

## Noninvasive detection of fetal trisomy 21 by sequencing of DNA in maternal blood: a study in a clinical setting

Mathias Ehrich, MD; Cosmin Deciu, MSc; Tricia Zwiefelhofer; John A. Tynan, DPhil; Lesley Cagasan, MSc; Roger Tim, DPhil; Vivian Lu; Ron McCullough, DPhil; Erin McCarthy; Anders O. H. Nygren, DPhil; Jarrod Dean; Lin Tang, DPhil; Don Hutchison, MSc; Tim Lu, DPhil; Huiquan Wang, DPhil; Vach Angkachatchai, DPhil; Paul Oeth, MSc; Charles R. Cantor, DPhil; Allan Bombard, MD; Dirk van den Boom, DPhil

**OBJECTIVE:** We sought to evaluate a multiplexed massively parallel shotgun sequencing assay for noninvasive trisomy 21 detection using circulating cell-free fetal DNA.

**STUDY DESIGN:** Sample multiplexing and cost-optimized reagents were evaluated as improvements to a noninvasive fetal trisomy 21 detection assay. A total of 480 plasma samples from high-risk pregnant women were employed.

**RESULTS:** In all, 480 prospectively collected samples were obtained from our third-party storage site; 13 of these were removed due to insufficient quantity or quality. Eighteen samples failed prespecified assay quality control parameters. In all, 449 samples remained: 39 tri-

some 21 samples were correctly classified; 1 sample was misclassified as trisomy 21. The overall classification showed 100% sensitivity (95% confidence interval, 89–100%) and 99.7% specificity (95% confidence interval, 98.5–99.9%).

**CONCLUSION:** Extending the scope of previous reports, this study demonstrates that plasma DNA sequencing is a viable method for noninvasive detection of fetal trisomy 21 and warrants clinical validation in a larger multicenter study.

**Key words:** circulating cell-free fetal DNA, massively parallel shotgun sequencing, maternal blood, NIPD, noninvasive prenatal diagnosis

Cite this article as: Ehrich M, Deciu C, Zwiefelhofer T, et al. Noninvasive detection of fetal trisomy 21 by sequencing of DNA in maternal blood: a study in a clinical setting. *Am J Obstet Gynecol* 2011;204:205.e1-11.

Trisomy 21 is the most common chromosomal aneuploidy in live born infants. The overall incidence of trisomy 21 is approximately 1 in 800 births in the general population,<sup>1</sup> but this risk increases to 1 in 35 term births for women 45 years of age.<sup>2-4</sup> Advanced maternal age is only one factor contributing to increased risk. When other factors, such as positive serum screening results, fetal ultrasound abnormality, or family history are included, the odds of being affected given a positive result of

### ★ EDITORS' CHOICE ★

Down syndrome can be as high as 1 in 9 using the integrated test.<sup>5</sup> For women in this high-risk group, an invasive diagnostic procedure is currently the only way to confirm the diagnosis of trisomy 21, commonly by means of a fetal karyotype. Although the safety of the invasive procedures, specifically genetic amniocentesis and chorionic villus sampling (CVS), has improved greatly since their introduction, there remains a well-rec-

ognized risk of iatrogenic fetal loss. Tests that could better identify those women who would most benefit from confirmatory invasive diagnostic tests are of great public health interest. Since the initial seminal work by Merkatz et al,<sup>6</sup> continuous efforts have focused on increasing the specificity of primary screening methods; eg, by including more serum protein markers or through the addition of ultrasound findings suggestive of fetal aneuploidy. Consequently, screening tests have greatly improved in their clinical sensitivity and specificity over the last 2 decades. These developments have also led to contemporary testing programs involving a variety of potential screening algorithms.<sup>7</sup> Accurate gestational dating through ultrasound is a critical element to achieve high accuracy, but may not be readily available to all pregnant women.

A new approach to detect fetal aneuploidy analyzes fetal DNA itself rather than the surrogate biochemical or ultrasound markers in current maternal serum screening protocols. In 1997, Lo et al<sup>8</sup> reported that circulating cell-free (ccf) fetal (cff) DNA is present in the

From Sequenom Inc (Drs Ehrich, Cantor, Bombard, and van den Boom), and Sequenom Center for Molecular Medicine LLC (Mr Oeth, Mr Deciu, Ms Zwiefelhofer, Ms Cagasan, Ms V. Lu, Ms McCarthy, Mr Dean, Mr Hutchison, and Drs Tynan, Tim, McCullough, Nygren, Tang, Lu, Wang, and Angkachatchai), San Diego, CA.

