Circulating Fetal Cell-Free DNA Fractions Differ in Autosomal Aneuploidies and Monosomy X

Richard P. Rava,1* Anupama Srinivasan,1 Amy J. Sehnert,1 and Diana W. Bianchi2

BACKGROUND: Noninvasive prenatal testing based on massively parallel sequencing (MPS) of cell-free DNA in maternal plasma has become rapidly integrated into clinical practice for detecting fetal chromosomal aneuploidy. We directly determined the fetal fraction (FF) from results obtained with MPS tag counting and examined the relationships of FF to such biological parameters as fetal karyotype and maternal demographics.

METHODS: FF was determined from samples previously collected for the MELISSA (Maternal Blood Is Source to Accurately Diagnose Fetal Aneuploidy) study. Samples were resequenced, analyzed blindly, and aligned to the human genome (assembly hg19). FF was calculated in pregnancies with male or aneuploid fetuses by means of an equation that incorporated the ratio of the tags in these samples to those of a euploid training set.

RESULTS: The mean (SD) FF from euploid male pregnancies was 0.126 (0.052) (n = 160). Weak but statistically significant correlations were found between FF and the maternal body mass index (r² = 0.18; P = 2.3 × 10⁻⁵) and between FF and gestational age (r² = 0.02; P = 0.047). No relationship with maternal ethnicity or age was observed. Mean FF values for trisomies 21 (n = 90), 18 (n = 38), and 13 (n = 16) and for monosomy X (n = 20) were 0.135 (0.051), 0.089 (0.039), 0.090 (0.062), and 0.106 (0.045), respectively.

CONCLUSIONS: MPS tag-count data can be used to determine FF directly and accurately. Compared with male euploid fetuses, the FF is higher in maternal plasma when the fetus has trisomy 21 and is lower when the fetus has trisomy 18, 13, or monosomy X. The different biologies of these aneuploidies have practical implications for the determination of cutoff values, which in turn will affect the diagnostic sensitivity and specificity of the test.

© 2013 American Association for Clinical Chemistry
required to improve the precision of FF measurement. Yet, even when the statistical mixture model with thousands of polymorphisms is used, the median degree of deviation is 6%–7% (range, 0.6%–22%) (11). Therefore, we sought alternative methods to ascertain this measurement and to understand FF in these samples.

Fan et al. initially suggested that FF in male pregnancies could be measured directly from the MPS tag-count data for the Y chromosome and from the aneuploid chromosome in aneuploid pregnancies (12). Subchromosomal analysis using benign or pathogenic copy number variations may allow extension of this approach for use with any sample from any pregnancy (13). We describe our application of this method to a much larger sample set of newly resequenced samples from our previous clinical study of pregnant women at high risk for fetal aneuploidy (14). We used count data for the X chromosome for male and aneuploid pregnancies (in euploid male fetuses or monosomy X fetuses) and used the specific chromosome present in 3 copies (chromosomes 13, 18, and 21) when fetal aneuploidy (13, 18, and 21) was present. Our primary objective was to examine the relationship between FF and such biological parameters as fetal karyotype, gestational age, maternal age, ethnicity, and body mass index (BMI). We evaluate the practical implications of the observed FF distributions with respect to the diagnostic sensitivity and specificity of detecting fetal aneuploidy noninvasively and propose an optimal approach for maximizing the diagnostic sensitivity and avoiding false-negative results.

Materials and Methods

SAMPLE PREPARATION AND SEQUENCING
To validate our method analytically, we first created artificial mixtures of sheared genomic DNA from a mother (Coriell no. NG09387) paired with DNA from a trisomy 21 male child (Coriell no. NG09394; Coriell Institute for Medical Research). Genomic DNA samples were sheared to a size of approximately 200 bp with the Covaris S2 sonicator (Covaris) in accordance with the manufacturer’s recommended protocols. DNA fragments <100 bp were removed with AMPure XP beads (Beckman Coulter Genomics). Sequencing libraries were generated with TruSeq DNA Sample Preparation v2.0 kits (Illumina) from sheared DNA of the following: maternal DNA only and mixtures of maternal and child DNA consisting of 2.5% to 17.5% child DNA by weight in 2.5% increments. We sequenced samples with single-ended 36 bp reads on an Illumina HiSeq 2000 instrument with TruSeq v3 sequencing chemistry and with 6 samples of the same percentage mixture in a single sequencing lane, each with a unique index bar code.

