#5462

Improved accuracy of somatic variant detection from High Throughput Sequencing (HTS) data via site-specific noise estimation

ESMO Congress 2022
9-13 September, 2022
Paris, France

Ekaterina Ignatova [1], Valentina Yakushina [2], Vladislav Mileyko [2], Maxim Ivanov [2,*]
1. Research centre for medical genetics, Moscow, Russian Federation 2. Atlas Oncodiagnostics, LLC, Moscow, Russian Federation
* Corresponding author – maksim.v.Ivanov@phystech.edu

Introduction

Clinical value

- Detection of subclonal mutations with high sensitivity is useful for various clinical applications, including residual disease monitoring, cfDNA analysis and detection of mutations associated with therapy resistance
- Despite high information yield, process of variant detection in Next Generation Sequencing (NGS) data posses low limit of detection compared with orthogonal methods like ddPCR

Bioinformatics value

- Amplicon-based enrichment technology generates high amount of false-positive variant calls mainly due to primer-insert interaction and edge-effects
- NGS data posses site-specific **[Figure 1]** and site-non-specific **[Figure 2]** sequencing noise. While the second one can be efficiently handled with base quality recalibration, the first one may produced numerous false-positive calls. Moreover, noise levels depend on library preparation protocols restricting use of universal pipelines for all data.

Results

- We have developed a software for detection of genetic variants employing site-specific noise level estimation with beta distribution approximation **[Figure 3]**. It uses iterative beta distribution approximates which allows to use samples with mutations as controls, depriving the need for generating control set of samples. As a consequence, sample set within any single run may used as study set and control set at once. Developed software tailored to amplicon data sequencing leveraging by-amplicon read grouping for reference and alternative allele count
- Variant calling corrected by site-specific noise level estimation allows to increase analytical sensitivity **[Figure 4]**
- Developed software allowed efficient variant calling even in complex regions, like homopolymer regions. Generating distribution of homopolymer length within control samples allows automatic filtration of false-positive calls and detects true variants (validated on CFTR c.2052dupA variant detected by developed software in 4/98 patients with Cystic Fibrosis and validated with Sanger sequencing, none of which could not be called with any other tools)

- We have applied Ion Torrent Variant Caller along with developed software for variant calling on a set of 103 samples sequenced with Cancer Hotspot Panel V2 (50 target genes; 22Kb total size of target regions) and 402 samples sequenced with Comprehensive Cancer Panel (409 target genes; 1.7Mb total size of target regions) sequenced employing Ion Torrent technology. Focusing on mutations in hotspots (Cosmic Count 10 and more), developed software was able to detect 11 additional subclonal mutations not detected via software from sequencing platform provider, including: two non-common BRAF mutations (L597Q **[Figure 5]** and G466A, both patients with lung cancer); 5 KRAS mutations (three patients with colorectal cancer and two patients with lung cancer); 1 NRAS mutation (colorectal cancer) and 3 PIK3CA mutations (one patient lung cancer and two patients with breast cancer)
- Variant detection across over 100,000 in silico generated sequencing NGS datasets demonstrated that use of developed software allows to increase limit of detection up to 0.5% compared to 2% characteristic of other variant detection tools **[Figure 4]**

Conclusion

- Here we demonstrate that employing additional bioinformatics approaches we may increase the rate of detection of clinically significant subclonal mutations and efficiently filtrate false positive variant

- Variant detection in NGS data supported by site-specific noise level estimation allows to increase limit of detection, analytical sensitivity as well as diagnostic specificity
- Developed software may be used for highly efficient variant detection in amplicon-based NGS data

Figure 3. Proposed approach for variant detection in NGS data

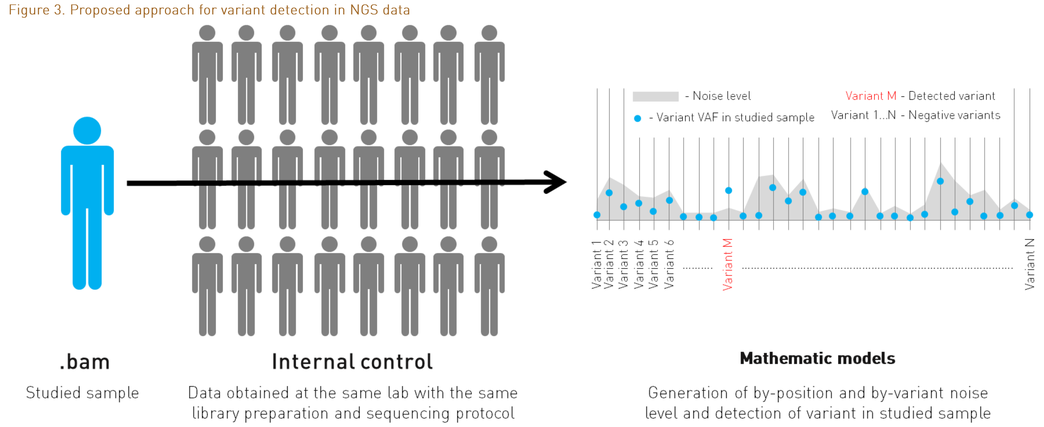


Figure 4. Analytical validation on a set of 100,000 in silico generated datasets with different mutations demonstrates high analytical sensitivity of developed tool

