bionano GENOMICS

Optical genome mapping capability expanded to enable detection of absence of heterozygosity

Aliz Raksi, Ernest T Lam, Javier A Velazquez-Muriel, Dong Zhang, Neil Miller, Andy WC Pang, Alex R Hastie, Alka Chaubey, Mark Oldakowski

Bionano Genomics, San Diego, USA

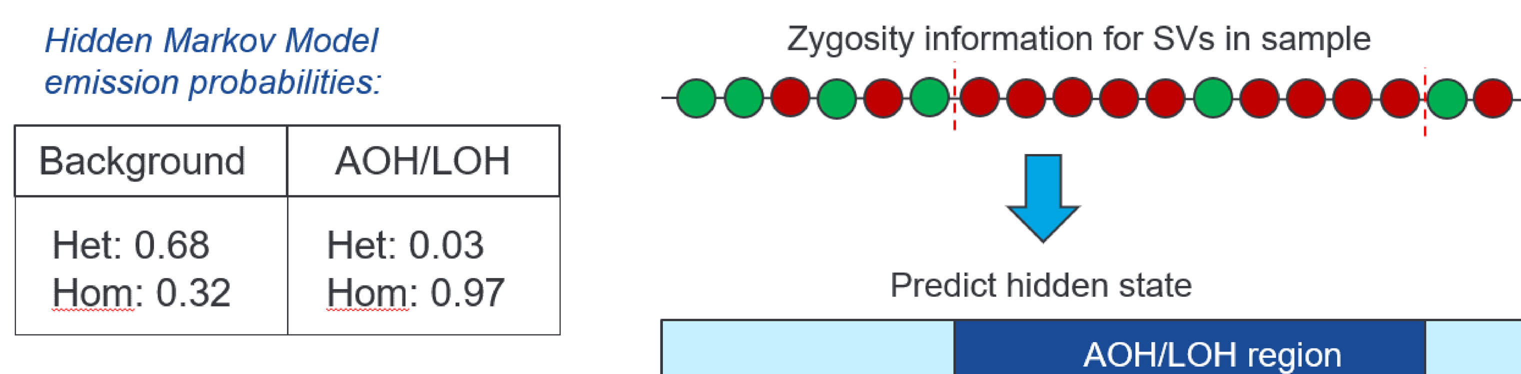
Background

Optical genome mapping (OGM) has been able to detect small and large structural variants (SVs), including deletions, insertions, inversions, aneuploidies and translocations, as well as complex rearrangements across the whole genome that are undetectable by traditional methods, such as DNA sequencing and cytogenetics.[1] OGM is now also able to detect copy-neutral absence of heterozygosity (AOH), which has traditionally been identified using chromosomal microarray (CMA) and microsatellite analysis.[2] Absence of heterozygosity refers to a specific type of genetic abnormality that can be a result of uniparental disomy or consanguinity, as a result of which there is an absence of one parental DNA contribution, and this leads to an increased susceptibility to recessive disease. Here we describe a method for AOH detection based on OGM data from the Bionano Genomics Saphyr platform. Regions of homozygosity are identified by a consistent decrease in heterozygous SV calls across a genomic region in the case sample compared to the level observed genome-wide in controls.

Methods

DNA was processed on the Bionano Genomics Saphyr instrument and analyzed using the Bionano Solve pipeline. After filtering SV calls for high-quality, informative sites, AOH events were simulated by splicing together SV calling datasets from 153 controls and 4 haploid samples, where regions derived from haploid genomes represented AOH events. A Hidden Markov Model (HMM) was used to model the spatial dependence between neighboring SVs of a given zygosity, and the model parameters were estimated by fitting the model to the simulated dataset. SVs were filtered for informative, high-confidence sites, and the input was the binary zygosity information: is the SV in this region heterozygous or homozygous? Improvements to the model would include using allele frequency information, refining SV filtering criteria, and including small previously-unreported indels in the analysis.

Figure 1 Summary of model used for AOH/LOH detection



Evaluating performance

1. Detected known AOH regions previously identified using microarrays
2. Simulated AOH regions by splicing known haploid regions into diploid control samples

Results

In samples with known AOH events, 4/4 large (>25 Mbp) events in confirmed constitutional samples were called with >60% overlap. In one sample with only small events (<25 Mbp), 3/4 events were still called. To summarize, we were able to detect high variant allele frequency events larger than 10 Mbp. False calls occurred below 20 Mbp.