The MELISSA (Maternal Blood Is Source to Accurately Diagnose Fetal Aneuploidy) trial was a registered clinical trial (http://clinicaltrials.gov; NCT01122524) designed to prospectively determine the accuracy of MPS for detecting whole-chromosome fetal aneuploidy in high-risk pregnancies. The study was conducted at 60 US medical centers under approval by the local institutional review boards. Samples were collected over a gestational age range of 10 to 23 weeks. Written informed consent was obtained from each participant. The enrollment criteria and results of this study have been published (14). Libraries in the MELISSA study were sequenced with TruSeq v2.5 sequencing chemistry on an Illumina HiSeq 2000 instrument, with single-end reads of 36 bp and with 6 samples per lane. For the present study, we resequenced all of the original-sample sequencing libraries from the MELISSA analysis set (n = 532) with TruSeq v3 sequencing chemistry on an Illumina HiSeq 2000 instrument, with single-end reads of 36 bp and with 6 samples per lane. Research laboratory personnel blinded to the fetal karyotype carried out the sequencing in the Verinata Health research laboratory. We collected samples from unaffected pregnancies from local clinics to develop a training set for the MPS TruSeq v3 chemistry used in the current study of the MELISSA clinical samples.

We used the calculations described below to measure FFs from the MPS sequence-tag data for different subgroups within the 532 sequenced samples in the MELISSA study population. These samples included samples from euploid males and all samples from pregnancies with monosomy X and pregnancies with trisomies 21, 18, and 13 (total for this study, 324). We also examined whether FF was significantly affected by various biological parameters, such as gestational age, maternal age, maternal ethnicity, and maternal BMI.

MAPPING, NORMALIZATION, AND ANALYSIS
Sequence reads were aligned to the human genome (assembly hg19, obtained from the UCSC database; http://hgdownload.cse.ucsc.edu/goldenPath/hg19/bigZips/). Alignments were carried out with the Bowtie short-read aligner (version 0.12.5), which allows up to 2 base mismatches during alignment. Only reads that unambiguously mapped to a single genomic location were included. Genomic sites where reads mapped uniquely were counted as tags.

Tag counts for each chromosome within and between sequencing runs were normalized by means of chromosome ratios, as previously described (14, 15). Any chromosome other than chromosomes 13, 18, 21, X, or Y could be used in the denominator in calculating chromosome ratios. The optimal chromosome ratios determined from an independent training set of n un-
affected (U) samples were those that minimized the standard deviation, $\sigma_{Ui}$, within and between sequencing runs for each chromosome, $i$, in the training set. The mean chromosome ratio from the $n$ unaffected (U) samples ($\bar{R}_{Ui}$) for chromosome $i$ is determined as follows:

$$R_{Ui} = \frac{1}{n} \sum_{j=1}^{n} R_{Uj}$$

For any sample that varies by ±1 chromosome (i.e., trisomy or monosomy) from 2 chromosomes (diploid), the ratios are expected to vary with FF according to Eq. 2 (2, 12):

$$R_{Ai} = \left(1 \pm \frac{FF}{2}\right) \bar{R}_{Ui}$$

where $R_{Ai}$ is the ratio for an affected (noneuploid) chromosome $i$. Thus, FF can be calculated directly from the ratios with Eq. 3:

$$FF = 2 \times \left|\frac{R_{Ai}}{R_{Ui}} - 1\right|.$$  

Fan et al. (12) presented a similar equation and used the small number of available Y chromosome tags in a limited number of samples to calculate FF. In the FF calculations presented and discussed below, we used tags from the X chromosome and $R_{AX}$ values to calculate FF for samples from male fetuses, because the number of available tags is much larger, leading to more precise values. $R_{Ai}$ values were used to calculate FFs for aneuploid samples, where $i = 21, 18, 13,$ or X.