Received Sept. 30, 2010; revised Dec. 28, 2010; accepted Dec. 28, 2010.

Reprints: Dirk van den Boom, DPhil, Sequenom Inc, 3595 John Hopkins Ct., San Diego, CA 92121. [dvandenboom@sequenom.com](mailto:dvandenboom@sequenom.com).

**DISCLOSURE:** All authors of this article are employees and shareholders of Sequenom, Inc. or its subsidiaries, and therefore, a potential conflict of interest exists.

0002-9378/free • © 2011 Mosby, Inc. All rights reserved. • doi: 10.1016/j.ajog.2010.12.060



For Editors' Commentary, see Table of Contents

origin ranges between 2% and 40% with a mean around 10% of the total ccf DNA across varying gestational ages.<sup>9-12</sup> The ccf DNA is cleared from the maternal bloodstream within hours after birth; thus, misdiagnosis from carryover contamination from a previous pregnancy is unlikely.<sup>13</sup> A noninvasive ccf method for prenatal Rhesus D testing in Europe has already been widely adopted.<sup>14</sup>

In comparison to Rhesus D testing, aneuploidy detection from ccf DNA is far more challenging. In principle, aneuploidy detection could be enabled through a variety of methods including the analysis of single nucleotide polymorphisms,<sup>15</sup> DNA methylation,<sup>16,17</sup> or fetally expressed RNA transcripts.<sup>18,19</sup> The most convincing data to date for a generally applicable test, however, have been generated through massively parallel shotgun sequencing (MPSS) of ccf DNA. Two groups have independently shown that MPSS can unambiguously identify plasma samples from women carrying a trisomy 21 fetus<sup>20-23</sup> compared to samples from women with euploid fetuses. These studies were performed with small numbers of clinical samples and, while these preliminary results are very promising, the true clinical performance remains to be established. As originally described in 2008, the overall cost of a sequencing-based test was prohibitive in terms of potential deployment in clinical practice. However, next-generation sequencing methods such as MPSS are rapidly evolving with concomitant declines in reagent and instrument costs.

We have implemented several process improvements in MPSS for noninvasive aneuploidy detection using ccf DNA. These modifications provide an affordable testing procedure with the potential for widespread utilization. Because such a test, first and foremost, has to be safe and efficacious we designed a blinded study that tested a total of 480 plasma samples collected from pregnant woman at high risk for fetal aneuploidy.

## MATERIALS AND METHODS

### Study design

The study was set up to include at least 40 trisomy 21 samples, a design chosen to

of 91% when all trisomy 21 cases are correctly identified. We matched trisomy 21 samples with euploid samples at a ~1:11 ratio, slightly higher than the more typical prevalence in a high-risk group of 1 in 15.

Patients at increased risk for fetal Down syndrome and other chromosomal aneuploidies were asked to participate in this prospective study. Risks included a positive serum biochemical screening test; advanced maternal age ( $\geq 35$  years at the estimated date of delivery); a fetal ultrasound finding suggestive of Down syndrome; or a personal/family history of Down syndrome. Patient informed consent was obtained for peripheral blood sampling and for the inclusion of karyotype results from an already scheduled, subsequent invasive diagnostic procedure. Fetal karyotypes or quantitative fluorescent PCR results were obtained as part of regular clinical care on either CVS or genetic amniocentesis samples. These data were unknown to the investigators prior to unblinding. The sample demographics were representative for pregnant women at high risk for fetal trisomy 21 (Table 1).

Samples were blinded to the investigators and prospectively collected, processed, and stored at an independent, contracted, third-party location (Biostorage Technologies Inc [BST], Indianapolis, IN). All information was kept within an independent, third-party database (Pharmaceutical Research Associates Inc [PRA], Raleigh, NC). A total of 480 samples were requested from PRA and provided by BST for analysis at Sequenom Center for Molecular Medicine, San Diego, CA. Karyotype results were unknown to the investigators and data analysts until after completion of all sample testing and submission for review. The MPSS results were sent to an independent, third-party biostatistician who had all clinical information including confirmatory karyotypes. The data were matched and unblinded by this third-party biostatistician and the concordance of the results was reported.