In simulated data, our method achieved high sensitivity and precision in detecting large AOH regions typical of the sizes that underpin uniparental disomy (40-50 Mbp range), with 92% sensitivity and 97% precision, where true positives were defined as an overlap of 80% between simulated and predicted regions.

Sample #	CNLOH identified using microarray	AOH/LOH detection results
1	7q11.22qter(71748536_159119707)x2 hmz, [CNLOH 87 Mb; EZH2]	7 76496409.0 143709638.0 (67.2 Mbps) 7 144329449.0 159345973 (15.0 Mbps) **Called as 2 different LOH events totaling 82.2 Mb (94.5% overlap)**
2	7q22.1qter(98411980_159119221)x2 hmz[0.7], [CNLOH 60.7 Mb; CUX1 and EZH2]	7 101355164.0 108191148.0 (6.8 Mbps) 7 124805963.0 143709638.0 (18.8 Mb) 7 148325410.0 159345973 (11.0 Mb) **Called as 3 different LOH events totaling 36.8 Mb (60.6% overlap)**
3	11q12.1qter(57994955_134942626)x2 hmz, [CNLOH 77 Mb; CBL] 13q13.1qter(32553841_115107733)x2 hmz [CNLOH 83 Mb]	11 56370292.0 99821083.0 (43.5 Mb) 11 121178985.0 129596627.0 (8.4 Mb) 11 133078807.0 135086622 (2.0 Mb, 1.9 Mb "correct") **Called as 3 different LOH events totaling 53.7 Mb (69.7% overlap)** 13 27998292.0 114364328 (86.4 Mb) **Called (98.6% overlap, reciprocal overlap is 94.7%)**
4	1q25.3q31.1(181,349,929-189,805,990)x2 hmz, 5q23.1q31.3(119,409,742-139,707,439) x2 hmz, 14q24.3q32.11(78,367,380-91,128,309)x2 hmz, 22q12.1q12.3(27,696,986-34,194,745)x2 hmz	1 182268699 207724864 (25.4 Mbps) 14 73748304 95896776 (22.1 Mbps) 22 27136865 39043969 (11.9 Mbps) All expected events < 25Mbp. 3 of 4 events called, two are <25 Mbp but near other call, one call filtered due to small size

Table 1 AOH detection results in 4 samples with known events. All 4 known AOH events greater than 25 Mbp (in bold) were detected.

