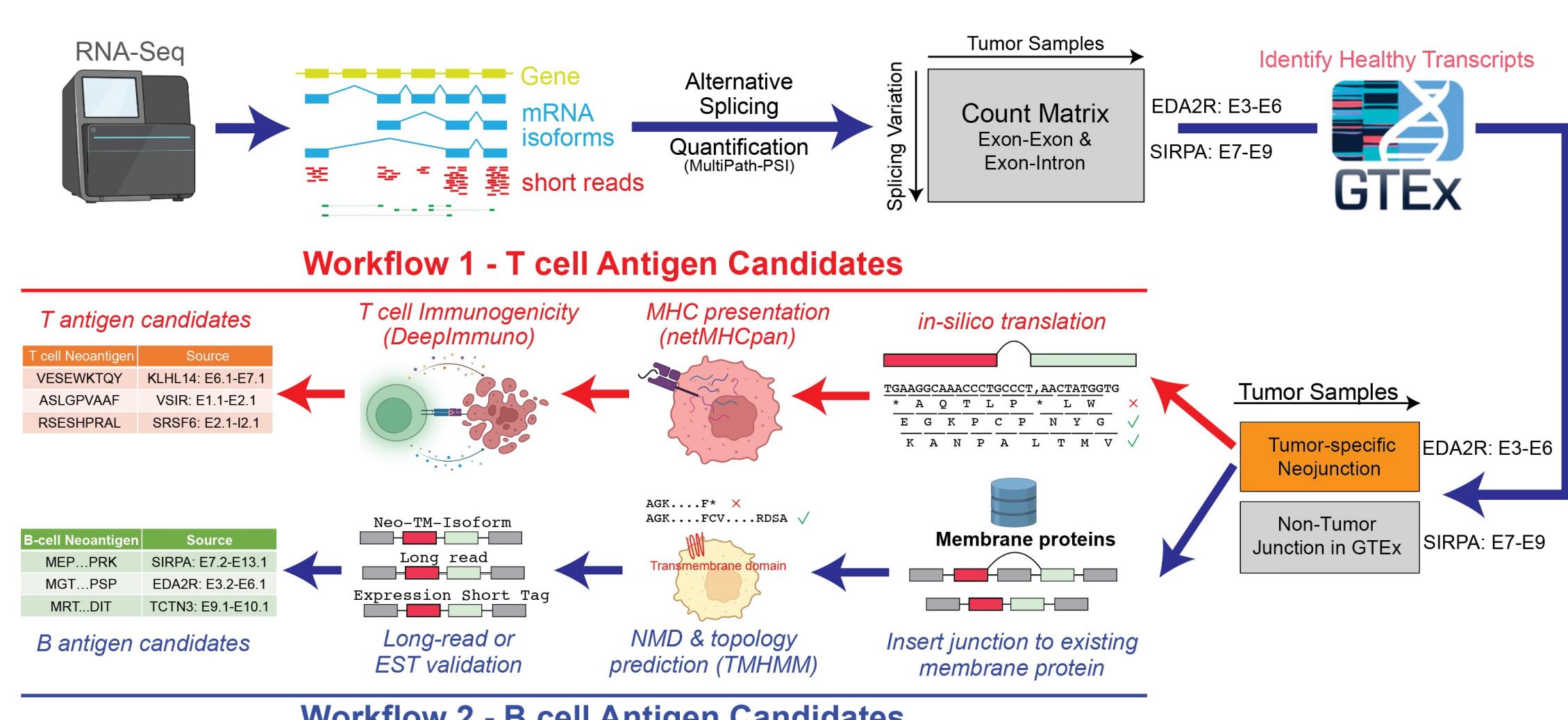
¹Department of Biomedical Informatics, University of Cincinnati College of Medicine, Cincinnati, OH, USA and ²Division of Biomedical Informatics, Cincinnati Children's Hospital Medical Center, Cincinnati, OH.