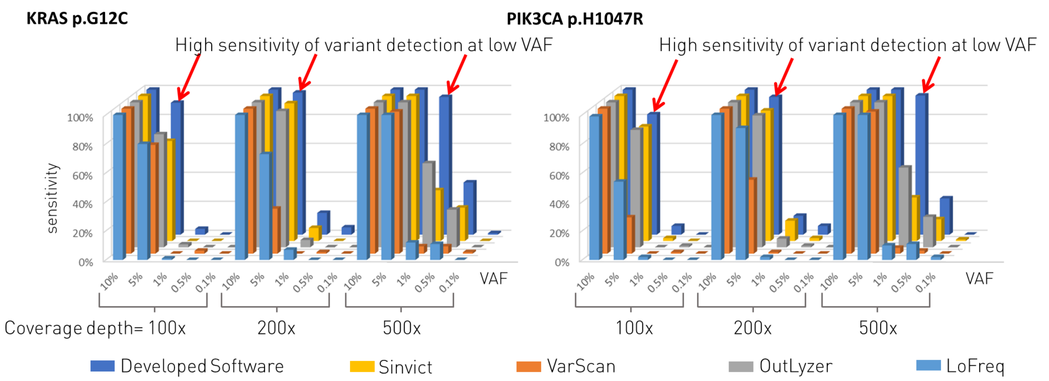


Figure 5. Employing developed software allowed to call subclonal non-common BRAF mutation, not detected with any conventional software

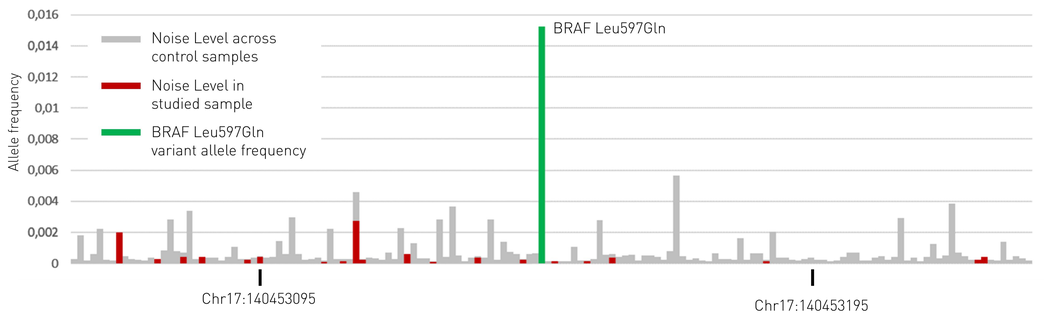


Figure 1. Site-specific errors produce false-positive calls due to high noise level in certain positions

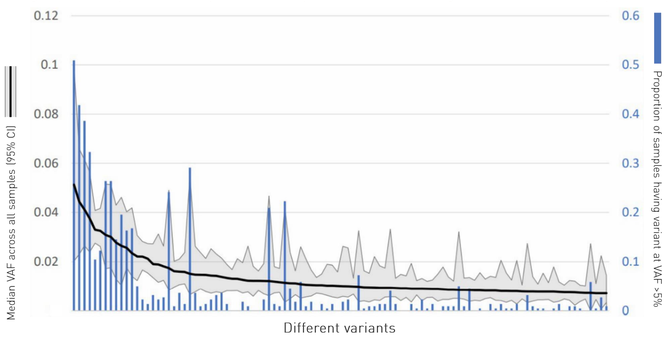
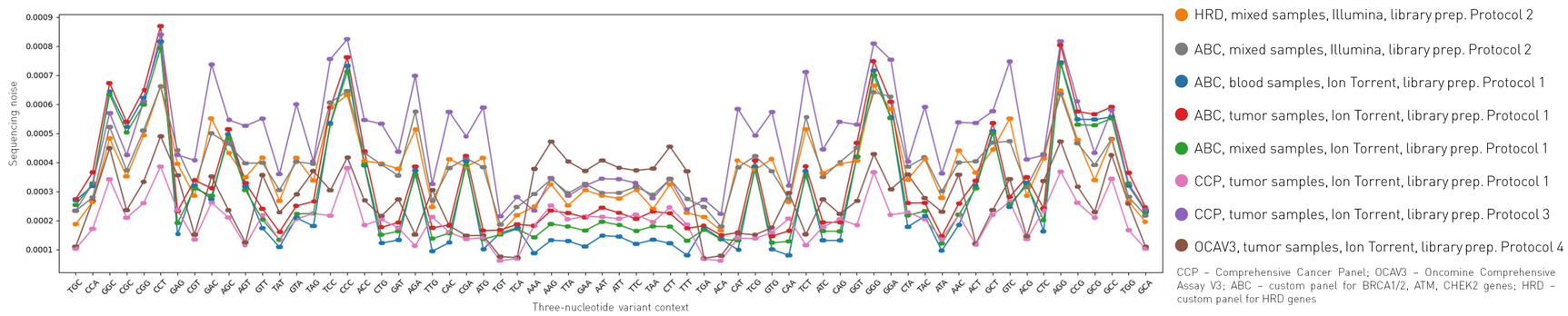


Figure 2. Despite base quality recalibration may be used to handle with site-non-specific errors, it depends on library preparation protocol and sequencing technology limiting use on for all pipeline



CCP – Comprehensive Cancer Panel; OCAV3 – Oncomine Comprehensive Assay V3; ABC – custom panel for BRCA1/2, ATM, CHEK2 genes; HRD – custom panel for HRD genes