### Sample collection

For the study presented here, samples

in the treatment of patients undergoing invasive prenatal diagnosis by CVS (first trimester) and genetic amniocentesis (second trimester) and, for some of the cases, from pregnancy termination centers. Eight samples were collected for research purposes under Food and Drug Administration approval (FDA Establishment Identifier no. 3005208435). All remaining samples were collected under institutional review board (IRB) approval (Western Institutional Review Board [WIRB] no. 20091396, WIRB no. 20080757, Compass IRB no. 00351). All samples, demographics, and karyotype results were completely blinded to the laboratory investigators by the third-party clinical research organization (PRA) and the BST facility. Patients were approached during their genetic counseling sessions and, if they gave their informed consent, the study protocol dictated that phlebotomy was to be performed prior to their invasive procedure. The vast majority of samples were collected after August 2009 and none were collected before May 2009; therefore, the oldest samples in the study were ~10 months old. Samples were all collected at specifically contracted processing centers operating under study-specific protocols. None of the samples were obtained and analyzed as fresh samples; ie, all were processed and frozen before shipment to the central, independent biostorage facility (<http://biostorage.com/>; a full description of the independent nature of this widely used biostorage company is detailed on their World Wide Web site).

All samples were collected and processed under the same protocol: 10 mL of maternal whole blood was drawn into an EDTA-K2 spray-dried Vacutainer (Becton Dickinson, Franklin Lakes, NJ), stored, and transported to the processing laboratory on wet ice. Within 6 hours of the blood draw, the maternal whole blood was centrifuged (Eppendorf 5810R plus swing out rotor) chilled (4°C) at 2500g for 10 minutes and the plasma was collected. The plasma was centrifuged a second time (Eppendorf 5810R plus fixed angle rotor) at 4°C at 15,000g for 10 minutes. After the second

pellet that formed at the bottom of the tube and distributed into 4-mL plasma bar-coded aliquots. In this study, only a single 4-mL plasma aliquot from each patient was used for DNA isolation.

### MPSS aneuploidy detection

DNA was prepared from 4 mL of maternal plasma. The short lengths of ccfDNA afford direct use in preparing the libraries of DNA fragments that were sequenced. In practice, 4 different libraries each containing a synthetic oligonucleotide sequence as a bar code were mixed and analyzed together (multiplexing). The bar code revealed which library each sequence read represented. Eight separate mixtures of 4 libraries were analyzed in parallel. One MPSS process required about 2 days and yielded 36 bases of sequence from each DNA fragment. Approximately 5 million 36-base fragments were sequenced from each library. These represented about 6% of the human genome in each sample. As is standard in MPSS, the 36-base reads were processed to exclude poor-quality data and then matched to a reference human genome to determine their chromosome origin. The fraction of reads is proportionate to chromosome size. Thus, typically 8.5% of all reads are from chromosome 1, while only about 1.2% are from chromosome 22.<sup>20,21</sup>

A fetus with trisomy 21 contributes additional genetic material to the total pool of ccfDNA. Consequently, in comparison to women carrying a euploid fetus, a slightly larger contribution of sequence reads mapping to chromosome 21 is observed in a plasma sample of a woman carrying a fetus with Down syndrome. CcfDNA in plasma from a pregnant woman with a euploid fetus shows an average 1.35% of all aligned sequence reads located on chromosome 21. A variety of analytical methods have been published to detect an overabundance of genetic material from chromosome 21 in trisomic pregnancies.<sup>20,21,23</sup> These use some form of normalization to calibrate the results against a known set of euploid reference samples. Contributions greater than the reference range are then indicative of additional genetic material from

be interpreted as a fetal trisomy 21. In this study, a modification of a method used by Chiu et al<sup>20</sup> was used for classification. Prior to the main study a set of known euploid reference samples was used to calculate the mean and standard deviation (SD) of the representation of chromosome 21 (percentage of reads obtained from chromosome 21). Then, for every test sample, the distance, measured in SD, from the mean in the euploid reference set was calculated. A fixed cutoff of 2.5 SD was used to identify samples with an overrepresentation of chromosome 21 material.