Motivation

- Immune Checkpoint Blockade (ICB) a crucial strategy to combat cancer, however, only a subset of patients benefit.
- Tumor Specific Antigens (Neoantigen) can train the immune system to recognize and kill cancer cells as cancer vaccines or in combination with ICB.
- While mutation-derived neoantigens have been used clinically with ICB, RNA-based variants (tumor-specific splice junctions) have not yet been clinically exploited.
- Splicing-neoantigen are more frequently detected than mutations in specific cancers, such as AML and are potentially associated with broad splicing disruptions.
- An alternative class of neoantigens are those which result in cancer-specific transmembrane proteins, that could be targets for new designed monoclonal antibodies (mAbs).
- No existing computational tools exist to solve the above challenges in a rigorous, precise and largely automated fashion from conventional genomics data.

Splicing Neo-Antigen Finder (SNAF)



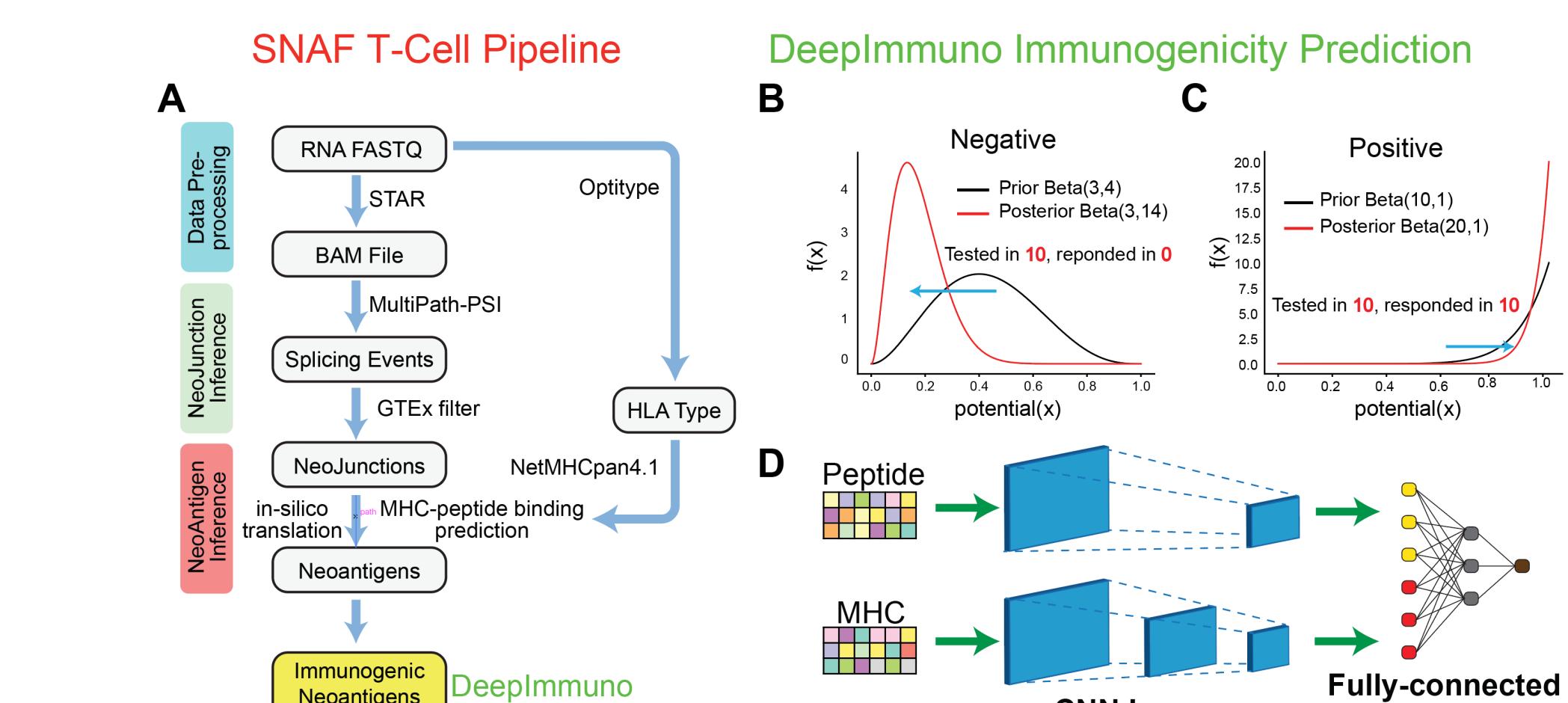
- SNAF-T identifies tumor specific short peptides bound with HLA that are predicted to elicit T cell responses.
- SNAF-B identifies tumor specific transmembrane protein isoforms with altered extracellular epitopes for detection.

