

# Noninvasive diagnosis of fetal aneuploidy by shotgun sequencing DNA from maternal blood

H. Christina Fan\*, Yair J. Blumenfeld†, Usha Chitkara‡, Louanne Hudgins‡, and Stephen R. Quake\*§

\*Department of Bioengineering, Stanford University and Howard Hughes Medical Institute, 318 Campus Drive, Clark Center, Room E300, Stanford, CA 94305; †Division of Maternal-Fetal Medicine, Department of Obstetrics and Gynecology, Stanford University, 300 Pasteur Drive, Room HH333, Stanford, CA 94305; and ‡Division of Medical Genetics, Department of Pediatrics, Stanford University, 300 Pasteur Drive, Stanford, CA 94305

Communicated by Leonard A. Herzenberg, Stanford University School of Medicine, Stanford, CA, August 22, 2008 (received for review July 13, 2008)

**We directly sequenced cell-free DNA with high-throughput shotgun sequencing technology from plasma of pregnant women, obtaining, on average, 5 million sequence tags per patient sample. This enabled us to measure the over- and underrepresentation of chromosomes from an aneuploid fetus. The sequencing approach is polymorphism-independent and therefore universally applicable for the noninvasive detection of fetal aneuploidy. Using this method, we successfully identified all nine cases of trisomy 21 (Down syndrome), two cases of trisomy 18 (Edward syndrome), and one case of trisomy 13 (Patau syndrome) in a cohort of 18 normal and aneuploid pregnancies; trisomy was detected at gestational ages as early as the 14th week. Direct sequencing also allowed us to study the characteristics of cell-free plasma DNA, and we found evidence that this DNA is enriched for sequences from nucleosomes.**

fetal DNA | next-generation sequencing | noninvasive prenatal diagnosis | Down syndrome | trisomy

**F**etal aneuploidy and other chromosomal aberrations affect 9 of 1,000 live births (1). The gold standard for diagnosing chromosomal abnormalities is karyotyping of fetal cells obtained via invasive procedures such as chorionic villus sampling and amniocentesis. These procedures impose small but potentially significant risks to both the fetus and the mother (2). Noninvasive screening of fetal aneuploidy using maternal serum markers and ultrasound are available but have limited reliability (3–5). There is therefore a desire to develop noninvasive genetic tests for fetal chromosomal abnormalities.

Since the discovery of intact fetal cells in maternal blood, there has been intense interest in trying to use them as a diagnostic window into fetal genetics (6–9). Although this has not yet moved into practical application (10), the later discovery that significant amounts of cell-free fetal nucleic acids also exist in maternal circulation has led to the development of new noninvasive prenatal genetic tests for a variety of traits (11, 12). However, measuring aneuploidy remains challenging because of the high background of maternal DNA; fetal DNA often constitutes <10% of total DNA in maternal cell-free plasma (13). Recently developed methods for aneuploidy detection focus on allelic variation between the mother and the fetus. Lo *et al.* (14) demonstrated that allelic ratios of placental-specific mRNA in maternal plasma could be used to detect trisomy 21 (T21) in certain populations. Similarly, they also showed the use of allelic ratios of imprinted genes in maternal plasma DNA to diagnose trisomy 18 (T18) (15). Dhallan *et al.* (16) used fetal-specific alleles in maternal plasma DNA to detect trisomy 21. However, these methods are limited to specific populations because they depend on the presence of genetic polymorphisms at specific loci. We and others argued that it should be possible, in principle, to use digital PCR to create a universal, polymorphism-independent test for fetal aneuploidy by using maternal plasma DNA (17–19), but because of technical challenges relating to the low fraction of fetal DNA, such a test has not yet been practically realized.

