May 13th, 11:30 AM - 1:30 PM

Performance Analysis of Three Bioinformatic Variant Callers Using a Somatic Reference Standard

Binu Porath
bporath@cmh.edu

Binu Porath

Performance Analysis of Three Bioinformatic Variant Callers Using a Somatic Reference Standard

Binu Porath[1], Byunggil Yoo[2], Neil Miller[2], Midhat Farooqi[1]

[1]Department of Pathology and Laboratory Medicine, [2]Center for Pediatric Genomic Medicine
Children's Mercy Hospital, Kansas City, MO

Introduction

- Clinical next-generation sequencing assays require rigorous testing of bioinformatics pipelines to ensure that they faithfully detect relevant variants and minimize false positive and negatives.
- Here we compare the performance of 3 different bioinformatic variant callers for somatic variant detection (SVD) using a cell line-derived reference standard (RS).

Methods

Reference Standard (RS) characteristics	• OncoSpan Reference Standard (RS) • Cell line-derived • 386 variants in 152 genes • Variants between 1-100% variant allele frequency (VAF)
Sequencing and alignment	• Illumina NovaSeq 6000 • Whole Genome Sequencing (WGS), 100X coverage • Reads aligned to GRCh37/hg19 using Illumina DRAGEN Bio-IT (DB) platform
Analysis parameters	• Mostly default parameters • Tumor-only analysis • Turned off "contamination filter" to account for admixed RS
Variant calling on BAM file	• DRAGEN in Germline mode (DBIG) • DRAGEN in Somatic mode (DBIS) • GATK4 Mutect2 (MT2)

Results

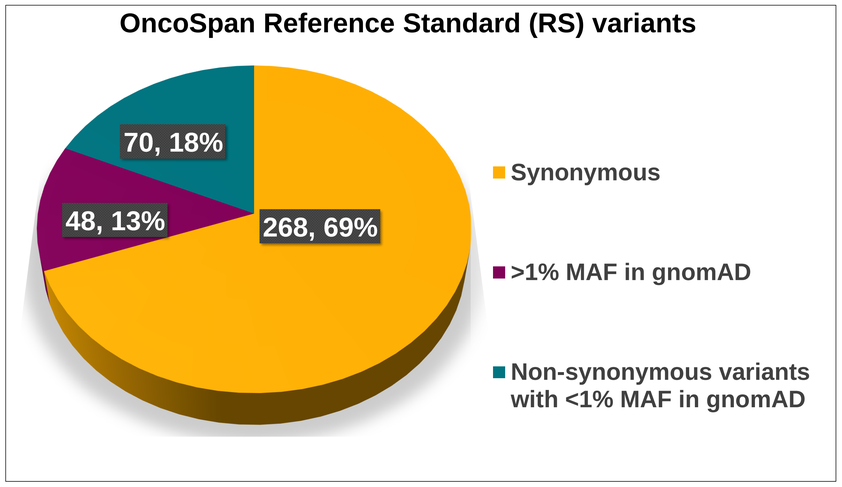


Figure 1: To focus on clinically relevant RS variants, we removed synonymous variants (n=268) and then any variant with a minor allele frequency (MAF) >1% in gnomAD (n=48). The remaining 70 variants were assessed.

Performance Comparison

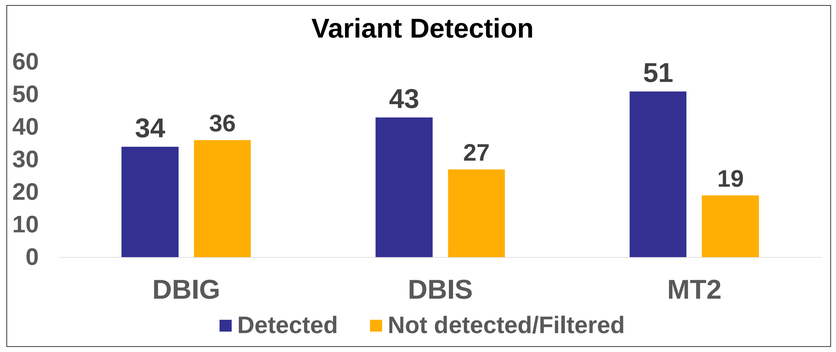


Figure 2
- DBIG detected only 34/70 variants with a variant allele frequency (VAF) of more than 30%.
- DBIG was removed from further analysis and we compared and contrasted DBIS and MT2.