	SNAF	ASNEO	NeoSplice	IRIS
Automated Program Features	in-silico translation Binding predictions Tumor-specific expression Immunogenicity predictions Tumor specificity score Advanced Visualization Integrated survival analysis Stand-alone python module Interactive web app Parallization Interface to Proteomics analysis Confirmation from long-read seq			
Possible feature (unpublished)				

- Deep-Learning predictions with improved accuracy over alternatives
- Diverse types of neoantigens and user-friendly and interactive

Deep Learning Model to Predict Immunogenicity

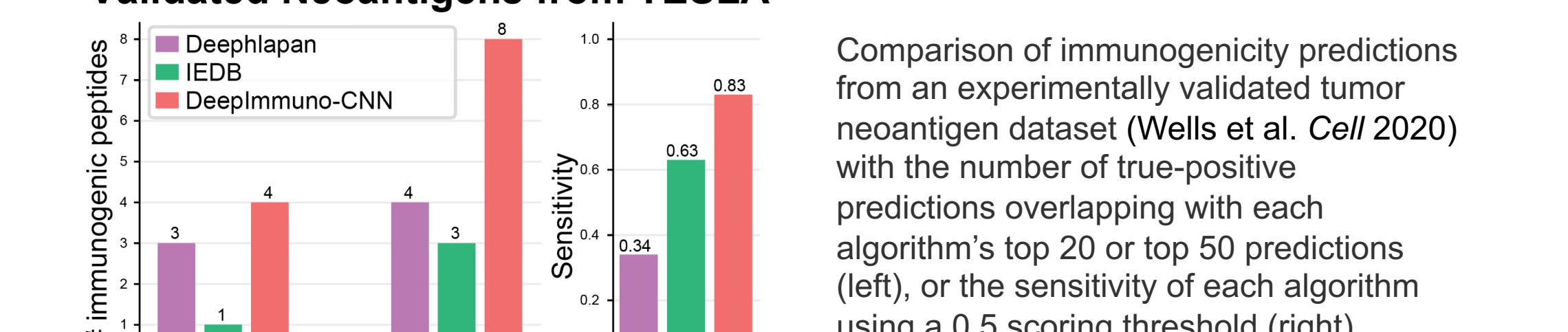
SNAF-T cell prediction pipeline: The accurate identification of splicing-neoantigens requires that predicted peptides are: 1) translated, 2) tumor-specific, 3) presented by HLA (patient specific) and are 4) immunogenic by patient T-cell receptors. SNAF leverages a deep learning approach called DeepImmuno (Li et al. Briefings in Bioinformatics, 2021), which uses a beta-binomial distribution approach to derive peptide immunogenic potential from sequence alone. Hence, we integrated DeepImmuno into SNAF.