An alternative method to achieve digital quantification of DNA is direct shotgun sequencing, followed by mapping to the chromo-

some of origin and enumeration of fragments per chromosome. Recent advances in DNA-sequencing technology allow massively parallel sequencing (20), producing tens of millions of short sequence tags in a single run and enabling a deeper sampling than can be achieved by digital PCR. By counting the number of sequence tags mapped to each chromosome, the over- or underrepresentation of any chromosome in maternal plasma DNA contributed by an aneuploid fetus can be detected. This method does not require the differentiation of fetal versus maternal DNA, and with large enough tag counts, it can be applied to arbitrarily small fractions of fetal DNA. We demonstrate here the successful use of shotgun sequencing to detect fetal trisomy 21 (Down syndrome), trisomy 18 (Edward syndrome), and trisomy 13 (T13) (Patau syndrome) noninvasively by using cell-free fetal DNA in maternal plasma. This forms the basis of a universal, polymorphism-independent noninvasive diagnostic test for fetal aneuploidy. The sequence data also allowed us to characterize plasma DNA in unprecedented detail, suggesting that it is enriched for nucleosome-bound fragments.

## Results

**Shotgun Sequencing of Cell-Free Plasma DNA.** Cell-free plasma DNA from 18 pregnant women and a male donor, as well as whole-blood genomic DNA from the same male donor, were sequenced on the Solexa/Illumina platform. We obtained on average  $\approx 10$  million 25-bp sequence tags per sample. Approximately 50% (i.e.,  $\approx 5$  million) of the reads mapped uniquely to the human genome with, at most, one mismatch against the human genome, covering  $\approx 4\%$  of the entire genome. An average of  $\approx 154,000$ ,  $\approx 135,000$ , and  $\approx 65,700$  sequence tags mapped to chromosomes 13, 18, and 21, respectively. The number of sequence tags for each sample is detailed in [supporting information \(SI\) Table S1](#).

We observed a nonuniform distribution of sequence tags across each chromosome. This pattern of intrachromosomal variation was common among all samples, including randomly sheared genomic DNA, indicating that the observed variation was most probably due to sequencing artifacts. We applied a sliding window of 50 kb across each chromosome and counted the number of tags falling within each window. The median count per 50-kb window for each chromosome was selected. The median of the autosomal values was used as a normalization constant to account for the differences in

Author contributions: H.C.F., Y.J.B., U.C., L.H., and S.R.Q. designed research; H.C.F. performed research; H.C.F. analyzed data; Y.J.B., U.C., and L.H. designed the IRB-approved clinical protocol and coordinated patient recruitment and enrollment; and H.C.F., Y.J.B., and S.R.Q. wrote the paper.

Conflict of interest statement: S.R.Q. is a founder, shareholder, and consultant of Fluidigm Corporation. S.R.Q. and H.C.F. have applied for a patent relating to the method described in this study. Other authors declare no conflict of interest.

Freely available online through the PNAS open access option.

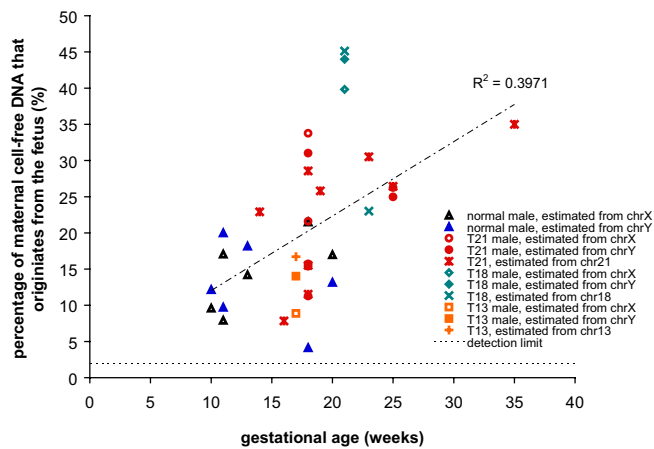
Data deposition: Sequence data have been deposited at the National Center for Biotechnology Information short read archive ([www.ncbi.nlm.nih.gov/Traces/sra/sra.cgi](http://www.ncbi.nlm.nih.gov/Traces/sra/sra.cgi)), accession no. SRA001174.