Filters Used

Panel of Normals (PoN)	• 82 normal samples used to account: • Misread bases • Sequencing artifacts • Misaligned reads • Position based filtering
Clustered events	• Filtered out variants if another variant was "nearby"- 15-100bp away
Bad haplotype/germline presence	• Filtered variants commonly present in healthy control databases • gnomAD

Undetected/Filtered Variants

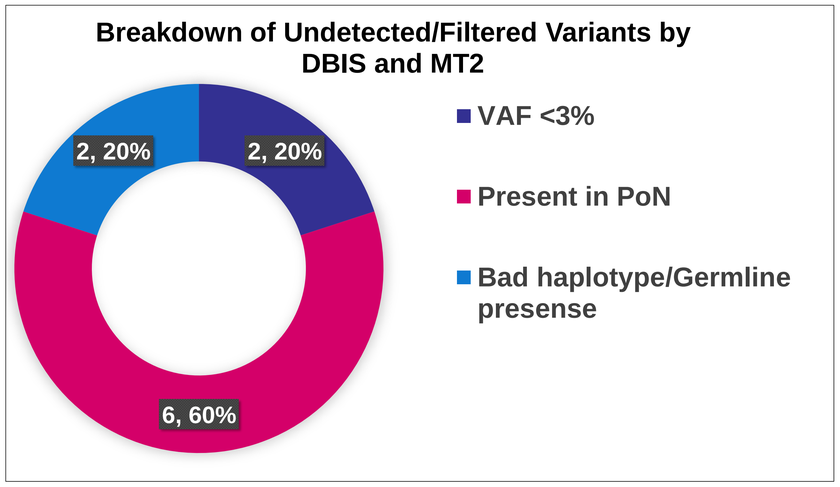


Figure 3: Both DBIS and MT2 did not detect 2 variants present at a VAF <3%, 6 variants 'present' in the PoN (filtered based on position, not nucleotide change), and appropriately filtered out 2 variants due to a bad haplotype and germline presence.

Improper Filtering

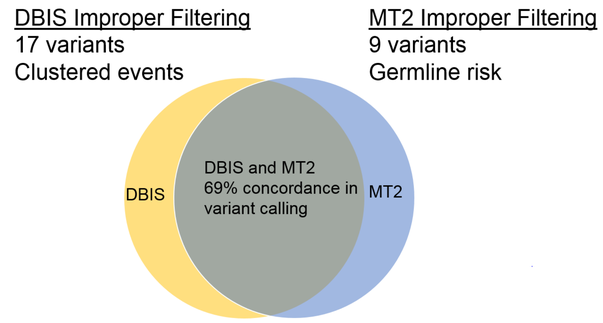
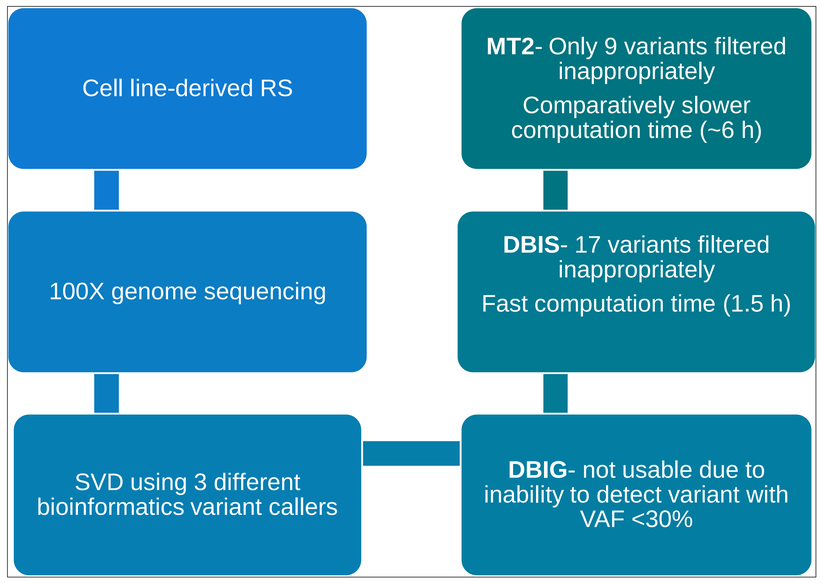


Figure 4
- MT2 inappropriately filtered out 9 variants because of "germline risk" even when the variant was absent in gnomAD.
- DBIS improperly filtered out 17 variants due to "clustered events" if another variant was 'nearby' (anywhere from 15 to 100 base pairs away).
- Overall, they were concordant for 69% of the variant calls.

Summary & Conclusions



- We included DBIG in our study as a control since it was clinically validated for germline testing, but quickly found that it was not adequate for SVD.
- Overall, MT2 called variants more accurately than DBIS, though fine-tuning of certain filters for both would result in better concordance and detection of the 70 RS variants.
- In terms of compute time, DBIS is faster than MT2 with the former taking about 1.5 hours (h) per case, compared to 48 to 72h for the latter (given above parameters).
- However, removal of certain filters can reduce MT2 processing time down to <6h.

References

- Miller NA et al. A 26-hour system of highly sensitive whole genome sequencing for emergency management of genetic diseases. Genome Med. 2015; 7:100
- https://software.broadinstitute.org/gatk/documentation/tooldocs/4.beta.4/org_broadinstitute_hellbender_tools_walkers_mutect_Mutect2.php