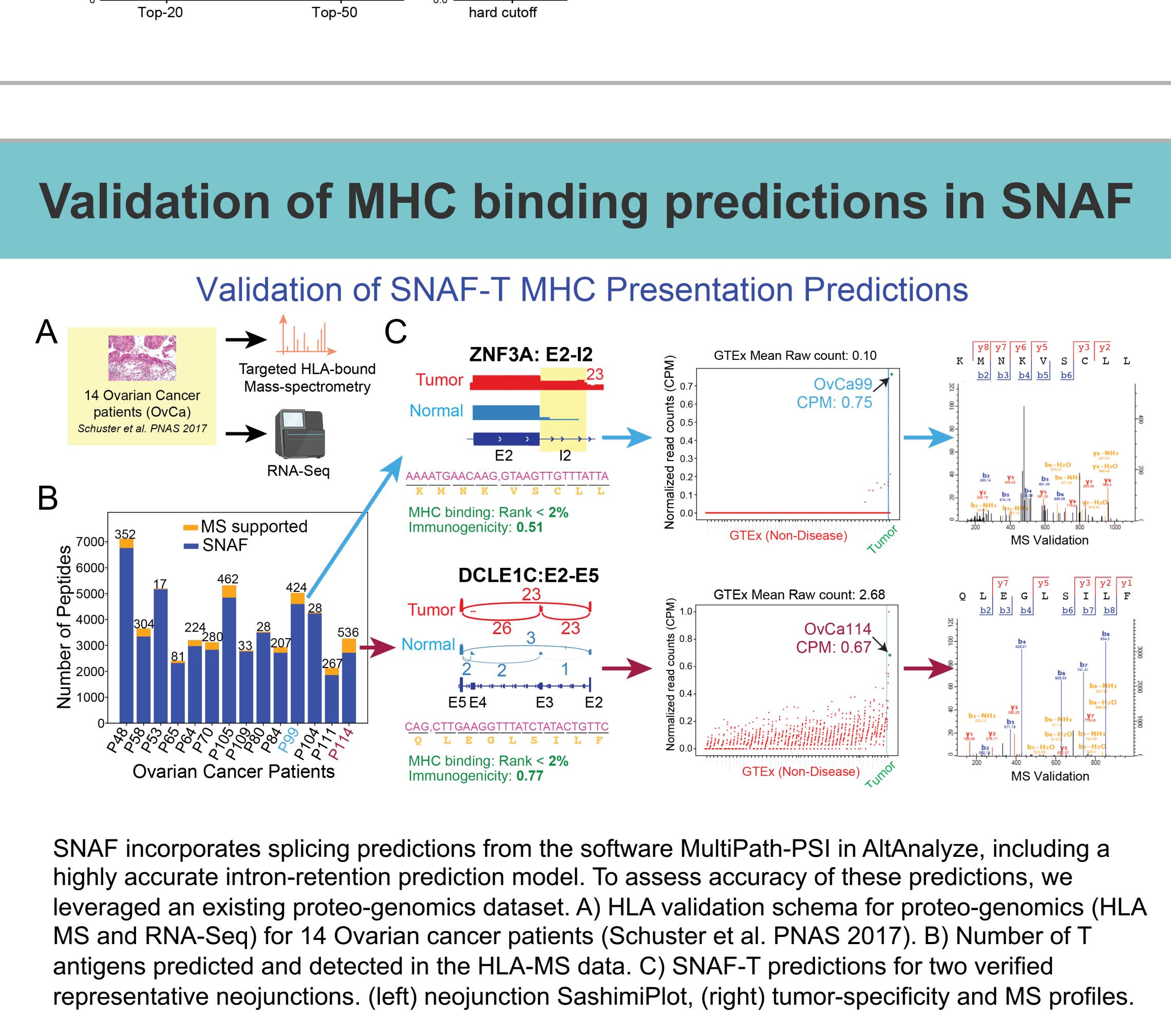
A) The SNAF T-cell prediction workflow. B) Immunogenic potential is computed by sampling from a posterior beta distribution of well-defined true-positive and true-negative immunogenic antigens to derive a continuous immunogenic score. The distribution is derived using a subset of T-cell immunogenic assay from the Immune Epitope Database (IEDB) and a prior beta distribution of either (A) negative or (B) positive assay results. (C) The DeepImmuno-CNN architecture to predict immunogenicity for each peptide-MHC complex. Each peptide-MHC pair is subjected to two consecutive convolutional layers, followed by two fully connected dense layers to output a predictive value for each pair.

Validation of neoantigen immunogenicity predictions: To verify DeepImmuno predictions, we evaluated leveraging the TESLA project collection of 608 tested predicted neoantigens. Our approach found an impressive 29 out of 35 (83%) immunogenic neoantigens, relative to IEDB which found 63%, and DeepHLApan which only found (34%).

