Fetal Fraction Evaluation in Non-Invasive Prenatal Screening (NIPS): Supplemental File 1

Sequencing Methods:

DNA was isolated from blood with QIAgen (Hilden, DE) QIAsymphony Circulating DNA Kits, libraries prepared with Bioo Scientific (Austin, TX, USA) custom NEXTflex HT barcodes and Life Technologies Europe BV (Bleiswijk, NL) 5500 SOLID Fragment Library Core Kits, and 51bp single-end sequencing performed on a HiSeq 4000 (Illumina Inc., San Diego, CA, USA: HiSeq 3000/4000 SR Cluster Kits and SBS Kits).

Bioinformatics Methods:

Reads were aligned to hg19 with BWA [1] aln (v0.7.10, options -n 0 -k 0), duplicates marked with Picard (v1.111, http://broadinstitute.github.io/picard/), and evaluated for aneuploidies with WISECONDOR (v2.0.1). FF were determined with DEFRAG (version supplied with WISECONDOR, whole and subset of Y chromosome methods), SANEFALCON (v.1.0), and SeqFF (including ENET and WRSC scores). Since DEFRAG and SANEFALCON use WISECONDOR supplied metrics that exclude duplicate reads and mapping quality <1, we similarly edited sam input files for SeqFF with SAMtools [2] (v0.1.18).

Creating a Titration Series:

To simulate variable FF, seqtk (v1.2, https://github.com/lh3/seqtk) was used to reduce the number of reads in negative controls (non-pregnant females) and ten randomly selected male fetus samples to the number of reads in the lowest depth sample (16 674 452 reads). Each control was randomly paired with a sample. For each pair we randomly mixed control reads and NIPS reads at increments of ten percent:

# Control Reads (% total)	+	# NIPS reads (% total)
16 674 452 (100%)		0 (0%)
15 007 007 (90%)		1 667 445 (10%)
13 339 562 (80%)		3 334 890 (20%)
11 672 116 (70%)		5 002 336 (30%)
10 004 671 (60%)		6 669 781 (40%)
8 337 226 (50%)		8 337 226 (50%)
6 669 781 (40%)		10 004 671 (60%)
5 002 336 (30%)		11 672 116 (70%)
3 334 890 (20%)		13 339 562 (80%)
1 667 445 (10%)		15 007 007 (90%)
0 (0%)		16 674 452 (100%)

Statistical Methods:

Statistics were performed in R [3] (v3.2.3). Comparisons of FF percentages by different bioinformatics tools were evaluated using Spearman correlations. Effects of gestational age, maternal age, weight, BMI, and shipping time on FF percentages were independently studied using logistic models and tested using ANOVA based upon a chi-square test. NIPS determined trisomy 21, 13, and 18 samples were individually compared to non-trisomy samples for FF percentages using logistic models.

Supplemental References:

1. Li H, Durbin R. Fast and accurate short read alignment with Burrows-Wheeler transform. Bioinformatics. 2009;25(14):1754-60.
2. Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, *et al.* The Sequence Alignment/Map format and SAMtools. Bioinformatics. 2009;25(16):2078-9.
3. R Core Team (2017). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL https://www.R-project.org/.