35P

Copy Number Variation detection from targeted amplicon-based high-throughput sequencing in clinical settings

MAP 2022
14-16 October 2022
Amsterdam, Netherlands

Yakushina V.D.[1], Mileyko V.A.[2], Ivanov M.V.[2]
[1] Laboratory of Epigenetics, Research Centre for Medical Genetics, Moscow, Russian Federation; [2] Atlas Oncodiagnostics LLC, Moscow, Russian Federation

Background

Accurate detection of Copy Number Variants (CNV) in homologous recombination repair (HRR) related genes is essential for HRR deficiency diagnostics and treatment administration such as PARP inhibitors. CNV detection from high-throughput sequencing (HTS) is expected to complete simultaneous comprehensive genetic markers evaluation, reduce cost and turnaround time and to increase sensitivity. It is still challenging to detect somatic CNV from small targeted panels that are widespread in clinical settings.

Aim: to evaluate clinical utility of somatic CNV detection from amplicon-based HTS panels.

Methods

Material
Tumor samples (FFPE), N=780
Tumor types: breast, ovarian, pancreatic & prostate

Sequencing
AmpliSeq targeted enrichment (custom panel)
Genes: ATM, BRCA1, BRCA2 and CHEK2
Number of amplicons: 409
Size: 36kbs
Sequencing platform: Illumina, MiSeq
Thermo Fisher, Ion S5 (+ 610 samples)

ML approach
Random Forest
Metrics (24 in total): score of VAF variation consistency*,
number of detactable SNPs,
score of SNP allelic imbalance consistency,
by-pool coverage drop,
data quality metrics
* score of VAF variation consistency – deviation of VAF** from expected for heterozygote
** Variant allele fraction (VAF) was calculated with amplicon-guided read counting algorithm (**AGRABAH**)

Accuracy assessment
Manual validation, based on by-eye curation of VAFs and coverage

Results

- 717 samples sequenced on MiSeq (91%) passed Quality Control (QC) assessed with developed ML algorithm
  Samples passed QC control were further analyzed for BRCA1 and BRCA2 deletions
- Developed ML – based approach allowed to predict
  BRCA1 deletion - in 24 samples (3%) and
  BRCA2 deletion - in 32 samples (4%)
  Of them, 4 samples had both, BRCA1 and BRCA2 deletion
- Accuracy of QC, BRCA1 & BRCA2 deletion detection with developed algorithm is above 0.979 (Table 1, Fig. 1).
- The accuracy for CNV prediction with developed algorithm is higher when allelic imbalance is calculated with amplicon-guided read counting algorithm AGRABAH (Fig. 1).
- AGRABAH for VAF calculation reduces amplicon dropout errors and PCR-enrichment artifacts, improves VAF & SNP allelic imbalance estimation (Fig. 3,4).

Conclusions

- Integrative analysis of allelic imbalance and amplicon coverage from targeted high-throughput sequencing with small panels allows to detect samples putative for CNV for further validation with orthogonal methods.
- This allows to increase information yield from small gene panel sequencing and may lead for PARP inhibitors indication for additional 7% of patients with breast, ovarian, pancreatic and prostate cancer

Results

Table 1. Accuracy of QC and CNV detection with developed algorithm

	accuracy	AUC	TPR	TNR	PPV	NPV
QC	0.979	0.958	98.17	93.75	99.53	78.95
BRCA1	0.981	0.986	100.00	98.10	55.56	100.00
BRCA2	0.986	0.999	91.67	99.01	84.62	99.50

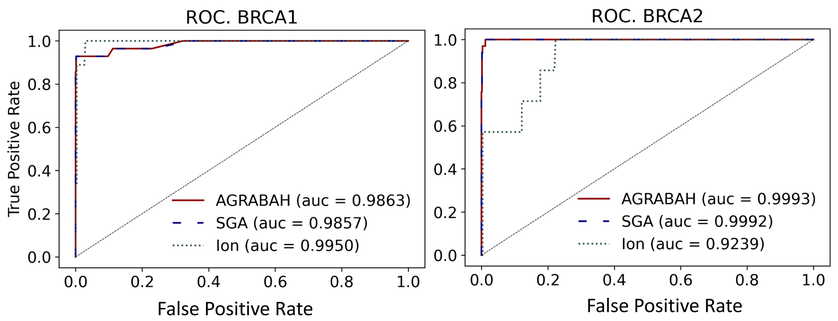

Figure 1. Accuracy of CNV detection with developed algorithm.
Area under ROC for CNV detection with developed ML-based approach reached 0.98 (BRCA1 deletions) and 0.99 (BRCA2 deletions).
Utilizing SGA* for VAF evaluation slightly decreases accuracy.
Applying of the developed algorithm (trained on Illumina MiSeq data) might give lower accuracy, as was shown for BRCA2 deletion.
* SGA - String Graph Assembler [Simpson J. T., 2012]