Validated Neoantigens from TESLA

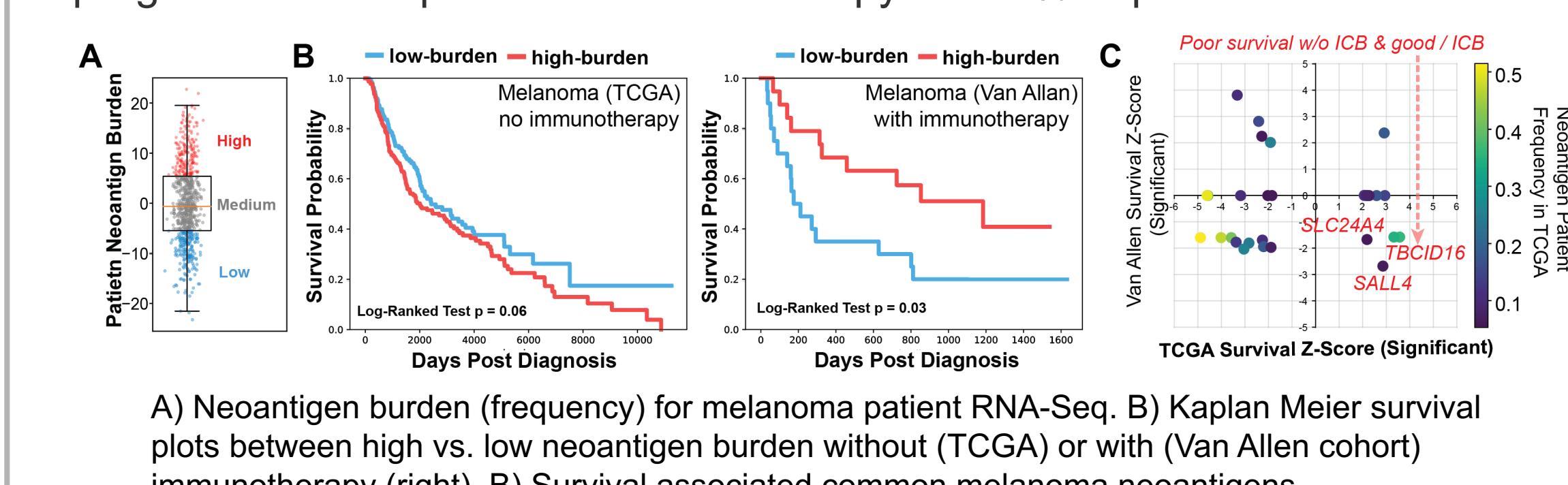


Number of immunogenic peptides: Top-20 vs Top-50. Sensitivity: hard cutoff vs soft cutoff.



