# Evaluating the Performance of Copy Number Variant Detection Tools in Clinical Applications: A Comparative Benchmark for Short-Read Whole Genome Sequencing

Francisco M. De La Vega,[1] Sean A. Irvine,[2] Pavana Anur,[1] Kelly Potts,[1] Lewis Kraft,[1] Raul Torres,[1] Sean Truong,[3] Yeonghun Lee,[3] Shunhua Han,[3] Vitor Onuchic,[3] James Han,[3] and Peter Kang[1]

[1]Tempus Labs, Chicago, IL, USA. [2]Real Time Genomics, Ltd., Hamilton, New Zealand, and [3]Illumina, Inc., San Diego, CA, USA

TEMPUS

## INTRODUCTION

- Whole Genome Sequencing (WGS) will soon be preferred over Whole Exome Sequencing (WES) and targeted sequencing in clinical settings due to its better CNV/SV detection, faster turnaround time, and dropping costs
- Current tools for short-read WGS CNV calling need to be evaluated for clinical settings where orthogonal confirmation of CNVs may be required, placing a higher priority on sensitivity over specificity/precision compared to research uses
- The aim of this study was to compare the performance of various CNV detection tools for short-read WGS data in a clinical context in reference cell lines with a comprehensive CNV truth set

## METHODS

### CNV calling tools evaluated

- Delly (v1.6) , CNVnator (v0.4.1), Lumpy (v.0.2.13), Parliament2, Cue (cue.v2.pt model), and the new DRAGEN 4.2 CNV caller that combines depth and breakpoint calls

### Data Sources

- PCR-free libraries of HG002 reference cell line sequenced to a mean depth of 50X using paired-end 2x150bp reads on Illumina NovaSeq 6000
- Reads were mapped to the GRCh37 human reference with the DRAGEN mapper
- As truth set, we used the Genome-in-a- Bottle SV truth set for HG002 v0.6. Analysis was confined to events (500bp-100kb) intersecting exons of canonical transcripts
- The truth set had 13 deletions and 4 duplications overlapping 45 and 8 exons, respectively

### Benchmarking analyses

- We focused our evaluation on protein-structure disrupting CNVs given our clinical application (hereditary gene panels)
- We thus identified CNVs in the truth set with overlaps with coding exons for the canonical transcript of all human genes in hg19
- We also simulated gene models on top of other variants in the truth set to increase the number of overlaps to evaluate. The simulation added 47 deletions and 6 duplications overlapping 94 and 19 exons, respectively
- For benchmarking we developed a custom tool that counted events intersecting an exon with dosage direction matching the truth set as true positives. Events not meeting this condition were deemed false positives. We adjusted for events spanning multiple exons to avoid double counting

## SIGNIFICANCE

- DRAGEN v4.2 integrated CNV caller showed the best balance between sensitivity and precision and offers a high sensitivity mode (HS) that achieves the highest sensitivity across the board
- We developed filters that significantly reduce false positives in DRAGEN HS output with little sacrifice on sensitivity, suitable for clinical settings where sensitivity is paramount and confirmatory tests are performed

## RESULTS

**Figure 1. Overall Performance**

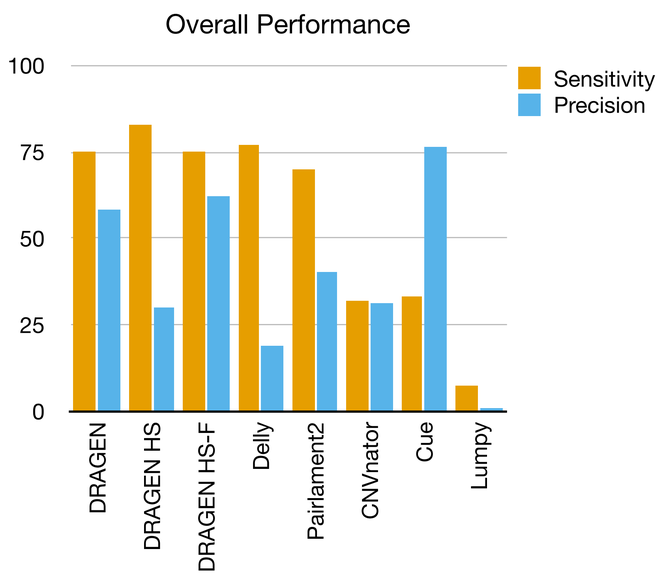


**Fig 1.** Overall performance of the CNV/SV callers benchmarked. The data shows that DRAGEN v4.2 CNV caller had the best balance of sensitivity/precision. The DRAGEN caller on high sensitivity mode (DRAGEN HS) had the best sensitivity, albeit at a lower precision. On the other hand, Cue had the best precision but low sensitivity. We developed a set of custom filters on top of DRAGEN HS (cf. DRAGEN HS-F) successfully improving precision with a small sensitivity cost

- Performance for deletions follows similar trend as described for overall metrics
- Across all CNV callers, sensitivity was much lower for duplications, DRAGEN HS having the highest sensitivity and Cue the best precision
- In general, all callers had lower sensitivity for smaller events, with some of them unable to call events in this range (e.g. CNVnator and Cue)
- Again, the sensitivity for duplications was poorer for all callers, but specifically for 1-5kb events
- DRAGEN had the best precision for small duplications, and surprisingly, many callers had very high sensitivity for duplications >5kb, although precision was low, except for Cue

