Detection of Fetal Subchromosomal Abnormalities by Sequencing Circulating Cell-Free DNA from Maternal Plasma

Chen Zhao,1 John Tynan,1 Mathias Ehrich,2 Gregory Hannum,1 Ron McCullough,1 Juan-Sebastian Saldívar,1 Paul Oeth,1 Dirk van den Boom,2* and Cosmin Deciu1*

BACKGROUND: The development of sequencing-based noninvasive prenatal testing (NIPT) has been largely focused on whole-chromosome aneuploidies (chromosomes 13, 18, 21, X, and Y). Collectively, they account for only 30% of all live births with a chromosome abnormality. Various structural chromosome changes, such as microdeletion/microduplication (MD) syndromes are more common but more challenging to detect. Recently, several publications have shown results on noninvasive detection of MDs by deep sequencing. These approaches demonstrated the proof of concept but are not economically feasible for large-scale clinical applications.

METHODS: We present a novel approach that uses low-coverage whole genome sequencing (approximately 0.2×) to detect MDs genome wide without requiring prior knowledge of the event’s location. We developed a normalization method to reduce sequencing noise. We then applied a statistical method to search for consistently increased or decreased regions. A decision tree was used to differentiate whole-chromosome events from MDs.

RESULTS: We demonstrated via a simulation study that the sensitivity difference between our method and the theoretical limit was <5% for MDs ≥9 Mb. We tested the performance in a blinded study in which the MDs ranged from 3 to 40 Mb. In this study, our algorithm correctly identified 17 of 18 cases with MDs and 156 of 157 unaffected cases.

CONCLUSIONS: The limit of detection for any given MD syndrome is constrained by 4 factors: fetal fraction, MD size, coverage, and biological and technical variability of the event region. Our algorithm takes these factors into account and achieved 94.4% sensitivity and 99.4% specificity.

© 2015 American Association for Clinical Chemistry
the individual bin method by applying a within-sample sliding window approach (12). Their method detected aberrations down to 20 Mb at relatively shallow sequencing coverage (0.15–1.66x). However, an attempt to further increase detection resolution resulted in an increased number of FPs. Another recent study by Chen et al. proposed a binary segmentation and dynamic threshold strategy to detect CNVs >10 Mb using low-coverage sequencing data (13). Out of the 1311 tested cases, 3 affected samples with known G-banding karyotype results were correctly identified. Although the method showed promising results on large CNVs, it is unclear how it will perform on CNVs smaller than 10 Mb. Most recently, Rampasek et al. presented a probabilistic method for detecting fetal CNVs from maternal plasma using a unified Hidden Markov model (14). With use of deep sequencing and a trio analysis, the authors of this study estimated that CNVs bigger than 400 kb could be detected at 90% sensitivity if the fetal fraction is sufficiently high (13%).

In this report, we present a novel approach which uses shallow whole-genome sequencing (approximately 0.2 X coverage) data to detect genome-wide microdeletions or microduplications (MDs) (also referred to in this report as “events”) without requiring a priori knowledge of an event’s location. Unlike previously published approaches, our focus is on improving the normalization method to remove as much sequencing “noise” as possible upfront. Following that, a rigorous statistical method is applied to detect consistently increased or decreased copy number regions using the likelihood ratio statistic, and it can pinpoint the change point precisely. Many studies have demonstrated its superior performance over other methods (21, 22). Although powerful, CBS tends to overly partition the genome when the signal-to-noise ratio is low. To compensate for oversegmentation we further applied a segment-merging algorithm similar to the one described by Willenbrock and Fridlyand (22).

CIRCULAR BINARY SEGMENTATION

Our goal is to develop an algorithm for detection of MDs regardless of their genomic coordinates. As such, the proposed algorithm needs to search for such events in a non-targeted fashion. To do so, we apply the circular binary segmentation (CBS) method which has been extensively used by the array comparative genomic hybridization (CGH) community for CNV detection (19, 20). CBS works by iteratively partitioning a chromosome into equal copy number regions using the likelihood ratio statistic, and it can pinpoint the change point precisely. Many studies have demonstrated its superior performance over other methods (21, 22). Although powerful, CBS tends to overly partition the genome when the signal-to-noise ratio is low. To compensate for oversegmentation we further applied a segment-merging algorithm similar to the one described by Willenbrock and Fridlyand (22).