### Assay design

Compared to previously published studies,<sup>20,21</sup> 3 important modifications were made to the sequencing protocol. We used custom purified enzymes in the library generation process to achieve a reduction in assay cost. We employed the latest sequencing biochemistry available for the GAIIX sequencer (Illumina Inc, San Diego, CA) in combination with the manufacturer's analysis software CASAVA version 1.6. These changes increased the number of sequence reads from approximately 13 to 20 million per lane. We also used indexing primers during library amplification to allow analysis of multiple samples in a single sequencing reaction ("multiplexing" vs "monoplexing"). In this study, 4 samples were analyzed per lane ("4-plex" or "tetraplexing"), which equates to approximately 3 to 5 million available sequence reads per sample. The combination of these modifications enabled 4 times higher throughput at about 4 times lower cost.

### DNA extraction

The Qiagen ccf nucleic acid kit (Qiagen, Hilden, Germany) was used according to the manufacturer's specifications. The resulting DNA was eluted in 55  $\mu$ L of buffer AVE (part of the Qiagen kit).

### Quality control of extracted DNA

The quantity of the extracted DNA was determined with an assay that uses simultaneous quantification of fetal and total ccfDNA. This fetal quantifier assay (FQA) was recently published<sup>10</sup> and uses

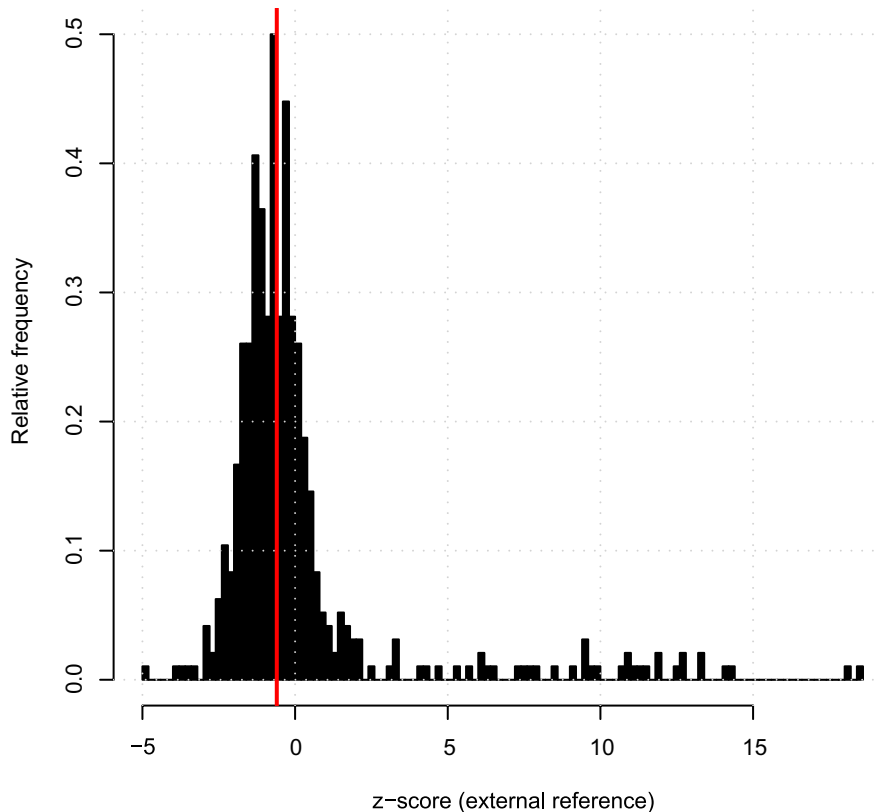
zymes to eliminate the maternal contribution of genomic regions that are methylated in fetal DNA and unmethylated in maternal DNA. The remaining nondigested fetal DNA is coamplified in the presence of a known amount of synthetic oligonucleotide to permit competitive polymerase chain reaction (PCR). This synthetic oligonucleotide has an identical sequence to the target genomic DNA, apart from 1 nucleotide that can be targeted by single-base extension and quantitative matrix-assisted laser desorption/ionization time-of-flight mass spectrometric analysis. To ensure accurate results, the assay comprises multiple markers in 4 different categories. Three markers are used to measure total DNA amounts. Three markers are used to measure chromosome Y copy numbers; 2 markers interrogate the efficiency of the methylation-specific digestion reaction, and 5 markers are used to measure fetal DNA amounts.

