Accurate and high-resolution copy number variant detection in clinical germline screening

Jiani Li, PhD; Kevin R. Haas, PhD; Sun Hae Hong, PhD
Myriad Women's Health

BACKGROUND

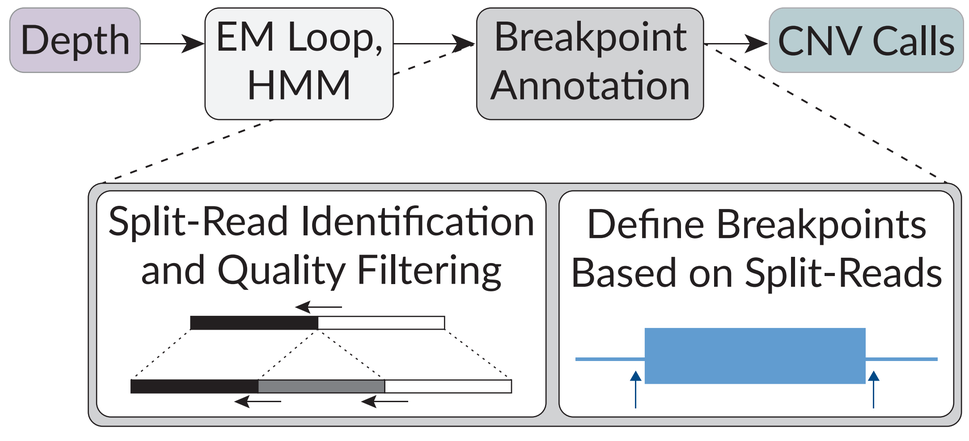
- Structural variants, especially copy number variants (CNVs), can cause genomic disorders.
- CNVs can bring genomic coordinates that are distant in the reference genome into contact in a sample genome, forming "breakpoints."
- Determining exact CNV breakpoint sequences (physical deletion or duplication boundaries) across individuals is crucial for clinical germline screening, and for associating genotype to phenotype.
- For targeted sequencing, the majority of the current tools rely only on the depth of coverage, which reports only approximate genomic coordinates.
- Here, we report a high-resolution CNV detection method that hierarchically combines read-depth signal and reads that align in split fashion to discontiguous loci in the reference genome (split-reads).

METHODS

- We used a Hidden Markov model (HMM) CNV caller that calls by observing sequencing depth (reflects the copy number) for ~100 base-pair segments tiled along the targeted region of interest.
- A CNV breakpoint annotator was applied to identify split-reads that have soft-clipped bases on the side on which the depth-based CNV is located and replaced the CNV breakpoints with breakpoints defined by supporting split-reads (Figure 1).
- Performance was validated by comparing to open-source CNV caller LUMPY[1] on 1,700 patient samples randomly-selected from research-allowed patients tested 8/18-1/19.

Figure 1. Schematics of the CNV Caller.

Breakpoint annotator applied to CNVs identified through depth-based caller. Breakpoints defined based on split-reads.
EM: Expectation-Maximization



RESULTS

- The HMM CNV caller identified 310 CNVs in 1,700 in-house samples. Among the 310 HMM CNVs, the in-house breakpoint annotator identified split-reads for 75 HMM CNVs and LUMPY identified split-reads for 67 HMM CNVs (Table 1).
  - 70 in-house breakpoint annotator CNVs and 67 LUMPY CNVs were confirmed by manual assessment.

Table 1. Performance Validation of In-House Breakpoint Annotator.

CNV Caller	Detected	Manually Confirmed
LUMPY	67	67
In-House Annotator	75	70

- We explored false positive CNV filters and found two parameters that can filter out false positive CNVs (Figure 2). A 'pass' split-read CNV should meet:
  - CNV size ratio: [(CNV size)/(HMM CNV size)] ≥30%
  - The number of split-reads: ≥2
- We compared coordinates of 67 CNVs identified by both LUMPY and in-house breakpoint annotator; 97.0% of CNV coordinates were concordant (Table 2).
  - Discordance was determined to be due to LUMPY also using paired-end read evidence to generate coordinates.

Figure 2. CNV Quality Distribution.

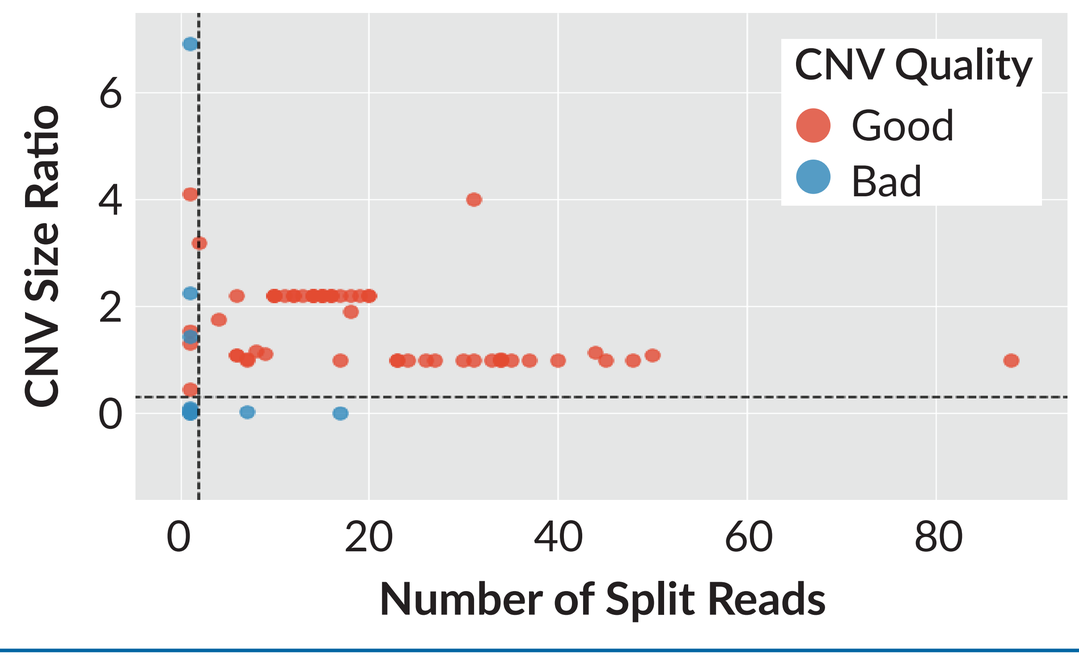


Table 2. Concordance of CNV Coordinates.

Position	Same	Different
Start	65	2
End	65	2

All posters available at
research.myriadwomenshealth.com

CONCLUSIONS

- Using a HMM CNV caller and in-house breakpoint annotator ensured accurate calling of CNVs and reduced the burden of manually refining the CNV breakpoints.
- These data suggest the in-house breakpoint annotator generates high-confidence CNV genome coordinates.

Reference: 1. Layer, RM, et al. *Genome Biol*. 2014; 15(6):R84

Presented at ASHG on October 16, 2019