QUANTIFYING THE STATISTICAL SIGNIFICANCE OF DETECTED EVENTS

After CBS, each chromosome was partitioned into regions of equal copy number. To quantify the statistical significance of the detected event, we developed 2 approaches. First, for each chromosome, the segment with the largest area was treated as a potential MD, and its \( z \)-score, \( z_{\text{CBS}} \), was calculated by comparing the summed bin count with respect to a reference set of samples in the same region (8). The whole-chromosome \( z \)-score, \( z_{\text{CHR}} \), was also calculated in the same way. Second, we calculated the log odds ratio (LOR) to quantify the likelihood of the event being true at the measured fetal fraction, \( f \) (see the LOR Section 1 in the Data Supplement that accompanies the online version of this article at http://www.clinchem.org/content/vol61/issue4).

DECISION TREE FOR GENOME-WIDE ANEUPLOIDY OR MD DETECTION

For each chromosome, 4 statistics can be inferred:

1. The \( z \)-score and LOR for the largest segment partitioned by the CBS process: \( z_{\text{CBS}} \) and LOR_{CBS}.

Methods

DATA NORMALIZATION

To remove sequencing bias, we developed a 2-step normalization approach. Briefly, reads were aligned to hg19, allowing for only perfect matches within the seed sequence using Bowtie 2 (15). The genome was then partitioned into 50-kbp nonoverlapping bins and the raw count for each bin was determined. After binning, regions with high variability or low mappability were excluded according to a previously established method (8).

To normalize the 50-kbp raw bin count, GC bias was removed by employing a LOESS-based correction for a sample-wise correction, similar to the one described in Alkan et al. (16). Then, principal component analysis (PCA) was used to remove higher-order artifacts for a population-based correction (17, 18). To train the PCA normalization step, we used a data matrix comprising LOESS normalized bin counts from 1000 female pregnancies and transformed the data matrix into the principal component space. Then for a new sample, we built a linear regression model between its LOESS normalized bin count and the top principal components from the training set. The residuals of this model serve as final normalized values for this given sample. Intuitively, the top principal components represent noise commonly seen in euploid samples, and therefore removing them can effectively improve normalization. The combination of the last 2 steps replaces the older, region-based normalization we previously published (8).

2. The whole-chromosome z score and LOR: $z_{\text{CHR}}$ and $\text{LOR}_{\text{CHR}}$.

Note that the largest segment is defined to be the segment with the largest area (height × length) out of all the segments on a given chromosome. For example, in Fig. 1B, there are 2 segments, and the second segment has the largest area. The latter represents the most significant finding on the chromosome of interest. On the basis of these 4 statistics, we propose the following decision tree method for aneuploidy or MD detection:

1. A chromosome is classified to be trisomy or monosomy if:
   
   \[ |z_{\text{CHR}}| \geq C, \text{LOR}_{\text{CHR}} > 0, \text{and } |z_{\text{CBS}}| \geq a|z_{\text{CBS}}| \]

2. A chromosome is classified to have microdeletion/microduplication if:

   It is not a trisomy or monosomy, and $|z_{\text{CBS}}| \geq C$, $\text{LOR}_{\text{CBS}} > 0$,

where $C$ is a predefined z-score cutoff that controls the tradeoff between sensitivity and specificity. In this report, we set $C$ to be 3 for chromosome 21 and 3.95 for all the other chromosomes. The threshold was selected on the basis of our previous study on trisomy detection, in which more than 99.9% sensitivity and specificity were achieved for all 3 major aneuploidies (8). The comparison $|z_{\text{CHR}}| \geq a|z_{\text{CBS}}|$ allows one to distinguish a whole chromosome event from a subchromosomal event; in practice, we find that $a$ between 0.6 and 0.8 works well. The value was determined by assessing its impact on the detection of in silico–created MDs at different sizes and fetal fractions. See the online Supplemental Material for detailed descriptions of the selection of $a$. Because there might be multiple MD events on the same chromosome, the decision tree process moves to the next-largest segment on the chromosome until there are no further significant segments.

**GENOMIC AND PLASMA DNA SAMPLE PREPARATION**

We conducted a blinded clinical evaluation study to test the performance of the proposed MD detection algorithm. The study comprised 2 parts: an analytical validation using genomic DNA (gDNA) mixtures and a clinical validation using plasma samples.