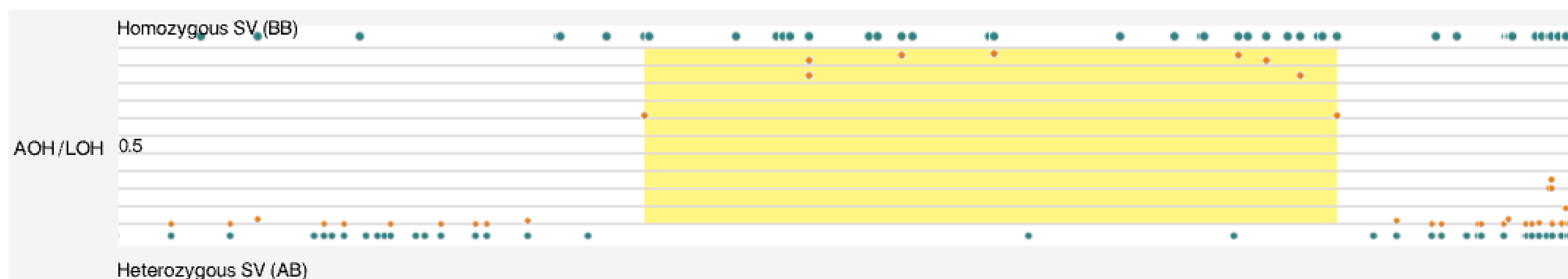
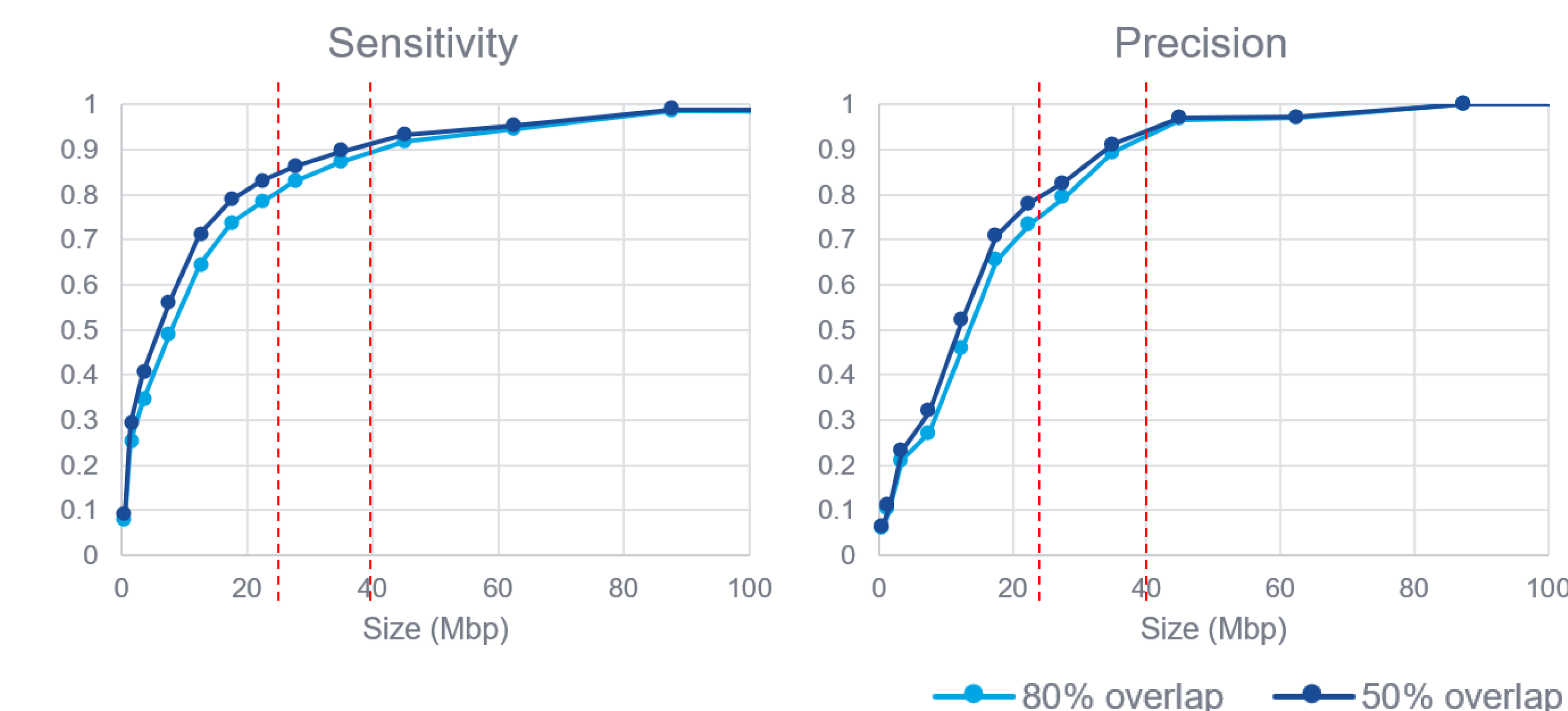


Figure 2 Visualization of an AOH region in Bionano Access v1.7. Teal dots at the top and bottom indicate the presence of homozygous and heterozygous SVs, respectively. Orange dots indicate the probability of the SV being in an AOH region, as calculated by the model. The AOH call is highlighted in yellow.

Figure 3 AOH detection performance using simulated datasets



Aberrations (Total=23000)	Sensitivity (Call rate %) 50% overlap	80% overlap
>40 Mb	96.6%	95.9%
30-40 Mb	89.6%	87.3%
25-30 Mb	86.3%	83.1%
20-25 Mb	83.1%	78.5%
15-20 Mb	78.9%	73.8%
	Precision (Accuracy)	
	50% overlap	80% overlap
>40 Mb	98.6%	98.4%
30-40 Mb	91.2%	89.4%
25-30 Mb	82.7%	79.2%
20-25 Mb	78.1%	73.3%
15-20 Mb	70.9%	65.4%

Table 2 AOH detection performance using simulated datasets

Conclusions

Our results show that it is possible to detect AOH regions using OGM alone. Future developments incorporating additional data will greatly improve the resolution to reliably detect events of smaller size, and those with low variant allele frequency, which would expand the utility of the method to analyzing tumor samples. The software is available as part of Bionano Access v1.7.

References:

1. Neveling K et al., 2020, Next generation cytogenetics: comprehensive assessment of 48 leukemia genomes by genome imaging
2. Zhou X et al., 2004, Whole genome loss of heterozygosity profiling on oral squamous cell carcinoma by high-density single nucleotide polymorphic allele (SNP) array

Corresponding author: Aliz Raksi, araksi@bionanogenomics.com