Strategies to avoid false positives caused by maternal copy number variants in noninvasive prenatal screening

Eerik Kaseniit* MEng, Gregory J. Hogan* PhD, Kevin D'Auria PhD, Carrie Haverty MS LCGC, Dale Muzzey PhD

Counsyl

If not addressed, maternal CNVs impact NIPS

Maternal copy number variants (mCNVs) are a major source of false positives in noninvasive prenatal screening (NIPS)[1-3]. Two recent studies of trisomies 13, 18, and 21 attributed one-third to one-half of NIPS false positives to maternal duplications[2-3]. Here, we characterize the impact of mCNVs on our NIPS analysis and outline a set of best practices for robust aneuploidy screening via whole-genome sequencing.

Strategies for NIPS analysis

A general whole genome sequencing (WGS) based NIPS analysis computes a z-score test statistic for the presence of an aneuploidy by considering the average amount of reads from a chromosome in comparison to the average of a reference population and the dispersion of the average of a reference population. A higher z-score indicates an enrichment in DNA suggestive of a trisomy. Maternal CNVs can change this average (e.g., a maternal duplication can increase the score), potentially causing false positives.

We modified our NIPS analysis pipeline to test the effectiveness of various general statistical approaches (e.g., using median instead of mean) and a directed strategy (filtering mCNV regions from analysis) for limiting mCNV-caused false calls.

	BASIC	ROBUST	mCNV FILTER
Mean & SD	+	-	-
Median & IQR	-	+	+
Remove outlier regions	-	+	+
Gaussian fit	-	+	+
Identify & filter mCNVs	-	-	+

Table 1. Strategies tested for NIPS analysis. SD: standard deviation estimated from raw data; IQR: standard deviation estimated from interquartile range of raw data.

mCNVs: identification and prevalence

mCNVs of ≥200kb were identified in the whole-genome sequencing data of 47,151 NIPS samples. Overall, 87% of samples contain at least one mCNV in the genome. 3% of samples contain a maternal duplication on chromosomes 13, 18, or 21.

	CNV size (% of chromosome)		
	<1%	1-10%	>10%
chr13	575 (1.2%)	59 (0.1%)	0 (<0.1%)
chr18	453 (1.0%)	73 (0.2%)	2 (<0.1%)
chr21	249 (0.5%)	147 (0.3%)	1 (0.1%)

Table 2. Number and frequency of mCNVs observed, binned by location and size.

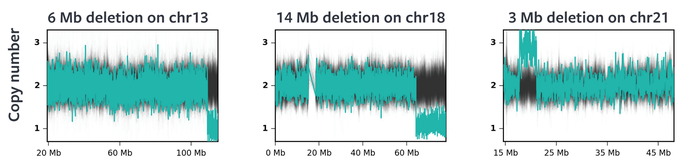

Figure 1. Examples of mCNVs identified from patient samples. Teal traces show estimated copy number of bins across the chromosomes; black traces are background samples.

Assessing performance using simulations

As large mCNVs spanning >10Mb are empirically rare in the healthy pregnant population, we used simulated maternal duplications to test the impact of mCNVs on NIPS.

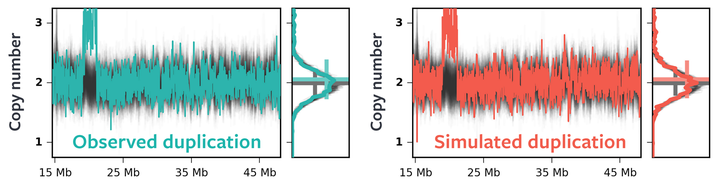

Figure 2. Example of a real (left) and simulated (right) duplication on chromosome 21. Smaller subplots show the histogram of the copy number estimate in bins along the chromosome. Black traces indicate normal background samples.

Impact of mCNVs on NIPS results

Ideally, mCNVs should not influence fetal z-score (the NIPS test statistic). Using the **BASIC** strategy, mCNVs substantially influence z-scores and specificity, with larger mCNVS having a bigger effect. The **ROBUST** strategy is an improvement against smaller mCNVs, whereas the **mCNV FILTER** strategy maintains high specificity even when large mCNVs are present (up to 62% of chromosome).