Forty-five gDNA samples (130 μL of 30 ng/μL) were processed for library preparation, and sequencing libraries were created as previously described by Jensen et al. (8). gDNA mixture models were created by normalizing gDNA and nonpregnant female plasma (NPP) DNA libraries to 1.6 nmol/L. A starting mix of 20% gDNA library was created in a background of NPP library and subsequently used to create 17.5%, 15%, 12.5%, 10, 7.5%, and 5% mixtures by serial dilution. These were then multiplexed, clustered, and sequenced on the HiSeq2000 for 36 + 7 cycles in a 12-plex format (see online Supplemental Material for detailed protocols).

A total of 183 clinical plasma samples were collected using Investigational Review Board–approved clinical protocols. Samples with abnormal G-band karyotype results collected through these clinical protocols were selected for sequencing to determine detection rates for MDs. Euploid samples were chosen on the basis of normal karyotype reports using either fluorescence in situ hybridization (FISH) or G-banding methods. Library preparation and sequencing were conducted similarly to that of gDNA samples.

Because the karyotype outcome was based on FISH or G-banding, we anticipated that some of the “euploid” samples contained MD events. In case of putative FP results at 12-plex sequencing, we further sequenced the library at uniplex to confirm its validity. If the results were confirmed by uniplex sequencing, we counted the results as true positives (TPs) instead of FPs. Previous
Results

NORMALIZATION
Sequencing bias can cause nonuniform coverage in data, which complicates downstream CNV analysis (23, 24). Fig. 1 shows the normalization result on a 12-plex sequencing sample. Without normalization, many regions would aberrantly appear as depletions or duplications (Fig. 1A). After normalization, the microduplication event became more evident (Fig. 1B).

MD LIMIT OF DETECTION
For subchromosomal abnormality detection, it is of fundamental importance to understand the limit of detection (LoD) for MDs. Four factors affect the LoD: fetal fraction ($f$), size of the event, coverage, and the biological and technical variability of the event region. The first 3 factors are easy to understand: it is easier to detect an event with higher fetal fraction, larger size, and higher sequencing coverage. The fourth factor is related to the fact that certain regions are more variable than others due to various factors (such as GC bias, repetitive elements, and mapping ability) and are harder to detect. Fig. 2 illustrates the impact of the 4 factors. Through a simulation study, we observed that at 0.2X coverage, the sensi-
tivity difference between the proposed decision tree method and the theoretical limit was <5% for MDs ≥9 Mb (see online Supplemental Fig. 1).

ANALYTICAL VALIDATION USING GENOMIC DNA MIXTURES
A blinded gDNA study was conducted to test the performance of the proposed decision tree method on several selected syndromes with high clinical relevance. Due to the rarity of these syndromes, we created a gDNA mixture model system for analytical validation. Forty-five gDNA samples from individuals with DiGeorge, Cri-du-chat, Prader-Willi, Angelman, or 1p36 deletion syndromes were obtained from Coriell Cell Repositories or from CombiMatrix Diagnostics. The source, disease, and karyotype results of the gDNAs are provided in online Supplemental Table 1. In total, 17 were positive for Prader-Willi or Angelman syndrome, 14 were positive for Cri-du-chat, 13 were positive for DiGeorge, and 1 was positive for 1p36. All 45 gDNA samples were sequenced once from a 20%–7.5% mixture, and they were sequenced twice at a 5% mixture. A total of 360 gDNA mixtures were sequenced.

Fetal fraction plays a crucial role in the detectability of any given MD. We measured the observed fetal fraction by chromosome Y on male samples to confirm the planned titration levels. Online Supplemental Fig. 2 shows that the measured values agreed well with the planned titrations, with slight overdilution. The median measured titration concentrations were 18.3%, 16.0%, 14.0%, 11.6%, 9.3%, 7.1%, and 4.7% respectively.

Before the gDNA unblinding, we predicted the sensitivity to detect each of the syndromes using the LoD framework. Briefly, the genomic location and size definition of each syndrome is acquired from the DECIPHER database (25). A syndrome-specific microdeletion or microduplication is created by sampling the observed titration values shown in online Supplemental Fig. 2. After unblinding, we found that the observed sensitivity for detection of each of these 3 syndromes fell within the 95% CI, with the exception of the 4.7% gDNA fractions (Fig. 3). Such high concordance suggests that the LoD framework is predictive of the actual performance. We note that the discrepancy at 4.7% can be mainly attributed to the size difference between the database size and the tested cases, and that such differences are most obvious at low fetal fractions. For example, the Cri-du-chat syndrome has a size of 12.52 Mb according to the DECIPHER database (25), and our testing cases ranged from 8 to 20 Mb, which made the detection easier at low fetal fractions.

