APPENDIX A.

Results of the Algorithm Training Set

Training Cohort

2,304 euploid samples were used to train the proprietary ploidy model, comprising a mixture of pregnant mothers with male and female fetuses with no selection for any specific fetal fraction range and male control samples.

The entire training cohort had cfDNA extracted, analyzed with the Progenity assay, barcoded for multiplexing, and sequenced using the Illumina NovaSeq 6000 with 272 samples per flow cell.

Data Analysis

Sequencing results are demultiplexed using Illumina's bcl2fastq v2.2 software [1] and aligned using Bowtie 2.0 [2] with customized settings. The analysis pipeline generates a genomic site-specific count file for each sample, which indicates the number of unique molecules counted for each specific site captured by the assay. The ploidy model regresses the site-specific counts against a design matrix composed of covariates obtained from the training data. This model results in estimates of ploidy for chromosomes 13, 18, and 21 (Figures S1-S3) and their standard errors. A T-value statistic is generated for these autosomes within each sample where the null hypothesis for ploidy is 2 (euploidy) for autosomes. This T-value is a test statistic for deviation from the null hypothesis of euploidy as obtained from our proprietary ploidy model. Under the alternative hypotheses of maternal or fetal nullisomy, trisomy, etc., the expectation of the ploidy is a linear function of the maternal ploidy, fetal ploidy and fetal fraction. The model T-value represents the deviation of the estimated ploidy from the euploid expectation divided by the standard error obtained from fitting a generalized linear model. Under certain simplifying assumptions, primarily the absence of overdispersion in the count data, the standard error is approximately proportional to the inverse of the number of reads aligned to each chromosome.

A T-value between 5 and 4 indicate that fetal ploidy is statistically indistinguishable from the euploid distribution. A T-value > 4 indicates that fetal ploidy is significantly higher than the euploid distribution and is classified as trisomy. Similar methods were used to test chromosome X and Y ploidy (data not shown).

Percent Fetal Fraction

The sequencing data from the Progenity assay also contains information from 1-2,000 SNPs with minor allele frequency >0.3%. The allele counts for these sites are used to estimate the percent fetal fraction (% FF_{SNP}). The reliability of the % FF_{SNP} approach was tested by comparing it to the ploidy estimates for ChrY in samples determined to have a male fetus (data not shown). 14/14 non-pregnant samples showed a % FF_{SNP} <1%. The T-values for trisomy cases with % FF_{SNP} displayed a high correlation with the % FF_{SNP} in the training set cohort (Figures S1-S3). Samples that deviate significantly from the expected T-value may indicate confined placental mosaicism, vanishing twin, or another source of heterogeneity in the fetal karyotype.

References

1. See https://support.illumina.com/downloads/bcl2fastq-conversion-software-v2-20.html
2. Langmead B, Salzberg S. Fast gapped-read alignment with Bowtie 2. Nature Methods. 2012, 9:357-359.