Editorial Board October 8, 2019  
NAR Genomics and Bioinformatics

Dear colleague:

I request your consideration of our attached manuscript, “BATCAVE: Calling somatic mutations with a tumor- and site-specific prior” for publication in *NAR Genomics and Bioinformatics*.

Processing next generation sequencing reads to accurately identify somatic mutations within tumors is key to understanding tumor evolution and developing treatment strategies. Consequently, many algorithms for somatic variant calling have been developed. Among the most widely used methods are those based on a statistical model of the sequencing process. But no existing algorithm models the biology of mutation generation within tumors. A key aspect of tumor biology is that each tumor has its own mutational profile, particular types of mutations that it preferentially generates, depend on which molecular processes have gone awry during tumorigenesis. These profiles differ dramatically even among tumors of the same cancer type, and they can place strong constraints on the mutational content of the tumor.

To leverage the biology of tumor mutation generation to improve somatic variant calling, we developed BATCAVE (Bayesian Analysis Tools for Context-Aware Variant Evaluation), which we describe in the present manuscript. Briefly, the BATCAVE algorithm uses high confidence variant calls to infer the individual tumor’s mutational profile and mutation rate. These inferences are then used in a prior to assess the evidence for lower-confidence variant calls. We implemented the BATCAVE algorithm as a filter on top of MuTect, one of the most widely used somatic variant callers. In tests on simulated data, BATCAVE substantially improves precision and recall of somatic variants when compared to MuTect. BATCAVE also substantially improves calibration; the posterior probabilities provided by BATCAVE are much more accurate than those provided by MuTect. In our manuscript, we also test on two real data sets, showing that BATCAVE also performs well on real data.

We believe our manuscript will be of broad interest to the readers of *NAR Genomics* *and Bioinformatics*. For users, BATCAVE is a lightweight addition to tumor variant calling pipelines that substantially improves results. For developers, our manuscript demonstrates the power that can be gained by incorporating tumor biology into the variant calling algorithm. The BATCAVE algorithm is general, so it could be easily added to callers other than MuTect. And other constraints on mutation generation could also be incorporated. We thus believe our manuscript will be of interest to a broad community of bioinformatics users and developers.

Sincerely,  
  
Ryan Gutenkunst, Ph.D.  
Associate Professor and Associate Department Head  
Department of Molecular and Cellular Biology  
University of Arizona