**Figure 2. Performance by CNV type**

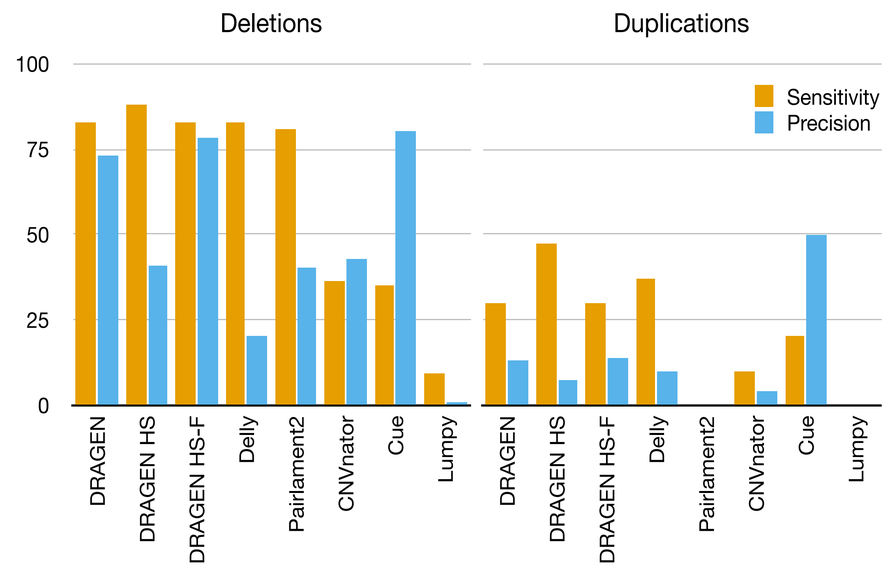


**Fig 2.** Performance exhibited by the callers stratified by type, either deletions or duplications

**Figure 3. Performance by CNV length**

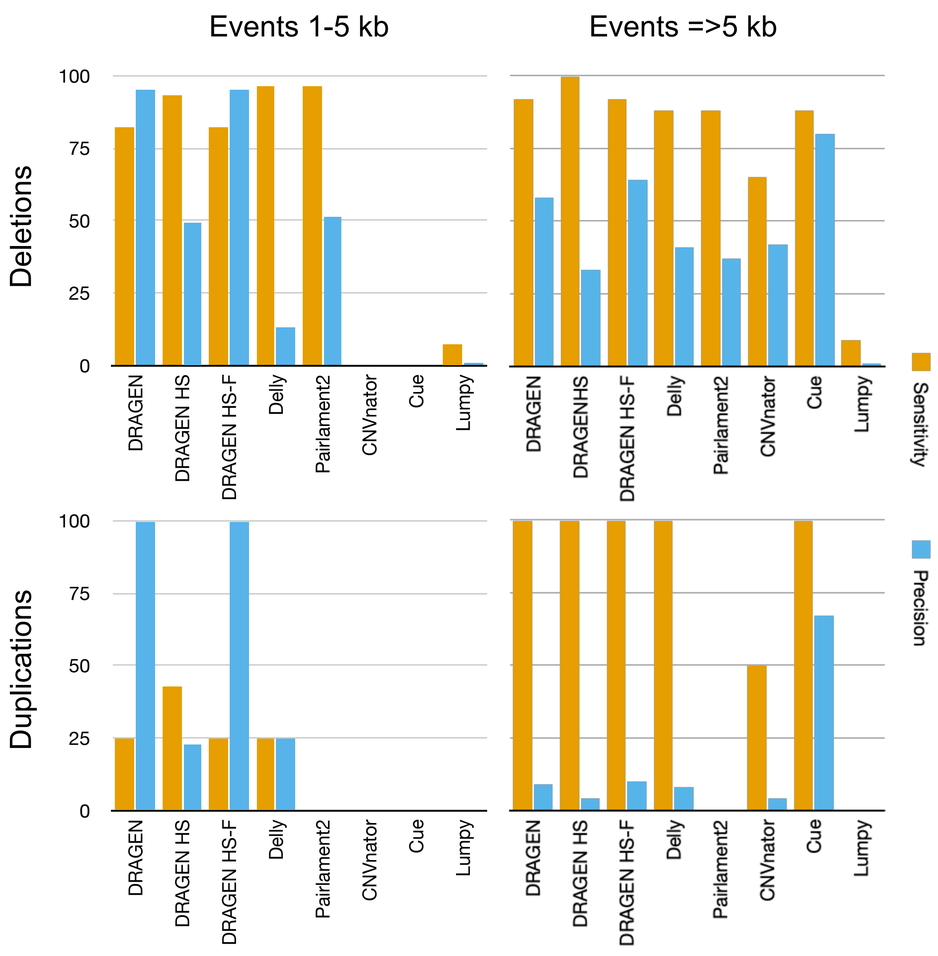


**Fig 3**. Results denoting performance further stratified by event size, either 1-5kb or larger than 5kb

**Acknowledgements:** We thank Amrita A. Iyer, Ph.D., from Tempus Scientific Communications for editorial poster review.

**Correspondence:** Francisco.DeLaVega@Tempus.com