Methylation-based DNA discrimination was performed using 10  $\mu$ L of eluted DNA per reaction. All reagents and apparatus were obtained from Sequenom Inc, San Diego, CA, unless stated otherwise. Digestion of plasma DNA was performed for 30 minutes at 41°C by adding 25  $\mu$ L of a mixture containing 3.5X PCR buffer, 2.22 mmol/L MgCl<sub>2</sub>, 10 U HhaI (New England Biolabs, Ipswich, MA), 10 U HpaII (New England Biolabs), and 10 U ExoI (New England Biolabs). Exonuclease was added to eliminate single-stranded DNA that would escape digestion and overestimate the fetal fraction. After the restriction was complete, the enzymes were inactivated and the DNA denatured by heating the mixture for 10 minutes at 98°C. All steps following the restriction reaction were performed according to Nygren et al.<sup>10</sup>

### Library preparation

The extracted ccfDNA was used for library preparation without further fragmentation or size selection, because ccfDNA is already naturally fragmented, having an average length of approximately 160 base pairs. Low binding Eppendorf tubes were used to store 55  $\mu$ L of DNA eluent at 4°C following extraction

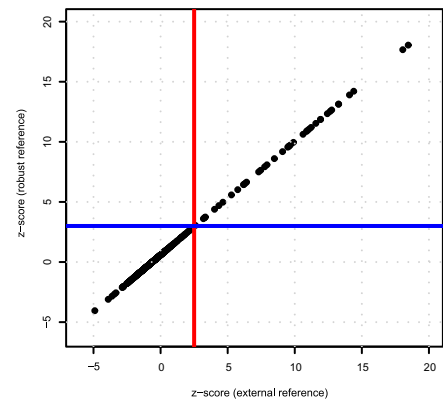
**FIGURE 1**  
Histogram of z-scores calculated using 24 reference samples



A red vertical line is drawn at  $z = -0.6$  and corresponds to the median z-score for all samples; its expected position, under the assumption that reference samples have same distribution of chromosome 21 representation as all euploid samples in this study, is at  $z = 0$ .

Ehrich. *Noninvasive detection of fetal trisomy 21. Am J Obstet Gynecol* 2011.

**FIGURE 2**  
Robust versus externally referenced z-scores



Relationship between z-scores calculated based on 24 external samples using mean and SD (x-axis) and using median and median absolute deviation (y-axis). The blue line indicates the empirically derived z-score cutoff of 2.5, which is comparable to the cutoff of 3 (red line) using the robust method.

Ehrich. *Noninvasive detection of fetal trisomy 21. Am J Obstet Gynecol* 2011.

Storage times ranged from 24-72 hours. The library preparation was carried out according to the manufacturer's specifications (Illumina Inc) with some modifications. Enzymes and buffers were sourced from Enzymatics (End Repair Mix -LC; dNTP Mix [25 mmol/L each]; Exo(-) Klenow polymerase; 10X Blue Buffer; 100 mmol/L dATP; T4 DNA Ligase; 2X Rapid Ligation Buffer) and New England Biolabs (Phusion MM). Adapter oligonucleotides, indexing oligonucleotides, and PCR primers were obtained from Illumina Inc.

Library preparation was initiated by taking 40  $\mu\text{L}$  of ccf DNA for end repair, retaining 15  $\mu\text{L}$  for QC by FQA. End repair was performed with a final concentration of 1X End Repair buffer, 24.5  $\mu\text{mol/L}$  each dNTPs, and 1  $\mu\text{L}$  of End Repair enzyme mix. The end repair reac-

ture for 30 minutes and the products were cleaned with Qiagen Qiaquick columns, eluting in 36  $\mu\text{L}$  of elution buffer (EB). 3' mono-adenylation of the end-repaired sample was performed by mixing it with a final concentration of 1X Blue Buffer, 192  $\mu\text{mol/L}$  dATP, and 5 U of Exo(-) Klenow Polymerase. The reaction was incubated at 37°C for 30 minutes and cleaned up with Qiagen MinElute columns, eluting the products in 14  $\mu\text{L}$  of EB. Adapters were ligated to the fragments by incubating for 10 minutes at room temperature with 1X Rapid Ligation buffer, 48.3 nmol/L Index PE Adapter Oligos, and 600 U T4 DNA Ligase. The ligation reaction was cleaned up with QiaQuick columns, and the sample eluted in 23  $\mu\text{L}$  of EB. The adapter-modified sample was enriched by amplifying with a high fidelity polymerase