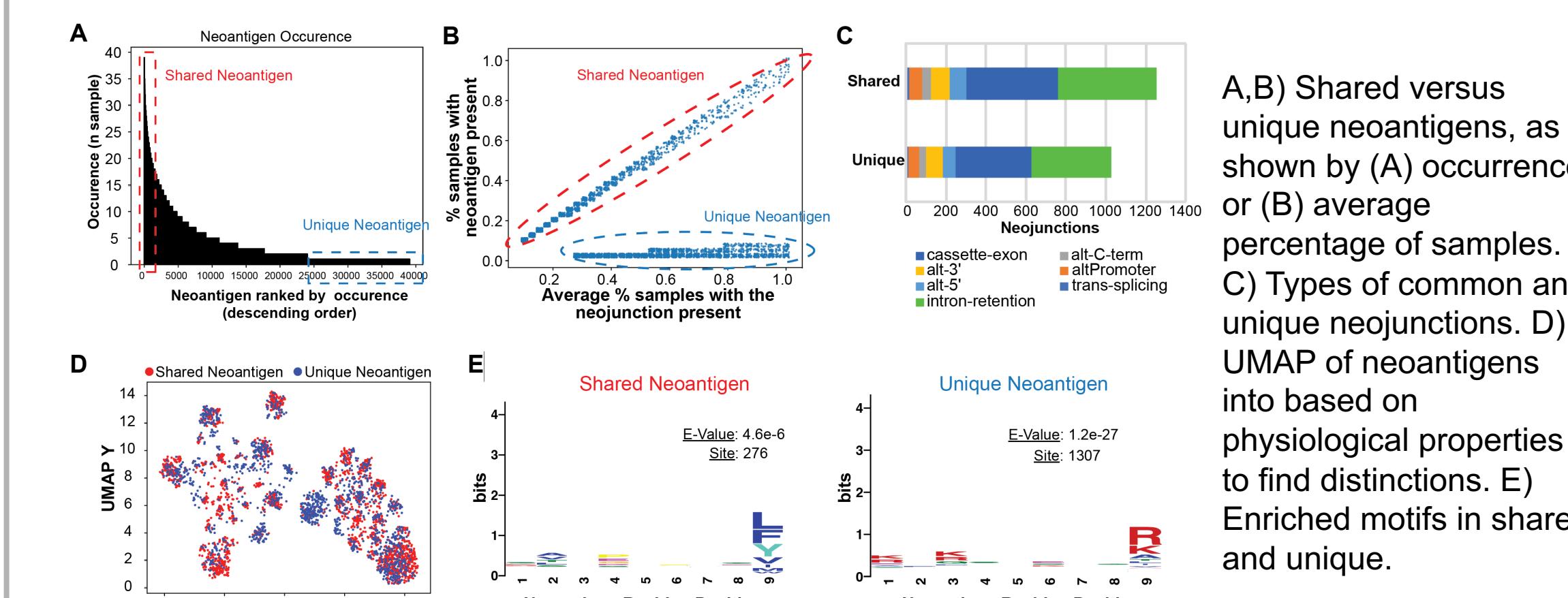
Melanoma Splice-Neoantigens Predict Response to Immunotherapy

Splicing neoantigens predict response to immunotherapy: Applied to existing patient bulk RNA-Seq with and without immunotherapy, we find patients with the most splicing-neoantigens have poor survival in the absence of immunotherapy, while patients with high burden have far improved survival, suggesting that splicing neoantigen burden is an indicator of patient response to ICB. Neoantigens can be further distinguished that predict poor or good prognosis and response to immunotherapy in >25% of patients.



