Sample-specific background correction leveraging vast historical patient cohort maximizes sensitivity of noninvasive prenatal screening

Dale Muzzey, Jeffrey Tratner, Kevin M. D'Auria, Kristjan E. Kaseniit, Dustin H. Hite, Gregory J. Hogan, Chuba B. Oyolu, Carrie Haverty, Kevin R. Haas

South San Francisco, California

Background

Noninvasive prenatal screening (NIPS) using whole-genome sequencing (WGS) identifies fetal chromosomal aneuploidies by measuring larger-than-expected deviations in the depth of sequenced cell-free DNA. A key determinant of test quality, therefore, is an accurate expectation of the disomic depth distribution, such that significant deviations are properly detected. Early implementations of the WGS method calculated expected disomic depth in a simple "one-dimensional" manner, using either the average depth of other "reference" chromosomes in the same sample or the average of the single chromosome of interest across the other 50-100 samples in the batch. These approaches have two drawbacks: (1) limited background observations (e.g., only 23 chromosomes), and (2) insufficient accounting for systematic biases—e.g., GC bias and total depth—between samples and chromosomes. Here we present a strategy for determining an accurate expected disomic depth that is multidimensional, accounts for sources of bias, uses all chromosomes, and utilizes tens of thousands of samples.

Methods

WGS data from 34,050 anonymized clinical NIPS samples was featurized by three values: the depth, the GC-bias profile, and the chromosome-bin-depth medians. Applying K-means clustering to the resulting dataset yielded 200 "background clusters", each with 500 samples. Each newly sequenced sample was next assigned to the closest background cluster (**Figure 1-3**). Aneuploidy calling used the chromosome's z-score, which reflects the extent of depth deviation from the background cluster. A regression model leveraging all chromosomes and all samples of the background cluster predicted the expected disomic depth for the chromosome of interest in the given sample (**Figure 4**). Sensitivity was calculated by developing a modeling framework that uses the fetal-fraction distribution and the relationship between z-scores and fetal fraction to predict the proportion of aneuploid samples expected to be called positive (**Figure 5-6**).

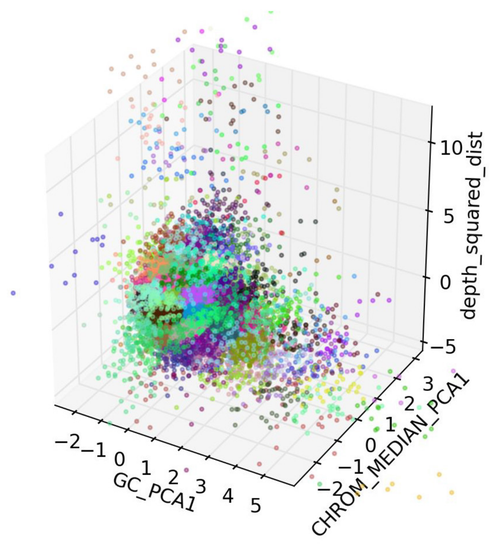

Figure 1: Three dimensional sample clustering.
K-means clustering with 200 centroids of 34,050 NIPT samples identifies groups with similar depth, GC bias, and median chromosome bin count features. Both GC bias and chromosome medians are vectors that were represented by their first principal component. Depth was reduced to a single number as the squared deviation from the population median.

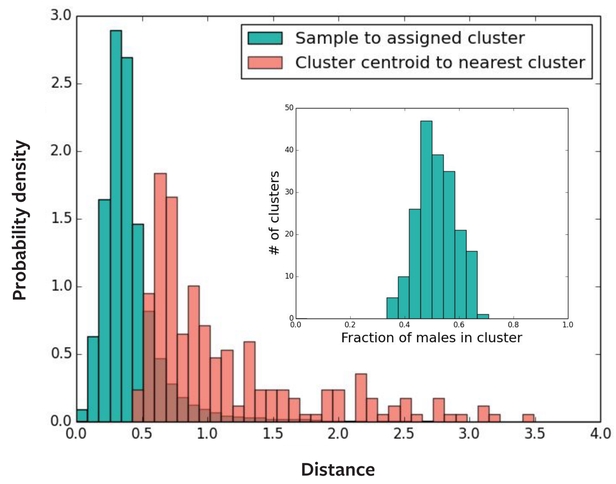

Figure 2: Clustering quality control.
During NIPS calling, samples are assigned to the nearest cluster centroid in 3D space (**Figure 1**), as the comparison to a highly similar background cohort is expected to decrease variance in the calculation of z-scores. The distance between a sample and its nearest centroid (**green**) is generally less than the distance between centroids of the clusters themselves (**red**), suggesting that the cluster centroids were well distributed across the 3D population. Further, clusters have similar representation of males and females, indicating that allosome differences were not a dominant feature or artifact during clustering.