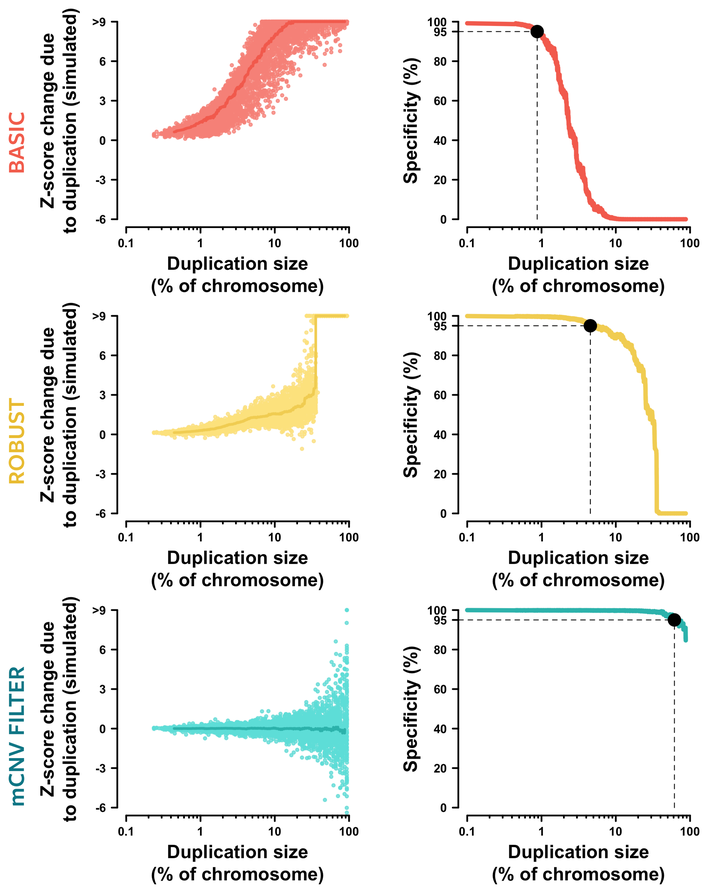

Figure 3. The relationship between mCNV size and impact on z-scores (left) and sample specificity (right), resulting from 6,000 simulated duplications in chromosomes 13, 18, and 21.

Expected false positives per 100,000 tested

BASIC

If not addressed, mCNVs can lead to a false positive for 1 in every 540 patients tested.

ROBUST

Using outlier-robust approaches, mCNV-caused false positives can be limited to 1 in 13,000 patients tested.

mCNV FILTER

When directly addressed, mCNV-caused false positives can be reduced to fewer than 1 in 100,000 patients tested.

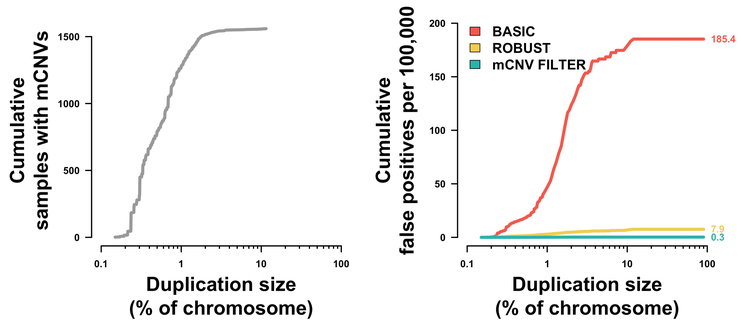

Figure 4. False Positives Expected per 100,000 Tests. Using the measured relationship between size of duplication and change in z-score (see Figure 3) along with the empirically observed frequency of maternal duplications in 47,151 NIPS samples (left), we estimated how many false positives are to be expected under different strategies for NIPS data analysis (right).

Recommended best practices

- Robust statistical approaches and a directed strategy to identify and filter mCNVs result in the lowest false-positive rate.
- Simulations allow thorough testing of NIPS analysis pipelines.
- An important consideration in selecting an NIPS laboratory is which strategy they use to address mCNVs and how they test its performance.

Conclusions

High specificity in NIPS can be achieved—even in the presence of mCNVs that range widely in size—using a suite of best practices, including algorithmic omission of outlying bins, fine-tuned quality-control metrics, and manual call review. Critically, the mechanisms that ensure robustness to mCNVs do not compromise detection of true aneuploidies, thereby preserving both high sensitivity and a low test-failure rate. High specificity, which enables higher positive predictive value (PPV), is critical to preserve clinical utility as NIPS adoption increases in the average-risk population.

REFERENCES: **1.** Snyder MW et al. Copy-number variation and false positive prenatal aneuploidy screening results. N Engl J Med. 2015; 372(17):1639-45. **2.** Strom CM, Maxwell MD, Owen R. Improving the Accuracy of Prenatal Screening with DNA Copy-Number Analysis. N Engl J Med. 2017; 376(2):188-189. **3.** Chudova DI, Sehnert AJ, Bianchi DW. Copy-Number Variation and False Positive Prenatal Screening Results. N Engl J Med. 2016; 375(1):97-8.

Q504C V1 817