was mixed with 1X Phusion MM, Illumina Inc PE 1.0 and 2.0 primers, and 1 of 12 index primers for a total PCR reaction volume of 50  $\mu\text{L}$ . The sample was amplified in a 0.65-mL PCR tube using a MJ Research (Bio-Rad, Hercules, CA) Model PTC-200 thermal cycler. The PCR conditions were an initial denaturation at 98°C for 30 seconds, 15 cycles of denaturation at 98°C for 10 seconds, annealing at 65°C for 30 seconds, and extension at 72°C for 30 seconds. A final extension at 72°C for 5 minutes was followed by a 4°C hold. The PCR products were cleaned with MinElute columns and the libraries eluted in 17  $\mu\text{L}$  of EB.

### Quality control of generated sequencing library

The libraries were quantified via SYBR Green quantitative PCR (qPCR) analysis as outlined by Meyer et al.<sup>24</sup> Each library was diluted 1:10<sup>8</sup> and quantified against a library standard using Power SYBR Green PCR Master Mix (ABI, Foster City, CA).

Each sample or standard was assayed in triplicate, including triplicate non-template control reactions. The sample



down to mix, then spun down. In the reaction 2  $\mu\text{L}$  of the  $1:10^8$  dilution are added to a reaction mix containing 9  $\mu\text{L}$  of Ultrapure Water, 12.5  $\mu\text{L}$  2x Power Mix, 0.5  $\mu\text{L}$  of each forward (GAT ACG GCG ACC ACC GAG AT) and reverse (CAA GCA GAA GAC GGC ATA CGA G) primer at 10  $\mu\text{mol/L}$ , and 0.5  $\mu\text{L}$  of 1 U/ $\mu\text{L}$  uracil-N-glycosylase. Amplification was performed on an ABI 7500 (Applied Biosystem, Foster City, CA). The cycling protocol began with a 2-minute uracil-N-glycosylase decontamination step at 50°C, this was followed by Hot-Start activation 95°C for 10 minutes. The program cycling was commenced and continued through 46 cycles of 15 seconds denaturation at 95°C followed by 1 minute of annealing/extension at 60°C. The final step was a 15-second denaturation at 95°C.