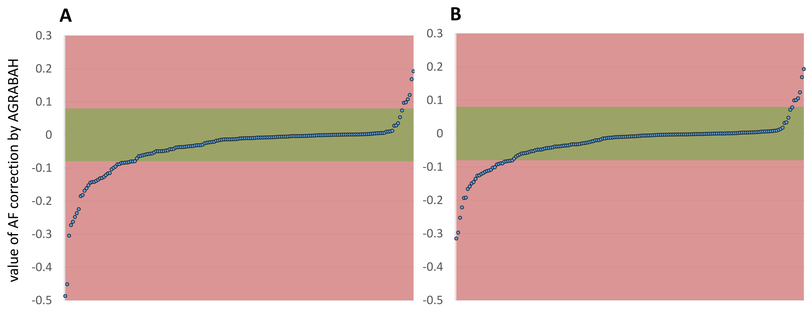

Figure 3. The value of VAF correction by AGRABAH for AFs determined by samtools mpileup (A) or SGA* (B).
Here is an assumption that for SNPs the VAF at around 0.5 is a prevalent.
For 181 SNPs, the difference of deviation from 0.5 for VAF evaluated with AGRABAH compared to VAF evaluated with mpileup/SGA is calculated.
SNPs with difference below green area – considered as corrected by AGRABAH.
SNPs with difference above green area – considered as miss estimated by AGRABAH.
In comparison with samtools mpileup: 36 SNPs (20%) were found to be corrected by AGRABAH & 6 SNPs (3%) were found to be miss estimated
In comparison with SGA: 30 SNPs (17%) were found to be corrected by AGRABAH & 6 SNPs (3%) were found to be miss estimated.
The assessment of VAF with amplicon-guided read counting algorithm AGRABAH is more precise.

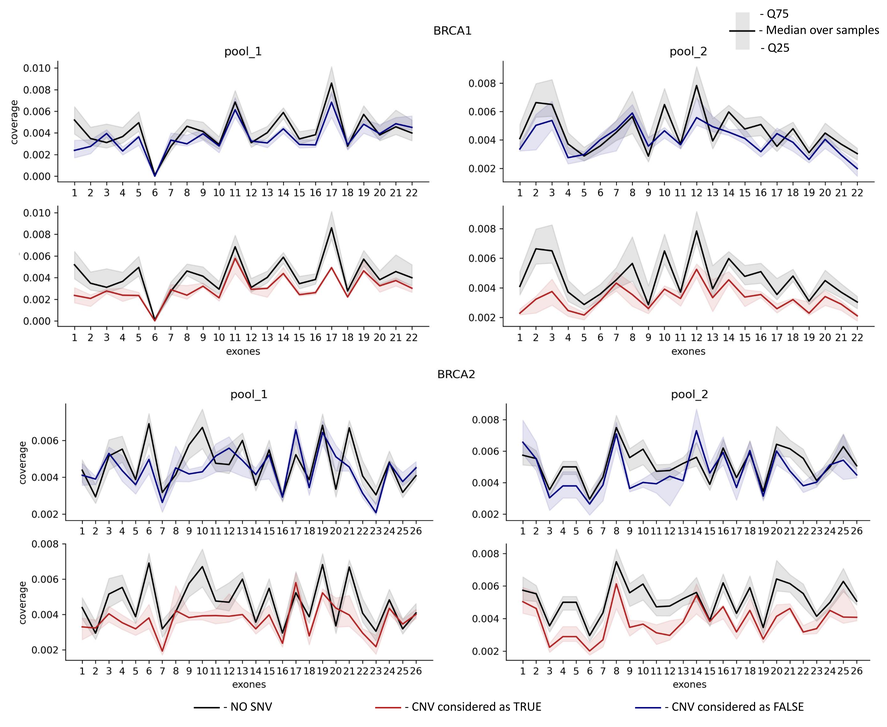

Figure 2. The coverage* of exons for BRCA1 and BRCA2.
The baseline of coverage distribution over exons is established (black line).
Samples predicted by the developed approach as putative for CNV and confirmed by manual curation (red line) have consistent drop in coverage.
For samples putative for CNV but failed with manual validation (blue line), significant crossing with baseline is observed.
* The coverage is normalized by total read number in pool.
* For exons covered with several amplicons in a pool mean over amplicons was calculated.

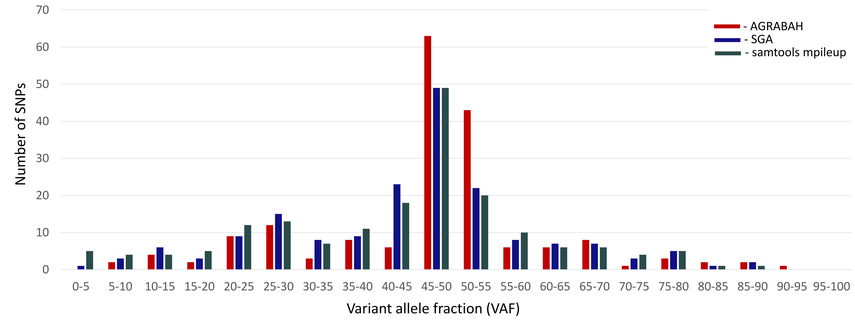

Figure 4. Distribution of allelic frequencies (VAF) determined by AGRABAH, SGA* & samtools mpileup for 181 SNPs.
amplicon-guided read counting algorithm AGRABAH determines VAF with highest frequency at 45-50%, followed by 50-55%. That outperforms SGA and samtools mpileup at these range. At ranges of AF below 40% & above 60%, AGRABAH has lower number of AFs

Conflict of interests disclosure:
no conflict of interests

Corresponding author:
vdyakushina@gmail.com