

PROFILING MICROSATELLITE INSTABILITY USING RNA SEQUENCING DATA

¹Sukhi Singh*, ²Janez Kokosar, ¹Jan Otonicar, ²Luka Ausec. ¹Genialis, Ljubljana, Slovenia; ²Genialis, Inc., Boston, MA, USA

Background Microsatellite instability (MSI) has emerged as an important biomarker for guiding treatment decisions in immuno-oncology. The FDA recently approved the use of pembrolizumab in patients with metastatic MSI-high or mismatch repair-deficient (dMMR) solid tumors. Approved assays for MSI, such as Foundation One, typically rely on measuring variants by targeted DNA sequencing. While DNA-based assays are the state of the art today, the precision oncology field is beginning to see new types of biomarkers based on complex signatures like those measured by RNA sequencing, which aim to increase predictive accuracy by better describing disease subtypes. If MSI status could be determined from RNA-seq data, this would enable combining MSI with other biomarker signatures to provide a more comprehensive portrait of the disease state, all from a single sequencing assay. In this study, we developed a RNA-seq variant calling pipeline and used it to characterize MSI in different cancer indications from tumor samples without the need for matched normal samples.

Methods Publicly available data sets that included MSI status were selected for a range of cancer types—colorectal, ovarian, endometrial, and gastric cancer. A bioinformatics pipeline was developed following GATK best practices for calling and quantifying single nucleotide variants (SNVs) and insertion-deletion mutations (INDELs) from RNA-seq data. This pipeline was used to characterize an inventory of frequently altered MSI hotspots in the human transcriptome by filtering for microsatellites from MSI-high patients with high frequency of INDELs across multiple cancers. The RNA-seq variant calling pipeline is available under Apache 3.0 open source license.

Results The RNA-seq pipeline was validated by comparing its outputs to the 1000 Genomes Project. Next, the RNA-seq workflow was used to predict MSI status in hundreds of tumor samples representing four cancer types based on tumor INDEL alteration at the cataloged hotspots. We observed >90% deletions than insertions in the MSI-high hotspots, consistent with previously published observations. This method showed comparable performance to established commercially available tests.

Conclusions This study demonstrated reliable prediction of MSI status using genomic variants called from RNA-seq data. Measuring variation at hundreds of hotspot loci present in different tissues and demographically distinct human cohorts may contribute to more robust and generalizable performance. Further, this method does not require a normal control to estimate the mutational load. Ongoing work aims to evaluate the potential as a pan-cancer diagnostic that can be combined readily with other gene signature biomarkers to maximize clinically actionable insights.

<http://dx.doi.org/10.1136/jitc-2023-SITC2023.0179>