For the purposes of detecting or classifying aneuploidy or the sex of the fetus, we evaluated a statistical measure, the normalized chromosome value (NCV$_{ki}$), for each chromosome, $i$, in each sample, $k$, according to Eq. 4 (15):

$$NCV_{ki} = \frac{R_{ki} - \bar{R}_{Ui}}{\sigma_{Ui}},$$

where the SD, $\sigma_{Ui}$, is obtained from a training set of unaffected samples for each chromosome. NCV calculates the number of SDs from the unaffected mean value for each value of $R_{ki}$. In our previous work (15), we demonstrated that NCVs follow a normal distribution, and thus NCV$_{ki}$ is equivalent to a statistical $z$-score. At a given sequencing depth, a fixed NCV cutoff for aneuploidy classification specifies a FF cutoff according to Eq. 5:

$$FF = 2 \times \left|\frac{NCV_{ki} \times \sigma_{Ui}}{R_{Ui}}\right|.$$  

Deeper sequencing (i.e., more tags per sample) can be used to lower the FF cutoff thresholds for each chromosome.

**Statistical Analysis**

All statistical analyses were carried out in Microsoft Excel® for Mac 2011. Population correlations were evaluated with Pearson correlation coefficients, with a $P$ value derived from the Student $t$ distribution of <0.05 considered as statistically significant.

**Results**

**Artificial Mixtures**

We tested the applicability of Eq. 3 on the artificial mixtures from 2.5% to 17.5% of trisomy 21 DNA by weight, in 2.5% increments. A training set was first used to optimize the normalizing chromosome denominators. The mean $R_{Ui}$ and $\sigma_{Ui}$ values were calculated from the data points obtained from the 6 mothers only (i.e., 0% FF). The values for $R_{AX}$ and $R_{A21}$ were then individually determined for each mixture. These values were used in Eq. 3 to compute FFs for the 2 different nondiploid chromosomes, X and 21. Fig. 1 in the Data Supplement (accompanying the online version of this article at http://www.clinchem.org/content/vol60/supplement) shows the plots of FF values calculated from Eq. 3 for the mixtures by using either $R_{AX}$ or $R_{A21}$, which were virtually identical. The excellent agreement between these measurements and the input FF values demonstrate the validity of Eq. 3 for determining the FF from the counted MPS tags.

**MELISSA Study Samples**

Table 1 presents the means and SDs of the FFs for the euploid male and monosomy X fetuses calculated from the X chromosome $R$ values for the MELISSA study samples. For the aneuploid samples, the FFs were calculated from the $R$ values for chromosomes 21, 18, and 13. Fig. 1 shows the relative-frequency histograms for the samples from women carrying euploid male fetuses and those carrying fetuses with trisomies 21 or 18. The histograms for trisomy 13 ($n = 16$) and monosomy X ($n = 20$) are not included because of the small numbers of available samples. As Table 1 shows, the mean FF for trisomy 21 was 9.3% higher than in euploid males. The mean FFs in cases of trisomy 18, trisomy 13, and monosomy X were 29.7%, 28.3%, and 15.9% lower, respectively, than in the euploid male samples.

In a separate analysis that used only the $R_{AX}$ values, the mean FFs calculated from the male aneuploid samples were 0.138 (0.053) for trisomy 21 ($n = 45$) and 0.085 (0.036) for trisomy 18 ($n = 16$). These results were virtually identical to the mean FFs calculated with
For the male fetuses with trisomy 13 (n = 8), the mean FF calculated was 0.046 (0.032), which is lower than the mean FF calculated with all of the trisomy 13 samples. Fig. 2 shows the correlation between the FFs calculated by RA and RA for all trisomy 21 samples and the FFs calculated by RA and RA for all of the trisomy 18 samples. Note that the line in this figure is the unity line (i.e., \( y = x \)), not a regression line. The slopes measured for the trisomy 21 and trisomy 18 samples are 1.01 (\( r^2 = 0.89 \)) and 0.77 (\( r^2 = 0.83 \)), respectively. These independent measurements of the FF and their associated strong correlation demonstrate the robustness of this approach for measuring FF.