§To whom correspondence should be addressed. E-mail: [quake@stanford.edu](mailto:quake@stanford.edu).

This article contains supporting information online at [www.pnas.org/cgi/content/full/0808319105/DCSupplemental](http://www.pnas.org/cgi/content/full/0808319105/DCSupplemental).

© 2008 by The National Academy of Sciences of the USA





**Fig. 2.** Fetal DNA fraction and gestational age. The fraction of fetal DNA in maternal plasma correlates with gestational age. Fetal DNA fraction was estimated in three different ways: (i) from the additional amount of chromosomes 13, 18, and 21 sequences for T13, T18, and T21 cases, respectively; (ii) from the depletion in amount of chromosome X sequences for male cases; (iii) from the amount of chromosome Y sequences present for male cases. The horizontal dashed line represents the estimated minimum fetal DNA fraction required for the detection of aneuploidy. For each sample, the values of fetal DNA fraction calculated from the data of different chromosomes were averaged. There is a statistically significant correlation between the average fetal DNA fraction and gestational age ( $P = 0.0051$ ). The dashed line represents the simple linear regression line between the average fetal DNA fraction and gestational age. The  $R^2$  value represents the square of the correlation coefficient.

Plasma DNA of pregnant women carrying T18 fetuses (two cases) and a T13 fetus (one case) were also directly sequenced. Overrepresentation was observed for chromosomes 18 and 13 in T18 and T13 cases, respectively (Fig. 1A). Although there were not enough positive samples to measure a representative distribution, it is encouraging that all of these three positives are outliers from the distribution of disomy values. The T18 are large outliers and are clearly statistically significant ( $P < 10^{-7}$ ), whereas the statistical significance of the single T13 case is marginal ( $P < 0.05$ ). Fetal DNA fraction was also calculated from the overrepresented chromosome as described above (Fig. 2 and Table S1).

**Fetal DNA Fraction in Maternal Plasma.** Using digital TaqMan PCR for a single locus on chromosome 1, we estimated the average cell-free DNA concentration in the sequenced maternal plasma samples to be  $\approx 360$  cell equivalents per milliliter of plasma (range: 57–761 cell equivalents per milliliter of plasma) (Table S1), in rough accordance with previously reported values (13). The cohort included 12 male pregnancies (6 normal cases, 4 T21 cases, 1 T18 case, and 1 T13 case) and 6 female pregnancies (5 T21 cases and 1 T18 case). DYS14, a multicopy locus on chromosome Y, was detectable in maternal plasma by real-time PCR in all these pregnancies but not in any of the female pregnancies (data not shown). The fraction of fetal DNA in maternal cell-free plasma DNA is usually determined by comparing the amount of fetal-specific locus (such as the SRY locus on chromosome Y in male pregnancies) to that of a locus on any autosome that is common to both the mother and the fetus by using quantitative real-time PCR (13, 22, 23). We applied a similar duplex assay on a digital PCR platform (see *Materials and Methods*) to compare the counts of the SRY locus and a locus on chromosome 1 in male pregnancies. SRY locus was not detectable in any plasma DNA samples from female pregnancies. We found with digital PCR that for the majority samples, fetal DNA constituted  $\leq 10\%$  of total DNA in maternal plasma (Table S1), agreeing with previously reported values (13).