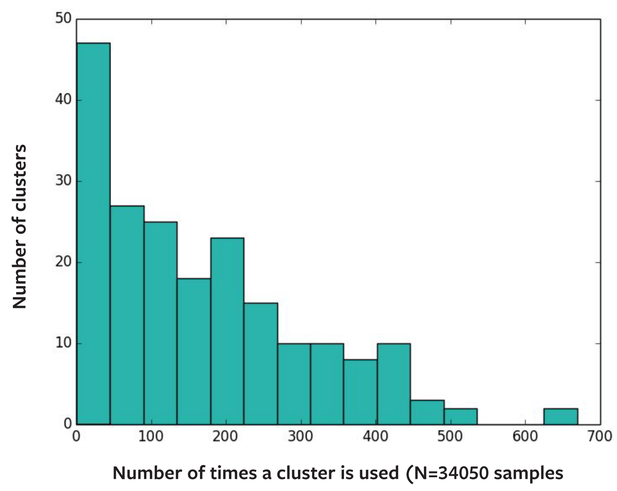

Figure 3: Cluster utilization.
In analysis of 34,050 samples, there is no evidence for artifactual overuse of a minority of clusters. Rather, most clusters had a few hundred samples assigned to them.

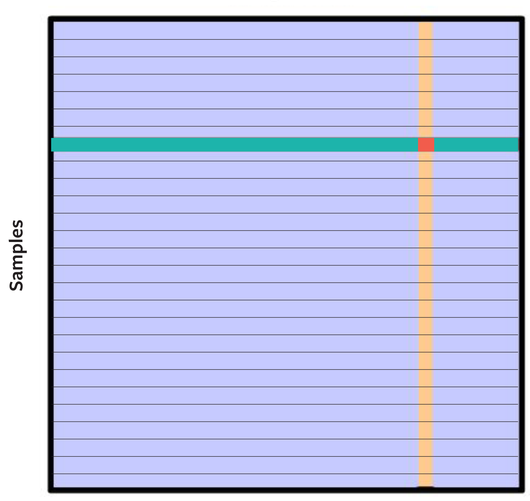

Figure 4: Regression for background prediction.
In z-score calculation for a sample and region of interest (red box), the disomic normalized depth can be predicted by (1) training a regression model where other sample's depth vectors (**purple**) are regressed onto the region of interest (**tan**), and (2) applying the regression to the sample's depth vector (**green**) to yield a predicted depth (**red**). To reduce risk of an aneuploid region in the depth vector compromising the regression, robust regression methods are preferred.

Results

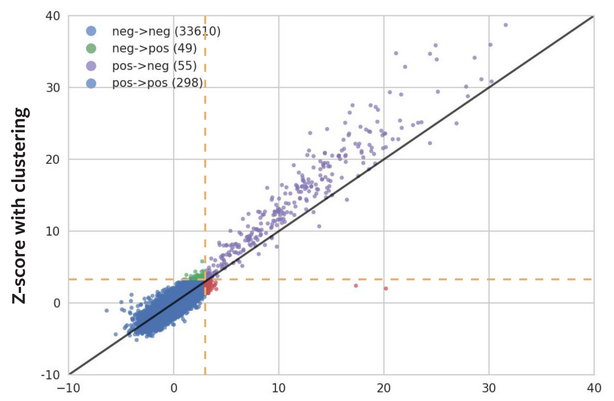

Figure 5: Chromosome 21 z-score improvement with clustering.
Using K-means clustering and robust regression, the z-scores for T21 positive samples increase without a conspicuous change in z-scores for euploid samples (further quantified in **Figure 6**; similar results were observed for chr13, chr18, and chrX). As z-scores scale linearly with fetal fraction, a slope greater than one in the scatter plot is consistent with higher sensitivity at low fetal fraction for the algorithm incorporating clustering and robust regression (y-axis).

Figure 6: Chromosome 21 analytical ROC.
Using an MCMC-based analysis of a Bayesian graphical model of fetal aneuploidies and NIPT, posterior z-score distributions of negative and positive samples were deduced. Scanning different z-score threshold values for emitting positive or negative calls produces true- and false-positive rates, yielding the above ROC curves for algorithms with or without the use of clustering and robust regression.

Conclusions

Our algorithm appreciably increased sensitivity relative to a less-sophisticated algorithm, notably yielding a substantial reduction in the expected false-negative rate for chromosomes 21 (for chr13 and chr18 as well, though not shown here). Though our work showcases sample-specific background correction for NIPS, we expect that harnessing multidimensional data from thousands of historical samples will improve other large-scale clinical genomic assays as well.

View all posters and research at
research.counsyl.com