For euploid male fetuses, the mean FFs for gestations of <100 days (n = 63) appeared to be slightly lower than for gestations of >100 days (n = 97) (Fig. 3). For euploid males, the respective means were 0.113 (0.046) and 0.134 (0.054), whereas the corresponding mean FFs for the trisomy 21 samples were 0.131 (0.052) (n = 53) and 0.142 (0.054) (n = 37). The FF values for samples from euploid males were weakly but signifi-

### Table 1. FFs obtained from maternal plasma samples in pregnant women carrying fetuses with different karyotypes.

<table>
<thead>
<tr>
<th>Fetal Karyotype</th>
<th>No. (All)</th>
<th>FF from RA</th>
<th>No. (males)</th>
<th>FF from RA</th>
</tr>
</thead>
<tbody>
<tr>
<td>Euploid Males</td>
<td>160</td>
<td>NA</td>
<td>160</td>
<td>0.126 (0.052)</td>
</tr>
<tr>
<td>Trisomy 21</td>
<td>90</td>
<td>0.135 (0.051)</td>
<td>45</td>
<td>0.138 (0.053)</td>
</tr>
<tr>
<td>Trisomy 18</td>
<td>38</td>
<td>0.089 (0.039)</td>
<td>16</td>
<td>0.085 (0.036)</td>
</tr>
<tr>
<td>Trisomy 13</td>
<td>16</td>
<td>0.090 (0.062)</td>
<td>8</td>
<td>0.046 (0.032)</td>
</tr>
<tr>
<td>Monosomy X</td>
<td>20</td>
<td>NA</td>
<td>20</td>
<td>0.106 (0.045)</td>
</tr>
</tbody>
</table>

* FF data are presented as the mean (SD).

b NA, not applicable; RA, ratio for an affected (A) (noneuploid) chromosome; RA, ratio for affected (noneuploid) chromosome X.

c The fetuses with monosomy X were not male. They had only a single X chromosome, however, so we used the relative deficiency of the X sequence compared with the reference to calculated FF.

---

**Fig. 1.** Histogram of the relative frequency of samples at a given FF for maternal plasma samples from women carrying euploid male, trisomy 21, and trisomy 18 fetuses.

Trisomy 13 and monosomy X are not shown because of the small numbers of these samples.
Fig. 2. FFs for all trisomy 21 samples in this study calculated from $R_{AX}$ and $R_{A21}$ values (A) and FFs for all trisomy 18 samples in this study calculated from $R_{AX}$ and $R_{A18}$ values (B). The horizontal data points along the axis for $y = 0.00$ are of samples for female fetuses. The solid diagonal line in both panels is the unity line (i.e., $y = x$). It is not a linear regression fit to the data.
cantly correlated with gestational age ($r^2 = 0.02; P = 0.047$). The trisomy 21 samples also showed a weak correlation of FF with gestational age, but it was not statistically significant ($r^2 = 0.04; P = 0.0068$).

Fig. 4 shows the relationship between maternal BMI and FF for euploid male fetuses. FF was weakly correlated with BMI but with a high statistical significance ($r^2 = 0.18; P = 2.3 \times 10^{-10}$). The results obtained with the trisomy 21 samples were also highly significant ($r^2 = 0.08; P = 0.0001$). The statistical significance of the trisomy 21 data set appears to have been driven primarily by the samples that were obtained at 100 days of gestation ($r^2 = 0.14; P = 0.007$), compared with those samples obtained at >100 days of gestation ($r^2 = 0.05; P = 0.184$). For the euploid male samples, however, both gestational-age subgroups were significantly dependent on the BMI (at <100 days, $r^2 = 0.26$, and $P = 1.64 \times 10^{-5}$; at >100 days, $r^2 = 0.18$, and $P = 1.33 \times 10^{-5}$). No other biological parameters, including maternal age and ethnicity, showed a significant dependence on FF.