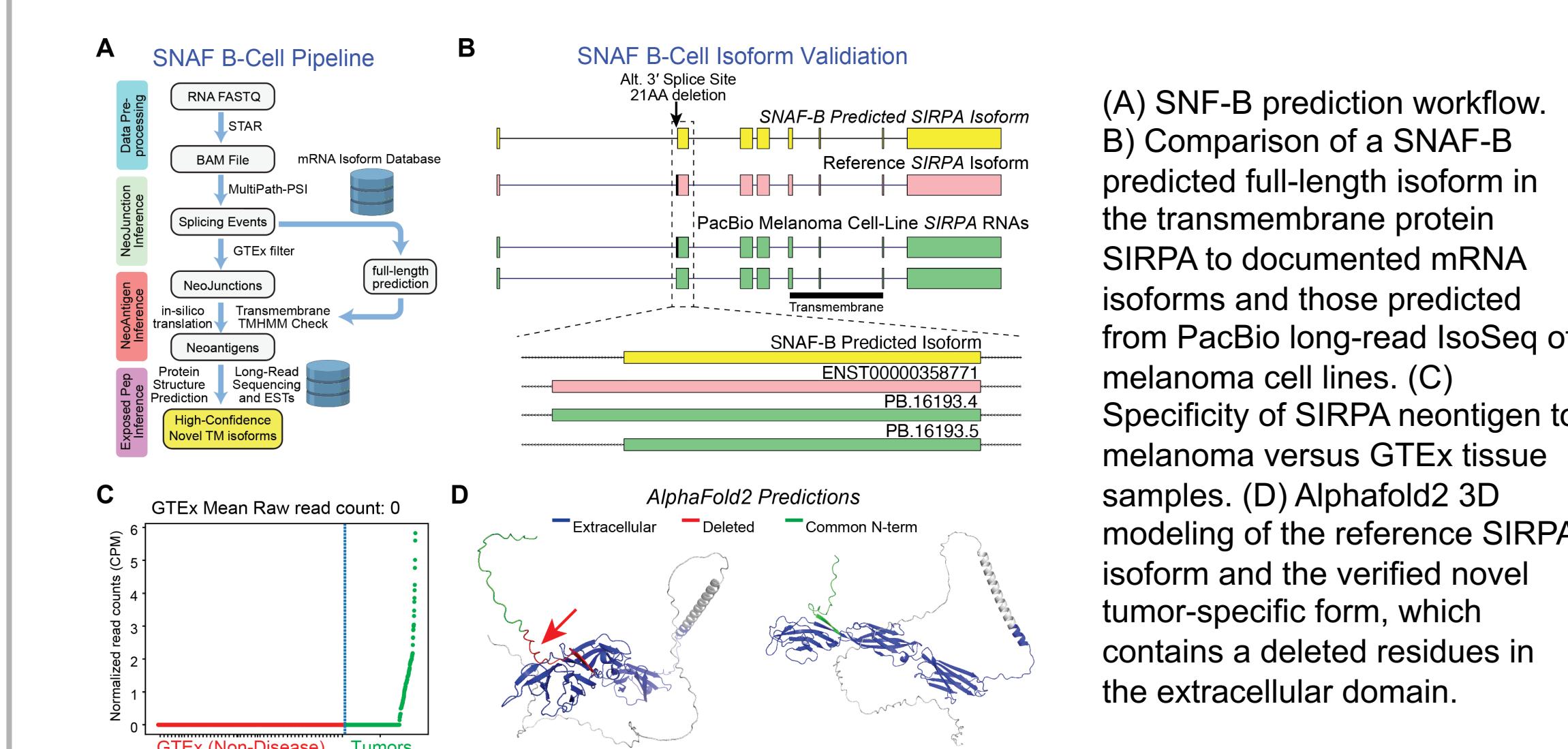
Shared Melanoma Splice-Neoantigens Demonstrate Selective Amino Acid Preferences

Shared splicing neoantigens represent common targets for therapy: Separation of SNAF-T predicted neoantigens into those shared among patients or unique finds that many are common, including those present in >80% of patients. Such neoantigens have stereotypical amino acid sequence preferences, suggesting a common mechanism of regulation/binding.