The percentage of fetal DNA among total cell-free DNA in maternal plasma can also be calculated from the density of se-

quence tags of the sex chromosomes for male pregnancies. By comparing the sequence tag density of chromosome Y of plasma DNA from male pregnancies to that of adult male plasma DNA, we estimated fetal DNA percentage to be, on average,  $\approx 19\%$  (range: 4–44%) for all male pregnancies (Table S1 and Fig. 2). Because human males have one fewer chromosome X than human females, the sequence tag density of chromosome X in male pregnancies should be  $(1 - \epsilon/2)$  of that of female pregnancies, where  $\epsilon$  is fetal DNA fraction (see *SI Appendix* for derivation). We indeed observed underrepresentation of chromosome X in male pregnancies as compared with that of female pregnancies (Fig. S2). Based on the data from chromosome X, we estimated fetal DNA percentage to be, on average,  $\approx 19\%$  (range: 8–40%) for all male pregnancies (Table S1 and Fig. 2). The fetal DNA percentage estimated from chromosomes X and Y for each male pregnancy sample correlated with each other ( $P = 0.0015$ ) (Fig. S3).

We plotted in Fig. 2 the fetal DNA fraction calculated from the overrepresentation of trisomic chromosome in aneuploid pregnancies and the underrepresentation of chromosome X and the presence of chromosome Y for male pregnancies against gestational age. The average fetal DNA fraction for each sample correlates with gestational age ( $P = 0.0051$ ), a trend that is also previously reported (13).

**Size Distribution of Cell-Free Plasma DNA.** We analyzed the sequencing libraries with a commercial lab-on-a-chip capillary electrophoresis system. There is a striking consistency in the peak fragment size, as well as the distribution around the peak, for all plasma DNA samples, including those from pregnant women and male donor. The peak fragment size was, on average, 261 bp (range: 256–264 bp) (Fig. S4). Subtracting the total length of the Solexa adaptors (92 bp) from 261 bp gives 169 bp as the actual peak fragment size. This size corresponds to the length of DNA wrapped in a chromosome, which is a nucleosome bound to a H1 histone (24). Because the library preparation includes an 18-cycle PCR, there are concerns that the distribution might be biased. To verify that the size distribution observed in the electropherograms is not an artifact of PCR, we also sequenced cell-free plasma DNA from a pregnant woman carrying a male fetus by using the 454 platform. The sample preparation for this system uses emulsion PCR, which does not require competitive amplification of the sequencing libraries and creates product that is largely independent of the amplification efficiency. The size distribution of the reads mapped to unique locations of the human genome resembled those of the Solexa sequencing libraries, with a predominant peak at 176 bp, after subtracting the length of 454 universal adaptors (Fig. 3 and Fig. S5). These findings suggest that the majority of cell-free DNA in the plasma is derived from apoptotic cells, in accordance with previous findings (22, 23, 25, 26).

Of particular interest is the size distribution of maternal and fetal DNA in maternal cell-free plasma. Two groups have previously shown that the majority of fetal DNA has size range of that of mononucleosome ( $< 200$ – $300$  bp), whereas maternal DNA is longer (22, 23). Because 454 sequencing has a targeted read length of 250 bp, we interpreted the small peak at  $\approx 250$  bp (Fig. 3 and Fig. S5) as the instrumentation limit from sequencing higher-molecular-mass fragments. We plotted the distribution of all reads and those mapped to Y chromosome (Fig. 3). We observed a slight depletion of Y-chromosome reads in the higher end of the distribution. Reads  $< 220$  bp constitute 94% of Y-chromosome and 87% of the total reads. Our results are not in complete agreement with previous findings in that we do not see as dramatic an enrichment of fetal DNA at short lengths (22, 23). Future studies will be needed to resolve this point and to eliminate any potential residual bias in the 454 sample preparation process, but it is worth noting that the ability to sequence single plasma samples permits one to measure the distribution in length enrichments across many individual



sequencing. In practice, a larger amount of DNA was used because there is sample loss during sequencing library preparation, but it may be possible to further reduce the amount of blood required for analysis.