For the 1p36 sample, we were able to detect at 9.3% titration and above. An example is shown in online Supplemental Fig. 3. We note that this 1p36 sample has a size of only approximately 3 Mb, which is much smaller than its database definition of 12.83 Mb (25). In real life applications, we expect the overall sensitivity to be higher.

Within this set of gDNA mixtures each gDNA had a combination of karyotype and microarray results confirming the presence of an MD event for 1 of the 5 syndromes queried in this study. We used this information to determine specificity for the nonaffected syndrome regions in each of the gDNA samples. There was no false-positive result by the decision tree algorithm for any of the 360 gDNA mixtures. An overall specificity of 100% (95% CI, 98.7%–100%) was achieved in the gDNA study. The high specificity value is paramount for good positive predictive values, given the low prevalence of these conditions.

CLINICAL VALIDATION USING PLASMA SAMPLES
To further test the performance of the proposed decision tree method in a genome-wide fashion, we conducted a blinded study using maternal plasma DNAs with
matched karyotype. For this study, 183 samples were analyzed using 12-plex sequencing. Fetal fractions were measured by the same method used in our current clinical practice (26, 27).

On the completion of sequencing, 1 (1) affected sample and 4 (4) euploid samples failed QC and were subsequently excluded from analysis. Hence, we had a total of 178 plasma samples with a median fetal fraction of 9.2% (range, 2.8%–24.9%). We did not impose any QC cutoff for fetal fraction. According to the karyotype table, the expected MD sizes ranged from 3 to 40 Mb (see online Supplemental Table 2).

The algorithm detected 13 of the 16 affected samples. Of the 3 supposed FNs, one was expected to show trisomy 8, one was expected to show a deletion of chromosome 4q34, and the third one was expected to show unbalanced translocation between chromosome 12 and chromosome 19. Subsequent detailed review of the first karyotype reports indicated the trisomy 8 sample was a low-level mosaic sample (approximately 8% cells affected). Because this low level of mosaicism would not have met the sample inclusion criteria, we excluded this sample from the remaining analysis. Review of the second karyotype report indicated that the mother also carried the same chromosome 4q34 deletion as the fetus. Our algorithm did successfully detect the event and correctly called it as a maternal event. Because we focused only on de novo fetal events, this sample was also excluded from the analysis. Review of the third sample did not reveal any complex karyotype information, and the negative algorithm outcome was probably a result of a low fetal fraction of 4.8%. We note that this FN sample was still nondetectable at uniplex sequencing.

For the remaining 162 presumed euploid samples, 5 MD events were detected. Additionally, in 2 of the 16 affected samples, we detected 2 events, whereas the karyotype report indicated only 1 event. Because the study collection protocol provided only the karyotyping result of amniocytes for these samples, no pure fetal tissue was available for sequencing or array confirmation. Therefore, to confirm the validity of our findings, we resequenced at higher coverage (uniplex) all the putative FP samples. Of these, 1 sample led to a different outcome based on the higher coverage sequencing and was deemed an FP result. The remaining 4 samples produced the same outcome when sequenced in uniplex. The MD events detected in these 4 samples included MDs of 2.85, 3.25, 22.15, 23.55, 25, 32.45, and 42.3 Mb. Fig. 4 illustrates the comparisons between 12-plex and uniplex results for (A) a confirmed TP sample by uniplex, (B) a confirmed FP by uniplex, and (C) the maternal deletion detected by both 12-plex and uniplex.

Table 1 summarizes the performance of our algorithm on 176 plasma samples (excluding the mosaic trisomy 8 and maternal chr4q34 deletion). Our algorithm attained a 94.4% sensitivity (95% CI, 70.6%–99.7%), and a 99.4% specificity (95% CI, 95.96%–99.97%) for the plasma samples. The CI was calculated on the basis of the proportion test with continuity correction (28). We want to emphasize again that 4 out of 5 putative FP samples were confirmed by uniplex sequencing and subsequently counted as TP samples. Because we did not have pure tissue samples for further invasive arrayCGH confirmation, the interpretation of specificity was limited to the signal detection perspective. We note that other researchers have also observed discrepancies between sequencing and the traditional karyotyping methods (10).