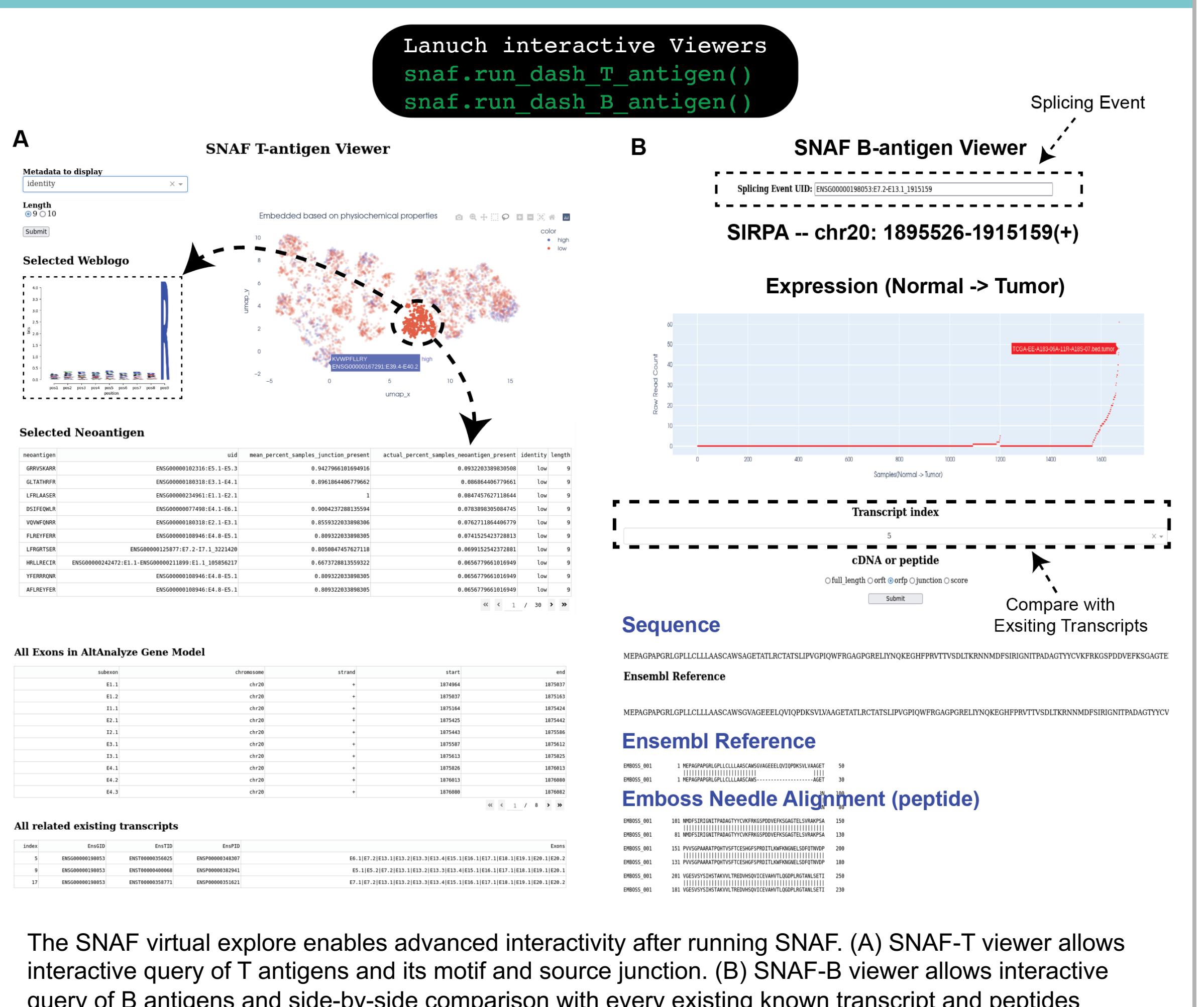


SNAF-B cell Accurately Predicts Full-Length RNAs and Stable-Proteoforms

SNAF-B finds transmembrane proteins. We assessed SNAF-B's ability to predict full-length stable proteins that could serve for mAbs.



Interactive Web Explorer



Conclusions and Future Work

- SNAF-T predictions are highly accurate and identify reproducible results between cancer cohorts (melanoma).
- Shared splicing-neoantigens represent novel prognostics for immunotherapy response and represent broad targets for cancer therapy.
- Shared neoantigens have sequence preferences that suggest a common mechanism of regulation or HLA/TCR binding.
- SNAF-B predictions can be verified from existing (EST) or new (full-length isoform) sequencing data that result in high confidence protein structure (AlphaFold2, ModFold8).
- SNAF is fast, easily installed, works with different genome versions and has interactivity to explore novel amino-acid bias and the diverse basis for its underlying predictions (algorithm scores, GTEx specificity, TMHMM).
- Future work is focusing on experimental validation, additional reporting metrics and integration with AltAnalyze version 3.
- We welcome new opportunities for collaboration in diverse cancers.

Acknowledgements and Availability

This work received support from an NIH R01 award from the National Cancer Institute (R01-CA226802) and associated supplement award. Additional support from a Cincinnati Children's Research Foundation Endowed Scholar award. We thank Dr. Surya Prasath for his supervision of DeepImmuno library development and Dr. Gloria Sheynkman for providing the long-read RNA-Seq PacBio melanoma data. No human subjects were recruited for these studies.

- GitHub: <https://github.com/franklgy/SNAF>
- Contact: li2g@mail.uc.edu
- Citation: Please cite the GitHub. Manuscript in submission.