We observed that certain chromosomes have large variations in the counts of sequenced fragments from sample to sample, and that this depends strongly on the GC content (Fig. S1A–C). It is unclear at this point whether this stems from PCR artifacts during sequencing library preparation or cluster generation or the sequencing process itself or whether it is a true biological effect relating to chromatin structure. We strongly suspect that it is an artifact because we also observe GC bias on genomic DNA control, and such bias on the Solexa sequencing platform has recently been reported (38, 39). It has a practical consequence because the sensitivity to aneuploidy detection will vary from chromosome to chromosome; fortunately the most common human aneuploidies (such as 13, 18, and 21) have low variation and therefore high detection sensitivity. Both this problem and the sample-volume limitations may possibly be resolved by the use of single-molecule sequencing technologies, which do not require the use of PCR for library preparation (40).

Plasma DNA samples used in this study were obtained ≈15–30 min after amniocentesis or chorionic villus sampling. Because these invasive procedures disrupt the interface between the placenta and maternal circulation, there have been discussions whether the amount of fetal DNA in maternal blood might increase after invasive procedures. Neither of the studies to date have observed a significant effect (41, 42). Our results support this conclusion, because using the digital PCR assay, we estimated that fetal DNA constituted ≤10% of total cell-free DNA in the majority of our maternal plasma samples. This is within the range of previously reported values in maternal plasma samples obtained before invasive procedures (13). It would be valuable to have a direct measurement addressing this point in a future study.

The average fetal DNA fraction estimated from sequencing data of sex chromosomes are higher than the values estimated from digital PCR data by an average factor of two ( $P < 0.005$ , paired  $t$  test on all male pregnancies that have complete set of data). One possible explanation for this is that the PCR step during Solexa library preparation preferentially amplifies shorter fragments, which others have found to be enriched for fetal DNA (22, 23). Our own measurements of length distribution on one sample do not support this explanation, but we also cannot reject it at this point. It should also be pointed out that using the sequence tags, we find some variation of fetal fraction even in the same sample depending on which chromosome we use to make the calculation (Fig. 2, Fig. S3 and Table S1). This is most likely because of artifacts and errors in the sequencing and mapping processes, which are substantial—recall that only half of the sequence tags map to the human genome with one error or less. Finally, it is also possible that the PCR measurements are biased because they are only sampling a tiny fraction of the fetal genome. These discrepancies will be sorted out in future studies as sequencing reliability improves, and our results show that they do not materially affect the ability to determine fetal aneuploidy.

Our sequencing data suggest that the majority of cell-free plasma DNA is of apoptotic origin and shares features of nucleosomal DNA. Because nucleosome occupancy throughout the eukaryotic genome is not necessarily uniform and depends on factors such as function, expression, or sequence of the region (30, 43), the representation of sequences from different loci in cell-free maternal plasma may not be equal, as one usually expects in genomic DNA extracted from intact cells. Thus, the quantity of a particular locus may not be representative of the quantity of the entire chromosome, and care must be taken when one designs assays for measuring gene dosage in cell-free maternal plasma DNA that target only a few loci.

Historically, because of risks associated with chorionic villus sampling and amniocentesis, invasive diagnosis of fetal aneuploidy

was primarily offered to women who were considered at risk of carrying an aneuploid fetus based on evaluation of risk factors such as maternal age, levels of serum markers, and ultrasonographic findings. Recently, an American College of Obstetricians and Gynecologists Practice Bulletin recommended that “invasive diagnostic testing for aneuploidy should be available to all women, regardless of maternal age” and that “pretest counseling should include a discussion of the risks and benefits of invasive testing compared with screening tests” (2). A noninvasive genetic test based on the results described here and in future large-scale studies would presumably carry the best of both worlds: minimal risk to the fetus while providing true genetic information. The costs of the assay are already fairly low; the sequencing cost per sample is approximately \$700, and the cost of sequencing is expected to continue to drop dramatically in the near future.

In conclusion, we demonstrated the use of massively parallel sequencing to detect fetal aneuploidy noninvasively with maternal cell-free plasma DNA. Shotgun sequencing can potentially reveal many more previously unknown features of cell-free nucleic acids such as plasma mRNA distributions, as well as epigenetic features of plasma DNA such as DNA methylation and histone modification, in fields including perinatology, oncology, and transplantation, thereby improving our understanding of the basic biology of pregnancy, early human development, and disease.

