21. Correcting neoantigens by accounting for proximal variants

Jasreet Hundal [1], Susanna Kiwala [1], Yang-Yang Feng [1], Connor J. Liu [1], Joshua McMichael [1], Christopher A. Miller [1,2,3], Huiming Xia [1], Jason Walker [1], Elaine R. Mardis [4], Obi L. Griffith [1,2,3,5], Malachi Griffith [1,2,3,5]

[1] *McDonnell Genome Institute, Washington University School of Medicine, Saint Louis, MO, USA*
[2] *Siteman Cancer Center, Washington University School of Medicine, Saint Louis, MO, USA*
[3] *Division of Oncology, Department of Internal Medicine, Washington University School of Medicine, Saint Louis, MO, USA*
[4] *Institute for Genomic Medicine, Nationwide Children's Hospital, Columbus, OH, USA*
[5] *Department of Genetics, Washington University School of Medicine, Saint Louis, MO, USA*

Recent efforts to design personalized cancer immunotherapies use predicted neoantigens. Typically, to evaluate neoantigens from genomic sequencing data, the raw reads are aligned to the Human Reference Genome, and somatic variants are identified by comparison of tumor to normal read alignments, followed by variant annotation and prediction of strong-binding neoantigenic peptides. This process typically assumes that the reference genome sequence surrounding each somatic variant is representative of the patient's genome, and does not account for the effect of nearby variants (somatic or germline) in the neoantigenic sequence. Because the accuracy of neoantigen identification has important implications for many clinical trials, there is a need for patient-specific inclusion of proximal variants to address this previously oversimplified assumption.

We evaluated somatic variants from 430 tumors to determine the impact of proximal alterations on neoantigens. Without incorporating proximal variant correction for MHC class I peptides, the overall FDR and FNR across peptides of lengths 8–11 were estimated as 0.069 (6.9%) and 0.026 (2.6%), respectively. Thus, for "uncorrected" neoantigen identification in 100 individuals, we can expect that approximately 51 individuals would receive a suboptimal vaccine due to receiving a neoantigen with an incorrect peptide sequence, 23 would receive a suboptimal vaccine due to missing a strong-binding neoantigen, and 62 would receive a suboptimal vaccine due to at least one of these causes.

We also added this improvement in our computational toolkit - *pVACtools,* that aids in neoantigen prediction from somatic alterations (pVACseq and pVACfuse), prioritization and selection using a graphical interface (pVACviz), and determination of optimal order of candidates in a DNA vaccine (pVACvector).

The results from pVACtools analyses are already being used in cancer immunology studies, and ongoing clinical trials. We anticipate that pVACtools will make such analyses more robust, reproducible, and facile.

doi: 10.1016/j.cancergen.2019.04.027

22. Standardization and systematization of somatic variant refinement using a standard operating procedure and deep learning

E.K. Barnell [1], B.A. Ainscough [1], K.M. Campbell [1], T.E. Rohan [2], R. Govindan [1], E.R. Mardis [3,4], S.J. Swamidass [1], M. Griffith [1], O.L. Griffith [1]

[1] *Washington University School of Medicine, Saint Louis, MO, USA*
[2] *Albert Einstein College of Medicine, New York, NY, USA*
[3] *Ohio State University College of Medicine, Columbus, OH, USA*
[4] *The Research Institute at Nationwide Children's Hospital, Columbus, OH, USA*

Introduction: During cancer genomic analysis, accurate identification of somatic variants typically requires manual review of aligned read sequences after automated variant calling. However, this process is costly, time-consuming, and poorly standardized. Here we describe a systematized method for somatic variant refinement using a manual review standard operating procedure (SOP) and an automated deep learning model (DeepSVR-https://github.com/griffithlab/DeepSVR).

Methods: First, the SOP was developed using 4 different calls and 19 tags. The SOP was validated by having 4 individual classify variants prior to, and after, reading the SOP. Second, the DeepSVR algorithm was developed using sequencing data from 41,000 variants derived from 440 samples. Performance was validated using internal cross-validation, orthogonal sequencing data, and independent testing sets.

Results: After reading the SOP, average accuracy in somatic variant identification increased by 16.7% (p-value = 0.0298) and average inter-reviewer agreement increased by 12.7% (p-value < 0.001). DeepSVR internal cross-validation showed a receiver operating characteristic (ROC) area under the curve (AUC) of 0.96. When employed on 212,158 variants (107 samples) with orthogonal sequencing, DeepSVR attained a ROC AUC of 0.94. When employed on 37 independent samples (17,356 variants), DeepSVR attained a ROC AUC range between 0.72-0.92. Reduced accuracy was recovered (ROC AUC > 0.9) after incorporating 100-250 variants from the independent testing set and retraining the model.

Conclusions: Combined use of the SOP and DeepSVR is recommended to standardize somatic variant refinement. These tools will hopefully improve variant refinement accuracy and reduce inter-reviewer variability for variant calling and annotation.

doi: 10.1016/j.cancergen.2019.04.028

23. Ultra-deep sequencing of classical Hodgkin lymphoma (cHL) reveals novel somatic mutations and exemplifies the utility of deep sequencing in the characterization of rare malignant cells

Felicia Gomez, Zachary Skidmore, Alina Schmidt, Fernanda Rodrigues-Martins, Kilannin Krysiak, Cody Ramirez, Matthew Mosior, Eric Duncavage, Grace Triska, Lee Trani, Nancy Bartlett, Amanda Cashen,