Discussion

The methodology described in this report for noninvasive detection of fetal subchromosomal abnormalities relies on a simple yet fundamental assumption: namely, that any plasma sample pertaining to a mixture of maternal DNA and fetal DNA in which the latter component has a genomic abnormality will lead to sequencing results similar to those of euploid samples over the entire genome except for the region where the fetus is affected. Moreover, for the affected region, the sequencing results will show a deviation in sequenced reads that is directly proportional to the fetal fraction but described by the same variability as observed in a euploid population. We have tested this hypothesis over a set of 16 plasma samples from women carrying affected fetuses and have proven it to be consistent with these assumptions. This result has warranted the exploration of the performance of this methodology over a specific set of genomic abnormalities that were experimentally created to mimic the fetal and maternal ccf DNA mixture as present in maternal plasma.

Compared to our previous clinical validation efforts (4, 6–8), in this study we have performed an analytical, genome-wide validation of this bioinformatics algorithm. None of the abnormalities investigated were statistically powered individually in a classical sense. Given the low prevalence of many of these abnormalities and given the fact this is a genome-wide assay, it would have been impossible to collect a sufficiently large number of affected sample for each possible abnormality. This kind of difficulty is not unlike the one encountered in the development of genome-wide chromosomal array CGH invasive prenatal diagnosis (29). The performance of the chromosomal microarray analysis has been evaluated through larger cohort studies, such as reported by Wapner et al. in 2012 (30). Based on the findings of this particular study, the American College of Obstetricians and Gynecologists and the Society for Maternal-Fetal Medicine have recently issued the recommendation that in cases where in which one or more major fetal structural abnormalities have been
identified by ultrasonography and the prospective mother is undergoing invasive prenatal diagnosis, chromosomal microarray analysis can replace the need for fetal karyotyping (31).

A complexity associated with the reporting of genome-wide abnormalities is the increased prevalence of maternal-only and maternal-and-fetal abnormalities. The analyte queried by noninvasive testing is always a mixture of maternal and fetal DNA and although the maternal copy number for chromosomes 21, 13, and 18 is highly likely to be known in advance, this is not necessarily the case for other genomic abnormalities. A recently published study (32) demonstrated how a significant fraction of the NIPT FP results for chromosome X are due to mosaic maternal abnormalities. The solution proposed in the aforementioned report is to process the maternal-only component (buffy coat) for many if not all the samples undergoing NIPT. Currently our algorithms detect and report any genomic abnormality (of sufficient signal-to-noise ratio), regardless of whether its origin is fetal, maternal, or a combination. An issue common to

**Table 1. Contingency table for the plasma study.**

<table>
<thead>
<tr>
<th>Known genome abnormality</th>
<th>No known genome abnormality</th>
</tr>
</thead>
<tbody>
<tr>
<td>TP</td>
<td>17</td>
</tr>
<tr>
<td>TN</td>
<td>1</td>
</tr>
<tr>
<td>Not reported</td>
<td>0</td>
</tr>
</tbody>
</table>

*The decision tree algorithm correctly detected 17/18 positives and 156/157 negatives. Note 4 out of 5 putative FP samples at 12-plex were subsequently confirmed at uniplex and were counted as TP samples here.*

Fig. 4. Comparison between 12-plex (left panel) and uniplex (right panel) results. The plots follow the same convention as Fig. 1. The uniplex results show significantly improved signal to noise ratio compared to the 12-plex results. (A), A TP event detected at 12-plex and confirmed by uniplex sequencing. (B), A FP event detected at 12-plex but not by uniplex sequencing. (C), The maternal deletion detected by both the 12-plex and uniplex sequencing.
Noninvasive Detection of Fetal Microdeletions

both the genome-wide array CGH invasive testing as well as sequencing-based NIPT is that both have the potential for findings of unknown significance. A conservative approach is to limit the findings to a subset of categories of high clinical relevance. This was indeed the strategy behind the selection of the 6 conditions we have investigated in our study by means of sheared gDNA + ccf DNA mixtures. This is only a temporary and intentional limitation: going forward, whole-genome sequencing–based noninvasive testing will have the benefit of the improved knowledge of the human genome as derived from the findings of microarray analysis.

References