## Materials and Methods

**Subject Enrollment.** The study was approved by the Institutional Review Board of Stanford University. Pregnant women at risk for fetal aneuploidy were recruited at the Lucile Packard Children’s Hospital Perinatal Diagnostic Center of Stanford University during the period of April 2007 to May 2008. Informed consent was obtained from each participant before the blood draw. Blood was collected 15–30 min after amniocentesis or chorionic villus sampling except for one sample that was collected during the third trimester. Karyotype analysis was performed via amniocentesis or chorionic villus sampling to confirm fetal karyotype. Nine T21, 2 T18, 1 T13, and 6 normal singleton pregnancies were included in this study. The gestational age of the subjects at the time of blood draw ranged from 10 to 35 weeks (Table S1). A blood sample from a male donor was obtained from the Stanford Blood Center.

**Sample Processing and DNA Quantification.** Seven to 15 ml of peripheral blood drawn from each subject and donor was collected in EDTA tubes. Blood was centrifuged at  $1,600 \times g$  for 10 min. Plasma was transferred to microcentrifuge tubes and centrifuged at  $16,000 \times g$  for 10 min to remove residual cells. The two centrifugation steps were performed within 24 h after blood collection. Cell-free plasma was stored at  $-80^\circ\text{C}$  until further processing and was frozen and thawed only once before DNA extraction. DNA was extracted from cell-free plasma by using the QIAamp DNA Micro kit (Qiagen) or the NucleoSpin Plasma kit (Macherey–Nagel) according to the manufacturers’ instructions. Genomic DNA was extracted from  $200 \mu\text{l}$  of whole blood of the donors by using the QIAamp DNA Blood Mini kit (Qiagen). Microfluidic digital PCR (Fluidigm) was used to quantify the amount of total and fetal DNA by using TaqMan assays targeting at the EIF2C1 locus on chromosome 1 (forward: 5’-GTTGGCTTTCACAGTCT-3’; reverse: 5’-CTCCATAGCTCTCCCACTC-3’; probe: 5’-HEX-CGCCCTGCCATGTGGAA-GAT-BHQ1-3’; amplicon size: 81 bp) and the SRY locus on chromosome Y (forward: 5’-CGCTTAACATAGCAGAAGCA-3’; reverse: 5’-AGTTTCGAACTCTGG-CACCT-3’; probe: 5’-FAM-TGTCGCACTCTCTTTTGTGACA-BHQ1-3’; amplicon size: 84 bp), respectively. A TaqMan assay targeted at DYS14 (forward: 5’-ATCGTCCATTCCAGAATCA-3’; reverse: 5’-GTTGACAGCCGTGGAATC-3’; probe: 5’-FAM-TGCCACAGACTGAACTGAATGATTTTC-BHQ1-3’; amplicon size: 84 bp); a multicopy locus on chromosome Y was used for the initial determination of fetal sex from cell-free plasma DNA with traditional real-time PCR. PCR reactions were performed with  $1 \times$  iQ Supermix (Bio-Rad), 0.1% Tween-20 (microfluidic digital PCR only), 300 nM primers, and 150 nM probes. The PCR thermal cycling protocol was  $95^\circ\text{C}$  for 10 min, followed by 40 cycles of  $95^\circ\text{C}$  for 15 s and  $60^\circ\text{C}$  for 1 min. Primers and probes were purchased from IDT.

**Sequencing.** A total of 19 cell-free plasma DNA samples, including 18 from pregnant women and 1 from a male blood donor, and genomic DNA sample from whole blood of the same male donor, were sequenced on the Solexa/Illumina platform. ≈1–8 ng of DNA fragments extracted from 1.3–3.2 ml of cell-free plasma was used for sequencing library preparation (Table S1). Library preparation was carried out according to the manufacturer’s protocol with slight mod-

# 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.