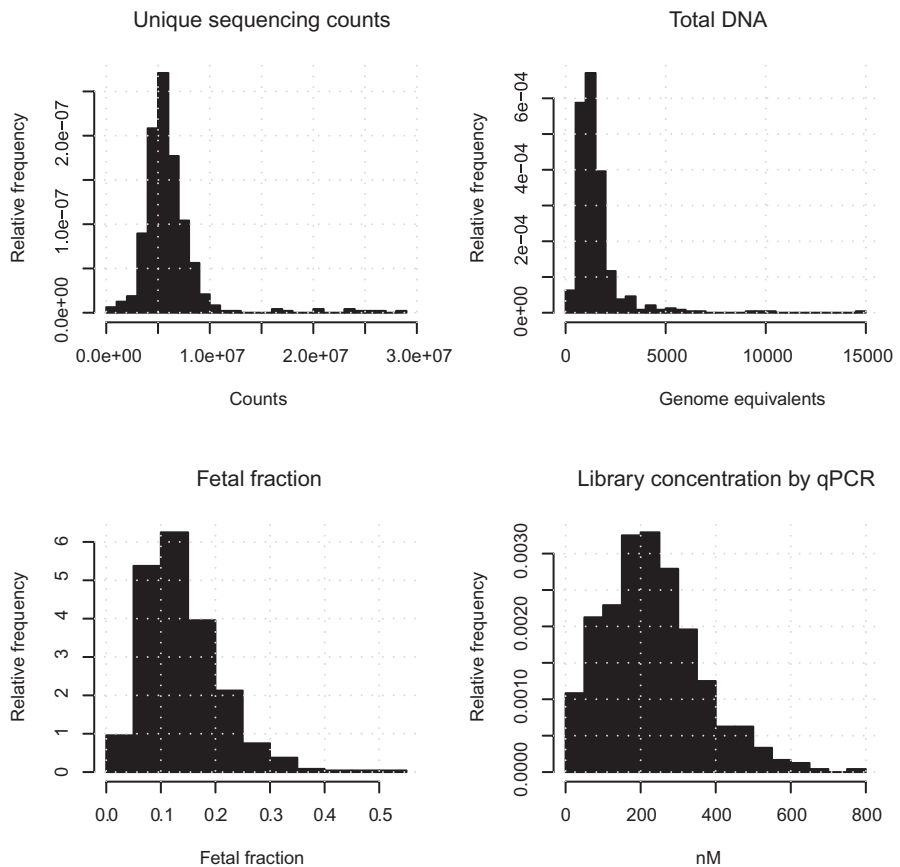
### Clustering and sequencing

Clustering and sequencing were performed according to standard Illumina Inc protocols. Individual libraries were normalized to a 5-nmol/L concentration and then clustered in 4-plex format to a final flow cell loading concentration of 1.75 pmol/L per sample or 7 pmol/L per flow cell lane. The cBOT instrument and v4 Single-Read cBOT reagent kit (Illumina Inc) were used. Thirty-six cycles of single-read multiplexed sequencing were performed on the Genome Analyzer IIx with Paired-End module using v4 SBS reagent kits and supplemental Multiplex Sequencing Primer kits (Illumina Inc). Image analysis and base calling were performed with RTA1.6/SCS2.6 software (Illumina Inc). Sequences were aligned to the UCSC hg19 human reference genome (nonrepeat-masked) using CASAVA version 1.6 (Illumina Inc).

### Data analysis

Sequence reads unique to a chromosome were counted, up to 1 mismatch (UI counts), and the chromosome 21-specific genomic representation was calculated based on these unique sequence reads. The fractional genomic representation of chromosome 21 (also referred to as the percentage of chromosome 21) was determined by dividing the number

**FIGURE 3**  
Distributions of quality control parameters



Histograms showing distribution of quality control parameters (each histogram has optimal number of bins, as calculated with Scott's rule). Very high values of unique sequence counts ( $>10$  million) are mainly obtained from samples that were analyzed in monoplex.

Ehrich. Noninvasive detection of fetal trisomy 21. *Am J Obstet Gynecol* 2011.

by all sequence reads excluding sequence reads from chromosomes X and Y. The fractional genomic representation was then standardized by subtracting the mean of a control group and dividing by the standard deviation (SD) of that same control group. Using a set of known euploid samples as a control group, this method determines the distance in SD of the tested sample to the mean of the euploid reference group. This metric, standardized fractional genomic representation (the so-called z-scores), is the metric used to classify samples as euploid or trisomy 21. Details of this procedure are outlined in Chiu et al.<sup>20</sup>

Ideally, the standardization process would be based on the true mean and true SD as calculated from a very large

of such a large set, a control group of 24 euploid samples from a previous experiment was used. Given both the limited sample size and the latent differences between the 2 different experiments, this control group may produce biased estimates of the true mean and true SD. In the design of the current study, similar to an expected clinical setting, it was anticipated that the majority of the samples would be euploid. Therefore the distribution of the z-scores should have a large normally distributed component centered on 0 and with SD close to 1. Any significant departure from this situation would be an indicator of improper standardization. Alternatively, a robust standardization can be employed given the data

# Explore Litigation Insights

Docket Alarm provides insights to develop a more informed litigation strategy and the peace of mind of knowing you're on top of things.

## Real-Time Litigation Alerts



Keep your litigation team up-to-date with **real-time alerts** and advanced team management tools built for the enterprise, all while greatly reducing PACER spend.

Our comprehensive service means we can handle Federal, State, and Administrative courts across the country.

## Advanced Docket Research



With over 230 million records, Docket Alarm's cloud-native docket research platform finds what other services can't. Coverage includes Federal, State, plus PTAB, TTAB, ITC and NLRB decisions, all in one place.

Identify arguments that have been successful in the past with full text, pinpoint searching. Link to case law cited within any court document via Fastcase.

## Analytics At Your Fingertips



Learn what happened the last time a particular judge, opposing counsel or company faced cases similar to yours.

Advanced out-of-the-box PTAB and TTAB analytics are always at your fingertips.

## API

Docket Alarm offers a powerful API (application programming interface) to developers that want to integrate case filings into their apps.

## LAW FIRMS

Build custom dashboards for your attorneys and clients with live data direct from the court.

Automate many repetitive legal tasks like conflict checks, document management, and marketing.

## FINANCIAL INSTITUTIONS

Litigation and bankruptcy checks for companies and debtors.

## E-DISCOVERY AND LEGAL VENDORS

Sync your system to PACER to automate legal marketing.