Classification for aneuploidy uses NCVs or $z$-scores to set a statistically relevant threshold. This threshold depends on the SD obtained for individual chromosomes, as illustrated by Eq. 4 with an associated FF threshold given by Eq. 5. The CV (i.e., $CV_{13} = \sigma_{13}/R_{13}$) relates $\sigma_{13}$ to the ratio of tags in an unaffected sample being measured and thus provides a relative measure that can be compared among chromosomes with different numbers of tags. Table 2 lists the CVs obtained for chromosomes 13, 18, 21, and X, as well as the FFs at an NCV of 3 for each of the individual chromosomes. An NCV of 3 is a typical cutoff value for classifying a sample as aneuploid. Note that because of the different CVs for each chromosome, different chromosomes have different FF thresholds for determining aneuploidy. Thus, the different FF thresholds have a potential role in the performance characteristics of noninvasive prenatal testing, which is discussed below.

**Discussion**

In this study, we have demonstrated that FF differs as a function of fetal karyotype. Compared with samples from pregnant women carrying euploid male fetuses, the FF is higher when the fetus has trisomy 21 and is lower when the fetus has trisomy 18, 13, or monosomy X. Our results are in agreement with the results for gestational ages of 11–13 weeks reported for an earlier study, which used an independent method (6). This study also showed higher FFs in cases of trisomy 21 and lower FFs in cases of trisomy 18. This finding may be yet another reason why the diagnostic sensitivity and specificity for detecting trisomy 21 is nearly 100% in most published studies. These observations are important, because FF plays an important role in determining
the diagnostic sensitivity and specificity for detecting aneuploidy. As Eq. 5 illustrates, for aneuploidy classification based on a fixed NCV or z-score threshold, a lower SD, allows a sample with a lower FF to cross the classification threshold and be identified as aneuploid (Table 2).

Although the relative FF is lower in trisomy 18 cases, it is offset by the lower CV for chromosome 18. This feature has the practical consequence of allowing smaller amounts of excess chromosome 18 sequence to be detected, which increases the diagnostic sensitivity for detecting mosaicism for trisomy 18, including confined placental mosaicism. Data from both our own clinical laboratory (16) and others (17) support the fact that more “false positives” are detected for chromosome 18 than for the other major aneuploidies. Furthermore, the literature is increasingly showing that some of these “false positive” test results are due to confined placental mosaicism, as proved by karyotyping of placental biopsies (18) or sequencing placental tissue DNA (19).

Given the distributions shown in Fig. 1, <1% of the samples in this study had FFs of <0.03. Thus, for chromosomes with low CVs, the false-negative rate is expected to be extremely small. The low false-negative rates observed in both the published clinical studies and the clinical experience in our commercial laboratory (16) are consistent with this expectation. For FF measurements made with polymorphism methods, larger errors for low FFs may play a role in mischaracterizing samples as having a putatively sufficient FF for clinical determination.

With regard to the demographic variables we have analyzed, our results show that FF is weakly but significantly correlated with maternal BMI. Several other studies have also suggested that the FF is lower in women with a higher BMI (5, 10); however, the weakness of the correlation ($r^2 = 0.18$) does not suggest that it is appropriate to reject samples on the basis of BMI criteria alone. The data shown in Fig. 4 reveal that many of the samples from women with lower BMIs have low FFs. In contrast to the results of Ashoor et al. (6), who reported that women of Afro-Caribbean origin have significantly lower FFs, we observed no dependence of FF on maternal ethnic background. This finding may be due either to differences between the 2 sample sets or to possible biases related to ethnicity in the loci chosen for the Ashoor et al. study.

In summary, we have shown that the FF can be calculated directly from MPS tag-count data with the X chromosome ratios in samples from pregnancies carrying male fetuses and/or with aneuploid chromosome ratios in samples from women carrying fetuses with a chromosome aneuploidy. The method was validated with artificial mixtures. Similar quantitative results were obtained with actual patient samples from a previously performed clinical study, even when the data from 2 different chromosomes were used independently to calculate FF. In contrast to methods that enrich for specific polymorphic loci, this approach does not require separate processing steps to determine FF. As implemented in this study, however, the method cannot be used as an FF quality control step for pregnancies carrying female euploid fetuses. The counting approach may be extendable to samples when the fetus is female by using slightly deeper sequencing to identify genome regions where the mother and fetus have clinically benign copy number differences (13). The major and novel conclusion of this study is that the different biology of each aneuploidy has practical implications for determining cutoff values for aneuploidy classification, which in turn affects the diagnostic sensitivity and specificity of the screening test.
References